



AWS ParallelCluster Benutzerhandbuch (v3)

AWS ParallelCluster



AWS ParallelCluster: AWS ParallelCluster Benutzerhandbuch (v3)

Copyright © 2024 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Die Handelsmarken und Handelsaufmachung von Amazon dürfen nicht in einer Weise in Verbindung mit nicht von Amazon stammenden Produkten oder Services verwendet werden, durch die Kunden irregeführt werden könnten oder Amazon in schlechtem Licht dargestellt oder diskreditiert werden könnte. Alle anderen Marken, die nicht im Besitz von Amazon sind, gehören den jeweiligen Besitzern, die möglicherweise mit Amazon verbunden sind oder von Amazon gesponsert werden.

Table of Contents

Was ist AWS ParallelCluster?	1
Preisgestaltung	1
Einrichten AWS ParallelCluster	2
Einrichten eines AWS-Konto	2
So melden Sie sich für ein AWS-Konto an	2
Einen Administratorbenutzer erstellen	3
Erstellen eines Schlüsselpaares	4
Installation der AWS ParallelCluster CLI	4
Installation AWS ParallelCluster in einer virtuellen Umgebung (empfohlen)	5
Installation AWS ParallelCluster in einer nicht-virtuellen Umgebung mit Pip	7
AWS ParallelCluster Als eigenständige Anwendung installieren	8
Nach der Installation zu ergreifende Schritte	9
Installation der AWS ParallelCluster Benutzeroberfläche	10
Installieren Sie die AWS ParallelCluster Benutzeroberfläche	11
Erstellen Sie eine benutzerdefinierte Domäne	14
Amazon Cognito Cognito-Benutzerpool-Optionen	16
Identifizieren Sie die AWS ParallelCluster und die AWS ParallelCluster UI-Version	19
Aktualisieren Sie die AWS ParallelCluster Benutzeroberfläche auf eine neue AWS ParallelCluster Version	19
AWS ParallelClusterDie Benutzeroberfläche kostet	20
Erste Schritte	20
Konfiguration und Erstellung eines Clusters mit der AWS ParallelCluster CLI	21
Konfigurieren und erstellen Sie einen Cluster mit der AWS ParallelCluster Benutzeroberfläche	31
Stellen Sie eine Verbindung zu einem Cluster her	33
Zugriff mehrerer Benutzer auf Cluster	34
Erstellen Sie ein Active Directory	35
Erstellen Sie einen Cluster mit einer AD-Domäne	35
Melden Sie sich bei einem Cluster an, der in eine AD-Domäne integriert ist	39
MPI-Jobs werden ausgeführt	40
Beispiel für LDAP (S) AWS Managed Microsoft AD -Clusterkonfigurationen	40
Bewährte Methoden	44
Bewährte Methoden: Auswahl des Hauptknoten-Instanztyps	44
Best Practices: Netzwerkleistung	45

Best Practices: Budgetwarnungen	47
Bewährte Methoden: Verschieben eines Clusters auf eine neueAWS ParallelCluster Minor- oder Patch-Version	47
Umstellung von AWS ParallelCluster 2.x auf 3.x	48
Benutzerdefinierte Bootstrap-Aktionen	48
AWS ParallelCluster2.x und 3.x verwenden unterschiedliche Syntax für Konfigurationsdateien	49
Inklusive Sprache	55
Scheduler-Unterstützung	55
AWS ParallelCluster CLI	56
IMDS-Konfigurationsupdate	59
Unterstützte Regionen für AWS ParallelCluster	59
Benutzen AWS ParallelCluster	61
AWS ParallelClusterUI	61
AWS LambdaVPC-Konfiguration inAWS ParallelCluster	64
AWS Identity and Access Management Berechtigungen in AWS ParallelCluster	65
AWS ParallelCluster EC2-Instanzrollen	66
AWS ParallelCluster Beispiele <code>pcluster</code> für Benutzerrichtlinien	67
AWS ParallelCluster Benutzerbeispielrichtlinien für die Verwaltung von IAM-Ressourcen	82
AWS ParallelCluster Konfigurationsparameter zur Verwaltung von IAM-Berechtigungen	88
Netzwerkkonfigurationen	103
AWS ParallelCluster in einem einzelnen öffentlichen Subnetz	105
AWS ParallelCluster mit zwei Subnetzen	107
AWS ParallelCluster in einem einzigen privaten Subnetz mit Verbindung über AWS Direct Connect	108
AWS ParallelClustermitAWS Batch Scheduler	109
AWS ParallelClusterin einem einzelnen Subnetz ohne Internetzugang ohne Internetzugang ohne Internetzugang ohne Internetzugang.	112
Login-Knoten	118
Benutzerdefinierte Bootstrap-Aktionen	122
Konfiguration	124
Argumente	128
Beispielcluster mit benutzerdefinierten Bootstrap-Aktionen	128
Beispiel für die Aktualisierung eines benutzerdefinierten Bootstrap-Skripts für IMDSv2	130
Beispiel für die Aktualisierung einer Konfiguration für IMDSv1	131
Arbeiten mit Amazon S3	132

Beispiele	132
Arbeiten mit Spot-Instances	133
Szenario 1: Spot-Instance ohne ausgeführte Aufgaben wird unterbrochen	134
Szenario 2: Spot-Instance mit Einzelknotenaufgaben wird unterbrochen	134
Szenario 3: Spot-Instance, auf der Aufgaben mit mehreren Knoten ausgeführt werden, wird unterbrochen	134
Scheduler werden unterstützt von AWS ParallelCluster	134
Slurm Workload Manager	135
AWS Batch	201
Gemeinsamer Speicher	209
Konfigurieren Sie gemeinsam genutzten Speicher	213
Arbeiten mit gemeinsam genutztem Speicher	216
Kontingente	220
Markierung	221
Überwachung AWS ParallelCluster und Protokolle	224
Integration mit Amazon CloudWatch Logs	225
CloudWatchAmazon-Dashboard	228
CloudWatchAmazon-Alarme für Cluster-Metriken	230
AWS ParallelClusterkonfigurierte Protokollrotation	233
pcClusterCLI-Protokolle	234
Ausgabeprotokolle der EC2-Konsole	235
AWS ParallelClusterUI- undAWS ParallelCluster Laufzeitprotokolle abrufen	236
Protokolle abrufen und aufbewahren	238
AWS CloudFormation benutzerdefinierte Ressource	241
Provider-Stack, gehostet von AWS ParallelCluster	242
Cluster-Ressource	244
Cluster-Operationen	247
Fehlerbehebung bei Stacks, die die AWS ParallelCluster benutzerdefinierte Ressource enthalten	248
Elastic Fabric Adapter	248
Intel MPI aktivieren	249
AWS ParallelCluster-API	251
AWS ParallelClusterAPI-Dokumentation	251
Bereitstellen mit AWS CLI	252
Aktualisierung der API	255
API aufrufen AWS ParallelCluster	256

Zugriff auf die API-Protokolle und -Metriken	258
Stellen Sie mit NICE DCV eine Connect zum Hauptknoten her	259
NICE DCV HTTPS-Zertifikat	259
Lizenzierung NICE DCV	259
Verwenden von <code>pcluster update-cluster</code>	260
Richtlinie aktualisieren: Definitionen	260
Beispiele für <code>pcluster update-cluster</code>	264
AWS ParallelCluster AMI-Anpassung	267
AWS ParallelCluster Überlegungen zur AMI-Anpassung	267
Führen Sie Validierungstests für benutzerdefinierte Komponenten durch	268
Überwachen Sie den Image Builder Builder-Prozess mit <code>pcluster</code> Befehlen, die beim Debuggen helfen	269
Weitere Überlegungen	269
Erstellen Sie Instances mit ODCR Starten Sie Instances	270
Verwenden von ODCR mit AWS ParallelCluster	270
AMI-Patching und Austausch von EC2-Instances	280
Aktualisierung oder Ersatz der Head-Knoteninstanz	281
Daten von kurzlebigen Laufwerken speichern	281
und Starten und Starten und Starten und Starten und Starten	282
Betriebssysteme	284
Überlegungen zum Betriebssystem	284
Referenz für AWS ParallelCluster	286
AWS ParallelClusterVersion 3 CLI-Befehle	286
<code>pcluster</code>	287
<code>pcluster3-config-converter</code>	331
Konfigurationsdateien	332
Cluster-Konfigurationsdatei	332
Image-Konfigurationsdateien erstellen	469
AWS ParallelCluster-API-Referenz	478
Bild erstellen	479
<code>createCluster</code>	484
Cluster löschen	490
<code>deleteClusterInstances</code>	493
Bild löschen	495
Cluster beschreiben	498
<code>describeClusterInstances</code>	506

describeComputeFleet	510
Beschreiben Sie das Bild	511
getClusterLogEreignisse	518
getClusterStackEreignisse	522
getImageLogEreignisse	526
getImageStackEreignisse	530
Cluster auflisten	534
listClusterLogStreams	538
listImageLogStröme	542
Bilder auflisten	546
listOfficialImages	549
Cluster aktualisieren	552
updateComputeFleet	558
AWS ParallelClusterPython-Bibliothek-API	561
AWS ParallelClusterAutorisierung der Python-Bibliothek	561
Installiere dasAWS ParallelClusterPython-Bibliothek	561
Cluster-API-Operationen	562
Berechne die Flotten-API-Operationen	566
Cluster- und Stack-Log-Operationen	568
Image-API-Operationen	570
Image- und Stack-Log-Operationen	573
Beispiel	576
AWS Lambda für dieAWS ParallelClusterPython-Bibliothek	577
Funktionsweise von AWS ParallelCluster	579
AWS ParallelCluster Prozesse	579
clustermgtd	579
clusterstatusmgtd	580
computemgtd	581
AWS Dienste verwendet von AWS ParallelCluster	581
Amazon API Gateway	582
AWS Batch	582
AWS CloudFormation	582
Amazon CloudWatch	583
CloudWatch Amazon-Veranstaltungen	583
CloudWatch Amazon-Protokolle	583
AWS CodeBuild	584

Amazon DynamoDB	584
Amazon Elastic Block Store	584
Amazon Elastic Compute Cloud	584
Amazon Elastic Container Registry	585
Amazon EFS	585
Amazon FSx für Lustre	585
Amazon FSx für ONTAP NetApp	586
Amazon FSx for OpenZFS	586
AWS Identity and Access Management	586
AWS Lambda	587
Amazon RDS	587
Amazon Route 53	587
Amazon Simple Notification Service	587
Amazon Simple Storage Service	587
Amazon VPC	588
Elastic Fabric Adapter	588
EC2 Image Builder	588
NICE DCV	589
AWS ParallelClusterInterne Verzeichnisse	589
Tutorials	590
Ausführen des ersten Auftrags in AWS ParallelCluster	590
Überprüfen der Installation	591
Erstellen Sie Ihren ersten Cluster	591
Loggen Sie sich in Ihren Hauptknoten ein	592
Führen Sie Ihren ersten Job mit Slurm aus	593
Erstellen eines benutzerdefinierten AWS ParallelCluster-AMIs	594
So passen Sie das AWS ParallelCluster-AMI an	595
Erstellen Sie ein benutzerdefiniertesAWS ParallelClusterAMI	596
Ändern eines AWS ParallelCluster-AMIs	603
Integrieren von Active Directory	605
Konfiguration der Verschlüsselung von gemeinsam genutztem Speicher mit einem AWS KMS Schlüssel	638
Erstellen der -Richtlinie	639
Konfiguration und Erstellung des Clusters	640
Ausführung von Jobs in einem Cluster mit mehreren Warteschlangen	641
Konfigurieren Sie Ihren Cluster	642

Erstellen Ihres -Clusters	644
Loggen Sie sich in den Hauptknoten ein	645
Job im Modus mit mehreren Warteschlangen ausführen	645
Verwenden der AWS ParallelCluster-API	649
Einen Cluster mit Slurm Accounting erstellen	664
Schritt 1: Erstellen Sie die VPC und die Subnetze für AWS ParallelCluster	665
Schritt 2: Erstellen Sie den Datenbank-Stack	665
Schritt 3: Erstellen Sie einen Cluster mit aktivierter Slurm Kontoführung	666
Zu einer früheren AWS Systems Manager-Dokumentversion zurückkehren	667
Zu einer früheren SSM-Dokumentversion zurückkehren	667
Einen Cluster erstellen mit AWS CloudFormation	670
Clustererstellung mit einem CloudFormation Schnellerstellungstapel	670
Clustererstellung mit der AWS CloudFormation Befehlszeilenschnittstelle (CLI)	673
CloudFormation Cluster-Ausgabe anzeigen	675
Greifen Sie auf Ihren Cluster zu	676
Bereinigen	676
AWS ParallelClusterUI-Integration mit Identity Center	677
IAM Identity Center aktivieren	677
Ihre Anwendung zum IAM Identity Center hinzufügen	681
Fehlerbehebung für AWS ParallelCluster	689
Es wird versucht, einen Cluster zu erstellen	690
failureCode ist OnNodeConfiguredExecutionFailure	690
failureCode ist OnNodeConfiguredDownloadFailure	691
failureCode ist OnNodeConfiguredFailure	691
failureCode ist OnNodeStartExecutionFailure	691
failureCode ist OnNodeStartDownloadFailure	692
failureCode ist OnNodeStartFailure	692
failureCode ist EbsMountFailure	693
failureCode ist EfsMountFailure	693
failureCode ist FsxMountFailure	693
failureCode ist RaidMountFailure	694
failureCode ist AmiVersionMismatch	694
failureCode ist InvalidAmi	694
failureCode lautet „failureReasonFehler HeadNodeBootstrapFailure beim Einrichten des Hauptknotens“	695

failureCode hat das Timeout HeadNodeBootstrapFailure bei der failureReason Clustererstellung überschritten.	695
failureCode lautet „failureReason Fehler HeadNodeBootstrapFailure beim Bootstrapping des Hauptknotens“.	697
failureCode ist ResourceCreationFailure	697
failureCode ist ClusterCreationFailure	697
WaitCondition timed out...Im CloudFormation Stapel sehen	698
Resource creation cancelledIm CloudFormation Stapel sehen	698
Sehen Failed to run cfn-init... oder andere Fehler im AWS CloudFormation Stapel	698
Sehen chef-client.log endet mit INFO: Waiting for static fleet capacity provisioning	698
Sehen Failed to run preinstall or postinstall in cfn-init.log	698
This AMI was created with xxx, but is trying to be used with xxx...Im CloudFormation Stapel sehen	699
This AMI was not baked by AWS ParallelCluster...Im CloudFormation Stapel sehen	699
Der pcluster create-cluster Befehl Seeing kann nicht lokal ausgeführt werden	699
Zusätzliche Unterstützung	699
Ich versuche, einen Job auszuführen	699
srunDer interaktive Job schlägt mit einem Fehler fehl srun: error:	
fwd_tree_thread: can't find address for <host>, check slurm.conf	699
Der Job steckt im CF Status mit dem squeue Befehl fest	700
Großaufträge ausführen und sehen nfsd: too many open connections, consider increasing the number of threads in /var/log/messages	700
Einen MPI-Job ausführen	701
Es wird versucht, einen Cluster zu aktualisieren	702
pcluster update-clusterDer Befehl kann nicht lokal ausgeführt werden	702
clusterStatusDas Sehen erfolgt UPDATE_FAILED mit einem pcluster describe-cluster Befehl	702
Das Cluster-Update hat das Zeitlimit überschritten	702
Ich versuche, auf Speicher zuzugreifen	703
Verwenden eines externen Amazon FSx for Lustre-Dateisystems	703
Verwenden eines externen Amazon Elastic File System-Dateisystems	703
Es wird versucht, einen Cluster zu löschen	703
Der pcluster delete-cluster Befehl kann nicht lokal ausgeführt werden	703

Der Cluster-Stack kann nicht gelöscht werden	703
Ich versuche, den AWS ParallelCluster API-Stack zu aktualisieren	703
Fehler bei der Initialisierung von Compute-Knoten werden angezeigt	704
Ich sehe rein <code>Node bootstrap errorclustermgtd.log</code>	704
Ich habe On-Demand-Kapazitätsreservierungen (ODCRs) oder zonale Reserved Instances konfiguriert	704
Ich sehe <code>An error occurred (VcpuLimitExceeded)slurm_resume.log</code> , wenn ich einen Job nicht ausführen kann, oder wenn ich keinen Cluster erstellen kann <code>clustermgtd.log</code>	706
Ich sehe <code>An error occurred (InsufficientInstanceCapacity)slurm_resume.log</code> , wenn ich einen Job nicht ausführen kann, oder wenn ich keinen Cluster erstellen kann <code>clustermgtd.log</code>	706
Ich sehe, dass sich die Knoten im DOWN Zustand von befinden Reason <code>(Code:InsufficientInstanceCapacity)...</code>	706
Seht rein <code>cannot change locale (en_US.utf-8) because it has an invalid nameslurm_resume.log</code>	706
Keines der vorherigen Szenarien trifft auf meine Situation zu	707
Anleitung <code>ster-ster-ster-ster-ster-ster-ster-ster-</code>	707
Diagramm mit Fehlern bei der Instanzbereitstellung anzeigen	708
Das Diagramm „Ungesunde Instance-Fehler“ anzeigen	710
Das Diagramm zur Leerlaufzeit der Rechenflotte anzeigen	712
Behebung von Problemen bei der Clusterbereitstellung	712
AWS CloudFormationEreignisse anzeigen auf <code>CREATE_FAILED</code>	713
Verwenden Sie die CLI, um Protokollstreams anzuzeigen	715
Erstellen Sie den ausgefallenen Cluster erneut mit <code>rollback-on-failure</code>	717
Behebung von Skalierungsproblemen	718
Wichtige Protokolle für das Debuggen	720
Es <code>InsufficientInstanceCapacity</code> wird ein Fehler angezeigt, <code>slurm_resume.log</code> wenn ich einen Job nicht ausführen kann oder <code>clustermgtd.log</code> wenn ich keinen Cluster erstellen kann	706
Behebung von Problemen bei der Knoteninitialisierung	722
Behebung unerwarteter Knotenersetzungen und -beendigungen	725
Ersetzen, Beenden oder Herunterfahren problematischer Instanzen und Knoten	726
Status der Warteschlange (Partition) <code>Inactive</code>	727
Behebung anderer bekannter Knoten- und Jobprobleme	727
Probleme beim Platzieren von Gruppen und beim Starten von Instances	727

Verzeichnisse, die nicht ersetzt werden können	727
Behebung von Problemen in NICE DCV	728
Logs für NICE DCV	728
Probleme mit Ubuntu NICE DCV	728
Behebung von Problemen in Clustern mit AWS Batch Integration	729
Probleme mit dem Hauptknoten	729
Probleme mit der Rechenleistung	729
Fehlschläge Job	730
Verbindungstimeout bei Endpunkt-URL-Fehler	730
Problembehandlung bei der Mehrbenutzerintegration mit Active Directory	730
Active Directory-spezifische Fehlerbehebung	731
Aktivieren Sie den Debug-Modus	732
Wie wechselt man von LDAPS zu LDAP	732
Wie deaktiviere ich die Überprüfung von LDAPS-Serverzertifikaten	732
Wie melde ich mich mit einem SSH-Schlüssel statt mit einem Passwort an	733
Wie setze ich ein Benutzerpasswort und abgelaufene Passwörter zurück	733
Wie verifiziert man die beigetretene Domäne	734
Wie behebt man Probleme mit Zertifikaten	734
Wie kann überprüft werden, ob die Integration mit Active Directory funktioniert	737
Wie behebt man Probleme bei der Anmeldung bei Rechenknoten	737
Bekannte Probleme mit SimCenter StarCCM+-Jobs in einer Mehrbenutzerumgebung	738
Bekannte Probleme bei der Benutzernamenauflösung	738
Wie löst man Probleme beim Erstellen von Home-Verzeichnissen	739
Behebung von Problemen mit benutzerdefinierten AMIs	740
Fehlerbehebung bei einem Timeout für ein Cluster-Update, wenn es nicht läuft cfn-hup	741
Fehlerbehebung im Netzwerk	742
Probleme mit Clustern in einem einzelnen öffentlichen Subnetz	742
Das Cluster-Update ist bei der benutzerdefinierten Aktion onNodeUpdated fehlgeschlagen	742
Fehler bei der benutzerdefinierten Slurm Konfiguration werden angezeigt	742
Cluster-Alarme	743
Zusätzliche Unterstützung	744
AWS ParallelCluster Unterstützungspolitik	745
Sicherheit	746
Sicherheitsinformationen für Dienste, die von AWS ParallelCluster	747
Datenschutz	747
Datenverschlüsselung	748

Weitere Informationen finden Sie auch unter	750
Identity and Access Management	750
Compliance-Validierung	751
Erzwingen von TLS 1.2	752
Ermitteln Ihrer derzeit unterstützten Protokolle	752
Kompilieren von OpenSSL und Python	754
Versionshinweise und Dokumentenverlauf	756
.....	dcclvi

Was ist AWS ParallelCluster?

AWS ParallelCluster ist ein von AWS unterstütztes Open-Source-Cluster-Verwaltungstool, mit dem Sie High Performance Computing (HPC)-Cluster in der AWS Cloud bereitstellen und verwalten können. Es richtet automatisch die erforderlichen Rechenressourcen, den Scheduler und das gemeinsame Dateisystem ein. Sie können AWS ParallelCluster mit AWS Batch und Slurm Scheduler verwenden.

Mit AWS ParallelCluster können Sie schnell Machbarkeitsnachweise und HPC-Produktionsumgebungen erstellen und bereitstellen. Sie können darüber hinaus auch einen übergeordneten Workflow erstellen und bereitstellen AWS ParallelCluster, z. B. ein Genomikportal, das einen gesamten DNA-Sequenzierungsworkflow automatisiert.

Sie können mit AWS ParallelCluster den folgenden Methoden darauf zugreifen:

- [AWS ParallelCluster Befehlszeilenschnittstelle \(CLI\)](#)
- [AWS ParallelCluster-API](#)
- [AWS ParallelClusterUI](#) (mit Version 3.5.0 hinzugefügt)
- [AWS ParallelCluster Python-Bibliothek-API](#) (mit Version 3.5.0 hinzugefügt)
- Als [AWS CloudFormation benutzerdefinierte Ressource](#) (hinzugefügt mit Version 3.6.0)

Preisgestaltung

Wenn Sie die AWS ParallelCluster Befehlszeilenschnittstelle (CLI) oder die API verwenden, zahlen Sie nur für die AWS Ressourcen, die beim Erstellen oder Aktualisieren von AWS ParallelCluster Images und Clustern erstellt oder aktualisiert werden. Weitere Informationen finden Sie unter [AWS Dienste verwendet von AWS ParallelCluster](#).

Die AWS ParallelCluster Benutzeroberfläche basiert auf einer serverlosen Architektur und Sie können sie in den meisten Fällen innerhalb der Kategorie AWS Free Tier verwenden. Weitere Informationen finden Sie unter [AWS ParallelCluster Die Benutzeroberfläche kostet](#).

Einrichten AWS ParallelCluster

Themen

- [Einrichten eines AWS-Konto](#)
- [Erstellen eines Schlüsselpaares](#)
- [Installation der AWS ParallelCluster Befehlszeilenschnittstelle \(CLI\)](#)
- [Nach der Installation zu ergreifende Schritte](#)
- [Installation der AWS ParallelCluster Benutzeroberfläche](#)
- [Erste Schritte mit AWS ParallelCluster](#)
- [Zugriff mehrerer Benutzer auf Cluster](#)
- [Bewährte Methoden](#)
- [Umstellung von AWS ParallelCluster 2.x auf 3.x](#)
- [Unterstützte Regionen für AWS ParallelCluster](#)

Einrichten eines AWS-Konto

Richten Sie ein zu AWS verwendendes Konto ein AWS ParallelCluster.

So melden Sie sich für ein AWS-Konto an

Wenn Sie kein AWS-Konto haben, führen Sie die folgenden Schritte zum Erstellen durch.

Anmeldung für ein AWS-Konto

1. Öffnen Sie <https://portal.aws.amazon.com/billing/signup>.
2. Folgen Sie den Online-Anweisungen.

Bei der Anmeldung müssen Sie auch einen Telefonanruf entgegennehmen und einen Verifizierungscode über die Tasten eingeben.

Wenn Sie sich für ein AWS-Konto anmelden, wird ein Root-Benutzer des AWS-Kontos erstellt. Der Root-Benutzer hat Zugriff auf alle AWS-Services und Ressourcen des Kontos. Als bewährte Sicherheitsmethode weisen Sie einem [Administratorbenutzer Administratorzugriff](#) zu und verwenden Sie nur den Root-Benutzer, um [Aufgaben auszuführen, die Root-Benutzerzugriff erfordern](#).

AWS sendet Ihnen eine Bestätigungs-E-Mail, sobald die Anmeldung abgeschlossen ist. Sie können jederzeit Ihre aktuelle Kontoaktivität anzeigen und Ihr Konto verwalten. Rufen Sie dazu <https://aws.amazon.com/> auf und klicken Sie auf Mein Konto.

Einen Administratorbenutzer erstellen

Nachdem Sie sich für einen angemeldet haben AWS-Konto, sichern Sie Ihren Root-Benutzer des AWS-Kontos AWS IAM Identity Center, aktivieren und erstellen Sie einen Administratorbenutzer, sodass Sie den Root-Benutzer nicht für alltägliche Aufgaben verwenden.

Schützen Ihres Root-Benutzer des AWS-Kontos

1. Melden Sie sich bei der [AWS Management Console](#) als Kontobesitzer an, indem Sie Root-Benutzer auswählen und Ihre AWS-Konto E-Mail-Adresse eingeben. Geben Sie auf der nächsten Seite Ihr Passwort ein.

Hilfe bei der Anmeldung mit dem Root-Benutzer finden Sie unter [Anmelden als Root-Benutzer](#) im AWS-Anmeldung Benutzerhandbuch zu .

2. Aktivieren Sie die Multi-Faktor-Authentifizierung (MFA) für den Root-Benutzer.

Anweisungen dazu finden Sie unter [Aktivieren eines virtuellen MFA-Geräts für den Root-Benutzer Ihres AWS-Konto \(Konsole\)](#) im IAM-Benutzerhandbuch.

Erstellen eines Administratorbenutzers

1. Aktivieren Sie IAM Identity Center.

Anweisungen finden Sie unter [Aktivieren AWS IAM Identity Center](#) im AWS IAM Identity Center Benutzerhandbuch.

2. Gewähren Sie in IAM Identity Center einem Administratorbenutzer Administratorzugriff.

Ein Tutorial zur Verwendung von IAM-Identity-Center-Verzeichnis als Identitätsquelle finden [Sie unter Benutzerzugriff mit der Standardeinstellung konfigurieren IAM-Identity-Center-Verzeichnis](#) im AWS IAM Identity Center Benutzerhandbuch.

Als Administratorbenutzer anmelden

- Um sich mit Ihrem IAM-Identity-Center-Benutzer anzumelden, verwenden Sie die Anmelde-URL, die an Ihre E-Mail-Adresse gesendet wurde, als Sie den IAM-Identity-Center-Benutzer erstellt haben.

Hilfe bei der Anmeldung mit einem IAM-Identity-Center-Benutzer finden Sie unter [Anmelden beim AWS-Zugangsportale](#) im AWS-Anmeldung Benutzerhandbuch zu.

Erstellen eines Schlüsselpaares

Um Cluster bereitzustellen, werden EC2-Instances AWS ParallelCluster gestartet, um den Cluster-Hauptknoten und die Rechenknoten zu erstellen. Um Cluster-Aufgaben wie das Ausführen und Überwachen von Jobs oder das Verwalten von Benutzern auszuführen, müssen Sie auf den Cluster-Hauptknoten zugreifen können. Um zu überprüfen, ob Sie mit SSH auf die Head-Node-Instance zugreifen können, müssen Sie ein EC2-Schlüsselpaar verwenden. Informationen zum Erstellen eines key pair finden Sie unter [Erstellen eines key pair](#) im Amazon Elastic Compute Cloud-Benutzerhandbuch für Linux-Instances.

Installation der AWS ParallelCluster Befehlszeilenschnittstelle (CLI)

AWS ParallelCluster wird als Python-Paket verteilt und mit dem pip Python-Paketmanager installiert. Anweisungen zur Installation von Python-Paketen finden Sie unter [Pakete installieren](#) im Python Packaging User Guide.

Möglichkeiten zur Installation AWS ParallelCluster:

- [Installation AWS ParallelCluster in einer virtuellen Umgebung \(empfohlen\)](#)
- [Installation AWS ParallelCluster in einer nicht-virtuellen Umgebung mit Pip](#)
- [AWS ParallelCluster Als eigenständige Anwendung installieren](#)

Die Versionsnummer der neuesten CLI finden Sie auf der [Releases-Seite unter GitHub](#). In diesem Handbuch wird bei den Befehlsbeispielen davon ausgegangen, dass Sie eine neuere Version von Python als Version 3.6 installiert haben. Die pip-Beispielbefehle verwenden die pip3-Version.

Verwalte sowohl AWS ParallelCluster 2 als auch AWS ParallelCluster 3

Für Kunden, die sowohl AWS ParallelCluster 2 als auch AWS ParallelCluster 3 verwenden und die CLIs für beide Pakete verwalten möchten, empfehlen wir, AWS ParallelCluster 2 und AWS ParallelCluster 3 in unterschiedlichen [virtuellen Umgebungen](#) zu installieren. Dadurch wird sichergestellt, dass Sie weiterhin jede Version von AWS ParallelCluster und alle zugehörigen Clusterressourcen verwenden können.

Installation AWS ParallelCluster in einer virtuellen Umgebung (empfohlen)

Wir empfehlen die Installation AWS ParallelCluster in einer virtuellen Umgebung, um Versionskonflikte mit anderen pip Paketen zu vermeiden.

Voraussetzungen

- AWS ParallelCluster benötigt Python 3.7 oder höher. Falls Sie es noch nicht installiert haben, [laden Sie eine kompatible Version](#) für Ihre Plattform von [python.org](#) herunter.

Zur Installation AWS ParallelCluster in einer virtuellen Umgebung

1. Falls `virtualenv` nicht installiert, installieren Sie `virtualenv` mit `pip3`. Wenn `python3 -m virtualenv help` Hilfeinformationen anzeigt, fahren Sie mit Schritt 2 fort.

```
$ python3 -m pip install --upgrade pip
$ python3 -m pip install --user --upgrade virtualenv
```

Führen Sie `exit` aus, um das aktuelle Terminalfenster zu verlassen, und öffnen Sie ein neues Terminalfenster, um Änderungen an der Umgebung zu übernehmen.

2. Erstellen Sie eine virtuelle Umgebung und benennen Sie sie.

```
$ python3 -m virtualenv ~/apc-ve
```

Alternativ können Sie mit der `-p`-Option eine bestimmte Version von Python angeben.

```
$ python3 -m virtualenv -p $(which python3) ~/apc-ve
```

3. Aktivieren Sie die neue virtuelle Umgebung.

```
$ source ~/apc-ve/bin/activate
```

4. Installieren Sie AWS ParallelCluster in Ihrer virtuellen Umgebung.

```
(apc-ve)~$ python3 -m pip install --upgrade "aws-parallelcluster"
```

5. Installieren Sie Node Version Manager und die neueste Version von Long-Term Support (LTS) Node.js. AWS Cloud Development Kit (AWS CDK) (AWS CDK) benötigt Node.js CloudFormation für die Vorlagengenerierung.

Note

Wenn Ihre Installation von Node.js auf Ihrer Plattform nicht funktioniert, können Sie eine LTS-Version vor der neuesten LTS-Version installieren. Weitere Informationen finden Sie im [Veröffentlichungszeitplan von Node.js](#) und in den [AWS CDK-Voraussetzungen](#).

Beispiel für einen Installationsbefehl für Node.js:

```
$ nvm install --lts=Hydrogen
```

```
$ curl -o- https://raw.githubusercontent.com/nvm-sh/nvm/v0.38.0/install.sh | bash
$ chmod ug+x ~/.nvm/nvm.sh
$ source ~/.nvm/nvm.sh
$ nvm install --lts
$ node --version
```

6. Stellen Sie sicher, dass AWS ParallelCluster es korrekt installiert ist.

```
$ pcluster version
{
  "version": "3.7.0"
}
```

Sie können den Befehl `deactivate` verwenden, um die virtuelle Umgebung zu beenden. Jedes Mal, wenn Sie eine Sitzung starten, müssen Sie [die Umgebung erneut aktivieren](#).

Um auf die neueste Version von zu aktualisieren AWS ParallelCluster, führen Sie den Installationsbefehl erneut aus.

```
(apc-ve)~$ python3 -m pip install --upgrade "aws-parallelcluster"
```

Installation AWS ParallelCluster in einer nicht-virtuellen Umgebung mit Pip

Voraussetzungen

- AWS ParallelCluster benötigt Python 3.7 oder höher. Falls Sie es noch nicht installiert haben, [laden Sie eine kompatible Version](#) für Ihre Plattform von [python.org](#) herunter.

Installieren AWS ParallelCluster

1. `pip` Zum Installieren verwenden AWS ParallelCluster.

```
$ python3 -m pip install "aws-parallelcluster" --upgrade --user
```

Wenn Sie den `--user` Switch verwenden, `pip` installiert er AWS ParallelCluster unter `~/local/bin`.

2. Installieren Sie Node Version Manager und die neueste Version von Long-Term Support (LTS) Node.js. AWS Cloud Development Kit (AWS CDK) (AWS CDK) benötigt Node.js CloudFormation für die Vorlagengenerierung.

Note

Wenn Ihre Installation von Node.js auf Ihrer Plattform nicht funktioniert, können Sie eine LTS-Version vor der neuesten LTS-Version installieren. Weitere Informationen finden Sie im [Veröffentlichungszeitplan von Node.js](#) und in den [AWS CDK-Voraussetzungen](#).

```
$ nvm install --lts=Gallium
```

```
$ curl -o- https://raw.githubusercontent.com/nvm-sh/nvm/v0.38.0/install.sh | bash
$ chmod ug+x ~/.nvm/nvm.sh
$ source ~/.nvm/nvm.sh
$ nvm install --lts
$ node --version
```

3. Stellen Sie sicher, dass es korrekt AWS ParallelCluster installiert wurde.

```
$ pcluster version
{
```

```
"version": "3.7.0"  
}
```

4. Führen Sie das Installationsprogramm erneut aus, um auf die neueste Version zu aktualisieren.

```
$ python3 -m pip install "aws-parallelcluster" --upgrade --user
```

AWS ParallelCluster Als eigenständige Anwendung installieren

AWS ParallelCluster Als eigenständige Anwendung in Ihrer Umgebung installieren. Folgen Sie den Anweisungen zur Installation AWS ParallelCluster auf einem verfügbaren Betriebssystem im folgenden Abschnitt.

Voraussetzungen

- Eine Umgebung mit einem Betriebssystem, das mit einer verfügbaren Version des Installationsprogramms kompatibel ist.

Note

AWS ParallelCluster benötigt NodeJS. AWS ParallelCluster Das Installationsprogramm enthält eine gebündelte Version von NodeJS (v18), die installiert wird, falls sie noch nicht existiert. Wenn Ihr System nicht mit NodeJS v18 kompatibel ist, sollten Sie NodeJS vor der Installation installieren. AWS ParallelCluster

Linux

Linux x86 (64-bit)

Installieren Sie es in Ihrer Umgebung. AWS ParallelCluster

1. Laden Sie das neueste [Pcluster-Installationsprogramm](#) herunter.
2. Entpacken Sie das Installationspaket und installieren Sie es mit AWS ParallelCluster den folgenden Befehlen:

```
$ unzip pcluster-installer-bundle-3.9.1.608-node-v18.17.1-Linux_x86_64-signed.zip  
-d pcluster-installer-bundle  
$ cd pcluster-installer-bundle
```

```
$ chmod +x install_pcluster.sh
```

3. Führen Sie das folgende Installationskript aus.

```
$ bash install_pcluster.sh
```

4. Stellen Sie sicher, dass AWS ParallelCluster es korrekt installiert ist.

```
$ pcluster version
{
  "version": "3.9.1"
}
```

Behebung von **pcluster** Installationsfehlern

- Wenn die AWS ParallelCluster Version in Schritt 4 nicht zurückgegeben wird, starten Sie das Terminal neu, `bash_profile` um `source` die `PATH` Variable so zu aktualisieren, dass sie das neue Binärverzeichnis enthält, wie im folgenden Beispiel gezeigt:

```
$ source ~/.bash_profile
```

- Wenn Sie Ihre `pcluster` Installation verwenden, um Cluster mit `CustomActions` angegebenen `HTTPS`-Ressourcen und nicht mit `S3-URIs` zu erstellen, wird möglicherweise eine `WARNING` Meldung angezeigt, dass diese Ressourcen möglicherweise nicht verifiziert wurden (`[SSL: CERTIFICATE_VERIFY_FAILED]`). Dies wird durch ein bekanntes Problem verursacht. Sie können diese Warnung ignorieren, wenn Sie der Authentizität der angegebenen Ressourcen vertrauen.

Frühere Versionen des Installationspakets

- None

Nach der Installation zu ergreifende Schritte

Sie können überprüfen, ob die AWS ParallelCluster Installation korrekt ist, indem Sie Folgendes ausführen [pcluster version](#).

```
$ pcluster version
```

```
{  
  "version": "3.7.0"  
}
```

AWS ParallelCluster wird regelmäßig aktualisiert. Um auf die neueste Version von zu aktualisieren AWS ParallelCluster, führen Sie den Installationsbefehl erneut aus. Weitere Informationen zur neuesten Version von AWS ParallelCluster finden Sie in den [AWS ParallelCluster Versionshinweisen](#).

```
$ pip3 install aws-parallelcluster --upgrade --user
```

Verwenden Sie zur Deinstallation AWS ParallelCluster `pip3 uninstall`.

```
$ pip3 uninstall aws-parallelcluster
```

Wenn Sie nicht über Python und `pip3` verfügen, befolgen Sie die Anleitung für Ihre Umgebung.

Installation der AWS ParallelCluster Benutzeroberfläche

Die AWS ParallelCluster Benutzeroberfläche ist eine webbasierte Benutzeroberfläche, die die AWS ParallelCluster `pccluster` CLI widerspiegelt und gleichzeitig ein konsolenähnliches Erlebnis bietet. Sie installieren und greifen auf die AWS ParallelCluster Benutzeroberfläche in Ihrem zu. AWS-Konto Wenn Sie es ausführen, greift die AWS ParallelCluster Benutzeroberfläche auf eine Instanz der API zu, die auf Amazon AWS ParallelCluster API Gateway in Ihrem AWS-Konto gehostet wird. Weitere Informationen zur AWS ParallelCluster Benutzeroberfläche finden Sie unter [AWS ParallelClusterUI](#).

Voraussetzungen:

- Ein AWS-Konto
- Zugriff auf die AWS Management Console

Themen

- [Installieren Sie die AWS ParallelCluster Benutzeroberfläche](#)
- [Erstellen Sie eine benutzerdefinierte Domäne](#)
- [Amazon Cognito Cognito-Benutzerpool-Optionen](#)
- [Identifizieren Sie die AWS ParallelCluster und die AWS ParallelCluster UI-Version](#)
- [Aktualisieren Sie die AWS ParallelCluster Benutzeroberfläche auf eine neue AWS ParallelCluster Version](#)

- [AWS ParallelClusterDie Benutzeroberfläche kostet](#)

Installieren Sie die AWS ParallelCluster Benutzeroberfläche

Um eine Instanz der AWS ParallelCluster Benutzeroberfläche zu installieren, wählen Sie einen AWS CloudFormation Schnellerstellungslink für die InstanzAWS-Region, in der Sie Cluster erstellen. Über die Schnellerstellungs-URL gelangen Sie zu einem Assistenten zum Erstellen von Stacks, in dem Sie Eingaben für die Schnellerstellung von Stack-Vorlagen eingeben und den Stack bereitstellen. Weitere Informationen zu CloudFormation Schnellerstellungsstapeln finden Sie im Benutzerhandbuch unter Schnellerstellungslinks für Stacks [erstellen](#). AWS CloudFormation

Note

Sie können Cluster oder Images nur mit derselben AWS ParallelCluster Version erstellen und bearbeiten, die Sie für die Installation der Benutzeroberfläche verwenden. AWS ParallelCluster

AWS ParallelClusterSchnellerstellung von Links auf der Benutzeroberfläche nach Regionen

Schnelle Erstellung von Links über die Benutzeroberfläche

[us-east-1](#)

[us-east-2](#)

[us-west-1](#)

[us-west-2](#)

[eu-west-1](#)

[eu-west-2](#)

[eu-west-3](#)

[eu-central-1](#)

Schnelle Erstellung von Links über die Benutzeroberfläche

[eu-north-1](#)

[me-south-1](#)

[sa-east-1](#)

[ca-central-1](#)

[ap-northeast-1](#)

[ap-northeast-2](#)

[ap-south-1](#)

[ap-southeast-1](#)

[ap-southeast-2](#)

[us-gov-west-1](#)


Verwenden Sie einen Link zur AWS CloudFormation Schnellerstellung, um einen AWS ParallelCluster UI-Stack mit verschachtelten Amazon Cognito-, API Gateway- und Amazon EC2 Systems Manager Manager-Stacks bereitzustellen.

1. Melden Sie sich an der AWS Management Console an.
2. Stellen Sie die AWS ParallelCluster Benutzeroberfläche bereit, indem Sie einen AWS-Region Schnellerstellungslink aus der Tabelle am Anfang dieses Abschnitts auswählen. Dadurch gelangen Sie zum CloudFormation Create Stack Wizard in der Konsole.
3. Geben Sie eine gültige E-Mail-Adresse für die E-Mail-Adresse des Administrators ein.

Nach erfolgreichem Abschluss der Bereitstellung sendet Ihnen die AWS ParallelCluster Benutzeroberfläche ein temporäres Passwort an diese E-Mail-Adresse. Sie verwenden das temporäre Passwort, um auf die AWS ParallelCluster Benutzeroberfläche zuzugreifen. Wenn Sie die E-Mail löschen, bevor Sie das temporäre Passwort speichern oder verwenden, müssen Sie den Stack löschen und die AWS ParallelCluster Benutzeroberfläche neu installieren.

4. Lassen Sie den Rest des Formulars leer oder geben Sie Werte für (optionale) Parameter ein, um den AWS ParallelCluster UI-Build anzupassen.

5. Notieren Sie sich den Namen des Stacks zur Verwendung in späteren Schritten.
6. Navigieren Sie zu Capabilities. Stimmen Sie den CloudFormation Funktionen zu.
7. Wählen Sie Create (Erstellen) aus. Es dauert etwa 15 Minuten, bis die AWS ParallelCluster API- und AWS ParallelCluster UI-Bereitstellung abgeschlossen ist.
8. Sehen Sie sich die Stack-Details an, während der Stack erstellt wird.
9. Öffnen Sie nach Abschluss der Bereitstellung die Administrator-E-Mail, die an die von Ihnen eingegebene Adresse gesendet wurde. Sie enthält ein temporäres Passwort, das Sie für den Zugriff auf die AWS ParallelCluster Benutzeroberfläche verwenden. Wenn Sie die E-Mail dauerhaft löschen und sich noch nicht bei der AWS ParallelCluster Benutzeroberfläche angemeldet haben, müssen Sie den von Ihnen erstellten AWS ParallelCluster UI-Stack löschen und die AWS ParallelCluster Benutzeroberfläche erneut installieren.
10. Wählen Sie in der AWS CloudFormation Konsolenliste der Stacks den Link zu dem Stack-Namen aus, den Sie sich in einem vorherigen Schritt notiert haben.
11. Wählen Sie in den Stack-Details die Option Outputs und dann den Link für den Schlüssel mit dem Namen **Stackname** URL aus, um die Benutzeroberfläche zu öffnen. AWS ParallelCluster **Stackname ist der Name**, den Sie sich in einem vorherigen Schritt notiert haben.
12. Geben Sie das temporäre Passwort ein. Folgen Sie den Schritten, um Ihr eigenes Passwort zu erstellen und sich anzumelden.
13. Sie befinden sich jetzt auf der Startseite der von AWS-Region Ihnen ausgewählten AWS ParallelCluster Benutzeroberfläche.
14. Informationen zu den ersten Schritten mit der AWS ParallelCluster Benutzeroberfläche finden Sie unter [Konfiguration und Erstellung eines Clusters mit der AWS ParallelCluster Benutzeroberfläche](#).

 Note

PCUI-Sitzungen haben eine Standarddauer von 5 Minuten. Dies ist der Mindestwert, den Cognito ab PCUI 2023.12.0 bereitstellt. Daher wird erwartet, dass ein Benutzer, der aus den Cognito-Benutzerpools entfernt wurde, weiterhin auf das System zugreifen kann, bis die Sitzung abläuft.

Erstellen Sie eine benutzerdefinierte Domäne

Erfahren Sie, wie Sie eine benutzerdefinierte Domain für die AWS ParallelCluster Benutzeroberfläche erstellen. Die Benutzeroberfläche wird auf Amazon API Gateway in Ihrem gehostetAWS-Konto. Sie können eine benutzerdefinierte Domain in der API Gateway Gateway-Konsole erstellen.

Voraussetzungen:

- Sie haben eineAWS-Konto.
- Sie haben eine AWS ParallelCluster UI-Instanz, auf die Sie zugreifen können.
- Sie besitzen eine Domain.
- Sie können die grundlegenden DNS-Einstellungen (Domain Name System) ändern.

Schritt 1: Erstellen Sie eine neue Domain im Amazon API Gateway

1. Navigieren Sie im zu [API Gateway AWS Management Console](#), wo Ihre AWS ParallelCluster UI-API aufgeführt ist.
2. Wählen Sie im Navigationsbereich Custom Domain Namens (Benutzerdefinierte Domainnamen).
3. Wählen Sie Create (Erstellen) aus.
4. Geben Sie unter Domaindetails Ihren Domainnamen ein.
5. Wählen Sie in der Endpunktkonfiguration ein vorhandenes ACM-Zertifikat aus, oder wählen Sie Neues ACM-Zertifikat erstellen aus.

(Optional) Erstellen Sie ein Zertifikat

- a. Wählen Sie in der ACM-Konsole die Option Request aus.
- b. Geben Sie unter Domainnamen Ihren Domainnamen ein.
- c. Wählen Sie unter Validierungsmethode eine Validierungsmethode aus.

Wenn Sie E-Mail-Validierung wählen, wird eine E-Mail an die E-Mail-Adresse gesendet, die beim Domain-Registrar hinterlegt ist.

- d. Wählen Sie Ich stimme zu, um das Zertifikat zu aktivieren.

Schritt 2: API-Zuordnungen einrichten

1. Wählen Sie unter [API Gateway](#) unter Benutzerdefinierte Domainnamen die Option API-Zuordnungen konfigurieren aus. your-domain-name
2. Wählen Sie benutzerdefinierte Domain-Namen aus.
3. Wählen Sie Neues Mapping hinzufügen aus.
4. Wählen Sie die AWS ParallelCluster UI-API, die \$default Stage und Speichern aus.
5. Kopieren Sie unter API Gateway Gateway-Domänenname den Wert, der in den nächsten Schritten verwendet werden soll.

Schritt 3: DNS einrichten

- Erstellen Sie eine DNS-CNAME-Regel, die Ihre Domain auf die API Gateway Gateway-Domain verweist. Geben Sie nur die Domain ein. Fügen Sie beispielsweise keine Phase hinzu, z. B. beta oder prod. Ersetzen Sie *abcde12345* durch Ihre API-Gateway-API-ID und ersetzen Sie *us-east-2* durch die API. AWS-Region

Regel	Quelle	Ziel
CNAME	<i>example.com</i>	<i>d- abcde1234 5 .execute-api. us- ost-2 .amazonaws.com</i>

Schritt 4: Fügen Sie die Domain zu Ihrem Amazon Cognito Cognito-Benutzerpool hinzu

1. Navigieren Sie zur [Amazon-Cognito-Konsole](#).
2. Wählen Sie Ihren Benutzerpool-Link aus.
3. Wählen Sie App-Integration.
4. Wählen Sie unter Domain die Optionen Aktionen, Benutzerdefinierte Domain erstellen aus.
5. Geben Sie Ihre benutzerdefinierte Domain ein und wählen Sie Ihr ACM-Zertifikat aus.
6. Wählen Sie Benutzerdefinierte Domain erstellen.

Schritt 5: Konfigurieren Sie die API Gateway Gateway-Callback-URL

1. Navigieren Sie zur [Amazon-Cognito-Konsole](#).
2. Wählen Sie in Ihrem Amazon Cognito Cognito-Benutzerpool App-Integration, App-Clients und Analysen den Anwendungslink aus.
3. Wählen Sie in der gehosteten Benutzeroberfläche die Option Bearbeiten aus.
4. Wählen Sie unter Zulässige Rückruf-URLs die Option Weitere URL hinzufügen aus und geben Sie eine Rückruf-URL ein, z. B. `example.com/login`

Schritt 6: Lambda-Funktion konfigurieren

1. Navigieren Sie zur [Lambda-Konsole](#).
2. Wählen Sie im Navigationsbereich Funktionen aus.
3. Filtern Sie die Liste der Funktionen, um den Link zu finden, `ParallelClusterUIFunction` und wählen Sie ihn aus.
4. Wählen Sie Konfiguration, Umgebungsvariablen.
5. Wählen Sie Bearbeiten aus.
6. Geben Sie für den `SITE_URL` Wert Ihre benutzerdefinierte Domain ein.
7. Navigieren Sie zu Ihrer Domain, z. B. `example.com`, und authentifizieren Sie sich, um eine Verbindung zur AWS ParallelCluster Benutzeroberfläche herzustellen.

Amazon Cognito Cognito-Benutzerpool-Optionen

Die folgenden Abschnitte beziehen sich auf CloudFormation Schnellerstellungslinks oder Schnellerstellungs-URLs. Über die Schnellerstellungs-URL gelangen Sie zu einem Assistenten zum Erstellen von Stacks, in dem Sie Eingaben für die Schnellerstellung von Stack-Vorlagen eingeben und den Stack bereitstellen. Weitere Informationen zu CloudFormation Schnellerstellungstapeln finden Sie im Benutzerhandbuch unter Schnellerstellungslinks für Stacks [erstellen](#). AWS CloudFormation

Um einen Amazon Cognito Cognito-Benutzerpool zu verwalten, den Sie mit mehreren AWS ParallelCluster UI-Instances verwenden können, sollten Sie die folgenden Optionen in Betracht ziehen:

- Verwenden Sie eine vorhandene AWS ParallelCluster UI-Instanz, die auf einen Amazon Cognito Cognito-Benutzerpool verweist, der aus einem verschachtelten CloudFormation Stack erstellt wurde. Dies wird erstellt, wenn Sie die AWS ParallelCluster Benutzeroberfläche mithilfe des Schnellerstellungslinks bereitstellen und alle Amazon Cognito Cognito-Parameter leer lassen.
- Verwenden Sie einen eigenständigen Amazon Cognito Cognito-Benutzerpool, der vor der Bereitstellung der AWS ParallelCluster Benutzeroberfläche bereitgestellt wird. Stellen Sie dann eine neue AWS ParallelCluster UI-Instanz bereit, die mit dem eigenständigen Amazon Cognito Cognito-Benutzerpool verknüpft ist, den Sie bereits bereitgestellt haben. Auf diese Weise trennen Sie die Amazon Cognito Cognito-Bereitstellung von der AWS ParallelCluster UI-Bereitstellung. Darüber hinaus sind nicht verschachtelte AWS ParallelCluster CloudFormation UI-Stacks einfacher zu aktualisieren.

Verwenden Sie einen vorhandenen Amazon Cognito Cognito-Benutzerpool mit einer neuen AWS ParallelCluster UI-Instanz

1. Wählen Sie in der CloudFormation Konsole den AWS ParallelCluster UI-Stack aus, der den Amazon Cognito Cognito-Benutzerpool enthält, den Sie mit mehreren AWS ParallelCluster UI-Instances verwenden möchten.
2. Navigieren Sie zu dem verschachtelten Stapel, der den Amazon Cognito Cognito-Benutzerpool erstellt hat.
3. Wählen Sie die Registerkarte Ausgaben aus.
4. Kopieren Sie die Werte der folgenden Parameter:
 - `UserPoolId`
 - `UserPoolAuthDomain`
 - `SNSRole`
5. Stellen Sie mithilfe des Schnellerstellungslinks eine neue AWS ParallelCluster UI-Instanz bereit und füllen Sie alle `External AWS ParallelCluster UI Amazon Cognito` Parameter mit den Ausgaben aus, die Sie kopiert haben. Dadurch wird verhindert, dass der neue AWS ParallelCluster UI-Stack einen neuen Pool erstellt, und verknüpft ihn mit dem vorhandenen Amazon Cognito Cognito-Benutzerpool, der aus einem verschachtelten Stack erstellt wurde. Sie können nachfolgende neue AWS ParallelCluster UI-Instances bereitstellen, die dieselben Parameterwerte haben, und Sie können sie mit dem Amazon Cognito Cognito-Benutzerpool verknüpfen.

Erstellen Sie einen eigenständigen Amazon Cognito Cognito-Benutzerpool

AWS ParallelClusterUI Amazon Cognito: Links schnell nach Region erstellen

UI Amazon Cognito Schnellverknüpfungen

[us-east-1](#)

[us-east-2](#)

[us-west-1](#)

[us-west-2](#)

[eu-west-1](#)

[eu-west-2](#)

[eu-west-3](#)

[eu-central-1](#)

[eu-north-1](#)

[me-south-1](#)

[sa-east-1](#)

[ca-central-1](#)

[ap-northeast-1](#)

[ap-northeast-2](#)

[ap-south-1](#)

[ap-southeast-1](#)

[ap-southeast-2](#)

[us-gov-west-1](#)

1. Starten Sie einen reinen Amazon Cognito-Stack, indem Sie einen Schnellerstellungslink mit derselben Bezeichnung auswählen, AWS-Region in dem Sie Ihre UI-Instances bereitstellen. AWS ParallelCluster Sehen Sie sich die Links zur Schnellerstellung am Anfang dieses Abschnitts an.
2. Wählen Sie nach Abschluss der Stapelerstellung die Registerkarte Ausgaben aus und kopieren Sie die Werte der folgenden Parameter:
 - `UserPoolId`
 - `UserPoolAuthDomain`
 - `SNSRole`
3. Stellen Sie eine neue AWS ParallelCluster UI-Instanz bereit, indem Sie einen AWS ParallelCluster UI-Schnellstart-Link auswählen und alle `External AWS ParallelCluster UI Amazon Cognito` Parameter mit den Werten ausfüllen, die Sie kopiert haben. Die neue AWS ParallelCluster UI-Instance ist mit dem eigenständigen Amazon Cognito Cognito-Benutzerpool verknüpft und erstellt weder einen verschachtelten Stack noch einen neuen Benutzerpool. Sie können nachfolgende neue AWS ParallelCluster UI-Instances bereitstellen, die dieselben Parameterwerte haben, und Sie können sie mit dem eigenständigen Amazon Cognito Cognito-Benutzerpool verknüpfen.

Identifizieren Sie die AWS ParallelCluster und die AWS ParallelCluster UI-Version

1. Wählen Sie in der CloudFormation Konsole einen AWS ParallelCluster UI-Stack aus.
2. Wählen Sie die Registerkarte Parameter aus.
3. Die AWS ParallelCluster Version ist der Wert des Parameters `Version`.
4. Die AWS ParallelCluster UI-Version steht am Ende des `PublicEcrImageUri` Werts. Wenn der Wert beispielsweise `istpublic.ecr.aws/pcui/parallelcluster-ui-awslambda:2023.02`, dann ist die Version `2023.02`.

Aktualisieren Sie die AWS ParallelCluster Benutzeroberfläche auf eine neue AWS ParallelCluster Version

Um die AWS ParallelCluster Benutzeroberfläche auf die neueste AWS ParallelCluster Version zu aktualisieren, starten Sie einen neuen Stack, indem Sie einen [Schnellerstellungslink](#) auswählen.

AWS ParallelCluster Die Benutzeroberfläche kostet

Die AWS ParallelCluster Benutzeroberfläche basiert auf einer serverlosen Architektur und kann in den meisten Fällen innerhalb der Kategorie „AWS-Kostenloses Kontingent“ verwendet werden. In der folgenden Tabelle sind die Komponenten aufgeführt, von AWS-Services denen die AWS ParallelCluster Benutzeroberfläche abhängt, sowie deren Limits für das kostenlose Nutzungskontingent. Die typische Nutzung kostet schätzungsweise weniger als einen Dollar pro Monat.

Service	AWS Free Tier
Amazon Cognito	50.000 aktive Nutzer pro Monat
Amazon API Gateway	1 Million Rest-API-Aufrufe
AWS Lambda	1 Million kostenlose Anfragen pro Monat und 400.000 GB-Sekunden Rechenzeit pro Monat
EC2 Image Builder	Keine Kosten, außer EC2
Amazon Elastic Compute Cloud	Einmaliges Erstellen eines Container-Images in 15 Minuten
AWS CloudFormation	5 GB Daten (Aufnahme, Archivierung und Daten, die mit Logs Insights-Abfragen gescannt wurden)

Erste Schritte mit AWS ParallelCluster

Beginnen Sie mit der Konfiguration und Erstellung eines Clusters mithilfe der AWS ParallelCluster Befehlszeilenschnittstelle (CLI) oder der webbasierten Benutzeroberfläche (UI). Die AWS ParallelCluster Benutzeroberfläche wurde in Version 3.5.0 hinzugefügt.

Themen

- [Konfigurieren und erstellen Sie einen Cluster mit der AWS ParallelCluster Befehlszeilenschnittstelle](#)
- [Konfiguration und Erstellung eines Clusters mit der AWS ParallelCluster Benutzeroberfläche](#)
- [Stellen Sie eine Verbindung zu einem Cluster her](#)

Konfigurieren und erstellen Sie einen Cluster mit der AWS ParallelCluster Befehlszeilenschnittstelle

Führen Sie nach der Installation AWS ParallelCluster die folgenden Konfigurationsschritte aus.

Stellen Sie sicher, dass Ihr AWS Konto über eine Rolle verfügt, die die für die Ausführung der `pcluster` CLI erforderlichen Berechtigungen umfasst. Weitere Informationen finden Sie unter [AWS ParallelCluster Beispiele `pcluster` für Benutzerrichtlinien](#).

Richten Sie Ihre AWS Anmeldeinformationen ein. Weitere Informationen finden Sie unter [Konfigurieren der AWS CLI](#) im AWS CLI -Benutzerhandbuch.

```
$ aws configure
AWS Access Key ID [None]: AKIAIOSFODNN7EXAMPLE
AWS Secret Access Key [None]: wJalrXUtnFEMI/K7MDENG/bPxrFiCYEXAMPLEKEY
Default region name [us-east-1]: us-east-1
Default output format [None]:
```

Der AWS-Region Ort, an dem der Cluster gestartet wird, muss mindestens ein Amazon EC2 EC2-Schlüsselpaar haben. Weitere Informationen dazu finden Sie unter [Amazon-EC2-Schlüsselpaare](#) im Amazon-EC2-Benutzerhandbuch für Linux-Instances.

Wenn Sie die AWS ParallelCluster Befehlszeilenschnittstelle (CLI) verwenden, zahlen Sie nur für die AWS Ressourcen, die beim Erstellen oder Aktualisieren von AWS ParallelCluster Images und Clustern erstellt werden. Weitere Informationen finden Sie unter [AWS Dienste verwendet von AWS ParallelCluster](#).

Konfigurieren und erstellen Sie Ihren ersten Cluster

Erstellen Sie Ihren ersten Cluster, indem Sie den `pcluster configure` CLI-Befehl verwenden, um einen Assistenten zu starten, der Sie zur Eingabe aller Informationen auffordert, die für die Konfiguration und Erstellung Ihres Clusters erforderlich sind. Die Einzelheiten der Sequenz unterscheiden sich bei der Verwendung AWS Batch als Scheduler von der Verwendung. Slurm

Slurm

```
$ pcluster configure --config config-file.yaml
```

Wählen Sie aus der Liste der gültigen AWS-Region Bezeichner den Ort aus, AWS-Region an dem Ihr Cluster ausgeführt werden soll.

Note

Die AWS-Regionen angezeigte Liste basiert auf der Partition Ihres Kontos und enthält nur die Partitionen, AWS-Regionen die für Ihr Konto aktiviert sind. Weitere Informationen zur Aktivierung AWS-Regionen für Ihr Konto finden Sie unter [Verwaltung AWS-Regionen](#) in der Allgemeine AWS-Referenz. Das gezeigte Beispiel stammt aus der AWS globalen Partition. Wenn sich Ihr Konto in der AWS GovCloud (US) Partition befindet, werden nur AWS-Regionen in dieser Partition aufgeführt (gov-us-east-1 und gov-us-west-1). Ebenso werden nur cn-north-1 und cn-northwest-1 angezeigt, wenn sich Ihr Konto in der AWS chinesischen Partition befindet. Eine vollständige Liste der AWS-Regionen unterstützten AWS ParallelCluster Programme finden Sie unter [Unterstützte Regionen für AWS ParallelCluster](#).

Allowed values for AWS-Region ID:

1. af-south-1
2. ap-east-1
3. ap-northeast-1
4. ap-northeast-2
5. ap-south-1
6. ap-southeast-1
7. ap-southeast-2
8. ca-central-1
9. eu-central-1
10. eu-north-1
11. eu-south-1
12. eu-west-1
13. eu-west-2
14. eu-west-3
15. me-south-1
16. sa-east-1
17. us-east-1
18. us-east-2
19. us-west-1
20. us-west-2

AWS-Region ID [ap-northeast-1]:

Das key pair wird aus den Schlüsselpaaren ausgewählt, die in der ausgewählten AWS-Region Version bei Amazon EC2 registriert sind. Wählen Sie das key pair:

Allowed values for EC2 Key Pair Name:

1. your-key-1
2. your-key-2

EC2 Key Pair Name [your-key-1]:

Wählen Sie den Scheduler aus, der mit dem Cluster verwendet werden soll.

Allowed values for Scheduler:

1. slurm
2. awsbatch

Scheduler [slurm]:

Wählen Sie das Betriebssystem aus.

Allowed values for Operating System:

1. alinux2
2. centos7
3. ubuntu2204
4. ubuntu2004
5. rhel8

Operating System [alinux2]:

Wählen Sie den Instanztyp des Hauptknotens aus:

Head node instance type [t2.micro]:

Wählen Sie die Warteschlangenkonfiguration. Hinweis: Der Instanztyp kann nicht für mehrere Rechenressourcen in derselben Warteschlange angegeben werden.

Number of queues [1]:

Name of queue 1 [queue1]:

Number of compute resources for queue1 [1]: 2

Compute instance type for compute resource 1 in queue1 [t2.micro]:

Maximum instance count [10]:

Ermöglichen Sie EFA, Anwendungen, die ein hohes Maß an Kommunikation zwischen den Instanzen erfordern, in großem Umfang ohne zusätzliche AWS Kosten auszuführen:

- Wählen Sie einen Instance-Typ, der [Elastic Fabric Adapter \(EFA\) unterstützt](#).


- Aktivieren Sie [EFA](#).
- Geben Sie einen Namen für eine bestehende [Platzierungsgruppe](#) an. Wenn Sie das Feld leer lassen, AWS ParallelCluster wird eine für Sie erstellt.

```
Compute instance type for compute resource 2 in queue1 [t2.micro]: c5n.18xlarge
Enable EFA on c5n.18xlarge (y/n) [y]: y
Maximum instance count [10]:
Placement Group name []:
```

Nachdem die vorherigen Schritte abgeschlossen sind, entscheiden Sie, ob Sie eine vorhandene VPC verwenden oder eine VPC für AWS ParallelCluster erstellen lassen möchten. Wenn Sie keine ordnungsgemäß konfigurierte VPC haben, AWS ParallelCluster können Sie eine neue für Sie erstellen. Es platziert entweder sowohl den Kopf- als auch den Rechenknoten im selben öffentlichen Subnetz oder nur den Hauptknoten in einem öffentlichen Subnetz mit allen Rechenknoten in einem privaten Subnetz. Wenn Sie eine VPC AWS ParallelCluster erstellen lassen, müssen Sie entscheiden, ob sich alle Knoten in einem öffentlichen Subnetz befinden sollen. Weitere Informationen finden Sie unter [Netzwerkkonfigurationen](#).

Wenn Sie Ihren Cluster für die Verwendung von Instance-Typen mit mehreren Netzwerkschnittstellen oder einer Netzwerkkarte konfigurieren, finden Sie weitere [Netzwerkkonfigurationen](#) Netzwerkanforderungen unter.

Es ist möglich, Ihr Kontingent für die Anzahl der in a AWS-Region zulässigen VPCs zu erreichen. Das Standardkontingent beträgt fünf VPCs für a. AWS-Region Weitere Informationen zu diesem Kontingent und dazu, wie Sie eine Erhöhung beantragen können, finden Sie unter [VPC und Subnetze](#) im Amazon VPC-Benutzerhandbuch.

 **Important**

VPCs, die von erstellt wurden, aktivieren standardmäßig AWS ParallelCluster keine VPC Flow Logs. Mit VPC Flow Logs können Sie Informationen über den IP-Verkehr erfassen, der zu und von Netzwerkschnittstellen in Ihren VPCs fließt. Weitere Informationen finden Sie unter [VPC-Flow-Protokolle](#) im Amazon-VPC-Benutzerhandbuch.

Wenn Sie eine VPC AWS ParallelCluster erstellen lassen, stellen Sie sicher, dass Sie entscheiden, ob sich alle Knoten in einem öffentlichen Subnetz befinden sollen.

Note

Wenn Sie möchten¹, Head node in a public subnet and compute fleet in a private subnet, AWS ParallelCluster wird ein NAT-Gateway erstellt, das zusätzliche Kosten verursacht, auch wenn Sie Ressourcen im kostenlosen Kontingent angeben.

```
Automate VPC creation? (y/n) [n]: y
Allowed values for Availability Zone:
1. us-east-1a
2. us-east-1b
3. us-east-1c
4. us-east-1d
5. us-east-1e
6. us-east-1f
Availability Zone [us-east-1a]:
Allowed values for Network Configuration:
1. Head node in a public subnet and compute fleet in a private subnet
2. Head node and compute fleet in the same public subnet
Network Configuration [Head node in a public subnet and compute fleet in a private
subnet]: 1
Beginning VPC creation. Please do not leave the terminal until the creation is
finalized
```

Wenn Sie keine neue VPC erstellen, müssen Sie eine vorhandene VPC auswählen.

Wenn Sie die VPC AWS ParallelCluster erstellen möchten, notieren Sie sich die VPC-ID, damit Sie sie später AWS CLI löschen können.

```
Automate VPC creation? (y/n) [n]: n
Allowed values for VPC ID:
# id name number_of_subnets
-----
1 vpc-0b4ad9c4678d3c7ad ParallelClusterVPC-20200118031893 2
2 vpc-0e87c753286f37eef ParallelClusterVPC-20191118233938 5
VPC ID [vpc-0b4ad9c4678d3c7ad]: 1
```

Nachdem die VPC ausgewählt wurde, entscheiden Sie, ob Sie vorhandene Subnetze verwenden oder neue erstellen möchten.

```
Automate Subnet creation? (y/n) [y]: y
```

```
Creating CloudFormation stack...  
Do not leave the terminal until the process has finished
```

AWS Batch

```
$ pcluster configure --config config-file.yaml
```

Wählen Sie aus der Liste der gültigen AWS-Region Kennungen den Ort aus, AWS-Region an dem Ihr Cluster ausgeführt werden soll.

Note

Die AWS-Regionen angezeigte Liste basiert auf der Partition Ihres Kontos. Es enthält nur AWS-Regionen die, die für Ihr Konto aktiviert sind. Weitere Informationen zur Aktivierung AWS-Regionen für Ihr Konto finden Sie unter [Verwaltung AWS-Regionen](#) in der Allgemeine AWS-Referenz. Das gezeigte Beispiel stammt aus der AWS globalen Partition. Wenn sich Ihr Konto in der AWS GovCloud (US) Partition befindet, werden nur AWS-Regionen in dieser Partition aufgeführt (gov-us-east-1 und gov-us-west-1). Ebenso werden nur cn-north-1 und cn-northwest-1 angezeigt, wenn sich Ihr Konto in der AWS chinesischen Partition befindet. Eine vollständige Liste der AWS-Regionen unterstützten AWS ParallelCluster Programme finden Sie unter [Unterstützte Regionen für AWS ParallelCluster](#).

Allowed values for AWS-Region ID:

1. af-south-1
2. ap-east-1
3. ap-northeast-1
4. ap-northeast-2
5. ap-south-1
6. ap-southeast-1
7. ap-southeast-2
8. ca-central-1
9. eu-central-1
10. eu-north-1
11. eu-south-1
12. eu-west-1

```
13. eu-west-2
14. eu-west-3
15. me-south-1
16. sa-east-1
17. us-east-1
18. us-east-2
19. us-west-1
20. us-west-2
AWS-Region ID [us-east-1]:
```

Das key pair wird aus den bei Amazon EC2 registrierten Schlüsselpaaren im ausgewählten Bereich ausgewählt AWS-Region. Wählen Sie das key pair:

```
Allowed values for EC2 Key Pair Name:
1. your-key-1
2. your-key-2
EC2 Key Pair Name [your-key-1]:
```

Wählen Sie den Scheduler aus, der mit dem Cluster verwendet werden soll.

```
Allowed values for Scheduler:
1. slurm
2. awsbatch
Scheduler [slurm]: 2
```

Wenn awsbatch als Scheduler ausgewählt wird, wird alinux2 als Betriebssystem verwendet. Der Instanztyp des Head-Knotens ist eingegeben:


```
Head node instance type [t2.micro]:
```

Wählen Sie die Warteschlangenkonfiguration. Der AWS Batch Scheduler enthält nur eine einzige Warteschlange. Die maximale Größe des Clusters von Rechenknoten wird eingegeben. Dies wird in vCPUs gemessen.

```
Number of queues [1]:
Name of queue 1 [queue1]:
Maximum vCPU [10]:
```


Entscheiden Sie, ob Sie bestehende VPCs verwenden oder VPCs für AWS ParallelCluster Sie erstellen lassen möchten. Wenn Sie nicht über eine ordnungsgemäß konfigurierte VPC verfügen,

kann AWS ParallelCluster eine neue erstellen. Es verwendet entweder sowohl den Head- als auch den Compute-Knoten im selben öffentlichen Subnetz oder nur den Head-Knoten in einem öffentlichen Subnetz mit allen Knoten in einem privaten Subnetz. Es ist möglich, Ihr Kontingent für die Anzahl der in einer Region zulässigen VPCs zu erreichen. Die Standard-VPC-Anzahl ist fünf. Weitere Informationen zu diesem Kontingent und dazu, wie Sie eine Erhöhung beantragen können, finden Sie unter [VPC und Subnetze](#) im Amazon VPC-Benutzerhandbuch.

 **Important**

VPCs, die von erstellt wurden, aktivieren standardmäßig AWS ParallelCluster keine VPC Flow Logs. Mit VPC Flow Logs können Sie Informationen über den IP-Verkehr erfassen, der zu und von Netzwerkschnittstellen in Ihren VPCs fließt. Weitere Informationen finden Sie unter [VPC-Flow-Protokolle](#) im Amazon-VPC-Benutzerhandbuch.

Wenn Sie eine VPC AWS ParallelCluster erstellen lassen, stellen Sie sicher, dass Sie entscheiden, ob sich alle Knoten in einem öffentlichen Subnetz befinden sollen.

 **Note**

Wenn Sie möchten¹. Head node in a public subnet and compute fleet in a private subnet, AWS ParallelCluster wird ein NAT-Gateway erstellt, das zusätzliche Kosten verursacht, auch wenn Sie Ressourcen im kostenlosen Kontingent angeben.

```
Automate VPC creation? (y/n) [n]: y
Allowed values for Availability Zone:
1. us-east-1a
2. us-east-1b
3. us-east-1c
4. us-east-1d
5. us-east-1e
6. us-east-1f
Availability Zone [us-east-1a]:
Allowed values for Network Configuration:
1. Head node in a public subnet and compute fleet in a private subnet
2. Head node and compute fleet in the same public subnet
Network Configuration [Head node in a public subnet and compute fleet in a private
subnet]: *1*
```

```
Beginning VPC creation. Please do not leave the terminal until the creation is
finalized
```

Wenn Sie keine neue VPC erstellen, müssen Sie eine vorhandene VPC auswählen.

Wenn Sie sich dafür entscheiden, die VPC AWS ParallelCluster erstellen zu lassen, notieren Sie sich die VPC-ID, damit Sie das AWS CLI oder verwenden können, um sie AWS Management Console später zu löschen.

```
Automate VPC creation? (y/n) [n]: n
Allowed values for VPC ID:
#  id                                     name                                     number_of_subnets
---  -----
 1  vpc-0b4ad9c4678d3c7ad  ParallelClusterVPC-20200118031893      2
 2  vpc-0e87c753286f37eef  ParallelClusterVPC-20191118233938      5
VPC ID [vpc-0b4ad9c4678d3c7ad]: 1
```

Stellen Sie nach der Auswahl der VPC sicher, dass Sie entscheiden, ob Sie vorhandene Subnetze verwenden oder neue erstellen möchten.

```
Automate Subnet creation? (y/n) [y]: y
```

```
Creating CloudFormation stack...
Do not leave the terminal until the process has finished
```

Wenn Sie die vorherigen Schritte abgeschlossen haben, wird ein einfacher Cluster in einer VPC gestartet. Die VPC verwendet ein vorhandenes Subnetz, das öffentliche IP-Adressen unterstützt. Die Routing-Tabelle für das Subnetz lautet. $0.0.0.0/0 \Rightarrow igw-xxxxxx$ Beachten Sie die folgenden Bedingungen:

- Die VPC muss DNS Resolution = yes und DNS Hostnames = yes lauten.
- Die VPC muss auch über DHCP-Optionen mit den richtigen domain-name für verfügen.
AWS-Region Der standardmäßige DHCP-Optionssatz gibt bereits das erforderliche AmazonProvidedDNS-Element an. Wenn Sie mehr als einen Domain-Namenserver angeben, finden Sie weitere Informationen unter [DHCP-Optionssätze](#) im Amazon VPC-Benutzerhandbuch. Wenn Sie private Subnetze verwenden, verwenden Sie ein NAT-Gateway oder einen internen Proxy, um den Webzugriff für Rechenknoten zu aktivieren. Weitere Informationen finden Sie unter [Netzwerkkonfigurationen](#).

Wenn alle diese Einstellungen gültige Werte enthalten, können Sie den Cluster starten, indem Sie den Befehl „create“ ausführen.

```
$ pcluster create-cluster --cluster-name test-cluster --cluster-configuration cluster-  
config.yaml  
{  
  "cluster": {  
    "clusterName": "test-cluster",  
    "cloudformationStackStatus": "CREATE_IN_PROGRESS",  
    "cloudformationStackArn": "arn:aws:cloudformation:eu-west-1:xxx:stack/test-cluster/  
abcdef0-f678-890a-5abc-021345abcdef",  
    "region": "eu-west-1",  
    "version": "3.7.0",  
    "clusterStatus": "CREATE_IN_PROGRESS"  
  },  
  "validationMessages": []  
}
```

Verfolgen Sie den Fortschritt des Clusters:

```
$ pcluster describe-cluster --cluster-name test-cluster
```

or

```
$ pcluster list-clusters --query 'clusters[?clusterName==`test-cluster`]'
```

Nachdem der Cluster den "clusterStatus": "CREATE_COMPLETE" Status erreicht hat, können Sie mithilfe Ihrer normalen SSH-Client-Einstellungen eine Verbindung zu ihm herstellen. Weitere Informationen zur Verbindung mit Amazon EC2 EC2-Instances finden Sie im [EC2-Benutzerhandbuch im Amazon EC2](#) EC2-Benutzerhandbuch für Linux-Instances. Oder Sie können den Cluster verbinden über

```
$ pcluster ssh --cluster-name test-cluster -i ~/path/to/keyfile.pem
```

Führen Sie den folgenden Befehl aus, um den Cluster zu löschen.

```
$ pcluster delete-cluster --region us-east-1 --cluster-name test-cluster
```

Nachdem der Cluster gelöscht wurde, können Sie die Netzwerkressourcen in der VPC löschen, indem Sie den CloudFormation Netzwerkstapel löschen. Der Name des Stacks beginnt mit

„parallelclusternetworking-“ und enthält die Erstellungszeit im Format „YYYYMMDDHHMMSS“. Sie können die Stacks mit dem Befehl auflisten. [list-stacks](#)

```
$ aws --region us-east-1 cloudformation list-stacks \
  --stack-status-filter "CREATE_COMPLETE" \
  --query "StackSummaries[].StackName" | \
  grep -e "parallelclusternetworking-"
"parallelclusternetworking-pubpriv-20191029205804"
```

Der Stapel kann mit dem [delete-stack](#) Befehl gelöscht werden.

```
$ aws --region us-east-1 cloudformation delete-stack \
  --stack-name parallelclusternetworking-pubpriv-20191029205804
```

Die VPC, die für Sie [pcluster configure](#) erstellt, wird nicht im CloudFormation Netzwerk-Stack erstellt. Sie können diese VPC manuell in der Konsole oder mithilfe von löschen. AWS CLI

```
$ aws --region us-east-1 ec2 delete-vpc --vpc-id vpc-0b4ad9c4678d3c7ad
```

Konfiguration und Erstellung eines Clusters mit der AWS ParallelCluster Benutzeroberfläche

Die AWS ParallelCluster Benutzeroberfläche ist eine webbasierte Benutzeroberfläche, die die AWS ParallelCluster `pcluster` CLI widerspiegelt und gleichzeitig ein konsolenähnliches Erlebnis bietet. Sie installieren und greifen auf die AWS ParallelCluster Benutzeroberfläche in Ihrem zu. AWS-Konto Wenn Sie es ausführen, greift die AWS ParallelCluster Benutzeroberfläche auf eine Instanz der API zu, die auf Amazon AWS ParallelCluster API Gateway in Ihrem AWS-Konto gehostet wird.

Note

Der AWS ParallelCluster UI-Assistent verfügt möglicherweise nicht über Benutzeroberflächenoptionen für alle unterstützten Funktionen in der neuesten unterstützten AWS ParallelCluster Version. Sie können die Konfigurationsdatei nach Bedarf manuell bearbeiten oder die AWS ParallelCluster CLI verwenden.

In diesem Abschnitt führen wir Sie durch die Konfiguration und Erstellung eines Clusters mithilfe der AWS ParallelCluster Benutzeroberfläche.

Voraussetzungen:

- Zugriff auf eine laufende AWS ParallelCluster UI-Instanz. Weitere Informationen finden Sie unter [Installation der AWS ParallelCluster Benutzeroberfläche](#).

Einen Cluster konfigurieren und erstellen

1. Wählen Sie in der Ansicht AWS ParallelCluster UI-Cluster die Option Cluster erstellen, Schritt für Schritt aus.
2. Geben Sie unter Cluster, Name einen Namen für Ihren Cluster ein.
3. Wählen Sie eine VPC mit einem öffentlichen Subnetz für Ihren Cluster und klicken Sie auf Weiter.
4. Wählen Sie im Hauptknoten die Option SSM-Sitzung hinzufügen und dann Weiter aus.
5. Wählen Sie unter Warteschlangen, Ressourcen berechnen die Option 1 für Statische Knoten aus.
6. Entfernen Sie unter Instanztyp den ausgewählten Standard-Instance-Typ, wählen Sie t2.micro und dann Next aus.
7. Wählen Sie unter Speicher die Option Weiter aus.
8. Überprüfen Sie in der Clusterkonfiguration die Cluster-Konfiguration YAML und wählen Sie Dry Run aus, um sie zu validieren.
9. Wählen Sie Create, um Ihren Cluster auf der Grundlage der validierten Konfiguration zu erstellen.
10. Nach einigen Sekunden führt Sie die AWS ParallelCluster Benutzeroberfläche automatisch zurück zu Clustern, wo Sie den Status der Cluster-Erstellung und die Stack-Ereignisse überwachen können.
11. Wählen Sie Details, um Cluster-Details wie Version und Status anzuzeigen.
12. Wählen Sie Instances aus, um die Liste der EC2-Instances und den Status anzuzeigen.
13. Wählen Sie Stack-Ereignisse, um Cluster-Stack-Ereignisse und einen AWS Management Console Link zu dem CloudFormation Stack anzuzeigen, der den Cluster erstellt.
14. Wählen Sie nach Abschluss der Clustererstellung unter Details die Option YAML anzeigen aus, um die YAML-Datei für die Cluster-Konfiguration anzuzeigen oder herunterzuladen.
15. Wählen Sie nach Abschluss der Clustererstellung Shell aus, um auf den Cluster-Hauptknoten zuzugreifen.

Note

Wenn Sie Shell wählen, AWS ParallelCluster öffnet eine Amazon EC2 Systems Manager Manager-Sitzung und fügt eine `ssm-user` zu `/etc/sudoers` hinzu. Weitere Informationen finden Sie unter [`ssm-user`Kontoverwaltungsberechtigungen aktivieren oder deaktivieren](#) im Amazon EC2 Systems Manager Manager-Benutzerhandbuch.

- Um zu bereinigen, wählen Sie in der Cluster-Ansicht den Cluster aus und klicken Sie auf Aktionen, Cluster löschen.

Stellen Sie eine Verbindung zu einem Cluster her

Bei Verwendung AWS ParallelCluster können Sie eine Verbindung zum Cluster-Hauptknoten herstellen, um Jobs auszuführen, Ergebnisse anzuzeigen, Benutzer zu verwalten und den Cluster- und Jobstatus zu überwachen. Stellen Sie mithilfe der folgenden Methoden eine Verbindung zur Cluster-Kopfknoteninstanz her:

- Melden Sie sich `ssh` mit einem [Schlüsselpaar](#) an. Geben Sie den privaten Schlüssel in [HeadNode/KeyName](#) in der Clusterkonfiguration an. Weitere Informationen finden Sie unter [Herstellen einer Verbindung mit Ihrer Linux-Instance mit SSH](#) im Amazon-EC2-Benutzerhandbuch für Linux-Instances.
- Melden Sie sich mit dem Befehl `pcluster ssh` Command Line Interface (CLI) an. Geben Sie den privaten Schlüssel in der Clusterkonfiguration an [HeadNode/KeyName](#). Weitere Informationen finden Sie unter [pcluster ssh](#).
- Stellen Sie mithilfe einer SSM-Sitzung eine Verbindung zum Cluster-Kopfknoten her. Sie müssen die AmazonSSMManagedInstanceCore verwaltete Richtlinie [AdditionalIamPolicies](#) in der Clusterkonfiguration zu [HeadNode](#)/hinzufügen, um mithilfe einer SSM-Sitzung eine Verbindung herzustellen. Weitere Informationen finden Sie unter [SSM Session Manager](#) im SSM-Benutzerhandbuch.
- Stellen Sie mithilfe von NICE DCV eine Verbindung zum Cluster-Kopfknoten her. Weitere Informationen finden Sie unter [Stellen Sie mit NICE DCV eine Connect zum Hauptknoten her](#).
- Wenn Sie die AWS ParallelCluster Benutzeroberfläche verwenden, können Sie mithilfe eines EC2 Connect-Befehls, den die Benutzeroberfläche bereitstellt, auch eine Verbindung zum Cluster-Kopfknoten herstellen.

Zugriff mehrerer Benutzer auf Cluster

Erfahren Sie, wie Sie den Zugriff mehrerer Benutzer auf einen einzelnen Cluster implementieren und verwalten.

In diesem Thema bezieht sich ein AWS ParallelCluster Benutzer auf einen Systembenutzer für Recheninstanzen. Ein Beispiel ist eine `ec2-user` für eine AWS EC2-Instance.

AWS ParallelCluster-Unterstützung für den Mehrbenutzerzugriff ist überall verfügbar, AWS-Regionen wo sie derzeit verfügbar AWS ParallelCluster ist. Es funktioniert mit anderen AWS-Services, einschließlich [Amazon FSx for Lustre](#) und [Amazon Elastic File System](#).

Sie können ein [AWS Directory Service for Microsoft Active Directory](#) oder [Simple AD](#) verwenden, um den Clusterzugriff zu verwalten. Stellen Sie sicher, dass Sie die [AWS-Region-Verfügbarkeit](#) dieser Dienste überprüfen. Um einen Cluster einzurichten, geben Sie eine [AWS ParallelCluster DirectoryService](#)-Konfiguration an. AWS Directory Service-Verzeichnisse können mit mehreren Clustern verbunden werden. Dies ermöglicht eine zentrale Verwaltung von Identitäten in mehreren Umgebungen und ein einheitliches Anmeldeerlebnis.

Wenn Sie den Zugriff AWS Directory Service für AWS ParallelCluster mehrere Benutzer verwenden, können Sie sich mit den im Verzeichnis definierten Benutzeranmeldedaten beim Cluster anmelden. Diese Anmeldeinformationen bestehen aus einem Benutzernamen und einem Passwort. Nachdem Sie sich zum ersten Mal beim Cluster angemeldet haben, wird automatisch ein Benutzer-SSH-Schlüssel generiert. Sie können ihn verwenden, um sich ohne Passwort anzumelden.

Sie können die Benutzer oder Gruppen eines Clusters erstellen, löschen und ändern, nachdem Ihr Verzeichnisdienst bereitgestellt wurde. Mit AWS Directory Service können Sie dies im AWS Management Console oder mithilfe des Tools Active Directory-Benutzer und -Computer tun. Auf dieses Tool kann von jeder EC2-Instance aus zugegriffen werden, die mit Ihrem Active Directory verknüpft ist. Weitere Informationen finden Sie unter [Installieren der Active-Directory-Verwaltungstools](#).

Wenn Sie die Verwendung AWS ParallelCluster in einem einzelnen Subnetz ohne Internetzugang planen, finden Sie [AWS ParallelCluster in einem einzelnen Subnetz ohne Internetzugang ohne Internetzugang ohne Internetzugang ohne Internetzugang](#). weitere Anforderungen unter.

Themen

- [Erstellen Sie ein Active Directory](#)
- [Erstellen Sie einen Cluster mit einer AD-Domäne](#)

- [Melden Sie sich bei einem Cluster an, der in eine AD-Domäne integriert ist](#)
- [MPI-Jobs werden ausgeführt](#)
- [Beispiel für LDAP \(S\) AWS Managed Microsoft AD -Clusterkonfigurationen](#)

Erstellen Sie ein Active Directory

Stellen Sie sicher, dass Sie ein Active Directory (AD) erstellen, bevor Sie Ihren Cluster erstellen. Informationen zur Auswahl des Active Directory-Typs für Ihren Cluster finden Sie im AWS Directory Service Administratorhandbuch unter [Welche Variante soll ausgewählt](#) werden.

Wenn das Verzeichnis leer ist, fügen Sie Benutzer mit Benutzernamen und Kennwörtern hinzu. Weitere Informationen finden Sie in der Dokumentation, die sich speziell auf [AWS Directory Service for Microsoft Active Directory Simple AD](#) bezieht.

Note

AWS ParallelCluster erfordert, dass sich jedes Active Directory-Benutzerverzeichnis im `/home/$user` Verzeichnis befindet.

Erstellen Sie einen Cluster mit einer AD-Domäne

Warning

In diesem einführenden Abschnitt wird die Einrichtung AWS ParallelCluster mit einem verwalteten Active Directory-Server (AD) über das Lightweight Directory Access Protocol (LDAP) beschrieben. LDAP ist ein unsicheres Protokoll. Für Produktionssysteme empfehlen wir dringend die Verwendung von TLS-Zertifikaten (LDAPS), wie im folgenden [Beispiel für LDAP \(S\) AWS Managed Microsoft AD -Clusterkonfigurationen](#) Abschnitt beschrieben.

Konfigurieren Sie Ihren Cluster für die Integration in ein Verzeichnis, indem Sie die entsprechenden Informationen im `DirectoryService` Abschnitt der Cluster-Konfigurationsdatei angeben. Weitere Informationen finden Sie im Abschnitt [DirectoryService](#) Konfiguration.

Sie können das folgende Beispiel verwenden, um Ihren Cluster AWS Managed Microsoft AD über das Lightweight Directory Access Protocol (LDAP) zu integrieren.

Spezifische Definitionen, die für eine AWS Managed Microsoft AD Over-LDAP-Konfiguration erforderlich sind:

- Sie müssen den `ldap_auth_disable_tls_never_use_in_production` Parameter auf `True` unter [DirectoryService/AdditionalSssdConfigs](#) setzen.
- Sie können entweder Controller-Hostnamen oder IP-Adressen für [DirectoryService/DomainAddr](#) angeben.
- [DirectoryService](#) Die [DomainReadOnlyUser](#) Syntax von/muss wie folgt lauten:

```
cn=ReadOnly,ou=Users,ou=CORP,dc=corp,dc=example,dc=com
```

Holen Sie sich Ihre AWS Managed Microsoft AD Konfigurationsdaten:

```
$ aws ds describe-directories --directory-id "d-abcdef01234567890"
```

```
{
  "DirectoryDescriptions": [
    {
      "DirectoryId": "d-abcdef01234567890",
      "Name": "corp.example.com",
      "DnsIpAddrs": [
        "203.0.113.225",
        "192.0.2.254"
      ],
      "VpcSettings": {
        "VpcId": "vpc-021345abcdef6789",
        "SubnetIds": [
          "subnet-1234567890abcdef0",
          "subnet-abcdef01234567890"
        ],
        "AvailabilityZones": [
          "region-idb",
          "region-idd"
        ]
      }
    }
  ]
}
```

Cluster-Konfiguration für ein AWS Managed Microsoft AD:

```
Region: region-id
Image:
  Os: alinux2
HeadNode:
  InstanceType: t2.micro
Networking:
  SubnetId: subnet-1234567890abcdef0
Ssh:
  KeyName: pcluster
Scheduling:
  Scheduler: slurm
SlurmQueues:
  - Name: queue1
    ComputeResources:
      - Name: t2micro
        InstanceType: t2.micro
        MinCount: 1
        MaxCount: 10
    Networking:
      SubnetIds:
        - subnet-abcdef01234567890
DirectoryService:
  DomainName: dc=corp,dc=example,dc=com
  DomainAddr: ldap://203.0.113.225,ldap://192.0.2.254
  PasswordSecretArn: arn:aws:secretsmanager:region-
id:123456789012:secret:MicrosoftAD.Admin.Password-1234
  DomainReadOnlyUser: cn=ReadOnly,ou=Users,ou=CORP,dc=corp,dc=example,dc=com
  AdditionalSssdConfigs:
    ldap_auth_disable_tls_never_use_in_production: True
```

Um diese Konfiguration für ein Simple AD zu verwenden, ändern Sie den **DomainReadOnlyUser** Eigenschaftswert im **DirectoryService** Abschnitt:

```
DirectoryService:
  DomainName: dc=corp,dc=example,dc=com
  DomainAddr: ldap://203.0.113.225,ldap://192.0.2.254
  PasswordSecretArn: arn:aws:secretsmanager:region-
id:123456789012:secret:SimpleAD.Admin.Password-1234
  DomainReadOnlyUser: cn=ReadOnlyUser,cn=Users,dc=corp,dc=example,dc=com
  AdditionalSssdConfigs:
```

```
ldap_auth_disable_tls_never_use_in_production: True
```

Überlegungen:

- Wir empfehlen, LDAP über TLS/SSL (oder LDAPS) und nicht nur LDAP zu verwenden. TLS/SSL stellt sicher, dass die Verbindung verschlüsselt ist.
- Der Wert der [DomainAddr](#)Eigenschaft [DirectoryService](#)/entspricht den Einträgen in der `DnsIpAddrs` Liste aus der `describe-directories` Ausgabe.
- Wir empfehlen, dass Ihr Cluster Subnetze verwendet, die sich in derselben Availability Zone befinden, auf die [DirectoryService/DomainAddr](#)verweist. Wenn Sie eine [benutzerdefinierte DHCP-Konfiguration \(Dynamic Host Configuration Protocol\)](#) verwenden, die für Verzeichnis-VPCs empfohlen wird, und sich Ihre Subnetze nicht in der [DirectoryService/DomainAddr](#)Availability Zone befinden, ist Querverkehr zwischen Availability Zones möglich. Die Verwendung benutzerdefinierter DHCP-Konfigurationen ist nicht erforderlich, um die AD-Integrationsfunktion für mehrere Benutzer zu verwenden.
- Der [DomainReadOnlyUser](#)Eigenschaftswert [DirectoryService](#)/gibt einen Benutzer an, der im Verzeichnis erstellt werden muss. Dieser Benutzer wird standardmäßig nicht erstellt. Wir empfehlen, diesem Benutzer keine Berechtigung zum Ändern von Verzeichnisdaten zu erteilen.
- Der [PasswordSecretArn](#)Eigenschaftswert [DirectoryService](#)/zeigt auf ein AWS Secrets Manager Geheimnis, das das Passwort des Benutzers enthält, den Sie für die [DomainReadOnlyUser](#)Eigenschaft [DirectoryService](#)/angegeben haben. Wenn sich das Passwort dieses Benutzers ändert, aktualisieren Sie den geheimen Wert und aktualisieren Sie den Cluster. Um den Cluster für den neuen geheimen Wert zu aktualisieren, müssen Sie die Rechenflotte mit dem `pcluster update-compute-fleet` Befehl beenden. Wenn Sie Ihren Cluster für die Verwendung konfiguriert haben [LoginNodes](#), beenden Sie [LoginNodes/Pool](#)s und aktualisieren Sie den Cluster, nachdem Sie [LoginNodes/Pool/Count](#)auf 0 gesetzt haben. Führen Sie dann den folgenden Befehl im Cluster-Kopfnoten aus.

```
sudo /opt/parallelcluster/scripts/directory_service/  
update_directory_service_password.sh
```

Ein anderes Beispiel finden Sie unter auch [Integrieren von Active Directory](#).

Melden Sie sich bei einem Cluster an, der in eine AD-Domäne integriert ist

Wenn Sie die Active Directory (AD) -Domänenintegrationsfunktion aktiviert haben, ist die Authentifizierung per Passwort auf dem Cluster-Hauptknoten aktiviert. Das Basisverzeichnis eines AD-Benutzers wird erstellt, wenn sich der erste Benutzer am Hauptknoten anmeldet oder wenn er zum ersten Mal zu dem AD-Benutzer auf dem Hauptknoten sudo-user wechselt.

Die Kennwortauthentifizierung ist für Cluster-Rechenknoten nicht aktiviert. AD-Benutzer müssen sich mit SSH-Schlüsseln bei Rechenknoten anmelden.

Standardmäßig werden SSH-Schlüssel bei der ersten SSH-Anmeldung am `/${HOME}/.ssh` Hauptknoten im AD-Benutzerverzeichnis eingerichtet. Dieses Verhalten kann deaktiviert werden, indem die [GenerateSshKeysForUsers](#)boolesche Eigenschaft [DirectoryService/false](#) in der Clusterkonfiguration auf gesetzt wird. Standardmäßig [GenerateSshKeysForUsers](#)ist [DirectoryService/auf](#) gesetzt. `true`

Wenn für eine AWS ParallelCluster Anwendung passwortloses SSH zwischen Clusterknoten erforderlich ist, stellen Sie sicher, dass die SSH-Schlüssel im Home-Verzeichnis des Benutzers korrekt eingerichtet sind.

AWS Managed Microsoft AD Passwörter laufen nach 42 Tagen ab. Weitere Informationen finden Sie [AWS Managed Microsoft AD im AWS Directory Service Administratorhandbuch unter Passwortsrichtlinien verwalten für](#). Wenn Ihr Passwort abläuft, muss es zurückgesetzt werden, um den Clusterzugriff wiederherzustellen. Weitere Informationen finden Sie unter [Wie setze ich ein Benutzerpasswort und abgelaufene Passwörter zurück](#).

Note

Wenn die AD-Integrationsfunktion nicht wie erwartet funktioniert, können die SSSD-Protokolle nützliche Diagnoseinformationen zur Behebung des Problems liefern. Diese Protokolle befinden sich im `/var/log/sss` Verzeichnis auf den Clusterknoten. Standardmäßig werden sie auch in der CloudWatch Amazon-Protokollgruppe eines Clusters gespeichert. Weitere Informationen finden Sie unter [Problembehandlung bei der Mehrbenutzerintegration mit Active Directory](#).

MPI-Jobs werden ausgeführt

Wie in SchedMD vorgeschlagen, starten Sie MPI-Jobs Slurm als MPI-Bootstrapping-Methode. Weitere Informationen finden Sie in der offiziellen [Slurm-Dokumentation](#) oder in der [offiziellen Dokumentation Ihrer MPI-Bibliothek](#).

In der [offiziellen Dokumentation von IntelMPI](#) erfahren Sie beispielsweise, dass Sie bei der Ausführung eines StarCCM-Jobs die Einstellung Slurm als Prozesskoordinator vornehmen müssen, indem Sie die Umgebungsvariable exportieren. `I_MPI_HYDRA_BOOTSTRAP=slurm`

Note

Bekanntes Problem

Falls Ihre MPI-Anwendung auf SSH als Mechanismus zur Erzeugung von MPI-Jobs angewiesen ist, ist es möglich, dass [in Slurm ein bekannter Bug](#) auftritt, der dazu führt, dass der Verzeichnisbenutzername falsch auf „nobody“ aufgelöst wird.

Konfigurieren Sie entweder Ihre Anwendung so, dass sie Slurm als MPI-Bootstrapping-Methode verwendet wird, oder finden Sie [Bekannte Probleme bei der Benutzernamenauflösung](#) im Abschnitt zur Fehlerbehebung weitere Informationen und mögliche Problemumgehungen.

Beispiel für LDAP (S) AWS Managed Microsoft AD -Clusterkonfigurationen

AWS ParallelCluster unterstützt den Zugriff mehrerer Benutzer durch Integration mit einem AWS Directory Service über das Lightweight Directory Access Protocol (LDAP) oder LDAP over TLS/SSL (LDAPS).

Die folgenden Beispiele zeigen, wie Clusterkonfigurationen für die Integration AWS Managed Microsoft AD über LDAP (S) erstellt werden.

AWS Managed Microsoft AD über LDAPS mit Zertifikatsüberprüfung

Sie können dieses Beispiel verwenden, um Ihren Cluster AWS Managed Microsoft AD über LDAPS mit Zertifikatsüberprüfung zu integrieren.

Spezifische Definitionen für eine Konfiguration AWS Managed Microsoft AD über LDAPS mit Zertifikaten:

- [DirectoryService/LdapTlsReqCert](#) muss für LDAPS mit Zertifikatsüberprüfung auf hard (Standard) gesetzt sein.
- [DirectoryService/LdapTlsCaCert](#) muss den Pfad zu Ihrem CA-Zertifikat (Certificate of Authority) angeben.

Das CA-Zertifikat ist ein Zertifikatspaket, das die Zertifikate der gesamten CA-Kette enthält, die Zertifikate für die AD-Domänencontroller ausgestellt hat.

Ihr CA-Zertifikat und Ihre Zertifikate müssen auf den Clusterknoten installiert sein.

- Die Hostnamen der Controller müssen für [DirectoryService/](#)angegeben werden [DomainAddr](#), nicht für IP-Adressen.
- [DirectoryService](#) Die [DomainReadOnlyUser](#) Syntax von/muss wie folgt lauten:

```
cn=ReadOnly,ou=Users,ou=CORP,dc=corp,dc=example,dc=com
```

Beispiel für eine Cluster-Konfigurationsdatei für die Verwendung von AD über LDAPS:

```
Region: region-id
Image:
  Os: alinux2
HeadNode:
  InstanceType: t2.micro
  Networking:
    SubnetId: subnet-1234567890abcdef0
  Ssh:
    KeyName: pcluster
  Iam:
    AdditionalIamPolicies:
      - Policy: arn:aws:iam::aws:policy/AmazonS3ReadOnlyAccess
  CustomActions:
    OnNodeConfigured:
      Script: s3://aws-parallelcluster/scripts/pcluster-dub-msad-ldaps.post.sh
Scheduling:
  Scheduler: slurm
  SlurmQueues:
    - Name: queue1
    ComputeResources:
```

```

- Name: t2micro
  InstanceType: t2.micro
  MinCount: 1
  MaxCount: 10
Networking:
  SubnetIds:
    - subnet-abcdef01234567890
Iam:
  AdditionalIamPolicies:
    - Policy: arn:aws:iam::aws:policy/AmazonS3ReadOnlyAccess
CustomActions:
  OnNodeConfigured:
    Script: s3://aws-parallelcluster-pcluster/scripts/pcluster-dub-msad-
ldaps.post.sh
DirectoryService:
  DomainName: dc=corp,dc=example,dc=com
  DomainAddr: ldaps://win-abcdef01234567890.corp.example.com,ldaps://win-
abcdef01234567890.corp.example.com
  PasswordSecretArn: arn:aws:secretsmanager:region-
id:123456789012:secret:MicrosoftAD.Admin.Password-1234
  DomainReadOnlyUser: cn=ReadOnly,ou=Users,ou=CORP,dc=corp,dc=example,dc=com
  LdapTlsCaCert: /etc/ldap/cacerts/corp.example.com.bundleca.cer
  LdapTlsReqCert: hard

```

Fügen Sie Zertifikate hinzu und konfigurieren Sie Domänencontroller im Post-Installationskript:

```

*#!/bin/bash*
set -e

AD_CERTIFICATE_S3_URI="s3://corp.example.com/bundle/corp.example.com.bundleca.cer"
AD_CERTIFICATE_LOCAL="/etc/ldap/cacerts/corp.example.com.bundleca.cer"

AD_HOSTNAME_1="win-abcdef01234567890.corp.example.com"
AD_IP_1="192.0.2.254"

AD_HOSTNAME_2="win-abcdef01234567890.corp.example.com"
AD_IP_2="203.0.113.225"

# Download CA certificate
mkdir -p $(dirname "${AD_CERTIFICATE_LOCAL}")
aws s3 cp "${AD_CERTIFICATE_S3_URI}" "${AD_CERTIFICATE_LOCAL}"
chmod 644 "${AD_CERTIFICATE_LOCAL}"

```

```
# Configure domain controllers reachability
echo "${AD_IP_1} ${AD_HOSTNAME_1}" >> /etc/hosts
echo "${AD_IP_2} ${AD_HOSTNAME_2}" >> /etc/hosts
```

Sie können die Hostnamen der Domänencontroller von Instanzen abrufen, die der Domäne angehören, wie in den folgenden Beispielen gezeigt.

Aus einer Windows-Instanz

```
$ nslookup 192.0.2.254
```

```
Server: corp.example.com
Address: 192.0.2.254

Name: win-abcdef01234567890.corp.example.com
Address: 192.0.2.254
```

Von der Linux-Instanz

```
$ nslookup 192.0.2.254
```

```
192.0.2.254.in-addr.arpa name = corp.example.com
192.0.2.254.in-addr.arpa name = win-abcdef01234567890.corp.example.com
```

AWS Managed Microsoft AD über LDAPS ohne Zertifikatsüberprüfung

Sie können dieses Beispiel verwenden, um Ihren Cluster mit einem Over-LDAPS ohne AWS Managed Microsoft AD Zertifikatsüberprüfung zu integrieren.

Spezifische Definitionen für eine Konfiguration AWS Managed Microsoft AD über LDAPS ohne Zertifikatsverifizierung:

- [DirectoryService/LdapTlsReqCert](#) muss auf gesetzt sein `never`.
- Für [DirectoryService/DomainAddr](#) können entweder Controller-Hostnamen oder IP-Adressen angegeben werden.
- [DirectoryService](#) Die [DomainReadOnlyUser](#) Syntax von/muss wie folgt lauten:

```
cn=ReadOnly,ou=Users,ou=CORP,dc=corp,dc=example,dc=com
```


Beispiel für eine Cluster-Konfigurationsdatei zur Verwendung AWS Managed Microsoft AD über LDAPS ohne Zertifikatsüberprüfung:

```
Region: region-id
Image:
  Os: alinux2
HeadNode:
  InstanceType: t2.micro
  Networking:
    SubnetId: subnet-1234567890abcdef0
  Ssh:
    KeyName: pcluster
Scheduling:
  Scheduler: slurm
  SlurmQueues:
    - Name: queue1
      ComputeResources:
        - Name: t2micro
          InstanceType: t2.micro
          MinCount: 1
          MaxCount: 10
      Networking:
        SubnetIds:
          - subnet-abcdef01234567890
DirectoryService:
  DomainName: dc=corp,dc=example,dc=com
  DomainAddr: ldaps://203.0.113.225,ldaps://192.0.2.254
  PasswordSecretArn: arn:aws:secretsmanager:region-
id:123456789012:secret:MicrosoftAD.Admin.Password-1234
  DomainReadOnlyUser: cn=ReadOnly,ou=Users,ou=CORP,dc=corp,dc=example,dc=com
  LdapTlsReqCert: never
```

Bewährte Methoden

Bewährte Methoden: Auswahl des Hauptknoten-Instanztyps

Auch wenn der Hauptknoten keinen Job ausführt, sind seine Funktionen und seine Größe entscheidend für die Gesamtleistung des Clusters. Beachten Sie bei der Auswahl des Instanztyps, der für Ihren Head-Node verwendet werden soll, die folgenden Merkmale:

Clustergröße: Der Hauptknoten orchestriert die Skalierungslogik des Clusters und ist dafür verantwortlich, neue Knoten an den Scheduler anzuhängen. Um einen Cluster mit einer großen

Anzahl von Knoten nach oben und unten zu skalieren, stellen Sie dem Hauptknoten zusätzliche Rechenkapazität zur Verfügung.

Gemeinsam genutzte Dateisysteme: Wenn Sie gemeinsam genutzte Dateisysteme verwenden, wählen Sie einen Instanztyp mit ausreichender Netzwerkbandbreite und ausreichender Amazon EBS-Bandbreite, um Ihre Workflows abzuwickeln. Stellen Sie sicher, dass der Hauptknoten in der Lage ist, sowohl genügend NFS-Serververzeichnisse für den Cluster verfügbar zu machen als auch die Artefakte zu verarbeiten, die von den Rechenknoten und dem Hauptknoten gemeinsam genutzt werden müssen.

Best Practices: Netzwerkleistung

Die Netzwerkleistung ist für High Performance Computing (HPC) -Anwendungen von entscheidender Bedeutung. Ohne zuverlässige Netzwerkleistung können diese Anwendungen nicht wie erwartet funktionieren. Um die Netzwerkleistung zu optimieren, sollten Sie die folgenden bewährte Methoden beachten.

- **Platzierungsgruppe:** Wenn Sie diese Gruppe verwenden Slurm, sollten Sie erwägen, jede Slurm Warteschlange so zu konfigurieren, dass sie eine Cluster-Platzierungsgruppe verwendet. Die Platzierungsgruppe eines Clusters ist eine logische Gruppierung von Instances innerhalb einer einzelnen Availability Zone. Weitere Informationen finden Sie unter [Platzierungsgruppen](#) im Amazon EC2 EC2-Benutzerhandbuch für Linux-Instances. Sie können [PlacementGroup](#) im [Networking](#) Abschnitt der Warteschlange angeben, dass jede Rechenressource der Platzierungsgruppe der Warteschlange zugewiesen wird. Wenn Sie [PlacementGroup](#) im [Networking](#) Abschnitt der Rechenressource angeben, wird diese spezifische Rechenressource dieser Platzierungsgruppe zugewiesen. Die Gruppenspezifikation für die Platzierung von Rechenressourcen hat Vorrang vor der Warteschlangenspezifikation für die Rechenressource. Weitere Informationen finden Sie unter [SlurmQueuesNetworking/PlacementGroup](#) und [SlurmQueues/ComputeResources/Networking/PlacementGroup](#).

```
Networking:
  PlacementGroup:
    Enabled: true
    Id: your-placement-group-name
```

Alternativ können AWS ParallelCluster Sie eine Praktikumsgruppe für Sie erstellen.

```
Networking:
```

```
PlacementGroup:  
  Enabled: true
```

AbAWS ParallelCluster Version 3.3.0 wurde die Erstellung und Verwaltung von Platzierungsgruppen geändert. Wenn Sie die zu aktivierende Platzierungsgruppe ohne `name` oder `id` in der Warteschlange angeben, wird jeder Rechenressource eine eigene verwaltete Platzierungsgruppe zugewiesen, anstatt einer verwalteten Gruppe für die gesamte Warteschlange. Dies trägt dazu bei, Fehler bei unzureichender Kapazität zu reduzieren. Wenn Sie eine Platzierungsgruppe für die gesamte Warteschlange benötigen, können Sie eine benannte Platzierungsgruppe verwenden.

[SlurmQueues/Networking/PlacementGroup/Name](#) wurde als bevorzugte Alternative zu [SlurmQueues//NetworkingPlacementGroup/](#) hinzugefügt `Id`.

Weitere Informationen finden Sie unter [Networking](#).

- **Verbessertes Netzwerk:** Erwägen Sie, einen Instanztyp zu wählen, der Enhanced Networking unterstützt. Diese Empfehlung gilt für alle [Instances der aktuellen Generation](#). Weitere Informationen finden Sie unter [EBS—im Amazon EC2 EC2-Benutzerhandbuch für Linux-Instances](#).
- **Elastic Fabric-Adapter:** Um ein hohes Maß an skalierbarer Kommunikation von Instanz zu Instanz zu unterstützen, sollten Sie die Wahl von EFA-Netzwerkschnittstellen für Ihr Netzwerk in Betracht ziehen. Die maßgeschneiderte Bypass-Hardware der EFA verbessert die Kommunikation zwischen Instances mit der On-Demand-Elastizität und -Flexibilität der AWS Cloud. Sie können jede [ComputeResource](#) zu Slurm verwendende Warteschlange konfigurieren [Efa](#). Weitere Informationen zur Verwendung von EFA mit AWS ParallelCluster finden Sie unter [Elastic Fabric Adapter](#).

```
ComputeResources:  
  - Name: your-compute-resource-name  
    Efa:  
      Enabled: true
```

Weitere Informationen zu EFA finden Sie unter [Elastic Fabric Adapter](#) im Amazon EC2 EC2-Benutzerhandbuch für Linux-Instances.

- **Instanzbandbreite:** Die Bandbreite skaliert mit der Instanzgröße. Informationen zu den verschiedenen Instance-Typen finden Sie unter [Amazon EBS-optimierte Instances](#) und [Amazon EBS-optimierte Instances](#) im Amazon EC2 EC2-Benutzerhandbuch für Linux-Instances.

Best Practices: Budgetwarnungen

Um die Ressourcenkosten in zu verwaltenAWS ParallelCluster, empfehlen wir, dass SieAWS Budgets Aktionen verwenden, um ein Budget zu erstellen. Sie können auch definierte Budgetschwellewarnungen für ausgewählteAWS Ressourcen erstellen. Weitere Informationen finden Sie im AWS BudgetsBenutzerhandbuch [unter Konfiguration einer Budgetaktion](#). In ähnlicher Weise können Sie Amazon auch verwenden CloudWatch , um einen Rechnungsalarm zu erstellen. Weitere Informationen finden Sie unter [Erstellen eines Rechnungsalarms zur Überwachung Ihrer geschätzten AWS-Gebühren](#).

Bewährte Methoden: Verschieben eines Clusters auf eine neueAWS ParallelCluster Minor- oder Patch-Version

Derzeit ist jedeAWS ParallelCluster Nebenversion zusammen mit ihrerpcCluster CLI in sich abgeschlossen. Um einen Cluster auf eine neue Minor- oder Patch-Version zu verschieben, müssen Sie den Cluster mithilfe der CLI der neuen Version neu erstellen.

Um den Prozess der Migration eines Clusters auf eine neue Minor- oder Patch-Version zu optimieren, empfehlen wir Ihnen, wie folgt vorzugehen:

- Speichern Sie persönliche Daten in externen Volumes, die außerhalb des Clusters erstellt wurden, z. B. Amazon EFS und FSx for Lustre. Auf diese Weise können Sie die Daten in future problemlos von einem Cluster in einen anderen verschieben.
- Erstellen Sie gemeinsam genutzte Speichersysteme mit den folgenden Typen. Sie können diese Systeme mit demAWS CLI oder erstellenAWS Management Console.
 - [SharedStorage](#) / [EbsSettings](#) / [VolumeId](#)
 - [SharedStorage](#) / [EfsSettings](#) / [FileSystemId](#)
 - [SharedStorage](#) / [FsxLustreSettings](#) / [FileSystemId](#)

Definieren Sie ein Dateisystem oder Volume in einer Clusterkonfiguration als vorhandenes Dateisystem oder Volume. Auf diese Weise bleiben sie erhalten, wenn Sie den Cluster löschen, und können an einen neuen Cluster angehängt werden.

Wir empfehlen, Amazon EFS oder FSx for Lustre Lustre-Dateisysteme zu verwenden. Beide Systeme können an mehrere Cluster gleichzeitig angeschlossen werden. Darüber hinaus können Sie eines dieser Systeme an einen neuen Cluster anhängen, bevor Sie Ihren vorhandenen Cluster löschen.

- Verwenden Sie [benutzerdefinierte Bootstrap-Aktionen](#), um Ihre Instances anzupassen, anstatt ein benutzerdefiniertes AMI zu verwenden. Wenn Sie stattdessen ein benutzerdefiniertes AMI verwenden, müssen Sie dieses AMI für jede neue Version löschen und neu erstellen.
- Wir empfehlen, dass Sie die vorherigen Empfehlungen in der folgenden Reihenfolge anwenden:
 1. Aktualisieren Sie die vorhandene Clusterkonfiguration, um vorhandene Dateisystemdefinitionen zu verwenden.
 2. Überprüfen Sie die `pc1uster` Version und aktualisieren Sie sie bei Bedarf.
 3. Erstellen und testen Sie den neuen Cluster. Wenn Sie den neuen Cluster testen, überprüfen Sie Folgendes:
 - Stellen Sie sicher, dass Ihre Daten im neuen Cluster verfügbar sind.
 - Stellen Sie sicher, dass Ihre Anwendung im neuen Cluster funktioniert.
 4. Nachdem Ihr neuer Cluster vollständig getestet und betriebsbereit ist und Sie den vorhandenen Cluster nicht mehr benötigen, löschen Sie ihn.

Umstellung von AWS ParallelCluster 2.x auf 3.x

Benutzerdefinierte Bootstrap-Aktionen

Mit AWS ParallelCluster 3 können Sie verschiedene benutzerdefinierte Bootstrap-Aktionsskripten für den Hauptknoten und die Rechenknoten angeben, indem Sie die Parameter `OnNodeStart` (`pre_install` in AWS ParallelCluster Version 2) und `OnNodeConfigured` (`post_install` in AWS ParallelCluster Version 2) in den Abschnitten [HeadNode](#) und [Scheduling/0](#) verwenden. `SlurmQueues` Weitere Informationen finden Sie unter [Benutzerdefinierte Bootstrap-Aktionen](#).

Benutzerdefinierte Bootstrap-Aktionsskripte, die für AWS ParallelCluster 2 entwickelt wurden, müssen so angepasst werden, dass sie in AWS ParallelCluster 3 verwendet werden können:

- Wir raten davon ab, zwischen Kopf `/etc/parallelcluster/cfnconfig` - und Rechenknoten `cfn_node_type` zu unterscheiden und zu verwenden. Stattdessen empfehlen wir, dass Sie zwei verschiedene Skripts in den Feldern [HeadNode](#) und [Scheduling/](#)angeben [SlurmQueues](#).
- Wenn Sie es vorziehen, weiter zu laden, `/etc/parallelcluster/cfnconfig` um es in Ihrem Bootstrap-Aktionsskript zu verwenden, beachten Sie, `cfn_node_type` dass der Wert von "MasterServer" in "HeadNode" geändert wurde (siehe: [Inklusive Sprache](#)).

- Bei AWS ParallelCluster 2 war das erste Eingabeargument für Bootstrap-Aktionsskripten die S3-URL zum Skript und war reserviert. In AWS ParallelCluster 3 werden nur die in der Konfiguration konfigurierten Argumente an die Skripte übergeben.

⚠ Warning

Die Verwendung interner Variablen, die in der `/etc/parallelcluster/cfnconfig` Datei bereitgestellt werden, wird offiziell nicht unterstützt. Diese Datei könnte im Rahmen einer zukünftigen Version entfernt werden.

AWS ParallelCluster 2.x und 3.x verwenden unterschiedliche Syntax für Konfigurationsdateien

AWS ParallelCluster Die 3.x-Konfiguration verwendet die YAML-Syntax. Die vollständige Referenz finden Sie unter [Konfigurationsdateien](#).

Zusätzlich zur Anforderung eines YAML-Dateiformats wurden in AWS ParallelCluster 3.x eine Reihe von Konfigurationsabschnitten, Einstellungen und Parameterwerten aktualisiert. In diesem Abschnitt finden Sie wichtige Änderungen an der AWS ParallelCluster Konfiguration sowie side-by-side Beispiele, die diese Unterschiede zwischen den einzelnen Versionen von AWS ParallelCluster veranschaulichen.

Beispiel für die Konfiguration mehrerer Scheduler-Warteschlangen mit aktiviertem und deaktiviertem Hyperthreading

AWS ParallelCluster 2:

```
[cluster default]
queue_settings = ht-enabled, ht-disabled
...

[queue ht-enabled]
compute_resource_settings = ht-enabled-i1
disable_hyperthreading = false

[queue ht-disabled]
compute_resource_settings = ht-disabled-i1
```

```

disable_hyperthreading = true

[compute_resource ht-enabled-i1]
instance_type = c5n.18xlarge
[compute_resource ht-disabled-i1]
instance_type = c5.xlarge

```

AWS ParallelCluster3:

```

...
Scheduling:
  Scheduler: slurm
  SlurmQueues:
    - Name: ht-enabled
      Networking:
        SubnetIds:
          - compute_subnet_id
      ComputeResources:
        - Name: ht-enabled-i1
          DisableSimultaneousMultithreading: true
          InstanceType: c5n.18xlarge
    - Name: ht-disabled
      Networking:
        SubnetIds:
          - compute_subnet_id
      ComputeResources:
        - Name: ht-disabled-i1
          DisableSimultaneousMultithreading: false
          InstanceType: c5.xlarge

```

Beispiel für eine neue FSx for Lustre-Dateisystemkonfiguration

AWS ParallelCluster2:

```

[cluster default]
fsx_settings = fsx
...

[fsx fsx]
shared_dir = /shared-fsx
storage_capacity = 1200
imported_file_chunk_size = 1024
import_path = s3://bucket

```

```
export_path = s3://bucket/export_dir
weekly_maintenance_start_time = 3:02:30
deployment_type = PERSISTENT_1
data_compression_type = LZ4
```

AWS ParallelCluster3:

```
...
SharedStorage:
  - Name: fsx
    MountDir: /shared-fsx
    StorageType: FsxLustre
    FsxLustreSettings:
      StorageCapacity: 1200
      ImportedFileChunkSize: 1024
      ImportPath: s3://bucket
      ExportPath: s3://bucket/export_dir
      WeeklyMaintenanceStartTime: "3:02:30"
      DeploymentType: PERSISTENT_1
      DataCompressionType: LZ4
```

Beispiel für eine Cluster-Konfiguration, die ein vorhandenes FSx for Lustre-Dateisystem einbindet

AWS ParallelCluster2:

```
[cluster default]
fsx_settings = fsx
...

[fsx fsx]
shared_dir = /shared-fsx
fsx_fs_id = fsx_fs_id
```

AWS ParallelCluster3:

```
...
SharedStorage:
  - Name: fsx
    MountDir: /shared-fsx
    StorageType: FsxLustre
    FsxLustreSettings:
      FileSystemId: fsx_fs_id
```


Beispiel für einen Cluster mit dem Software-Stack für die Intel HPC Platform Specification

AWS ParallelCluster2:

```
[cluster default]
enable_intel_hpc_platform = true
...
```

AWS ParallelCluster3:

```
...
AdditionalPackages:
  IntelSoftware:
    IntelHpcPlatform: true
```

Hinweise:

- Die Installation der Intel HPC Platform Specification-Software unterliegt den Bedingungen der geltenden [Intel-Endbenutzer-Lizenzvereinbarung](#).

Beispiel für benutzerdefinierte IAM-Konfigurationen, einschließlich: Instanzprofil, Instanzrolle, zusätzliche Richtlinien für Instances und die Rolle für die Lambda-Funktionen, die dem Cluster zugeordnet sind

AWS ParallelCluster2:

```
[cluster default]
additional_iam_policies = arn:aws:iam::aws:policy/
AmazonS3ReadOnlyAccess,arn:aws:iam::aws:policy/AmazonDynamoDBReadOnlyAccess
ec2_iam_role = ec2_iam_role
iam_lambda_role = lambda_iam_role
...
```

AWS ParallelCluster3:

```
...
Iam:
  Roles:
    CustomLambdaResources: lambda_iam_role
HeadNode:
  ...
```

```

Iam:
  InstanceRole: ec2_iam_role
Scheduling:
  Scheduler: slurm
  SlurmQueues:
    - Name: queue1
      ...
      Iam:
        InstanceProfile: iam_instance_profile
    - Name: queue2
      ...
      Iam:
        AdditionalIamPolicies:
          - Policy: arn:aws:iam::aws:policy/AmazonS3ReadOnlyAccess
          - Policy: arn:aws:iam::aws:policy/AmazonDynamoDBReadOnlyAccess

```

Hinweise:

- Für AWS ParallelCluster 2 werden die IAM-Einstellungen auf alle Instanzen eines Clusters angewendet und `additional_iam_policies` können nicht zusammen mit `ec2_iam_role` verwendet werden.
- Für AWS ParallelCluster 3 können Sie unterschiedliche IAM-Einstellungen für Kopf- und Rechenknoten verwenden und sogar unterschiedliche IAM-Einstellungen für jede Rechenwarteschlange angeben.
- Für AWS ParallelCluster 3 können Sie ein IAM-Instanzprofil als Alternative zu einer IAM-Rolle verwenden. `InstanceProfile`, `InstanceRole` oder `AdditionalIamPolicies` kann nicht zusammen konfiguriert werden.

Beispiel für benutzerdefinierte Bootstrap-Aktionen

AWS ParallelCluster2:

```

[cluster default]
s3_read_resource = arn:aws:s3:::bucket_name/*
pre_install = s3://bucket_name/scripts/pre_install.sh
pre_install_args = 'R curl wget'
post_install = s3://bucket_name/scripts/post_install.sh
post_install_args = "R curl wget"
...

```

AWS ParallelCluster3:

```

...
HeadNode:
  ...
  CustomActions:
    OnNodeStart:
      Script: s3://bucket_name/scripts/pre_install.sh
      Args:
        - R
        - curl
        - wget
    OnNodeConfigured:
      Script: s3://bucket_name/scripts/post_install.sh
      Args: ['R', 'curl', 'wget']
  Iam:
    S3Access:
      - BucketName: bucket_name
Scheduling:
  Scheduler: slurm
  SlurmQueues:
    - Name: queue1
    ...
    CustomActions:
      OnNodeStart:
        Script: s3://bucket_name/scripts/pre_install.sh
        Args: ['R', 'curl', 'wget']
      OnNodeConfigured:
        Script: s3://bucket_name/scripts/post_install.sh
        Args: ['R', 'curl', 'wget']
    Iam:
      S3Access:
        - BucketName: bucket_name

```

Beispiel für einen Cluster mit Lese- und Schreibzugriff auf die S3-Bucket-Ressourcen

AWS ParallelCluster2:

```

[cluster default]
s3_read_resource = arn:aws:s3:::bucket/read_only/*
s3_read_write_resource = arn:aws:s3:::bucket/read_and_write/*
...

```

AWS ParallelCluster3:

```
...
HeadNode:
  ...
  Iam:
    S3Access:
      - BucketName: bucket_name
        KeyName: read_only/
        EnableWriteAccess: False
      - BucketName: bucket_name
        KeyName: read_and_write/
        EnableWriteAccess: True
  Scheduling:
    Scheduler: slurm
    SlurmQueues:
      - Name: queue1
      ...
    Iam:
      S3Access:
        - BucketName: bucket_name
          KeyName: read_only/
          EnableWriteAccess: False
        - BucketName: bucket_name
          KeyName: read_and_write/
          EnableWriteAccess: True
```

Inklusive Sprache

AWS ParallelCluster3 verwendet die Wörter „Hauptknoten“ an Stellen, an denen in AWS ParallelCluster 2 „Master“ verwendet wurde. Diese umfasst die folgenden Funktionen:

- Die in der AWS Batch Arbeitsumgebung exportierte Variable wurde geändert: von MASTER_IP bis PCLUSTER_HEAD_NODE_IP.
- Alle AWS CloudFormation Ausgänge wurden von Master* bis geändertHeadNode*.
- Alle NodeType Tags wurden von Master bis geändertHeadNode.

Scheduler-Unterstützung

AWS ParallelCluster3.x unterstützt die Scheduler Son of Grid Engine (SGE) und Torque nicht.

Die AWS Batch Befehle `awsbhosts`, `awsbkill`, `awsbout`, `awsbqueues`, `awsbstat`, und `awsbsub` werden als separates `aws-parallelcluster-awsbatch-cli` PyPI-Paket verteilt. Dieses Paket

wird von AWS ParallelCluster auf dem Hauptknoten installiert. Sie können diese AWS Batch Befehle weiterhin vom Hauptknoten des Clusters aus verwenden. Wenn Sie jedoch AWS Batch Befehle von einem anderen Ort als dem Hauptknoten aus verwenden möchten, müssen Sie zuerst das `aws-parallelcluster-awsbatch-cli` PyPI-Paket installieren.

AWS ParallelCluster CLI

Die AWS ParallelCluster Befehlszeilenschnittstelle (CLI) wurde geändert. Die neue Syntax ist in beschrieben [AWS ParallelCluster-CLI-Befehle](#). Das Ausgabeformat für die CLI ist eine [JSON-Zeichenfolge](#).

Konfiguration eines neuen Clusters

Der `pcluster configure` Befehl enthält in AWS ParallelCluster 3 andere Parameter als in AWS ParallelCluster 2. Weitere Informationen finden Sie unter [pcluster configure](#).

Beachten Sie auch, dass sich die Syntax der Konfigurationsdatei von AWS ParallelCluster 2 geändert hat. Eine vollständige Referenz zu den Clusterkonfigurationseinstellungen finden Sie unter [Cluster-Konfigurationsdatei](#).

Einen neuen Cluster erstellen

AWS ParallelClusterDer `pcluster create` Befehl von 2 wurde durch den [pcluster create-cluster](#) Befehl ersetzt.

Beachten Sie, dass das Standardverhalten in AWS ParallelCluster 2.x ohne die `-nw` Option darin besteht, auf Ereignisse zur Clustererstellung zu warten, während der AWS ParallelCluster 3.x-Befehl sofort zurückkehrt. Der Fortschritt der Clustererstellung kann mithilfe von überwacht werden [pcluster describe-cluster](#).

Eine AWS ParallelCluster 3-Konfigurationsdatei enthält eine einzige Cluster-Definition, sodass der `-t` Parameter nicht mehr benötigt wird.

Im Folgenden finden Sie ein Beispiel für eine Konfigurationsdatei.

```
# AWS ParallelCluster v2
$ pcluster create \
  -r REGION \
  -c V2_CONFIG_FILE \
  -nw \
```

```
-t CLUSTER_TEMPLATE \  
CLUSTER_NAME  
  
# AWS ParallelCluster v3  
$ pcluster create-cluster \  
  --region REGION \  
  --cluster-configuration V3_CONFIG_FILE \  
  --cluster-name CLUSTER_NAME
```

Cluster auflisten

Der Befehl `pcluster list` AWS ParallelCluster 2.x muss durch einen [pcluster list-clusters](#) Befehl ersetzt werden.

Hinweis: Sie benötigen AWS ParallelCluster v2-CLI, um Cluster aufzulisten, die mit 2.x-Versionen von erstellt wurden. AWS ParallelCluster Informationen [Installation AWS ParallelCluster in einer virtuellen Umgebung \(empfohlen\)](#) zur Installation mehrerer Versionen von AWS ParallelCluster Using Virtual Environments finden Sie unter.

```
# AWS ParallelCluster v2  
$ pcluster list -r REGION  
  
# AWS ParallelCluster v3  
$ pcluster list-clusters --region REGION
```

Einen Cluster starten und stoppen

Die Befehle `pcluster start` und `pcluster stop` AWS ParallelCluster 2.x müssen durch [pcluster update-compute-fleet](#) Befehle ersetzt werden.

Eine Rechenflotte starten:

```
# AWS ParallelCluster v2  
$ pcluster start \  
  -r REGION \  
  CLUSTER_NAME  
  
# AWS ParallelCluster v3 - Slurm fleets  
$ pcluster update-compute-fleet \  
  --region REGION \  
  --cluster-name CLUSTER_NAME \  
  --status START_REQUESTED
```

```
# AWS ParallelCluster v3 - AWS Batch fleets
$ pcluster update-compute-fleet \
  --region REGION \
  --cluster-name CLUSTER_NAME \
  --status ENABLED
```

Stoppen einer Rechenflotte:

```
# AWS ParallelCluster v2
$ pcluster stop \
  -r REGION \
  CLUSTER_NAME

# AWS ParallelCluster v3 - Slurm fleets
$ pcluster update-compute-fleet \
  --region REGION \
  --cluster-name CLUSTER_NAME \
  --status STOP_REQUESTED

# AWS ParallelCluster v3 - AWS Batch fleets
$ pcluster update-compute-fleet \
  --region REGION \
  --cluster-name CLUSTER_NAME \
  --status DISABLED
```

Verbindung zu einem Cluster herstellen

Der `pcluster ssh` AWS ParallelCluster 2.x-Befehl hat in AWS ParallelCluster 3.x unterschiedliche Parameternamen. Siehe [pcluster ssh](#).

Verbindung zu einem Cluster herstellen:

```
# AWS ParallelCluster v2
$ pcluster ssh \
  -r REGION \
  CLUSTER_NAME \
  -i ~/.ssh/id_rsa

# AWS ParallelCluster v3
$ pcluster ssh \
  --region REGION \
  --cluster-name CLUSTER_NAME \
```

```
-i ~/.ssh/id_rsa
```

IMDS-Konfigurationsupdate

Ab Version 3.0.0 wurde standardmäßig Unterstützung für die Beschränkung des Zugriffs auf das IMDS des Hauptknotens (und die Anmeldeinformationen des Instanzprofils) auf eine Untergruppe von Superbenutzern AWS ParallelCluster eingeführt. Weitere Informationen finden Sie unter [Imds-Eigenschaften](#).

Unterstützte Regionen für AWS ParallelCluster

AWS ParallelCluster Version 3 ist in den folgenden Versionen verfügbar AWS-Regionen:

Name der Region	Region
US East (Ohio)	us-east-2
US East (N. Virginia)	us-east-1
US West (N. California)	us-west-1
US West (Oregon)	us-west-2
Afrika (Kapstadt)	af-south-1
Asien-Pazifik (Hongkong)	ap-east-1
Asien-Pazifik (Mumbai)	ap-south-1
Asia Pacific (Seoul)	ap-northeast-2
Asia Pacific (Singapore)	ap-southeast-1
Asia Pacific (Sydney)	ap-southeast-2
Asia Pacific (Tokyo)	ap-northeast-1
Canada (Central)	ca-central-1
China (Beijing)	cn-north-1

Name der Region	Region
China (Ningxia)	cn-northwest-1
Europe (Frankfurt)	eu-central-1
Europe (Ireland)	eu-west-1
Europe (London)	eu-west-2
Europa (Mailand)	eu-south-1
Europa (Paris)	eu-west-3
Europe (Stockholm)	eu-north-1
Naher Osten (Bahrain)	me-south-1
South America (São Paulo)	sa-east-1
AWS GovCloud (US-Ost)	us-gov-east-1
AWS GovCloud (US-West)	us-gov-west-1
Israel (Tel Aviv)	il-central-1

Benutzen AWS ParallelCluster

Themen

- [AWS ParallelClusterUI](#)
- [AWS LambdaVPC-Konfiguration inAWS ParallelCluster](#)
- [AWS Identity and Access Management Berechtigungen in AWS ParallelCluster](#)
- [Netzwerkkonfigurationen](#)
- [Login-Knoten](#)
- [Benutzerdefinierte Bootstrap-Aktionen](#)
- [Arbeiten mit Amazon S3](#)
- [Arbeiten mit Spot-Instances](#)
- [Scheduler werden unterstützt von AWS ParallelCluster](#)
- [Gemeinsamer Speicher](#)
- [AWS ParallelClusterRessourcen und Tagging](#)
- [Überwachung AWS ParallelCluster und Protokolle](#)
- [AWS CloudFormation benutzerdefinierte Ressource](#)
- [Elastic Fabric Adapter](#)
- [Intel MPI aktivieren](#)
- [AWS ParallelCluster-API](#)
- [Stellen Sie mit NICE DCV eine Connect zum Hauptknoten her](#)
- [Verwenden von pcluster update-cluster](#)
- [AWS ParallelCluster AMI-Anpassung](#)
- [Erstellen Sie Instances mit ODCR Starten Sie Instances](#)
- [AMI-Patching und Austausch von EC2-Instances](#)
- [Betriebssysteme](#)

AWS ParallelClusterUI

DieAWS ParallelCluster Benutzeroberfläche ist eine webbasierte Benutzeroberfläche, die als Dashboard für die Erstellung, Überwachung und Verwaltung von Clustern dient. Sie installieren

und greifen auf die AWS ParallelCluster Benutzeroberfläche in Ihrem zu AWS-Konto. Die AWS ParallelCluster Benutzeroberfläche wurde mit AWS ParallelCluster Version 3.5.0 hinzugefügt.

Informationen zur Installation der AWS ParallelCluster Benutzeroberfläche und zum Einstieg finden Sie unter [Installation der AWS ParallelCluster Benutzeroberfläche](#) und [Konfiguration und Erstellung eines Clusters mit der AWS ParallelCluster Benutzeroberfläche](#).

The screenshot displays the AWS ParallelCluster console interface. At the top, the header shows 'AWS ParallelCluster' and the user 'user@domain.com' in the 'eu-west-1' region. The main content area is titled 'Clusters (2)' and includes a search bar with the text 'Find clusters'. Below the search bar is a table listing clusters:

Name	Status	Version
hpc-cluster-1	CREATE COMPLETE	3.5.0
hpc-cluster-2	DELETE IN PROGRESS	3.5.0

Below the table, the details for 'Cluster: hpc-cluster-1' are shown. The 'Details' tab is active, displaying the following properties:

- Cluster configuration:** [VIEW](#)
- SSH command:** `ssh ec2-user@54.78.245.22`
- EC2 Instance Connect:** `mssh -r eu-west-1 ec2-user@i-0b14dc1a2f5dc048e`
- Cluster status:** CREATE COMPLETE
- Compute fleet status:** RUNNING
- Version:** 3.5.0
- Region:** eu-west-1
- Created time:** March 10, 2023 at 09:39 (UTC+1:00)
- Latest update time:** March 10, 2023 at 09:39 (UTC+1:00)


Die AWS ParallelCluster Benutzeroberfläche unterstützt die folgenden -Funktionen:

- Zeigt Folgendes an:
 - Die Liste der Cluster, die Sie in Ihrem AWS-Konto mit erstellt haben AWS ParallelCluster.
 - Der verfügbare Status und die verfügbaren Details für Ihre aufgelisteten Cluster.
 - CloudFormation stapeln Sie Ereignisse und AWS ParallelCluster Protokolle, die Sie für die Überwachung verwenden können.

- Der Status von Jobs, die in Ihren Clustern ausgeführt werden.
- Die Liste der benutzerdefinierten Images, mit denen Sie Cluster erstellen können.
- Die Liste der offiziellen Images, die die Benutzeroberfläche zum Erstellen von Clustern verwendet.
- Die Liste der Benutzer, mit denen Sie auf die AWS ParallelCluster Benutzeroberfläche zugreifen können. Sie können Benutzer hinzufügen oder entfernen.
- Enthält step-by-step Anleitungen zum Erstellen und Bearbeiten (Aktualisieren) eines Clusters und zur Auswahl unterstützter Cluster-Funktionen zum Hinzufügen, Bearbeiten oder Entfernen. Unzugängliche Eingabefelder können für die Clusterkonfiguration, die gerade bearbeitet wird, nicht geändert werden. Sie haben die Möglichkeit, vor der Cluster-Bereitstellung eine Testvalidierung Ihrer Clusterkonfiguration durchzuführen.
- Enthält direkte Shell-Links für den Zugriff auf den Hauptknoten in der Cluster-Ansicht. Wählen Sie in der step-by-step Anleitung die Option SSM-Sitzung hinzufügen aus, um den direkten Shell-Zugriff und die SSM Managed Instance Core-Richtlinie auf dem Hauptknoten hinzuzufügen.

Beachten Sie Folgendes, wenn Sie die AWS ParallelCluster Benutzeroberfläche verwenden, um Ihre Cluster zu erstellen und zu verwalten:

- Sie können Cluster nur mit derselben AWS ParallelCluster Version erstellen und bearbeiten oder Images erstellen, die für die Erstellung der AWS ParallelCluster Benutzeroberfläche verwendet wurden. Cluster oder Bilder früherer Versionen können nur angezeigt werden. Wenn Sie mehrere Versionen von Clustern und Images verwalten, empfehlen wir, eine AWS ParallelCluster UI-Instanz zu erstellen, die jede Version unterstützt.
- Die AWS ParallelCluster Benutzeroberfläche ist so konzipiert, dass sie die `pccluster` CLI-Funktionalität widerspiegelt. Es gibt einige Unterschiede. Wenn Sie sich an die step-by-step Anleitung halten, verwenden Sie alle unterstützten Funktionen. Vor der Bereitstellung haben Sie die Möglichkeit, die Cluster- oder Image-Konfiguration manuell zu bearbeiten. Wenn Sie dies tun, empfehlen wir Ihnen, die Konfiguration zu überprüfen, indem Sie Trockenlauf wählen, um sicherzustellen, dass Ihre Änderungen vollständig unterstützt werden.

 Note

AWS ParallelClusterUI unterstützt nicht AWS Batch.

AWS LambdaVPC-Konfiguration inAWS ParallelCluster

AWS ParallelCluster wird verwendet AWS Lambda, um Operationen während des Lebenszyklus des Clusters auszuführen. Eine [AWS Lambda Funktion läuft immer in einer VPC](#), die dem Lambda-Dienst gehört. Diese Lambda-Funktion kann auch mit privaten Subnetzen in einer Virtual Private Cloud (VPC) verbunden werden, um auf private Ressourcen herzustellen.

Note

Lambda-Funktionen können keine direkte Verbindung mit einer VPC mit Dedicated-Instance-Tenancy herstellen. Zum Herstellen einer Verbindung mit Ressourcen in einer dedizierten VPC verbinden Sie sie mit einer zweiten VPC mit Standard-Tenancy, die eine Verbindung mit einer dedizierten VPC herzustellen.

Weitere Informationen finden Sie unter [Dedicated Instances](#) im EC2-Benutzerhandbuch für Linux-Instances und [Wie verbinde ich eine Lambda-Funktion mit einer dedizierten VPC?](#) aus dem AWS Knowledge Center.

Lambda-Funktionen, die von erstellt wurden, AWS ParallelCluster können mit einer privaten VPC verbunden werden. Auf diese Lambda-Funktionen muss zugegriffen AWS-Services werden. Sie können den Zugriff über das Internet oder VPC-Endpunkte bereitstellen, indem Sie die folgenden Methoden verwenden.

- Zugang zum Internet

Zum Herstellen eines Zugriffs auf das Internet und AWS-Services erfordert eine Lambda-Funktion Network Address Translation (NAT). Leiten Sie ausgehenden Datenverkehr von Ihrem privaten Subnetz an ein [NAT-Gateway](#) in einem öffentlichen Subnetz weiter.

- VPC-Endpunkte

Mehrere AWS-Services bieten [VPC-Endpunkte](#). Sie können für die Verbindung zu VPC-Endpunkte verwenden, um Verbindungen zu AWS-Services VPC-Endpunkte herzustellen. Eine Liste der AWS ParallelCluster VPC-Endpunkte finden Sie unter [Networking](#).

Note

Jede Kombination von Subnetzen und Sicherheitsgruppen muss den Zugriff auf AWS-Services auf eine dieser Methoden ermöglichen. Subnetze und Sicherheitsgruppen müssen sich in derselben VPC befinden.

Weitere Informationen finden Sie unter [VPC-Endpoints im Amazon Virtual Private Cloud Cloud-Benutzerhandbuch](#) und [Internet- und Servicezugriff für mit VPC verbundene Funktionen](#) im AWS Lambda-Entwicklerhandbuch.

Informationen zur Konfiguration der Verwendung von Lambda-Funktionen und VPCs finden Sie unter [DeploymentSettings/LambdaFunctionsVpcConfig](#) für Cluster oder [DeploymentSettings/LambdaFunctionsVpcConfig](#) für Images.

AWS Identity and Access Management Berechtigungen in AWS ParallelCluster

AWS ParallelCluster verwendet IAM-Berechtigungen, um den Zugriff auf Ressourcen bei der Erstellung und Verwaltung von Clustern zu steuern.

Um Cluster in einem AWS Konto zu erstellen und zu verwalten, sind Berechtigungen auf zwei Ebenen erforderlich:

- Berechtigungen, die der `pcluster` Benutzer benötigt, um die `pcluster` CLI-Befehle zum Erstellen und Verwalten von Clustern aufzurufen.
- Berechtigungen, die die Clusterressourcen für die Ausführung von Clusteraktionen benötigen.

AWS ParallelCluster verwendet ein [EC2-Instanzprofil und eine EC2-Instanzrolle](#), um Clusterressourcenberechtigungen bereitzustellen. Für die Verwaltung von Cluster-Ressourcenberechtigungen sind AWS ParallelCluster auch Berechtigungen für IAM-Ressourcen erforderlich. Weitere Informationen finden Sie unter [AWS ParallelCluster Benutzerbeispielrichtlinien für die Verwaltung von IAM-Ressourcen](#).

`pcluster` Benutzer benötigen IAM-Berechtigungen, um die `pcluster` CLI zum Erstellen und Verwalten eines Clusters und seiner Ressourcen zu verwenden. Diese Berechtigungen sind in IAM-Richtlinien enthalten, die einem Benutzer oder einer Rolle hinzugefügt werden können. Weitere

Informationen zu IAM-Rollen finden Sie unter [Erstellen einer Benutzerrolle](#) im AWS Identity and Access Management Benutzerhandbuch.

Sie können auch [AWS ParallelCluster Konfigurationsparameter zur Verwaltung von IAM-Berechtigungen](#) verwenden.

Die folgenden Abschnitte enthalten die erforderlichen Berechtigungen mit Beispielen.

Um die Beispielrichtlinien zu verwenden `<REGION><AWS ACCOUNT ID>`, ersetzen Sie die Zeichenfolgen und ähnliche Zeichenfolgen durch die entsprechenden Werte.

Die folgenden Beispielrichtlinien beinhalten Amazon Resource Names (ARNs) für die Ressourcen. Wenn Sie in den Partitionen AWS GovCloud (US) oder AWS China arbeiten, müssen die ARNs geändert werden. Insbesondere müssen sie von „arn:aws“ auf „arn:aws-us-gov“ für die AWS GovCloud (US) Partition oder „arn:aws-cn“ für die China-Partition geändert werden. AWS Weitere Informationen finden Sie unter [Amazon Resource Names \(ARNs\) in AWS GovCloud \(US\) Regionen](#) im AWS GovCloud (US) Benutzerhandbuch und [ARNs für AWS Services in China](#) unter Erste Schritte mit AWS Services in China.

Sie können Änderungen an den Beispielrichtlinien in der [AWS ParallelCluster Dokumentation](#) unter nachverfolgen. GitHub

Themen

- [AWS ParallelCluster EC2-Instanzrollen](#)
- [AWS ParallelCluster Beispiele pcluster für Benutzerrichtlinien](#)
- [AWS ParallelCluster Benutzerbeispielrichtlinien für die Verwaltung von IAM-Ressourcen](#)
- [AWS ParallelCluster Konfigurationsparameter zur Verwaltung von IAM-Berechtigungen](#)

AWS ParallelCluster EC2-Instanzrollen

Wenn Sie einen Cluster mit den Standardkonfigurationseinstellungen erstellen, AWS ParallelCluster verwendet [EC2-Instance-Profile](#), um automatisch eine [Standard-Cluster-EC2-Instance-Rolle](#) zu erstellen, die die für die Erstellung und Verwaltung des Clusters und seiner Ressourcen erforderlichen Berechtigungen bereitstellt.

Alternativen zur Verwendung der Standard-Instanzrolle AWS ParallelCluster

Anstelle der AWS ParallelCluster Standard-Instanzrolle können Sie die InstanceRole Cluster-Konfigurationseinstellung verwenden, um Ihre eigene bestehende IAM-Rolle für EC2 anzugeben.

Weitere Informationen finden Sie unter [AWS ParallelCluster Konfigurationsparameter zur Verwaltung von IAM-Berechtigungen](#). In der Regel geben Sie vorhandene IAM-Rollen an, um die EC2 erteilten Berechtigungen vollständig zu kontrollieren.

[Wenn Sie beabsichtigen, der Standard-Instance-Rolle zusätzliche Richtlinien hinzuzufügen, empfehlen wir Ihnen, die zusätzlichen IAM-Richtlinien mithilfe der AdditionalIamPoliciesKonfigurationseinstellung anstelle der InstanceProfile Einstellungen oder zu übergeben. InstanceRole](#) Sie können aktualisieren, AdditionalIamPolicies wenn Sie Ihren Cluster aktualisieren, jedoch nicht, InstanceRole wenn Sie Ihren Cluster aktualisieren.

AWS ParallelCluster Beispiele **pcluster** für Benutzerrichtlinien

Die folgenden Beispiele zeigen die Benutzerrichtlinien, die für die Erstellung AWS ParallelCluster und Verwaltung der zugehörigen Ressourcen mithilfe der `pcluster` CLI erforderlich sind. Sie können Richtlinien an einen Benutzer oder eine Rolle anhängen.

Themen

- [AWS ParallelCluster pclusterGrundlegende Benutzerrichtlinie](#)
- [Zusätzliche AWS ParallelCluster pcluster Benutzerrichtlinie bei der Verwendung des AWS Batch Schedulers](#)
- [Zusätzliche AWS ParallelCluster pcluster Benutzerrichtlinie bei der Verwendung von Amazon FSx for Lustre](#)
- [AWS ParallelCluster Benutzerrichtlinie zur Image-Erstellung pcluster](#)

AWS ParallelCluster **pcluster**Grundlegende Benutzerrichtlinie

Die folgende Richtlinie zeigt die Berechtigungen, die zum Ausführen von AWS ParallelCluster `pcluster` Befehlen erforderlich sind.

Die letzte in der Richtlinie aufgeführte Aktion dient der Überprüfung aller in der Clusterkonfiguration angegebenen Geheimnisse. Beispielsweise wird ein AWS Secrets Manager Geheimnis verwendet, um die [DirectoryService](#)Integration zu konfigurieren. In diesem Fall wird ein Cluster nur erstellt, wenn ein gültiges Geheimnis in der vorhanden ist [PasswordSecretArn](#). Wenn diese Aktion ausgelassen wird, wird die geheime Überprüfung übersprungen. Um Ihre Sicherheitslage zu verbessern, empfehlen wir, dass Sie diese Richtlinienerklärung einschränken und nur die in Ihrer Clusterkonfiguration angegebenen Geheimnisse hinzufügen.

Note

Wenn bestehende Amazon EFS-Dateisysteme die einzigen Dateisysteme sind, die in Ihrem Cluster verwendet werden, können Sie die Amazon EFS-Beispielrichtlinien auf die spezifischen Dateisysteme beschränken, auf die in [SharedStorage Abschnitt](#) der Cluster-Konfigurationsdatei verwiesen wird.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Action": [
        "ec2:Describe*"
      ],
      "Resource": "*",
      "Effect": "Allow",
      "Sid": "EC2Read"
    },
    {
      "Action": [
        "ec2:AllocateAddress",
        "ec2:AssociateAddress",
        "ec2:AttachNetworkInterface",
        "ec2:AuthorizeSecurityGroupEgress",
        "ec2:AuthorizeSecurityGroupIngress",
        "ec2:CreateFleet",
        "ec2:CreateLaunchTemplate",
        "ec2:CreateLaunchTemplateVersion",
        "ec2:CreateNetworkInterface",
        "ec2:CreatePlacementGroup",
        "ec2:CreateSecurityGroup",
        "ec2:CreateSnapshot",
        "ec2:CreateTags",
        "ec2>DeleteTags",
        "ec2:CreateVolume",
        "ec2>DeleteLaunchTemplate",
        "ec2>DeleteNetworkInterface",
        "ec2>DeletePlacementGroup",
        "ec2>DeleteSecurityGroup",
        "ec2>DeleteVolume",
        "ec2:DisassociateAddress",
```

```

        "ec2:ModifyLaunchTemplate",
        "ec2:ModifyNetworkInterfaceAttribute",
        "ec2:ModifyVolume",
        "ec2:ModifyVolumeAttribute",
        "ec2:ReleaseAddress",
        "ec2:RevokeSecurityGroupEgress",
        "ec2:RevokeSecurityGroupIngress",
        "ec2:RunInstances",
        "ec2:TerminateInstances"
    ],
    "Resource": "*",
    "Effect": "Allow",
    "Sid": "EC2Write"
},
{
    "Action": [
        "dynamodb:DescribeTable",
        "dynamodb:ListTagsOfResource",
        "dynamodb:CreateTable",
        "dynamodb>DeleteTable",
        "dynamodb:GetItem",
        "dynamodb:PutItem",
        "dynamodb:UpdateItem",
        "dynamodb:Query",
        "dynamodb:TagResource"
    ],
    "Resource": "arn:aws:dynamodb:*:<AWS ACCOUNT ID>:table/parallelcluster-*",
    "Effect": "Allow",
    "Sid": "DynamoDB"
},
{
    "Action": [
        "route53:ChangeResourceRecordSets",
        "route53:ChangeTagsForResource",
        "route53:CreateHostedZone",
        "route53>DeleteHostedZone",
        "route53:GetChange",
        "route53:GetHostedZone",
        "route53:ListResourceRecordSets",
        "route53:ListQueryLoggingConfigs"
    ],
    "Resource": "*",
    "Effect": "Allow",
    "Sid": "Route53HostedZones"
}

```

```

    },
    {
      "Action": [
        "cloudformation:*"
      ],
      "Resource": "*",
      "Effect": "Allow",
      "Sid": "CloudFormation"
    },
    {
      "Action": [
        "cloudwatch:PutDashboard",
        "cloudwatch:ListDashboards",
        "cloudwatch>DeleteDashboards",
        "cloudwatch:GetDashboard",
        "cloudwatch:PutMetricAlarm",
        "cloudwatch>DeleteAlarms",
        "cloudwatch:DescribeAlarms"
        "cloudwatch:PutCompositeAlarm"
      ],
      "Resource": "*",
      "Effect": "Allow",
      "Sid": "CloudWatch"
    },
    {
      "Action": [
        "iam:GetRole",
        "iam:GetRolePolicy",
        "iam:GetPolicy",
        "iam:SimulatePrincipalPolicy",
        "iam:GetInstanceProfile"
      ],
      "Resource": [
        "arn:aws:iam::<AWS ACCOUNT ID>:role/*",
        "arn:aws:iam::<AWS ACCOUNT ID>:policy/*",
        "arn:aws:iam::aws:policy/*",
        "arn:aws:iam::<AWS ACCOUNT ID>:instance-profile/*"
      ],
      "Effect": "Allow",
      "Sid": "IamRead"
    },
    {
      "Action": [
        "iam:CreateInstanceProfile",

```

```

        "iam:DeleteInstanceProfile",
        "iam:AddRoleToInstanceProfile",
        "iam:RemoveRoleFromInstanceProfile"
    ],
    "Resource": [
        "arn:aws:iam::<AWS ACCOUNT ID>:instance-profile/parallelcluster/*"
    ],
    "Effect": "Allow",
    "Sid": "IamInstanceProfile"
},
{
    "Condition": {
        "StringEqualsIfExists": {
            "iam:PassedToService": [
                "lambda.amazonaws.com",
                "ec2.amazonaws.com",
                "spotfleet.amazonaws.com"
            ]
        }
    },
    "Action": [
        "iam:PassRole"
    ],
    "Resource": [
        "arn:aws:iam::<AWS ACCOUNT ID>:role/parallelcluster/*"
    ],
    "Effect": "Allow",
    "Sid": "IamPassRole"
},
{
    "Action": [
        "lambda:CreateFunction",
        "lambda>DeleteFunction",
        "lambda:GetFunctionConfiguration",
        "lambda:GetFunction",
        "lambda:InvokeFunction",
        "lambda:AddPermission",
        "lambda:RemovePermission",
        "lambda:UpdateFunctionConfiguration",
        "lambda:TagResource",
        "lambda>ListTags",
        "lambda:UntagResource"
    ],
    "Resource": [

```

```

        "arn:aws:lambda:*:<AWS ACCOUNT ID>:function:parallelcluster-*",
        "arn:aws:lambda:*:<AWS ACCOUNT ID>:function:pcluster-*"
    ],
    "Effect": "Allow",
    "Sid": "Lambda"
},
{
    "Action": [
        "s3:*"
    ],
    "Resource": [
        "arn:aws:s3:::parallelcluster-*",
        "arn:aws:s3:::aws-parallelcluster-*"
    ],
    "Effect": "Allow",
    "Sid": "S3ResourcesBucket"
},
{
    "Action": [
        "s3:Get*",
        "s3:List*"
    ],
    "Resource": "arn:aws:s3:::*-aws-parallelcluster*",
    "Effect": "Allow",
    "Sid": "S3ParallelClusterReadOnly"
},
{
    "Action": [
        "elasticfilesystem:*"
    ],
    "Resource": [
        "arn:aws:elasticfilesystem:*:<AWS ACCOUNT ID>:*"
    ],
    "Effect": "Allow",
    "Sid": "EFS"
},
{
    "Action": [
        "logs:DeleteLogGroup",
        "logs:PutRetentionPolicy",
        "logs:DescribeLogGroups",
        "logs:CreateLogGroup",
        "logs:TagResource",
        "logs:UntagResource",

```

```

        "logs:FilterLogEvents",
        "logs:GetLogEvents",
        "logs:CreateExportTask",
        "logs:DescribeLogStreams",
        "logs:DescribeExportTasks",
        "logs:DescribeMetricFilters",
        "logs:PutMetricFilter",
        "logs>DeleteMetricFilter"
    ],
    "Resource": "*",
    "Effect": "Allow",
    "Sid": "CloudWatchLogs"
},
{
    "Action": [
        "resource-groups:ListGroupResources"
    ],
    "Resource": "*",
    "Effect": "Allow",
    "Sid": "ResourceGroupRead"
},
{
    "Sid": "AllowDescribingFileCache",
    "Effect": "Allow",
    "Action": [
        "fsx:DescribeFileCaches"
    ],
    "Resource": "*"
},
{
    "Action": "secretsmanager:DescribeSecret",
    "Resource": "arn:aws:secretsmanager:<REGION>:<AWS ACCOUNT ID>:secret:<SECRET
NAME>",
    "Effect": "Allow"
}
]
}

```

Zusätzliche AWS ParallelCluster **pcluster** Benutzerrichtlinie bei der Verwendung des AWS Batch Schedulers

Falls Sie einen Cluster mit AWS Batch Scheduler erstellen und verwalten müssen, ist die folgende zusätzliche Richtlinie erforderlich.

```

{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Condition": {
        "StringEqualsIfExists": {
          "iam:PassedToService": [
            "ecs-tasks.amazonaws.com",
            "batch.amazonaws.com",
            "codebuild.amazonaws.com"
          ]
        }
      },
      "Action": [
        "iam:PassRole"
      ],
      "Resource": [
        "arn:aws:iam::<AWS ACCOUNT ID>:role/parallelcluster/*"
      ],
      "Effect": "Allow",
      "Sid": "IamPassRole"
    },
    {
      "Condition": {
        "StringEquals": {
          "iam:AWSServiceName": [
            "batch.amazonaws.com"
          ]
        }
      },
      "Action": [
        "iam:CreateServiceLinkedRole",
        "iam>DeleteServiceLinkedRole"
      ],
      "Resource": [
        "arn:aws:iam::<AWS ACCOUNT ID>:role/aws-service-role/
batch.amazonaws.com/*"
      ],
      "Effect": "Allow"
    },
    {
      "Action": [
        "codebuild:*"
      ]
    }
  ]
}

```

```
    ],
    "Resource": "arn:aws:codebuild:*:<AWS ACCOUNT ID>:project/pcluster-*",
    "Effect": "Allow"
  },
  {
    "Action": [
      "ecr:*"
    ],
    "Resource": "*",
    "Effect": "Allow",
    "Sid": "ECR"
  },
  {
    "Action": [
      "batch:*"
    ],
    "Resource": "*",
    "Effect": "Allow",
    "Sid": "Batch"
  },
  {
    "Action": [
      "events:*"
    ],
    "Resource": "*",
    "Effect": "Allow",
    "Sid": "AmazonCloudWatchEvents"
  },
  {
    "Action": [
      "ecs:DescribeContainerInstances",
      "ecs:ListContainerInstances"
    ],
    "Resource": "*",
    "Effect": "Allow",
    "Sid": "ECS"
  }
]
}
```


Zusätzliche AWS ParallelCluster **pcluster** Benutzerrichtlinie bei der Verwendung von Amazon FSx for Lustre

Falls Sie einen Cluster mit Amazon FSx for Lustre erstellen und verwalten müssen, ist die folgende zusätzliche Richtlinie erforderlich.

Note

Wenn bestehende Amazon FSx-Dateisysteme die einzigen Dateisysteme sind, die in Ihrem Cluster verwendet werden, können Sie die Amazon FSx-Beispielrichtlinien auf die spezifischen Dateisysteme beschränken, auf die in [SharedStorage Abschnitt](#) der Cluster-Konfigurationsdatei verwiesen wird.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Condition": {
        "StringEquals": {
          "iam:AWSServiceName": [
            "fsx.amazonaws.com",
            "s3.data-source.lustre.fsx.amazonaws.com"
          ]
        }
      },
      "Action": [
        "iam:CreateServiceLinkedRole",
        "iam>DeleteServiceLinkedRole"
      ],
      "Resource": "*",
      "Effect": "Allow"
    },
    {
      "Action": [
        "fsx:*"
      ],
      "Resource": [
        "arn:aws:fsx:*:<AWS ACCOUNT ID>:*"
      ],
      "Effect": "Allow",
    }
  ]
}
```

```

    "Sid": "FSx"
  },
  {
    "Action": [
      "iam:CreateServiceLinkedRole",
      "iam:AttachRolePolicy",
      "iam:PutRolePolicy"
    ],
    "Resource": "arn:aws:iam::<AWS ACCOUNT ID>:role/aws-service-role/s3.data-
source.lustre.fsx.amazonaws.com/*",
    "Effect": "Allow"
  },
  {
    "Action": [
      "s3:Get*",
      "s3:List*",
      "s3:PutObject"
    ],
    "Resource": "arn:aws:s3:::<S3 NAME>",
    "Effect": "Allow"
  }
]
}

```

AWS ParallelCluster Benutzerrichtlinie zur Image-Erstellung **pcluster**

Benutzer, die beabsichtigen, benutzerdefinierte EC2-Images mit zu erstellen, AWS ParallelCluster müssen über die folgenden Berechtigungen verfügen.

```

{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Action": [
        "ec2:DescribeImages",
        "ec2:DescribeInstanceTypeOfferings",
        "ec2:DescribeInstanceTypes",
        "ec2:DeregisterImage",
        "ec2:DeleteSnapshot"
      ],
      "Resource": "*",
      "Effect": "Allow",
      "Sid": "EC2"
    }
  ]
}

```

```
    },
    {
      "Action": [
        "iam:CreateInstanceProfile",
        "iam:AddRoleToInstanceProfile",
        "iam:GetRole",
        "iam:GetRolePolicy",
        "iam:GetInstanceProfile",
        "iam:RemoveRoleFromInstanceProfile"
      ],
      "Resource": [
        "arn:aws:iam::<AWS ACCOUNT ID>:instance-profile/parallelcluster/*",
        "arn:aws:iam::<AWS ACCOUNT ID>:instance-profile/ParallelClusterImage*",
        "arn:aws:iam::<AWS ACCOUNT ID>:role/parallelcluster/*"
      ],
      "Effect": "Allow",
      "Sid": "IAM"
    },
    {
      "Condition": {
        "StringEquals": {
          "iam:PassedToService": [
            "lambda.amazonaws.com",
            "ec2.amazonaws.com"
          ]
        }
      },
      "Action": [
        "iam:PassRole"
      ],
      "Resource": [
        "arn:aws:iam::<AWS ACCOUNT ID>:instance-profile/parallelcluster/*",
        "arn:aws:iam::<AWS ACCOUNT ID>:role/parallelcluster/*"
      ],
      "Effect": "Allow",
      "Sid": "IAMPassRole"
    },
    {
      "Action": [
        "logs:CreateLogGroup",
        "logs:TagResource",
        "logs:UntagResource",
        "logs>DeleteLogGroup"
      ],
```

```

    "Resource": [
      "arn:aws:logs:*:<AWS ACCOUNT ID>:log-group:/aws/imagebuilder/
ParallelClusterImage-*",
      "arn:aws:logs:*:<AWS ACCOUNT ID>:log-group:/aws/lambda/
ParallelClusterImage-*"
    ],
    "Effect": "Allow",
    "Sid": "CloudWatch"
  },
  {
    "Action": [
      "cloudformation:DescribeStacks",
      "cloudformation:CreateStack",
      "cloudformation>DeleteStack"
    ],
    "Resource": [
      "arn:aws:cloudformation:*:<AWS ACCOUNT ID>:stack/*"
    ],
    "Effect": "Allow",
    "Sid": "CloudFormation"
  },
  {
    "Action": [
      "lambda:CreateFunction",
      "lambda:GetFunction",
      "lambda:AddPermission",
      "lambda:RemovePermission",
      "lambda>DeleteFunction",
      "lambda:TagResource",
      "lambda:ListTags",
      "lambda:UntagResource"
    ],
    "Resource": [
      "arn:aws:lambda:*:<AWS ACCOUNT ID>:function:ParallelClusterImage-*"
    ],
    "Effect": "Allow",
    "Sid": "Lambda"
  },
  {
    "Action": [
      "imagebuilder:Get*"
    ],
    "Resource": "*",
    "Effect": "Allow",

```

```

    "Sid": "ImageBuilderGet"
  },
  {
    "Action": [
      "imagebuilder:CreateImage",
      "imagebuilder:TagResource",
      "imagebuilder:CreateImageRecipe",
      "imagebuilder:CreateComponent",
      "imagebuilder:CreateDistributionConfiguration",
      "imagebuilder:CreateInfrastructureConfiguration",
      "imagebuilder>DeleteImage",
      "imagebuilder>DeleteComponent",
      "imagebuilder>DeleteImageRecipe",
      "imagebuilder>DeleteInfrastructureConfiguration",
      "imagebuilder>DeleteDistributionConfiguration"
    ],
    "Resource": [
      "arn:aws:imagebuilder:*:<AWS ACCOUNT ID>:image/parallelclusterimage-*",
      "arn:aws:imagebuilder:*:<AWS ACCOUNT ID>:image-recipe/
parallelclusterimage-*",
      "arn:aws:imagebuilder:*:<AWS ACCOUNT ID>:component/
parallelclusterimage-*",
      "arn:aws:imagebuilder:*:<AWS ACCOUNT ID>:distribution-configuration/
parallelclusterimage-*",
      "arn:aws:imagebuilder:*:<AWS ACCOUNT ID>:infrastructure-configuration/
parallelclusterimage-*"
    ],
    "Effect": "Allow",
    "Sid": "ImageBuilder"
  },
  {
    "Action": [
      "s3:CreateBucket",
      "s3:ListBucket",
      "s3:ListBucketVersions"
    ],
    "Resource": [
      "arn:aws:s3:::parallelcluster-*"
    ],
    "Effect": "Allow",
    "Sid": "S3Bucket"
  },
  {
    "Action": [

```

```

        "sns:GetTopicAttributes",
        "sns:TagResource",
        "sns:CreateTopic",
        "sns:Subscribe",
        "sns:Publish",
        "SNS:DeleteTopic",
        "SNS:Unsubscribe"
    ],
    "Resource": [
        "arn:aws:sns:*:<AWS ACCOUNT ID>:ParallelClusterImage-*"
    ],
    "Effect": "Allow",
    "Sid": "SNS"
},
{
    "Action": [
        "s3:PutObject",
        "s3:GetObject",
        "s3:GetObjectVersion",
        "s3:DeleteObject",
        "s3:DeleteObjectVersion"
    ],
    "Resource": [
        "arn:aws:s3:::parallelcluster-*/*"
    ],
    "Effect": "Allow",
    "Sid": "S3Objects"
},
{
    "Action": "iam:CreateServiceLinkedRole",
    "Effect": "Allow",
    "Resource": "arn:aws:iam::*:role/aws-service-role/
imagebuilder.amazonaws.com/AWSServiceRoleForImageBuilder",
    "Condition": {
        "StringLike": {
            "iam:AWSServiceName": "imagebuilder.amazonaws.com"
        }
    }
}
]
}

```

AWS ParallelCluster Benutzerbeispielrichtlinien für die Verwaltung von IAM-Ressourcen

Bei der Erstellung von AWS ParallelCluster Clustern oder benutzerdefinierten AMIs müssen IAM-Richtlinien bereitgestellt werden, die Berechtigungen enthalten, um Komponenten die erforderlichen Berechtigungen zu AWS ParallelCluster gewähren. Diese IAM-Ressourcen können entweder automatisch von einem Cluster oder einem benutzerdefinierten Image erstellt werden AWS ParallelCluster oder als Eingabe bei der Erstellung eines Clusters oder eines benutzerdefinierten Images bereitgestellt werden.

Sie können die folgenden Modi verwenden, um dem AWS ParallelCluster Benutzer die für den Zugriff auf IAM-Ressourcen erforderlichen Berechtigungen zu gewähren, indem Sie zusätzliche IAM-Richtlinien in der Konfiguration verwenden.

Themen

- [Privilegierter IAM-Zugriffsmodus](#)
- [Eingeschränkter IAM-Zugriffsmodus](#)
- [Modus PermissionsBoundary](#)

Privilegierter IAM-Zugriffsmodus

In diesem Modus AWS ParallelCluster werden automatisch alle erforderlichen IAM-Ressourcen erstellt. Diese IAM-Richtlinien sind so begrenzt, dass sie nur den Zugriff auf Clusterressourcen ermöglichen.

Um den privilegierten IAM-Zugriffsmodus zu aktivieren, fügen Sie der Benutzerrolle die folgende Richtlinie hinzu.

Note

Wenn Sie die [AdditionalPolicies](#)Parameter [HeadNode/Iam/AdditionalPolicies](#) oder [Scheduling/SlurmQueues/Iam/](#) konfigurieren, müssen Sie dem AWS ParallelCluster Benutzer die Erlaubnis erteilen, Rollenrichtlinien für jede weitere Richtlinie anzuhängen und zu trennen, wie in der folgenden Richtlinie dargestellt. Fügen Sie die zusätzlichen Richtlinien-ARNs zur Bedingung für das Anhängen und Trennen von Rollenrichtlinien hinzu.

⚠ Warning

In diesem Modus erhält der Benutzer IAM-Administratorrechte in AWS-Konto

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Action": [
        "iam:CreateServiceLinkedRole",
        "iam:DeleteRole",
        "iam:TagRole"
      ],
      "Resource": [
        "arn:aws:iam::<AWS ACCOUNT ID>:role/parallelcluster/*"
      ],
      "Effect": "Allow",
      "Sid": "IamRole"
    },
    {
      "Action": [
        "iam:CreateRole"
      ],
      "Resource": [
        "arn:aws:iam::<AWS ACCOUNT ID>:role/parallelcluster/*"
      ],
      "Effect": "Allow",
      "Sid": "IamCreateRole"
    },
    {
      "Action": [
        "iam:PutRolePolicy",
        "iam>DeleteRolePolicy"
      ],
      "Resource": "arn:aws:iam::<AWS ACCOUNT ID>:role/parallelcluster/*",
      "Effect": "Allow",
      "Sid": "IamInlinePolicy"
    },
    {
      "Condition": {
        "ArnLike": {
```



```

    "iam:PolicyARN": [
      "arn:aws:iam::<AWS ACCOUNT ID>:policy/parallelcluster*",
      "arn:aws:iam::<AWS ACCOUNT ID>:policy/parallelcluster/*",
      "arn:aws:iam::aws:policy/CloudWatchAgentServerPolicy",
      "arn:aws:iam::aws:policy/AmazonSSMManagedInstanceCore",
      "arn:aws:iam::aws:policy/AWSBatchFullAccess",
      "arn:aws:iam::aws:policy/AmazonS3ReadOnlyAccess",
      "arn:aws:iam::aws:policy/service-role/AWSBatchServiceRole",
      "arn:aws:iam::aws:policy/service-role/
AmazonEC2ContainerServiceforEC2Role",
      "arn:aws:iam::aws:policy/service-role/
AmazonECSTaskExecutionRolePolicy",
      "arn:aws:iam::aws:policy/service-role/
AmazonEC2SpotFleetTaggingRole",
      "arn:aws:iam::aws:policy/EC2InstanceProfileForImageBuilder",
      "arn:aws:iam::aws:policy/service-role/
AWSLambdaBasicExecutionRole"
    ]
  },
  "Action": [
    "iam:AttachRolePolicy",
    "iam:DetachRolePolicy"
  ],
  "Resource": "arn:aws:iam::<AWS ACCOUNT ID>:role/parallelcluster/*",
  "Effect": "Allow",
  "Sid": "IamPolicy"
}
]
}

```

Eingeschränkter IAM-Zugriffsmodus

Wenn dem Benutzer keine zusätzlichen IAM-Richtlinien gewährt werden, müssen IAM-Rollen, die für Cluster oder benutzerdefinierte Image-Builds erforderlich sind, manuell von einem Administrator erstellt und als Teil der Clusterkonfiguration übergeben werden.

Bei der Erstellung eines Clusters sind die folgenden Parameter erforderlich:

- [Iam / Roles / LambdaFunctionsRole](#)
- [HeadNode / Iam / InstanceRole | InstanceProfile](#)
- [Scheduling / SlurmQueues / Iam / InstanceRole | InstanceProfile](#)

Beim Erstellen eines benutzerdefinierten Images sind die folgenden Parameter erforderlich:

- [Build](#) / [Iam](#) / [InstanceRole](#) | [InstanceProfile](#)
- [Build](#) / [Iam](#) / [CleanupLambdaRole](#)

Die IAM-Rollen, die als Teil der oben aufgeführten Parameter übergeben wurden, müssen mit dem `/parallelcluster/` Pfadpräfix erstellt werden. Wenn dies nicht möglich ist, muss die Benutzerrichtlinie aktualisiert werden, um `iam:PassRole` Berechtigungen für die spezifischen benutzerdefinierten Rollen zu gewähren, wie im folgenden Beispiel gezeigt.

```
{
  "Condition": {
    "StringEqualsIfExists": {
      "iam:PassedToService": [
        "ecs-tasks.amazonaws.com",
        "lambda.amazonaws.com",
        "ec2.amazonaws.com",
        "spotfleet.amazonaws.com",
        "batch.amazonaws.com",
        "codebuild.amazonaws.com"
      ]
    }
  },
  "Action": [
    "iam:PassRole"
  ],
  "Resource": [
    <list all custom IAM roles>
  ],
  "Effect": "Allow",
  "Sid": "IamPassRole"
}
```

Warning

Derzeit erlaubt dieser Modus nicht die Verwaltung von AWS Batch Clustern, da nicht alle IAM-Rollen in der Clusterkonfiguration übergeben werden können.

Modus **PermissionsBoundary**

Dieser Modus delegiert an AWS ParallelCluster die Erstellung von IAM-Rollen, die an die konfigurierten IAM-Berechtigungsgrößen gebunden sind. Weitere Informationen zu den IAM-Berechtigungsgrößen finden Sie unter [Berechtigungsgrößen für IAM-Entitäten im IAM-Benutzerhandbuch](#).

Die folgende Richtlinie muss der Benutzerrolle hinzugefügt werden.

Ersetzen Sie in der Richtlinie `< permissions-boundary-arn >` durch den ARN der IAM-Richtlinie, der als Berechtigungsgröße durchgesetzt werden soll.

Warning

Wenn Sie die [AdditionalPolicies](#)Parameter [HeadNode//AdditionalPolicies](#) oder [Iam/Scheduling/SlurmQueuesIam](#)/konfigurieren, müssen Sie dem Benutzer die Berechtigung erteilen, Rollenrichtlinien für jede weitere Richtlinie anzuhängen und zu trennen, wie in der folgenden Richtlinie dargestellt. Fügen Sie die zusätzlichen Richtlinien-ARNs zur Bedingung für das Anhängen und Trennen von Rollenrichtlinien hinzu.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Action": [
        "iam:CreateServiceLinkedRole",
        "iam>DeleteRole",
        "iam:TagRole"
      ],
      "Resource": [
        "arn:aws:iam::<AWS ACCOUNT ID>:role/parallelcluster/*"
      ],
      "Effect": "Allow",
      "Sid": "IamRole"
    },
    {
      "Condition": {
        "StringEquals": {
          "iam:PermissionsBoundary": [
            <permissions-boundary-arn>
          ]
        }
      }
    }
  ]
}
```

```

    ]
  }
},
"Action": [
  "iam:CreateRole"
],
"Resource": [
  "arn:aws:iam::<AWS ACCOUNT ID>:role/parallelcluster/*"
],
"Effect": "Allow",
"Sid": "IamCreateRole"
},
{
  "Condition": {
    "StringEquals": {
      "iam:PermissionsBoundary": [
        <permissions-boundary-arn>
      ]
    }
  },
  "Action": [
    "iam:PutRolePolicy",
    "iam>DeleteRolePolicy"
  ],
  "Resource": "arn:aws:iam::<AWS ACCOUNT ID>:role/parallelcluster/*",
  "Effect": "Allow",
  "Sid": "IamInlinePolicy"
},
{
  "Condition": {
    "StringEquals": {
      "iam:PermissionsBoundary": [
        <permissions-boundary-arn>
      ]
    },
    "ArnLike": {
      "iam:PolicyARN": [
        "arn:aws:iam::<AWS ACCOUNT ID>:policy/parallelcluster*",
        "arn:aws:iam::<AWS ACCOUNT ID>:policy/parallelcluster/*",
        "arn:aws:iam::aws:policy/CloudWatchAgentServerPolicy",
        "arn:aws:iam::aws:policy/AmazonSSMManagedInstanceCore",
        "arn:aws:iam::aws:policy/AWSBatchFullAccess",
        "arn:aws:iam::aws:policy/AmazonS3ReadOnlyAccess",
        "arn:aws:iam::aws:policy/service-role/AWSBatchServiceRole",

```

```

        "arn:aws:iam::aws:policy/service-role/
AmazonEC2ContainerServiceforEC2Role",
        "arn:aws:iam::aws:policy/service-role/
AmazonECSTaskExecutionRolePolicy",
        "arn:aws:iam::aws:policy/service-role/
AmazonEC2SpotFleetTaggingRole",
        "arn:aws:iam::aws:policy/EC2InstanceProfileForImageBuilder",
        "arn:aws:iam::aws:policy/service-role/
AWSLambdaBasicExecutionRole"
    ]
  },
  "Action": [
    "iam:AttachRolePolicy",
    "iam:DetachRolePolicy"
  ],
  "Resource": "arn:aws:iam::<AWS ACCOUNT ID>:role/parallelcluster/*",
  "Effect": "Allow",
  "Sid": "IamPolicy"
}
]
}

```

Wenn dieser Modus aktiviert ist, müssen Sie den ARN für die Berechtigungsgrenze im [PermissionsBoundary](#) Konfigurationsparameter [Iam/](#) angeben, wenn Sie einen Cluster erstellen oder aktualisieren, und im [PermissionBoundary](#) Parameter [Build/Iam/](#), wenn Sie ein benutzerdefiniertes Image erstellen.

AWS ParallelCluster Konfigurationsparameter zur Verwaltung von IAM-Berechtigungen

AWS ParallelCluster stellt eine Reihe von Konfigurationsoptionen zur Anpassung und Verwaltung der IAM-Berechtigungen und -Rollen bereit, die in einem Cluster oder bei der Erstellung benutzerdefinierter AMIs verwendet werden.

Themen

- [Cluster-Konfiguration](#)
- [Benutzerdefinierte Image-Konfiguration](#)

Cluster-Konfiguration

Themen

- [IAM-Rolle für den Hauptknoten](#)
- [Amazon S3 S3-Zugriff](#)
- [Zusätzliche IAM-Richtlinien](#)
- [AWS Lambda Funktionen, Rolle](#)
- [IAM-Rolle für Rechenknoten](#)
- [Berechtigungsgrenze](#)

IAM-Rolle für den Hauptknoten

[HeadNode](#) / [Iam](#) / [InstanceRole](#) | [InstanceProfile](#)

Mit dieser Option überschreiben Sie die Standard-IAM-Rolle, die dem Hauptknoten des Clusters zugewiesen ist. Weitere Informationen finden Sie in der [InstanceProfile](#)Referenz.

Im Folgenden finden Sie die Mindestanzahl an Richtlinien, die im Rahmen dieser Rolle verwendet werden müssen, wenn der Scheduler Slurm ist:

- `arn:aws:iam::aws:policy/CloudWatchAgentServerPolicy` verwaltete IAM-Richtlinie. Weitere Informationen finden Sie im [CloudWatch Amazon-Benutzerhandbuch](#) unter [Erstellen von IAM-Rollen und -Benutzern zur Verwendung mit dem CloudWatch Agenten](#).
- `arn:aws:iam::aws:policy/AmazonSSMManagedInstanceCore` verwaltete IAM-Richtlinie. Weitere Informationen finden Sie [AWS Systems Manager](#) im [AWS Systems Manager Benutzerhandbuch](#) unter [AWS Verwaltete Richtlinien für](#).
- Zusätzliche IAM-Richtlinie:

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Action": [
        "s3:GetObject",
        "s3:GetObjectVersion"
      ],
      "Resource": [
        "arn:aws:s3:::<REGION>-aws-parallelcluster/*",
        "arn:aws:s3:::dcv-license.<REGION>/*",
      ]
    }
  ]
}
```

```

        "arn:aws:s3:::parallelcluster-*~v1-do-not-delete/*"
    ],
    "Effect": "Allow"
},
{
    "Action": [
        "dynamodb:GetItem",
        "dynamodb:PutItem",
        "dynamodb:UpdateItem",
        "dynamodb:BatchWriteItem",
        "dynamodb:BatchGetItem"
    ],
    "Resource": "arn:aws:dynamodb:<REGION>:<AWS ACCOUNT ID>:table/
parallelcluster-*",
    "Effect": "Allow"
},
{
    "Condition": {
        "StringEquals": {
            "ec2:ResourceTag/parallelcluster:node-type": "Compute"
        }
    },
    "Action": "ec2:TerminateInstances",
    "Resource": "*",
    "Effect": "Allow"
},
{
    "Action": [
        "ec2:RunInstances",
        "ec2:CreateFleet"
    ]
    "Resource": "*",
    "Effect": "Allow"
},
{
    "Condition": {
        "StringEquals": {
            "iam:PassedToService": [
                "ec2.amazonaws.com"
            ]
        }
    },
    "Action": [
        "iam:PassRole"
    ]
}

```

```

    ],
    "Resource": [
        "arn:aws:iam::<AWS ACCOUNT ID>:role/parallelcluster/*",
        "arn:aws:iam::<AWS ACCOUNT ID>:instance-profile/parallelcluster/*"
    ],
    "Effect": "Allow"
},
{
    "Action": [
        "ec2:DescribeInstances",
        "ec2:DescribeInstanceStatus",
        "ec2:DescribeVolumes",
        "ec2:DescribeInstanceAttribute",
        "ec2:DescribeCapacityReservations"
    ],
    "Resource": "*",
    "Effect": "Allow"
},
{
    "Action": [
        "ec2:CreateTags",
        "ec2:AttachVolume"
    ],
    "Resource": [
        "arn:aws:ec2:<REGION>:<AWS ACCOUNT ID>:instance/*",
        "arn:aws:ec2:<REGION>:<AWS ACCOUNT ID>:volume/*"
    ],
    "Effect": "Allow"
},
{
    "Action": [
        "cloudformation:DescribeStacks",
        "cloudformation:DescribeStackResource",
        "cloudformation:SignalResource"
    ],
    "Resource": "*",
    "Effect": "Allow"
},
{
    "Action": [
        "route53:ChangeResourceRecordSets"
    ],
    "Resource": "*",
    "Effect": "Allow"
}

```



```

    },
    {
      "Action": "secretsmanager:GetSecretValue",
      "Resource": "arn:aws:secretsmanager:<REGION>:<AWS ACCOUNT
ID>:secret:<SECRET_ID>",
      "Effect": "Allow"
    }
  ]
}

```

Beachten Sie, dass, falls [SchedulingSlurmQueues//Iam](#) verwendet [InstanceRole](#) wird, um die Compute-IAM-Rolle außer Kraft zu setzen, die oben angegebene Headnode-Richtlinie diese Rolle in den Resource Abschnitt der `iam:PassRole` Berechtigung aufnehmen muss.

Im Folgenden finden Sie die Mindestanzahl an Richtlinien, die als Teil dieser Rolle verwendet werden können, wenn der Scheduler: AWS Batch

- `arn:aws:iam::aws:policy/CloudWatchAgentServerPolicy` verwaltete IAM-Richtlinie. Weitere Informationen finden Sie im [CloudWatch Amazon-Benutzerhandbuch](#) unter [Erstellen von IAM-Rollen und -Benutzern zur Verwendung mit dem CloudWatch Agenten](#).
- `arn:aws:iam::aws:policy/AmazonSSMManagedInstanceCore` verwaltete IAM-Richtlinie. Weitere Informationen finden Sie [AWS Systems Manager](#) im [AWS Systems Manager Benutzerhandbuch](#) unter [AWS Verwaltete Richtlinien für](#).
- Zusätzliche IAM-Richtlinie:

```

{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Action": [
        "s3:GetObject",
        "s3:PutObject",
        "s3:GetObjectVersion"
      ],
      "Resource": [
        "arn:aws:s3:::parallelcluster-*-*v1-do-not-delete/*"
      ],
      "Effect": "Allow"
    },
    {
      "Action": "s3:GetObject",

```

```

    "Resource": [
      "arn:aws:s3:::dcv-license.<REGION>/*",
      "arn:aws:s3:::<REGION>-aws-parallelcluster/*"
    ],
    "Effect": "Allow"
  },
  {
    "Condition": {
      "StringEquals": {
        "iam:PassedToService": [
          "batch.amazonaws.com"
        ]
      }
    },
    "Action": [
      "iam:PassRole"
    ],
    "Resource": [
      "arn:aws:iam::<AWS ACCOUNT ID>:role/parallelcluster/*",
      "arn:aws:iam::<AWS ACCOUNT ID>:instance-profile/parallelcluster/*"
    ],
    "Effect": "Allow"
  },
  {
    "Action": [
      "batch:DescribeJobQueues",
      "batch:DescribeJobs",
      "batch:ListJobs",
      "batch:DescribeComputeEnvironments"
    ],
    "Resource": "*",
    "Effect": "Allow"
  },
  {
    "Action": [
      "batch:SubmitJob",
      "batch:TerminateJob",
      "logs:GetLogEvents",
      "ecs:ListContainerInstances",
      "ecs:DescribeContainerInstances"
    ],
    "Resource": [
      "arn:aws:logs:<REGION>:<AWS ACCOUNT ID>:log-group:/aws/batch/job:log-
stream:PclusterJobDefinition*"

```

```

        "arn:aws:ecs:<REGION>:<AWS ACCOUNT ID>:container-instance/AWSBatch-
PclusterComputeEnviron*",
        "arn:aws:ecs:<REGION>:<AWS ACCOUNT ID>:cluster/AWSBatch-Pcluster*",
        "arn:aws:batch:<REGION>:<AWS ACCOUNT ID>:job-queue/
PclusterJobQueue*",
        "arn:aws:batch:<REGION>:<AWS ACCOUNT ID>:job-definition/
PclusterJobDefinition*:*",
        "arn:aws:batch:<REGION>:<AWS ACCOUNT ID>:job/*"
    ],
    "Effect": "Allow"
},
{
    "Action": [
        "ec2:DescribeInstances",
        "ec2:DescribeInstanceStatus",
        "ec2:DescribeVolumes",
        "ec2:DescribeInstanceAttribute"
    ],
    "Resource": "*",
    "Effect": "Allow"
},
{
    "Action": [
        "ec2:CreateTags",
        "ec2:AttachVolume"
    ],
    "Resource": [
        "arn:aws:ec2:<REGION>:<AWS ACCOUNT ID>:instance/*",
        "arn:aws:ec2:<REGION>:<AWS ACCOUNT ID>:volume/*"
    ],
    "Effect": "Allow"
},
{
    "Action": [
        "cloudformation:DescribeStackResource",
        "cloudformation:DescribeStacks",
        "cloudformation:SignalResource"
    ],
    "Resource": "*",
    "Effect": "Allow"
},
{
    "Action": "secretsmanager:GetSecretValue",

```

```
    "Resource": "arn:aws:secretsmanager:<REGION>:<AWS ACCOUNT ID>:secret:<SECRET_ID>",
    "Effect": "Allow"
  }
]
```

Amazon S3 S3-Zugriff

[HeadNode/Iam/S3Access](#) oder [Scheduling/SlurmQueues/S3Access](#)

In diesen Konfigurationsabschnitten können Sie den Amazon S3 S3-Zugriff anpassen, indem Sie den IAM-Rollen, die dem Hauptknoten oder den Rechenknoten des Clusters zugeordnet sind, zusätzliche Amazon S3 S3-Richtlinien zuweisen, wenn solche Rollen von AWS ParallelCluster erstellt werden. Weitere Informationen finden Sie in der Referenzdokumentation zu den einzelnen Konfigurationsparametern.

Dieser Parameter kann nur verwendet werden, wenn der Benutzer mit [Privilegierter IAM-Zugriffsmodus](#) oder konfiguriert ist [Modus PermissionsBoundary](#).

Zusätzliche IAM-Richtlinien

[HeadNode/Iam/AdditionalIamPolicies](#) oder [SlurmQueues//IamAdditionalIamPolicies](#)

Verwenden Sie diese Option, um zusätzliche verwaltete IAM-Richtlinien an die IAM-Rollen anzuhängen, die dem Hauptknoten oder den Rechenknoten des Clusters zugeordnet sind, wenn solche Rollen von erstellt werden. AWS ParallelCluster

Warning

Um diese Option verwenden zu können, stellen Sie sicher, dass dem [AWS ParallelCluster Benutzer](#) die `iam:DetachRolePolicy` erforderlichen Berechtigungen für die IAM-Richtlinien erteilt wurden `iam:AttachRolePolicy`.

AWS Lambda Funktionen, Rolle

[Iam / Roles / LambdaFunctionsRole](#)

Diese Option setzt die Rolle außer Kraft, die allen AWS Lambda Funktionen zugewiesen ist, die während der Clustererstellung verwendet werden. AWS Lambda muss so konfiguriert werden, dass der Principal die Rolle übernehmen darf.

Note

Falls [DeploymentSettings](#)/gesetzt [LambdaFunctionsVpcConfigist](#), [LambdaFunctionsRole](#) muss er die [AWS Lambda Rollenberechtigung zum Einstellen](#) der VPC-Konfiguration beinhalten.

Im Folgenden finden Sie die Mindestanzahl an Richtlinien, die als Teil dieser Rolle verwendet werden können:

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Action": [
        "route53:ListResourceRecordSets",
        "route53:ChangeResourceRecordSets"
      ],
      "Resource": "arn:aws:route53::hostedzone/*",
      "Effect": "Allow"
    },
    {
      "Action": ["logs:CreateLogStream", "logs:PutLogEvents"],
      "Effect": "Allow",
      "Resource": "arn:aws:logs:<REGION>:<AWS ACCOUNT ID>:log-group:/aws/lambda/
pcluster-*"
    },
    {
      "Action": "ec2:DescribeInstances",
      "Effect": "Allow",
      "Resource": "*"
    },
    {
      "Action": "ec2:TerminateInstances",
      "Condition": {
        "StringEquals": {
          "ec2:ResourceTag/parallelcluster:node-type": "Compute"
        }
      }
    }
  ]
}
```

```

    },
    "Effect": "Allow",
    "Resource": "*"
  },
  {
    "Action": [
      "s3:DeleteObject",
      "s3:DeleteObjectVersion",
      "s3:ListBucket",
      "s3:ListBucketVersions"
    ],
    "Effect": "Allow",
    "Resource": [
      "arn:aws:s3:::parallelcluster-*-*v1-do-not-delete",
      "arn:aws:s3:::parallelcluster-*-*v1-do-not-delete/*"
    ]
  }
]
}

```

IAM-Rolle für Rechenknoten

[Scheduling](#) / [SlurmQueues](#) / [Iam](#) / [InstanceRole](#) | [InstanceProfile](#)

Diese Option ermöglicht es, die IAM-Rolle zu überschreiben, die den Rechenknoten des Clusters zugewiesen ist. Weitere Informationen finden Sie unter [InstanceProfile](#).

Im Folgenden finden Sie die Mindestanzahl an Richtlinien, die als Teil dieser Rolle verwendet werden können:

- `arn:aws:iam::aws:policy/CloudWatchAgentServerPolicy` verwaltete IAM-Richtlinie. Weitere Informationen finden Sie im [CloudWatch Amazon-Benutzerhandbuch](#) unter [Erstellen von IAM-Rollen und -Benutzern zur Verwendung mit dem CloudWatch Agenten](#).
- `arn:aws:iam::aws:policy/AmazonSSMManagedInstanceCore` verwaltete IAM-Richtlinie. Weitere Informationen finden Sie [AWS Systems Manager im AWS Systems Manager Benutzerhandbuch](#) unter [AWS Verwaltete Richtlinien für](#).
- Zusätzliche IAM-Richtlinie:

```

{
  "Version": "2012-10-17",
  "Statement": [

```

```

    {
      "Action": [
        "dynamodb:Query",
        "dynamodb:UpdateItem",
        "dynamodb:PutItem",
        "dynamodb:GetItem"
      ],
      "Resource": "arn:aws:dynamodb:<REGION>:<AWS ACCOUNT ID>:table/
parallelcluster-*",
      "Effect": "Allow"
    },
    {
      "Action": "s3:GetObject",
      "Resource": [
        "arn:aws:s3:::<REGION>-aws-parallelcluster/*"
      ],
      "Effect": "Allow"
    },
    {
      "Action": "ec2:DescribeInstanceAttribute",
      "Resource": "*",
      "Effect": "Allow"
    },
    {
      "Action": "cloudformation:DescribeStackResource",
      "Resource":
        [ "arn:aws:cloudformation:<REGION>:<AWS ACCOUNT ID>:stack/*/*" ],
      "Effect": "Allow"
    }
  ]
}

```

Berechtigungsgrenze

[Iam / PermissionsBoundary](#)

Dieser Parameter erzwingt AWS ParallelCluster, die angegebene IAM-Richtlinie allen IAM-Rollen `PermissionsBoundary` zuzuweisen, die im Rahmen einer Clusterbereitstellung erstellt wurden.

Eine Liste der Richtlinien, die der Benutzer benötigt, wenn diese Einstellung definiert ist, finden Sie unter [Modus PermissionsBoundary](#).

Benutzerdefinierte Image-Konfiguration

Themen

- [Instanzrolle für EC2 Image Builder](#)
- [AWS Lambda Rolle beim Aufräumen](#)
- [Zusätzliche IAM-Richtlinien](#)
- [Berechtigungsgrenze](#)

Instanzrolle für EC2 Image Builder

[Build](#) / [Iam](#) / [InstanceRole](#) | [InstanceProfile](#)

Mit dieser Option überschreiben Sie die IAM-Rolle, die der von EC2 Image Builder gestarteten EC2-Instance zugewiesen ist, um ein benutzerdefiniertes AMI zu erstellen.

Im Folgenden finden Sie die Mindestanzahl an Richtlinien, die als Teil dieser Rolle verwendet werden können:

- `arn:aws:iam::aws:policy/AmazonSSMManagedInstanceCore` verwaltete IAM-Richtlinie. Weitere Informationen finden Sie AWS Systems Manager im AWS Systems Manager Benutzerhandbuch unter [AWS Verwaltete Richtlinien für](#).
- `arn:aws:iam::aws:policy/EC2InstanceProfileForImageBuilder` verwaltete IAM-Richtlinie. Weitere Informationen finden Sie unter [EC2InstanceProfileForImageBuilderRichtlinie](#) im Image Builder Benutzerhandbuch.
- Zusätzliche IAM-Richtlinie:

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Action": [
        "ec2:CreateTags",
        "ec2:ModifyImageAttribute"
      ],
      "Resource": "arn:aws:ec2:<REGION>::image/*",
      "Effect": "Allow"
    }
  ]
}
```



```
}
```

AWS Lambda Rolle beim Aufräumen

[Build](#) / [Iam](#) / [CleanupLambdaRole](#)

Diese Option setzt die Rolle außer Kraft, die allen AWS Lambda Funktionen zugewiesen ist, die während der Erstellung eines benutzerdefinierten Images verwendet werden. AWS Lambda muss so konfiguriert werden, dass der Principal die Rolle übernehmen darf.

Note

Falls [DeploymentSettings](#)/gesetzt [LambdaFunctionsVpcConfigist](#), CleanupLambdaRole muss er die [AWS Lambda Rollenberechtigung zum Einstellen](#) der VPC-Konfiguration beinhalten.

Im Folgenden finden Sie die Mindestanzahl an Richtlinien, die als Teil dieser Rolle verwendet werden können:

- `arn:aws:iam::aws:policy/service-role/AWSLambdaBasicExecutionRole` verwaltete IAM-Richtlinie. Weitere Informationen finden Sie im AWS Lambda Developer Guide unter [AWS Verwaltete Richtlinien für Lambda-Funktionen](#).
- Zusätzliche IAM-Richtlinie:

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Action": [
        "iam:DetachRolePolicy",
        "iam>DeleteRole",
        "iam>DeleteRolePolicy"
      ],
      "Resource": "arn:aws:iam::<AWS ACCOUNT ID>:role/parallelcluster/*",
      "Effect": "Allow"
    },
    {
      "Action": [
        "iam>DeleteInstanceProfile",
```

```

        "iam:RemoveRoleFromInstanceProfile"
    ],
    "Resource": "arn:aws:iam::<AWS ACCOUNT ID>:instance-profile/
parallelcluster/*",
    "Effect": "Allow"
},
{
    "Action": "imagebuilder:DeleteInfrastructureConfiguration",
    "Resource": "arn:aws:imagebuilder:<REGION>:<AWS ACCOUNT
ID>:infrastructure-configuration/parallelclusterimage-*",
    "Effect": "Allow"
},
{
    "Action": [
        "imagebuilder:DeleteComponent"
    ],
    "Resource": [
        "arn:aws:imagebuilder:<REGION>:<AWS ACCOUNT ID>:component/
parallelclusterimage-*/*"
    ],
    "Effect": "Allow"
},
{
    "Action": "imagebuilder:DeleteImageRecipe",
    "Resource": "arn:aws:imagebuilder:<REGION>:<AWS ACCOUNT ID>:image-recipe/
parallelclusterimage-*/*",
    "Effect": "Allow"
},
{
    "Action": "imagebuilder:DeleteDistributionConfiguration",
    "Resource": "arn:aws:imagebuilder:<REGION>:<AWS ACCOUNT ID>:distribution-
configuration/parallelclusterimage-*",
    "Effect": "Allow"
},
{
    "Action": [
        "imagebuilder:DeleteImage",
        "imagebuilder:GetImage",
        "imagebuilder:CancelImageCreation"
    ],
    "Resource": "arn:aws:imagebuilder:<REGION>:<AWS ACCOUNT ID>:image/
parallelclusterimage-*/*",
    "Effect": "Allow"
},
},

```

```

    {
      "Action": "cloudformation:DeleteStack",
      "Resource": "arn:aws:cloudformation:<REGION>:<AWS ACCOUNT ID>:stack/*/*",
      "Effect": "Allow"
    },
    {
      "Action": "ec2:CreateTags",
      "Resource": "arn:aws:ec2:<REGION>::image/*",
      "Effect": "Allow"
    },
    {
      "Action": "tag:TagResources",
      "Resource": "*",
      "Effect": "Allow"
    },
    {
      "Action": [
        "lambda:DeleteFunction",
        "lambda:RemovePermission"
      ],
      "Resource": "arn:aws:lambda:<REGION>:<AWS ACCOUNT ID>:function:ParallelClusterImage-*",
      "Effect": "Allow"
    },
    {
      "Action": "logs:DeleteLogGroup",
      "Resource": "arn:aws:logs:<REGION>:<AWS ACCOUNT ID>:log-group:/aws/lambda/ParallelClusterImage-*:*",
      "Effect": "Allow"
    },
    {
      "Action": [
        "SNS:GetTopicAttributes",
        "SNS:DeleteTopic",
        "SNS:GetSubscriptionAttributes",
        "SNS:Unsubscribe"
      ],
      "Resource": "arn:aws:sns:<REGION>:<AWS ACCOUNT ID>:ParallelClusterImage-*",
      "Effect": "Allow"
    }
  ]
}

```

Zusätzliche IAM-Richtlinien

[Build](#) / [Iam](#) / [AdditionalIamPolicies](#)

Sie verwenden diese Option, um zusätzliche verwaltete IAM-Richtlinien an die Rolle anzuhängen, die der EC2-Instance zugeordnet ist, die von EC2 Image Builder zur Erstellung des benutzerdefinierten AMI verwendet wird.

Warning

Um diese Option zu verwenden, stellen Sie sicher, dass dem [AWS ParallelCluster Benutzer](#) `iam:DetachRolePolicy` Berechtigungen für die IAM-Richtlinien erteilt wurden und `iam:AttachRolePolicy`, die angehängt werden müssen.

Berechtigungsgrenze

[Build](#) / [Iam](#) / [PermissionsBoundary](#)

Dieser Parameter erzwingt AWS ParallelCluster, die angegebene IAM-Richtlinie als an alle IAM-Rollen anzuhängen, die im Rahmen eines benutzerdefinierten AMI-Builds erstellt wurden. `PermissionsBoundary`

Eine Liste der Richtlinien, die [Modus PermissionsBoundary](#) für die Verwendung dieser Funktionen erforderlich sind, finden Sie unter.

Netzwerkkonfigurationen

AWS ParallelCluster verwendet Amazon Virtual Private Cloud (VPC) für Netzwerke. VPC bietet eine flexible und konfigurierbare Netzwerkplattform, auf der Sie Cluster bereitstellen können.

Die VPC muss über `DNS Resolution = yes-`, `DNS Hostnames = yes-` und DHCP-Optionen mit dem richtigen Domännennamen für die Region verfügen. Der standardmäßige DHCP-Optionsatz gibt bereits den erforderlichen `AmazonProvidedDNS` an. Wenn Sie mehr als einen Domainnamenserver angeben, lesen Sie die [DHCP-Optionssätze](#) im Amazon VPC-Benutzerhandbuch.

AWS ParallelCluster unterstützt die folgenden allgemeinen Konfigurationen:

- Ein Subnetz für Kopf- und Rechenknoten.

- Zwei Subnetze, mit dem Hauptknoten in einem öffentlichen Subnetz und Rechenknoten in einem privaten Subnetz. Bei den Subnetzen kann es sich entweder um neue oder um bestehende Subnetze handeln.

Alle diese Konfigurationen können mit oder ohne öffentliche IP-Adressierung betrieben werden. AWS ParallelCluster kann auch eingesetzt werden, um einen HTTP-Proxy für alle AWS Anfragen zu verwenden. Die Kombinationen dieser Konfigurationen bedeuten, dass viele Bereitstellungsszenarien möglich sind. Sie können beispielsweise ein einzelnes öffentliches Subnetz mit vollständigem Zugriff über das Internet konfigurieren. Oder Sie können ein vollständig privates Netzwerk mit AWS Direct Connect einem HTTP-Proxy für den gesamten Datenverkehr konfigurieren.

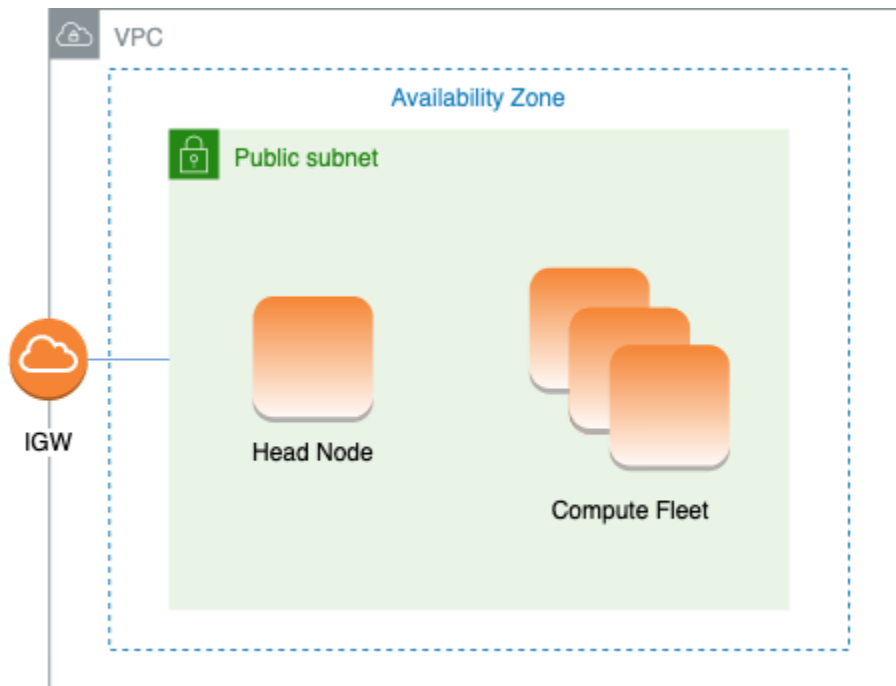
Ab AWS ParallelCluster 3.0.0 ist es möglich `SecurityGroups`, `AdditionalSecurityGroups` für jede Warteschlange unterschiedliche `PlacementGroup` Einstellungen zu konfigurieren. Weitere Informationen finden Sie unter [HeadNode/Networking](#) und [SlurmQueues/Networking](#) und [AwsBatchQueues/Networking](#).

Illustrationen einiger Netzwerkszenarien finden Sie in den folgenden Architekturdiagrammen.

Themen

- [AWS ParallelCluster in einem einzelnen öffentlichen Subnetz](#)
- [AWS ParallelCluster mit zwei Subnetzen](#)
- [AWS ParallelCluster in einem einzigen privaten Subnetz mit Verbindung über AWS Direct Connect](#)
- [AWS ParallelCluster mit AWS Batch Scheduler](#)
- [AWS ParallelCluster in einem einzelnen Subnetz ohne Internetzugang ohne Internetzugang ohne Internetzugang ohne Internetzugang.](#)

AWS ParallelCluster in einem einzelnen öffentlichen Subnetz



Die Konfiguration für diese Architektur erfordert die folgenden Einstellungen:

```
# Note that all values are only provided as examples
HeadNode:
  ...
  Networking:
    SubnetId: subnet-12345678 # subnet with internet gateway
    #ElasticIp: true | false | eip-12345678
Scheduling:
  Scheduler: slurm
  SlurmQueues:
    - ...
      Networking:
        SubnetIds:
          - subnet-12345678 # subnet with internet gateway
        #AssignPublicIp: true
```

In dieser Konfiguration muss allen Instanzen des Clusters eine öffentliche IP zugewiesen werden, um Internetzugang zu erhalten. Um dies zu erreichen, gehen Sie wie folgt vor:

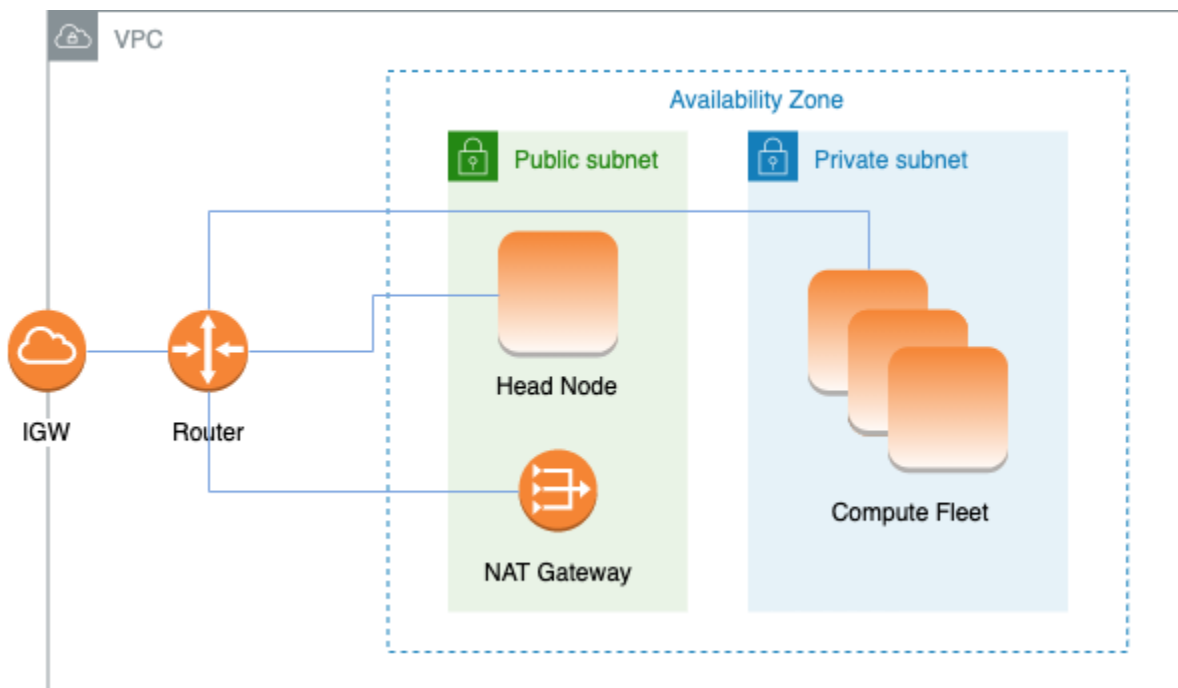
- Stellen Sie sicher, dass dem Hauptknoten eine öffentliche IP-Adresse zugewiesen wird, indem Sie entweder die Einstellung „Automatische Zuweisung öffentlicher IPv4-Adresse aktivieren“

für das in [HeadNode/Networking](#)/verwendete Subnetz aktivieren [SubnetId](#) oder indem Sie in [HeadNodeNetworking](#) eine Elastic IP zuweisen [ElasticIp](#).

- Stellen Sie sicher, dass den Rechenknoten eine öffentliche IP-Adresse zugewiesen wird, indem Sie entweder die Einstellung „Automatische Zuweisung öffentlicher IPv4-Adresse aktivieren“ für das in [Scheduling//SlurmQueues/Networking](#) verwendete Subnetz aktivieren [SubnetIds](#) oder indem Sie in [Scheduling/SlurmQueues](#) den Wert true setzen [AssignPublicIpNetworking](#).
- Wenn Sie einen `p4d` Instanztyp oder einen anderen Instanztyp definieren, der mehrere Netzwerkschnittstellen oder eine Netzwerkschnittstellenkarte für den Hauptknoten hat, müssen Sie [HeadNode/Networking](#) auf setzen, [ElasticIp](#) `true` um den öffentlichen Zugriff zu ermöglichen. AWS Öffentliche IPs können nur Instanzen zugewiesen werden, die mit einer einzigen Netzwerkschnittstelle gestartet wurden. In diesem Fall empfehlen wir die Verwendung eines [NAT-Gateways](#), um den öffentlichen Zugriff auf die Cluster-Rechenknoten bereitzustellen. Weitere Informationen über IP-Adressen finden Sie unter [Zuweisen einer öffentlichen IPv4-Adresse beim Start einer Instance](#) im Amazon EC2 EC2-Benutzerhandbuch für Linux-Instances.
- Sie können keinen Instanztyp `p4d` oder einen anderen `hp6id` Instanztyp mit mehreren Netzwerkschnittstellen oder eine Netzwerkschnittstellenkarte zur Berechnung von Knoten definieren, da AWS öffentliche IPs nur Instances zugewiesen werden können, die mit einer einzigen Netzwerkschnittstelle gestartet wurden. Weitere Informationen über IP-Adressen finden Sie unter [Zuweisen einer öffentlichen IPv4-Adresse beim Start einer Instance](#) im Amazon EC2 EC2-Benutzerhandbuch für Linux-Instances.

Weitere Informationen finden Sie unter [Aktivieren des Internetzugangs](#) im Amazon VPC Benutzerhandbuch.

AWS ParallelCluster mit zwei Subnetzen



Die Konfiguration zur Verwendung eines vorhandenen privaten Subnetzes für Compute-Instances erfordert die folgenden Einstellungen:

```
# Note that all values are only provided as examples
HeadNode:
  ...
  Networking:
    SubnetId: subnet-12345678 # subnet with internet gateway
    #ElasticIp: true | false | eip-12345678
Scheduling:
  Scheduler: slurm
  SlurmQueues:
    - ...
      Networking:
        SubnetIds:
          - subnet-23456789 # subnet with NAT gateway
        #AssignPublicIp: false
```

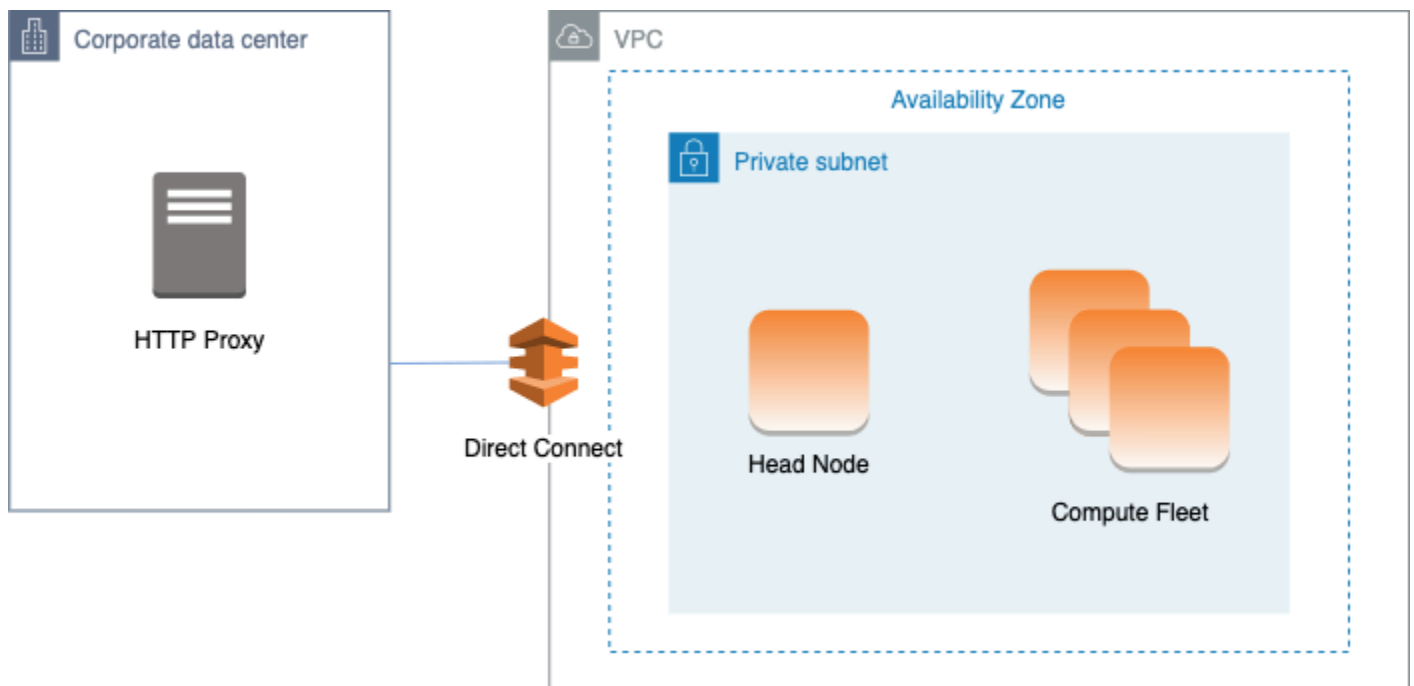
In dieser Konfiguration muss nur dem Hauptknoten des Clusters eine öffentliche IP zugewiesen werden. Sie können dies erreichen, indem Sie entweder die Einstellung „Automatische Zuweisung öffentlicher IPv4-Adresse aktivieren“ für das in [HeadNode](#)//verwendete Subnetz aktivieren

[SubnetId](#) oder indem Sie in [Networking/HeadNodeNetworking](#) eine Elastic IP zuweisen [ElasticIp](#).

Wenn Sie einen p4d-Instanztyp oder einen anderen Instanztyp definieren, der mehrere Netzwerkschnittstellen oder eine Netzwerkschnittstellenkarte zum Hauptknoten hat, müssen Sie [HeadNode/Networking](#) auf `true` setzen, [ElasticIp](#) um den öffentlichen Zugriff zu ermöglichen. AWS Öffentliche IPs können nur Instanzen zugewiesen werden, die mit einer einzigen Netzwerkschnittstelle gestartet wurden. Weitere Informationen über IP-Adressen finden Sie unter [Zuweisen einer öffentlichen IPv4-Adresse beim Start einer Instance](#) im Amazon EC2 EC2-Benutzerhandbuch für Linux-Instances.

Für diese Konfiguration ist ein [NAT-Gateway](#) oder ein interner Proxy in dem für die Warteschlangen verwendeten Subnetz erforderlich, um den Recheninstanzen den Internetzugang zu ermöglichen.

AWS ParallelCluster in einem einzigen privaten Subnetz mit Verbindung über AWS Direct Connect



Die Konfiguration für diese Architektur erfordert die folgenden Einstellungen:

```
# Note that all values are only provided as examples
HeadNode:
  ...
  Networking:
```

```
SubnetId: subnet-34567890 # subnet with proxy
Proxy:
  HttpProxyAddress: http://proxy-address:port
Ssh:
  KeyName: ec2-key-name
Scheduling:
  Scheduler: slurm
SlurmQueues:
- ...
Networking:
  SubnetIds:
    - subnet-34567890 # subnet with proxy
  AssignPublicIp: false
  Proxy:
    HttpProxyAddress: http://proxy-address:port
```

Wenn [Scheduling](#)/[SlurmQueues](#)/[Networking](#)/auf gesetzt [AssignPublicIp](#)istfalse, müssen die Subnetze korrekt eingerichtet sein, damit der Proxy für den gesamten Datenverkehr verwendet werden kann. Webzugriff ist sowohl für Head- als auch für Compute Nodes erforderlich.

AWS ParallelClustermitAWS Batch Scheduler

Wenn Sie `awsbatch` als Scheduler-Typ nutzen, erstellt AWS ParallelCluster eine AWS Batch-verwaltete Datenverarbeitungsumgebung. DieAWS Batch Umgebung verwaltet Amazon Elastic Container Service (Amazon ECS) Container-Instances. Diese Instanzen werden in dem Subnetz gestartet, das im [SubnetIds](#)Parameter [AwsBatchQueues](#)/[Networking](#)/konfiguriert ist. Um korrektAWS Batch zu funktionieren, benötigen Amazon ECS-Service-Instances externen Netzwerkzugriff, um mit dem Amazon ECS-Service-Endpoint zugreifen zu können. Daraus ergeben sich die folgenden Szenarien:

- Die für die Warteschlange angegebene Subnetz-ID verwendet ein [NAT-Gateway](#) für den Zugriff auf das Internet. Wir haben diesen Ansatz empfohlen.
- Instances, die im Queue-Subnetz gestartet werden, haben öffentliche IP-Adressen und können das Internet über ein Internet Gateway erreichen.

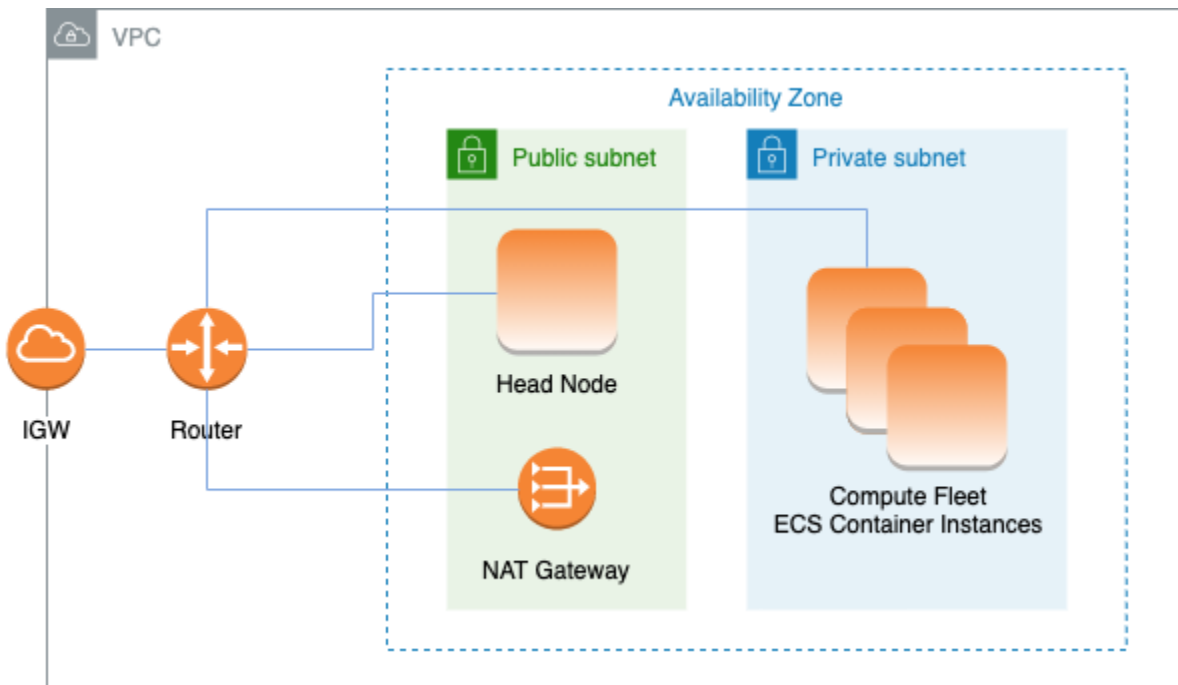
Wenn Sie außerdem an parallel Jobs mit mehreren Knoten interessiert sind (aus den [AWS BatchDokumenten](#)):

AWS Batchparallel Jobs mit mehreren Knoten verwenden den `Amazonawsipc` ECS-Netzwerkmodus. Dadurch erhalten Ihre parallel Auftragscontainer mit mehreren Knoten dieselben

Netzwerkeigenschaften wie Amazon-EC2-Instances. Jeder Container eines parallelen Auftrags mit mehreren Knoten erhält seine eigene Elastic Network-Schnittstelle, eine primäre private IP-Adresse und einen internen DNS-Hostnamen. Die Netzwerkschnittstelle wird im selben Amazon VPC-Subnetz wie die Host-Rechenressource erstellt. Alle Sicherheitsgruppen, die auf Ihre Datenverarbeitungsressourcen angewendet werden, werden auch darauf angewendet.

Wenn Sie Amazon ECS Task Networking verwenden, bietet `deraws_vpc` Netzwerkmodus keine elastischen Netzwerkschnittstellen mit öffentlichen IP-Adressen für Aufgaben, die den Amazon EC2 EC2-Starttyp verwenden. Um auf das Internet zuzugreifen, müssen Aufgaben, die den Amazon EC2 EC2-Starttyp verwenden, in einem privaten Subnetz gestartet werden, das für die Verwendung eines NAT-Gateways konfiguriert ist.

Sie müssen ein [NAT-Gateway](#) konfigurieren, damit der Cluster parallel Jobs mit mehreren Knoten ausführen kann.



Alle vorherigen Konfigurationen und Überlegungen gelten auch für AWS Batch. Es folgt ein Beispiel für eine AWS Batch Netzwerkkonfiguration.

```
# Note that all values are only provided as examples
HeadNode:
  ...
  Networking:
    SubnetId: subnet-12345678 # subnet with internet gateway, NAT gateway or proxy
```

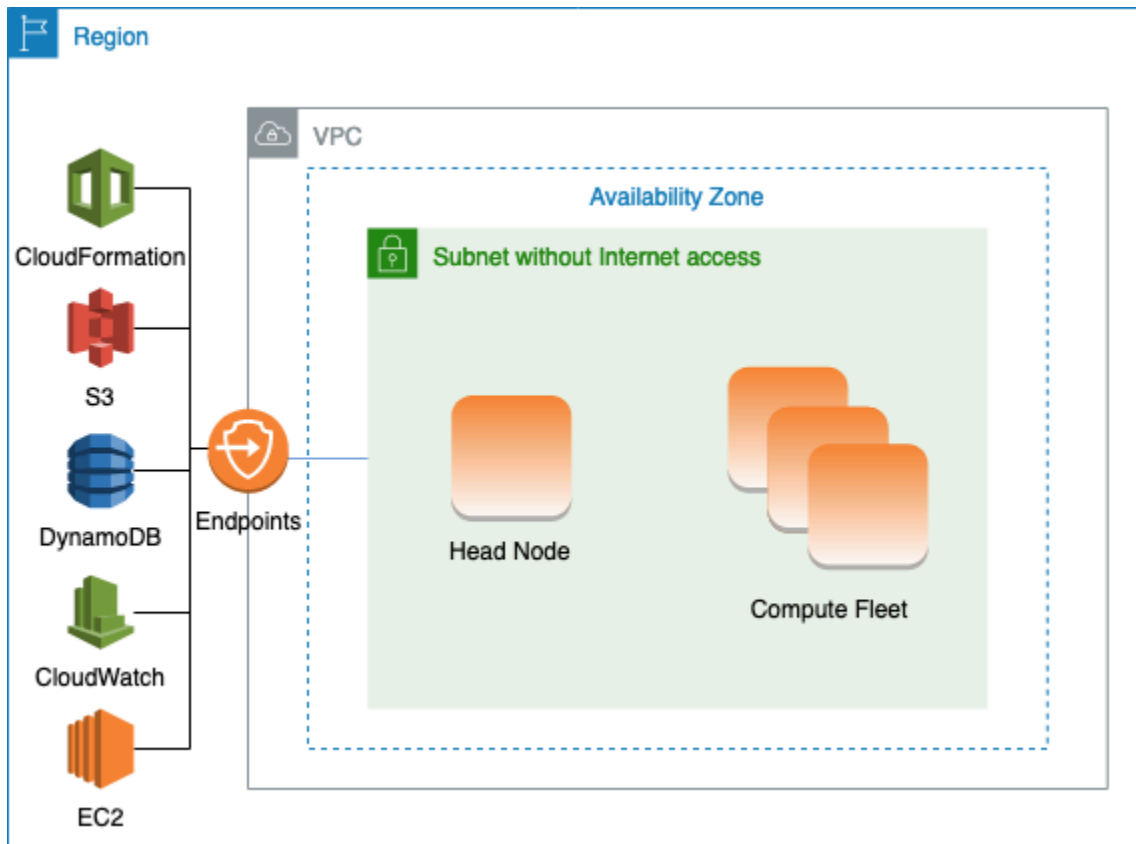
```
#ElasticIp: true | false | eip-12345678
#Proxy:
  #HttpProxyAddress: http://proxy-address:port
Ssh:
  KeyName: ec2-key-name
Scheduling:
  Scheduler: awsbatch
  AwsBatchQueues:
    - ...
  Networking:
    SubnetIds:
      - subnet-23456789 # subnet with internet gateway, NAT gateway or proxy
    #AssignPublicIp: true | false
```

Im [Networking](#) Abschnitt [Scheduling/AwsBatchQueues/SubnetIds](#) ist das ein Listentyp, aber derzeit wird nur ein Subnetz unterstützt.

Weitere Informationen finden Sie unter den folgenden Themen:

- [AWS Batchverwaltete Computerumgebungen](#)
- [AWS Batchparallel -Aufträge mit mehreren Knoten](#)
- [Amazon ECS-Aufgabenvernetzung mit dem awsvpc-Netzwerkmodus](#)

AWS ParallelCluster in einem einzelnen Subnetz ohne Internetzugang ohne Internetzugang ohne Internetzugang ohne Internetzugang.



Ein Subnetz ohne Internetzugang erlaubt keine eingehenden oder ausgehenden Verbindungen zum Internet. Diese AWS ParallelCluster Konfiguration kann sicherheitsbewussten Kunden helfen, die Sicherheit ihrer AWS ParallelCluster Ressourcen weiter zu verbessern. AWS ParallelCluster Knoten bestehen aus AWS ParallelCluster AMIs, die die gesamte Software enthalten, die für den Betrieb eines Clusters ohne Internetzugang erforderlich ist. Auf diese Weise AWS ParallelCluster können Cluster mit Knoten erstellt und verwaltet werden, die keinen Internetzugang haben.

In diesem Abschnitt erfahren Sie, wie Sie den Cluster konfigurieren. Sie erfahren auch, welche Einschränkungen beim Betrieb von Clustern ohne Internetzugang bestehen.

Konfiguration von VPC-Endpunkten

Um das reibungslose Funktionieren des Clusters zu gewährleisten, müssen die Clusterknoten in der Lage sein, mit einer Reihe von AWS Diensten zu interagieren.

Erstellen und konfigurieren Sie die folgenden [VPC-Endpunkte](#), damit Clusterknoten ohne Internetzugang mit den AWS Diensten interagieren können:

Commercial and AWS GovCloud (US) partitions

Service	Service-Name	Typ
Amazon CloudWatch	com.amazonaws. <i>Region-ID</i> .logs	Schnittstelle
AWS CloudFormation	com.amazonaws. <i>regions-id</i> .cloudformation	Schnittstelle
Amazon EC2	com.amazonaws. <i>region-id</i> .ec2	Schnittstelle
Amazon S3	com.amazonaws. <i>Region-ID</i> s.3	Gateway
Amazon DynamoDB	com.amazonaws. <i>Region-ID</i> .dynamodb	Gateway
AWS Secrets Manager**	com.amazonaws. <i>regions-id</i> .secretsmanager	Schnittstelle

China partition

Service	Service-Name	Typ
Amazon CloudWatch	com.amazonaws. <i>Region-ID</i> .logs	Schnittstelle
AWS CloudFormation	cn.com.amazonaws. <i>regions-id</i> .cloudformation	Schnittstelle
Amazon EC2	cn.com.amazonaws. <i>region-id</i> .ec2	Schnittstelle
Amazon S3	com.amazonaws. <i>Region-ID</i> s.3	Gateway

Service	Service-Name	Typ
Amazon DynamoDB	com.amazonaws. <i>Region-ID</i> .dynamodb	Gateway
AWS Secrets Manager**	com.amazonaws. <i>regions-id</i> .secretsmanager	Schnittstelle

** Dieser Endpunkt ist nur erforderlich, wenn er aktiviert [DirectoryService](#) ist, andernfalls ist er optional.

Alle Instanzen in der VPC müssen über die richtigen Sicherheitsgruppen verfügen, um mit den Endpunkten kommunizieren zu können. Sie können dies tun, indem Sie Sicherheitsgruppen [AdditionalSecurityGroups](#) unter [HeadNode](#) und [AdditionalSecurityGroups](#) unter den [SlurmQueues](#) Konfigurationen hinzufügen. Wenn die VPC-Endpunkte erstellt werden, ohne eine Sicherheitsgruppe explizit anzugeben, wird den Endpunkten die Standardsicherheitsgruppe zugeordnet. Indem Sie die Standardsicherheitsgruppe hinzufügen [AdditionalSecurityGroups](#), aktivieren Sie die Kommunikation zwischen dem Cluster und den Endpunkten.

Note

Wenn Sie IAM-Richtlinien verwenden, um den Zugriff auf VPC-Endpoints einzuschränken, müssen Sie dem Amazon S3 S3-VPC-Endpoint Folgendes hinzufügen:

```
PolicyDocument:
  Version: 2012-10-17
  Statement:
    - Effect: Allow
      Principal: "*"
      Action:
        - "s3:PutObject"
      Resource:
        - !Sub "arn:${AWS::Partition}:s3:::cloudformation-waitcondition-
          ${AWS::Region}/*"
```

Deaktivieren Sie Route 53 und verwenden Sie EC2-Hostnamen

Beim Erstellen eines Slurm-Clusters wird eine private Route 53-HostzoneAWS ParallelCluster erstellt, die zum Auflösen der benutzerdefinierten Compute-Node-Hostnamen verwendet wird, z. {queue_name}-{st|dy}-{compute_resource}-{N} B. Da Route 53 keine VPC-Endpunkte unterstützt, muss diese Funktion deaktiviert werden. Darüber hinausAWS ParallelCluster muss es so konfiguriert werden, dass es die standardmäßigen EC2-Hostnamen verwendet, ip-1-2-3-4 z. Wenden Sie die folgenden Einstellungen auf Ihre Clusterkonfiguration an:

```
...
Scheduling:
  ...
  SlurmSettings:
    Dns:
      DisableManagedDns: true
      UseEc2Hostnames: true
```

Warning

Bei Clustern, die mit [SlurmSettingsDns](#) erstellt [DisableManagedDns](#) und auf [UseEc2Hostnames](#) gesetzt sind `true`, wird der `SlurmNodeName` nicht vom DNS aufgelöst. Benutze `NodeHostName` stattdessen den Slurm.

Note

Dieser Hinweis ist abAWS ParallelCluster Version 3.3.0 nicht relevant.
FürAWS ParallelCluster unterstützte Version 3.3.3.3. Version 3.3.3. Version 3.3.3.3.
Wenn auf `UseEc2Hostnames` ist `true`, wird die Slurm-Konfigurationsdatei mit den `epilog` SkriptenAWS ParallelCluster `prolog` und `epilog` gesetzt:

- `prolog` wird ausgeführt, um Knoteninformationen/`etc/hosts` zu den Rechenknoten hinzuzufügen, wenn jeder Job zugewiesen wird.
- `epilog` läuft, um Inhalte zu reinigen, die von `prolog` geschrieben wurden.

Um benutzerdefinierte `epilog` Skripts `prolog` oder Skripts hinzuzufügen, fügen Sie sie jeweils zu den `/opt/slurm/etc/pcluster/epilog.d/` Ordnern `/opt/slurm/etc/pcluster/prolog.d/` oder hinzu.

Cluster-Konfiguration

Erfahren Sie, wie Sie Ihren Cluster so konfigurieren, dass er in einem Subnetz ohne Verbindung zum Internet ausgeführt wird.

Die Konfiguration für diese Architektur erfordert die folgenden Einstellungen:

```
# Note that all values are only provided as examples
...
HeadNode:
  ...
  Networking:
    SubnetId: subnet-1234567890abcdef0 # the VPC of the subnet needs to have VPC
    endpoints
    AdditionalSecurityGroups:
      - sg-abcdef01234567890 # optional, the security group that enables the
      communication between the cluster and the VPC endpoints
Scheduling:
  Scheduler: slurm # Cluster in a subnet without internet access is supported only when
  the scheduler is Slurm.
  SlurmSettings:
    Dns:
      DisableManagedDns: true
      UseEc2Hostnames: true
  SlurmQueues:
    - ...
    Networking:
      SubnetIds:
        - subnet-1234567890abcdef0 # the VPC of the subnet needs to have VPC
        endpoints attached
      AdditionalSecurityGroups:
        - sg-1abcdef01234567890 # optional, the security group that enables the
        communication between the cluster and the VPC endpoints
```

- [SubnetId\(s\)](#): Das Subnetz ohne Internetzugang.

Um die Kommunikation zwischen AWS ParallelCluster und AWS Services zu ermöglichen, müssen die VPC-Endpunkte an die VPC des Subnetzes angeschlossen sein. Bevor Sie Ihren Cluster erstellen, stellen Sie sicher, dass die [automatische Zuweisung der öffentlichen IPv4-Adresse im Subnetz deaktiviert ist](#), um sicherzustellen, dass die `pccluster` Befehle Zugriff auf den Cluster haben.

- [AdditionalSecurityGroups](#): Die Sicherheitsgruppe, die die Kommunikation zwischen dem Cluster und den VPC-Endpunkten ermöglicht.

Fakultativ:

- Wenn die VPC-Endpunkte erstellt werden, ohne eine Sicherheitsgruppe explizit anzugeben, wird der VPC die Standardsicherheitsgruppe der VPC zugeordnet. Geben Sie daher die Standardsicherheitsgruppe für `anAdditionalSecurityGroups`.
- Wenn benutzerdefinierte Sicherheitsgruppen bei der Erstellung des Clusters und/oder der VPC-Endpoints verwendet werden, `AdditionalSecurityGroups` ist dies nicht erforderlich, solange die benutzerdefinierten Sicherheitsgruppen die Kommunikation zwischen dem Cluster und den VPC-Endpunkten ermöglichen.
- [Scheduler](#): Der Cluster-Scheduler.

`slurm` ist der einzige gültige Wert. Nur der Slurm-Scheduler unterstützt einen Cluster in einem Subnetz ohne Internetzugang.

- [SlurmSettings](#): Die Slurm-Einstellungen.

Weitere Informationen finden Sie im vorherigen Abschnitt. Deaktivieren Sie Route53 und verwenden Sie EC2-Hostnamen.

Einschränkungen

- Verbindung zum Hauptknoten über SSH oder NICE DCV herstellen: Wenn Sie eine Verbindung zu einem Cluster herstellen, stellen Sie sicher, dass der Client der Verbindung den Hauptknoten des Clusters über seine private IP-Adresse erreichen kann. Wenn sich der Client nicht in derselben VPC wie der Head-Knoten befindet, verwenden Sie eine Proxy-Instance in einem öffentlichen Subnetz der VPC. Diese Anforderung gilt sowohl für SSH- als auch für DCV-Verbindungen. Auf die öffentliche IP eines Headknotens kann nicht zugegriffen werden, wenn das Subnetz keinen Internetzugang hat. Die `dcv-connect` Befehle `pc1uster ssh` und verwenden die öffentliche IP, falls sie existiert, oder die private IP. Bevor Sie Ihren Cluster erstellen, stellen Sie sicher, dass die [automatische Zuweisung der öffentlichen IPv4-Adresse im Subnetz deaktiviert ist](#), um sicherzustellen, dass die `pc1uster` Befehle Zugriff auf den Cluster haben.

Das folgende Beispiel zeigt, wie Sie eine Verbindung zu einer DCV-Sitzung herstellen können, die im Hauptknoten Ihres Clusters ausgeführt wird. Sie stellen eine Verbindung über eine EC2-Proxy-Instance her. Die Instanz fungiert als NICE-DCV-Server für Ihren PC und als Client für den Hauptknoten im privaten Subnetz.

Stellen Sie eine Connect über DCV über eine Proxyinstanz in einem öffentlichen Subnetz her:

1. Erstellen Sie eine EC2-Instance in einem öffentlichen Subnetz, das sich in derselben VPC wie das Subnetz des Clusters befindet.
 2. Stellen Sie sicher, dass der `nlde.on-dcv-client` und `-server` auf Ihrer EC2-Instance installiert sind.
 3. Hängen Sie eine [AWS ParallelCluster Benutzerrichtlinie](#) an die EC2-Proxy-Instance an. Weitere Informationen finden Sie unter [AWS ParallelCluster Beispiele `pc1uster` für Benutzerrichtlinien](#).
 4. Installieren Sie `aws-parallelcluster` auf der Proxy-EC2-Instance.
 5. Stellen Sie über DCV eine Verbindung zur Proxy-EC2-Instance her.
 6. Verwenden Sie den `pc1uster dcv-connect` Befehl auf der Proxyinstanz, um eine Verbindung zum Cluster innerhalb des Subnetzes ohne Internetzugang herzustellen.
- Interaktion mit anderen AWS Diensten: Nur Dienste, die von unbedingt erforderlich sind, `aws-parallelcluster` sind oben aufgeführt. Wenn Ihr Cluster mit anderen Diensten interagieren muss, erstellen Sie die entsprechenden VPC-Endpunkte.

Login-Knoten

Ab Version 3.7.0, `aws-parallelcluster` Clusteradministratoren können Anmeldeknoten bereitstellen, die Benutzern Zugriff auf die Ausführung von Jobs gewähren, anstatt direkt auf den Cluster-Hauptknoten zuzugreifen. Clusterbenutzer mit entsprechenden Berechtigungen können Active Directory oder ihre SSH-Anmeldeinformationen verwenden, um sich anzumelden, ihre Jobs einzureichen und zu verwalten. Dadurch kann die Clusterverwaltung verbessert und die Wahrscheinlichkeit minimiert werden, dass die Ressourcen des Hauptknotens aufgebraucht werden, die Slurm zur Verwaltung des Clusters benötigt. Angemeldete Benutzer haben außerdem Zugriff auf den gesamten gemeinsam genutzten Speicher des Clusters, der auf den Anmeldeknoten installiert ist. Wenn ein Anmeldeknoten gestoppt werden muss, werden angemeldete Benutzer im Voraus über die aktive Shell-Sitzung, die sie verwenden, benachrichtigt.

Anmeldeknoten werden als Pools spezifiziert, wobei ein Pool eine Gruppe von Anmeldeknoten definiert, die dieselbe Ressourcenkonfiguration haben. Alle Anmeldeknoten in einem Pool sind so konfiguriert, dass sie Teil eines sind [Netzwerk-Loadbalancer](#) das ermöglicht die Verteilung von Sitzungen auf Login-Knoten im Round-Robin-Modus. Die aktuelle Implementierung ermöglicht die Angabe eines Pools von Anmeldeknoten, der mehrere Anmeldeknoten enthält.

Sicherheit

Anmeldeknoten erben die `AllowedIPs` Einstellungen [AllowedIps](#) vom Kopfknoten aus. Auf diese Weise können Clusteradministratoren den Sicherheitsstatus des Clusters einschränken, indem sie den Quell-CIDR oder eine Präfixliste angeben, von der aus SSH-Verbindungen zulässig sind.

In der aktuellen Implementierung wird der Zugriff auf den Hauptknoten nicht automatisch eingeschränkt, wenn Anmeldeknoten aktiviert werden. Bei Bedarf kann ein Clusteradministrator diesen Zugriff einschränken und die SSH-Konfiguration der Hauptknoten mithilfe von Linux-Standardbefehlen aktualisieren. Dies kann auch erreicht werden, indem benutzerdefinierte Sicherheitsgruppen auf dem Hauptknoten angegeben werden. Verwenden Sie dazu `AdditionalSecurityGroups` Einstellung im Kopfknotenbereich des ParallelCluster YAML-Datei, um Verbindungen von nicht autorisierten Benutzern zu verweigern.

Netzwerkfunktionen

Anmeldeknoten erhalten eine einzige Verbindungsadresse für den Network Load Balancer, der für den Pool von Anmeldeknoten konfiguriert ist. Die Konnektivitätseinstellungen der Adresse basieren auf dem Subnetztyp, der in der Pool-Konfiguration für Anmeldeknoten angegeben ist.

- Wenn das Subnetz privat ist, ist die Adresse privat. Um Zugriff auf die Anmeldeknoten zu gewähren, muss der Clusteradministrator einen Bastion-Host bereitstellen.
- Wenn das Subnetz öffentlich ist, ist die Adresse öffentlich

Alle Verbindungsanfragen werden vom Network Load Balancer mithilfe von Round-Robin-Routing verwaltet.

Speicherung

Der gesamte gemeinsam genutzte Speicher wurde auf dem Cluster konfiguriert mit ParallelCluster einschließlich des verwalteten Speichers wird auf allen Anmeldeknoten bereitgestellt.

Rufen Sie die Informationen zu den Anmeldeknoten ab

Um die Adresse der einzelnen Verbindung abzurufen, die für den Zugriff auf die Anmeldeknoten bereitgestellt wurde, kann der Clusteradministrator den folgenden Befehl ausführen [describe-cluster](#) Befehl. Der Befehl bietet auch weitere Informationen über den Status der Anmeldeknoten.

Anmeldeknoten sind ein neuer Knotentyp, der unterstützt wird von ParallelCluster das kann mit dem spezifiziert werden [describe-cluster-instances](#) Befehl bei der Abfrage des Status eines bestimmten Knotentyps.

Die Verfügbarkeit einer einzigen Verbindungsadresse für den Pool der Anmeldeknoten verhindert nicht den direkten Zugriff auf einen bestimmten Anmeldeknoten. Es wird jedoch nicht empfohlen, die direkte Verbindung zu verwenden, um Warnungen vom SSH-Client zu vermeiden. Der SSH-Client speichert Host-Identifikatoren lokal für jede Zieladresse. Da die Host-ID für jeden Pool spezifisch ist, kann die Verwendung verschiedener IPs und/oder der einzelnen Verbindungsadresse dazu führen, dass dieselbe Host-ID unterschiedlichen Zieladressen zugeordnet ist. Dies kann zu einer Warnung durch den SSH-Client führen, da dieselbe Host-ID mehreren Zielen zugeordnet ist.

IMDS-Eigenschaften

Der Zugriff auf das IMDS des Anmeldeknotens (und die Anmeldeinformationen für das Instanzprofil) ist auf Root-Benutzer und Cluster-Administratoren beschränkt (`pc-cluster-admin` standardmäßig) und betriebssystemspezifischer Standardbenutzer (`ec2-user` auf Amazon Linux 2 und RedHat, `ubuntu` auf Ubuntu 18.04, `centos` auf CentOS 7).

Um den IMDS-Zugriff einzuschränken, AWS ParallelCluster verwaltet eine Kette von `iptables`.

Note

Jede Anpassung von `iptables` oder `ip6tables` Regeln können den Mechanismus stören, der zur Beschränkung des IMDS-Zugriffs auf den Anmeldeknoten verwendet wird. Siehe auch [Imds property setting](#).

Lebenszyklus von Login-Knoten

Derzeit gibt es keinen speziellen Befehl zum Stoppen und Starten der Anmeldeknoten in einem Pool. Um die Anmeldeknoten in einem Pool zu stoppen, muss der Clusteradministrator die Clusterkonfiguration aktualisieren und bei der Anzahl der Anmeldeknoten Null angeben (`Count: 0`) und führe dann eine aus [pcluster.update-cluster-v3](#) Befehl.

Note

Angemeldete Benutzer werden über die Kündigung der jeweiligen Instanz und über die damit verbundene Nachfrist informiert. Während der Gracetime-Periode sind

keine neuen Verbindungen erlaubt, außer denen von [Cluster-Standardbenutzer](#). Die angezeigte Meldung kann vom Clusteradministrator vom Hauptknoten oder von einem Anmeldeknoten aus angepasst werden, der die Datei bearbeitet `opt/parallelcluster/shared_login_nodes/loginmgt_config.json`.

Um den Anmeldeknotenpool zu starten, muss der Clusteradministrator den vorherigen `wiederherstellenCountWert` in der Clusterkonfiguration eingeben und dann eine ausführen [update-cluster](#) Befehl.

Zum Ausführen des Anmeldeknotenpools sind Berechtigungen erforderlich

Um den Pool der Anmeldeknoten verwalten zu können, muss der Clusteradministrator über die folgenden zusätzlichen Berechtigungen verfügen:

```
- Action:
  - autoscaling:DeleteAutoScalingGroup
  - autoscaling:DeleteLifecycleHook
  - autoscaling:Describe*
  - autoscaling:PutLifecycleHook
  - autoscaling:UpdateAutoScalingGroup
  - elasticloadbalancing:CreateListener
  - elasticloadbalancing:CreateTargetGroup
  - elasticloadbalancing>DeleteListener
  - elasticloadbalancing>DeleteLoadBalancer
  - elasticloadbalancing>DeleteTargetGroup
  - elasticloadbalancing:Describe*
  - elasticloadbalancing:ModifyLoadBalancerAttributes
Resource: '*'
Condition:
  ForAllValues:StringEquals:
    aws:TagKeys: [ "parallelcluster:cluster-name" ]
- Action:
  - autoscaling:CreateAutoScalingGroup
  - elasticloadbalancing:AddTags
  - elasticloadbalancing:CreateLoadBalancer
Resource: '*'
Effect: Allow
```

Benutzerdefinierte Bootstrap-Aktionen

Wenn Sie die [OnNodeStart](#)Konfigurationseinstellungen [HeadNode/CustomActions](#)/definieren, AWS ParallelCluster führt unmittelbar nach dem Start des Knotens beliebigen Code aus. Wenn Sie die [OnNodeConfigured](#)Konfigurationseinstellungen [HeadNode/CustomActions](#)/definieren, AWS ParallelCluster wird der Code ausgeführt, nachdem die Knotenkonfiguration korrekt abgeschlossen wurde.

Ab AWS ParallelCluster Version 3.4.0 kann der Code nach dem Update des Kopfknotens ausgeführt werden, wenn Sie die [OnNodeUpdated](#)Konfigurationseinstellungen [HeadNode/CustomActions](#)/definieren.

In den meisten Fällen wird dieser Code in Amazon Simple Storage Service (Amazon S3) gespeichert und der Zugriff erfolgt über eine HTTPS-Verbindung. Der Code wird in jeder Skriptsprache ausgeführt, die vom Cluster-Betriebssystem unterstützt wird, `root` und kann in dieser Sprache sein. Oft ist der Code in Bash oder Python.

Note

Ab AWS ParallelCluster Version 3.7.0 lautet die [ImdsSupport](#)Standardeinstellung `cluster Imds/.v2.0`

Wenn Sie einen neuen Cluster für ein Upgrade auf Version 3.7.0 und spätere Versionen erstellen, aktualisieren Sie entweder Ihre benutzerdefinierten Bootstrap-Aktionsskripte, sodass sie mit IMDSv2 kompatibel sind, oder setzen Sie `v1.0` in Ihrer [Imds](#)Cluster-Konfigurationsdatei [ImdsSupport](#)auf/.

Warning

Sie sind für die Konfiguration der benutzerdefinierten Skripts und Argumente verantwortlich, wie im Modell der gemeinsamen Verantwortung beschrieben. Stellen Sie sicher, dass Ihre benutzerdefinierten Bootstrap-Skripts und Argumente aus Quellen stammen, denen Sie vertrauen, dass sie vollen Zugriff auf Ihre Clusterknoten haben.

⚠ Warning

AWS ParallelCluster unterstützt nicht die Verwendung interner Variablen, die über die `/etc/parallelcluster/cfnconfig` Datei bereitgestellt werden. Diese Datei wird möglicherweise als Teil einer future Version entfernt.

`OnNodeStart` Aktionen werden aufgerufen, bevor eine Bootstrap-Aktion zur Knotenbereitstellung gestartet wird, z. B. die Konfiguration von NAT, Amazon Elastic Block Store (Amazon EBS) oder des Schedulers. `OnNodeStart` Bootstrap-Aktionen können das Ändern von Speicher, das Hinzufügen zusätzlicher Benutzer und das Hinzufügen von Paketen umfassen.

ℹ Note

Wenn Sie ein `HeadNodeCustomActionsOnNodeStart`-Skript für Ihren Cluster konfigurieren `DirectoryService`, AWS ParallelCluster konfiguriert `DirectoryService` und startet das `neussd`, bevor es das Skript ausführt. `OnNodeStart`

`OnNodeConfigured` Aktionen werden aufgerufen, nachdem die Bootstrap-Prozesse des Knotens abgeschlossen sind. `OnNodeConfigured` Aktionen dienen den letzten Aktionen, die ausgeführt werden, bevor eine Instanz als vollständig konfiguriert und abgeschlossen betrachtet wird. Einige `OnNodeConfigured` Aktionen umfassen das Ändern von Scheduler-Einstellungen, das Ändern des Speichers und das Ändern von Paketen. Sie können Argumente an Skripts übergeben, indem Sie sie bei der Konfiguration angeben.

`OnNodeUpdated` Aktionen werden aufgerufen, nachdem die Aktualisierung des Hauptknotens abgeschlossen ist und der Scheduler und der gemeinsam genutzte Speicher an die neuesten Änderungen der Cluster-Konfiguration angepasst wurden.

Wenn `OnNodeStart` unsere `OnNodeConfigured` benutzerdefinierten Aktionen erfolgreich sind, wird der Erfolg mit dem Exit-Code Null (0) angezeigt. Jeder andere Exit-Code weist darauf hin, dass der Instanz-Bootstrap fehlgeschlagen ist.

Wenn `OnNodeUpdated` benutzerdefinierte Aktionen erfolgreich sind, wird der Erfolg mit dem Exit-Code Null (0) signalisiert. Jeder andere Exit-Code weist darauf hin, dass das Update fehlgeschlagen ist.

Note

Wenn Sie die Konfiguration [OnNodeUpdated](#) vornehmen, müssen Sie die `OnNodeUpdated` Aktionen bei fehlgeschlagenen Updates manuell auf den vorherigen Status zurücksetzen. Wenn eine `OnNodeUpdated` benutzerdefinierte Aktion fehlschlägt, wird das Update auf den vorherigen Status zurückgesetzt. Die `OnNodeUpdated` Aktion wird jedoch nur zur Aktualisierungszeit und nicht zur Stack-Rollback-Zeit ausgeführt.

In den [CustomActions](#) Konfigurationsabschnitten/und [HeadNode](#)//können Sie unterschiedliche Skripten für den Hauptknoten [CustomActions](#) und [Scheduling](#) für jede Warteschlange angeben. [SlurmQueues OnNodeUpdated](#) kann nur im `HeadNode` Abschnitt konfiguriert werden.

Note

Vor AWS ParallelCluster Version 3.0 war es nicht möglich, unterschiedliche Skripte für Head- und Compute-Knoten anzugeben. Weitere Informationen finden Sie unter [Umstellung von AWS ParallelCluster 2.x auf 3.x](#).

Themen

- [Konfiguration](#)
- [Argumente](#)
- [Beispielcluster mit benutzerdefinierten Bootstrap-Aktionen](#)
- [Beispiel für die Aktualisierung eines benutzerdefinierten Bootstrap-Skripts für IMDSv2](#)
- [Beispiel für die Aktualisierung einer Konfiguration für IMDSv1](#)

Konfiguration

Die folgenden Konfigurationseinstellungen werden verwendet, um [HeadNode/CustomActions/OnNodeStart& OnNodeConfigured& OnNodeUpdated](#) und [Scheduling/CustomActions/OnNodeStart& OnNodeConfigured](#) Aktionen und Argumente zu definieren.

```
HeadNode:
  [...]
  CustomActions:
```

```
OnNodeStart:
  # Script URL. This is run before any of the bootstrap scripts are run
  Script: s3://bucket-name/on-node-start.sh
  Args:
    - arg1
OnNodeConfigured:
  # Script URL. This is run after all the bootstrap scripts are run
  Script: s3://bucket-name/on-node-configured.sh
  Args:
    - arg1
OnNodeUpdated:
  # Script URL. This is run after the head node update is completed.
  Script: s3://bucket-name/on-node-updated.sh
  Args:
    - arg1
# Bucket permissions
Iam:
  S3Access:
    - BucketName: bucket_name
      EnableWriteAccess: false
Scheduling:
  Scheduler: slurm
  [...]
SlurmQueues:
  - Name: queue1
  [...]
CustomActions:
  OnNodeStart:
    Script: s3://bucket-name/on-node-start.sh
    Args:
      - arg1
  OnNodeConfigured:
    Script: s3://bucket-name/on-node-configured.sh
    Args:
      - arg1
Iam:
  S3Access:
    - BucketName: bucket_name
      EnableWriteAccess: false
```

Verwenden der Sequence Einstellung (in AWS ParallelCluster Version 3.6.0 hinzugefügt):

```
HeadNode:
```

```
[...]
CustomActions:
  OnNodeStart:
    # Script URLs. The scripts are run in the same order as listed in the
    configuration, before any of the bootstrap scripts are run.
    Sequence:
      - Script: s3://bucket-name/on-node-start1.sh
        Args:
          - arg1
      - Script: s3://bucket-name/on-node-start2.sh
        Args:
          - arg1
    [...]
  OnNodeConfigured:
    # Script URLs. The scripts are run in the same order as listed in the
    configuration, after all the bootstrap scripts are run.
    Sequence:
      - Script: s3://bucket-name/on-node-configured1.sh
        Args:
          - arg1
      - Script: s3://bucket-name/on-node-configured2.sh
        Args:
          - arg1
    [...]
  OnNodeUpdated:
    # Script URLs. The scripts are run in the same order as listed in the
    configuration, after the head node update is completed.
    Sequence:
      - Script: s3://bucket-name/on-node-updated1.sh
        Args:
          - arg1
      - Script: s3://bucket-name/on-node-updated2.sh
        Args:
          - arg1
    [...]
# Bucket permissions
Iam:
  S3Access:
    - BucketName: bucket_name
      EnableWriteAccess: false
Scheduling:
  Scheduler: slurm
  [...]
  SlurmQueues:
```

```

- Name: queue1
  [...]
  CustomActions:
    OnNodeStart:
      # Script URLs. The scripts are run in the same order as listed in the
      configuration, before any of the bootstrap scripts are run
      Sequence:
        - Script: s3://bucket-name/on-node-start1.sh
          Args:
            - arg1
        - Script: s3://bucket-name/on-node-start2.sh
          Args:
            - arg1
        [...]
    OnNodeConfigured:
      # Script URLs. The scripts are run in the same order as listed in the
      configuration, after all the bootstrap scripts are run
      Sequence:
        - Script: s3://bucket-name/on-node-configured1.sh
          Args:
            - arg1
        - Script: s3://bucket-name/on-node-configured2.sh
          Args:
            - arg1
        [...]
  Iam:
    S3Access:
      - BucketName: bucket_name
        EnableWriteAccess: false

```

Die Sequence Einstellung wird ab AWS ParallelCluster Version 3.6.0 hinzugefügt. Wenn Sie angebenSequence, können Sie mehrere Skripts für eine benutzerdefinierte Aktion auflisten. AWS ParallelClusterunterstützt weiterhin die Konfiguration einer benutzerdefinierten Aktion mit einem einzigen Skript, ohne dies einzuschließenSequence.

AWS ParallelClusterunterstützt nicht, sowohl ein einzelnes Skript als auch Sequence dieselbe benutzerdefinierte Aktion einzubeziehen. Schlägt beispielsweise AWS ParallelCluster fehl, wenn Sie die folgende Konfiguration angeben.

```

[...]
CustomActions:
  OnNodeStart:
    # Script URL. This is run before any of the bootstrap scripts are run

```

```
Script: s3://bucket-name/on-node-start.sh
  Args:
    - arg1
  # Script URLs. The scripts are run in the same order as listed in the
  # configuration, before any of the bootstrap scripts are run.
  Sequence:
    - Script: s3://bucket-name/on-node-start1.sh
      Args:
        - arg1
    - Script: s3://bucket-name/on-node-start2.sh
      Args:
        - arg1
[...]
```

Argumente

Note

In AWS ParallelCluster 2.x waren die \$1 Argumente reserviert, um die URL des benutzerdefinierten Skripts zu speichern. Wenn Sie die für AWS ParallelCluster 2.x erstellten benutzerdefinierten Bootstrap-Skripte mit AWS ParallelCluster 3.x wiederverwenden möchten, müssen Sie sie anpassen, indem Sie die Verschiebung der Argumente berücksichtigen. Weitere Informationen finden Sie unter [Umstellung von AWS ParallelCluster 2.x auf 3.x](#).

Beispielcluster mit benutzerdefinierten Bootstrap-Aktionen

Mit den folgenden Schritten wird ein einfaches Skript erstellt, das nach der Konfiguration des Knotens ausgeführt wird und das die wget Pakete R, curl und in den Knoten des Clusters installiert.

1. Erstellen Sie ein Skript.

```
#!/bin/bash
echo "The script has $# arguments"
for arg in "$@"
do
    echo "arg: ${arg}"
done
yum -y install "${@:1}"
```

2. Laden Sie das Skript mit den richtigen Berechtigungen auf Amazon S3 hoch. Wenn öffentliche Leseberechtigungen für Sie nicht geeignet sind, verwenden [HeadNode](#) Sie die [SlurmQueues](#) Konfigurationsabschnitte [Scheduling](#)/[S3Access](#) und [Iam](#). Weitere Informationen finden Sie unter [Arbeiten mit Amazon S3](#).

```
$ aws s3 cp --acl public-read /path/to/myscript.sh s3://<bucket-name>/myscript.sh
```

Important

Wenn das Skript unter Windows bearbeitet wurde, müssen die Zeilenenden von CRLF in LF geändert werden, bevor das Skript auf Amazon S3 hochgeladen wird.

3. Aktualisieren Sie die AWS ParallelCluster Konfiguration, sodass sie die neue OnNodeConfigured Aktion enthält.

```
CustomActions:
OnNodeConfigured:
  Script: https://<bucket-name>.s3.<region>.amazonaws.com/myscript.sh
  Args:
    - "R"
    - "curl"
    - "wget"
```

Wenn der Bucket nicht öffentlich lesbar ist, verwenden Sie ihn s3 als URL-Protokoll.

```
CustomActions:
OnNodeConfigured:
  Script: s3://<bucket-name>/myscript.sh
  Args:
    - "R"
    - "curl"
    - "wget"
```

4. Starten Sie den Cluster.

```
$ pcluster create-cluster --cluster-name mycluster \
  --region <region> --cluster-configuration config-file.yaml
```

5. Überprüfen Sie die Ausgabe.

- Wenn Sie der HeadNode Konfiguration benutzerdefinierte Aktionen hinzugefügt haben, melden Sie sich beim Hauptknoten an und überprüfen Sie die `cfn-init.log` Datei unter, `/var/log/cfn-init.log` indem Sie den folgenden Befehl ausführen:

```
$ less /var/log/cfn-init.log
2021-09-03 10:43:54,588 [DEBUG] Command run
postinstall output: The script has 3 arguments
arg: R
arg: curl
arg: wget
Loaded plugins: dkms-build-requires, priorities, update-motd, upgrade-helper
Package R-3.4.1-1.52.amzn1.x86_64 already installed and latest version
Package curl-7.61.1-7.91.amzn1.x86_64 already installed and latest version
Package wget-1.18-4.29.amzn1.x86_64 already installed and latest version
Nothing to do
```

- Wenn Sie der SlurmQueues Einstellung benutzerdefinierte Aktionen hinzugefügt haben, überprüfen Sie die Option, die sich `/var/log/cloud-init.log` in einem Rechenknoten `cloud-init.log` befindet. CloudWatch dient zum Anzeigen dieser Protokolle.

Sie können diese beiden Protokolle in der CloudWatch Amazon-Konsole einsehen. Weitere Informationen finden Sie unter [Integration mit Amazon CloudWatch Logs](#).

Beispiel für die Aktualisierung eines benutzerdefinierten Bootstrap-Skripts für IMDSv2

Im folgenden Beispiel aktualisieren wir ein benutzerdefiniertes Bootstrap-Aktionsskript, das mit IMDSv1 für die Verwendung mit IMDSv2 verwendet wurde. Das IMDSv1-Skript ruft AMI-ID-Metadaten der EC2-Instanz ab.

```
#!/bin/bash
AMI_ID=$(curl http://169.254.169.254/latest/meta-data/ami-id)
echo $AMI_ID >> /home/ami_id.txt
```

Im Folgenden wird das benutzerdefinierte Bootstrap-Aktionsskript gezeigt, das so geändert wurde, dass es mit IMDSv2 kompatibel ist.

```
#!/bin/bash
```

```
AMI_ID=$(TOKEN=`curl -X PUT "http://169.254.169.254/latest/api/token" -H "X-aws-ec2-
metadata-token-ttl-seconds: 21600"` \
    && curl -H "X-aws-ec2-metadata-token: $TOKEN" -v http://169.254.169.254/
latest/meta-data/ami-id)
echo $AMI_ID >> /home/ami_id.txt
```

Weitere Informationen finden Sie unter [Instance-Metadaten abrufen](#) im EC2-Benutzerhandbuch für Linux-Instances.

Beispiel für die Aktualisierung einer Konfiguration für IMDSv1

Im Folgenden finden Sie ein Beispiel für eine Clusterkonfiguration, die IMDSv1 unterstützt, wenn AWS ParallelCluster Versionen 3.7.0 und älter verwendet werden.

```
Region: us-east-1
Imds:
  ImdsSupport: v1.0
Image:
  Os: alinux2
HeadNode:
  InstanceType: t2.micro
  Networking:
    SubnetId: subnet-abcdef01234567890
  Ssh
    KeyName: key-name
  CustomActions:
    OnNodeConfigured:
      Script: Script-path
Scheduling:
  Scheduler: slurm
  SlurmQueues:
    - Name: queue1
      CustomActions:
        OnNodeConfigured:
          Script: Script-path
  ComputeResources:
    - Name: t2micro
      Instances:
        - InstanceType: t2.micro
      MinCount: 11
  Networking:
    SubnetIds:
```



```
- subnet-abcdef01234567890
```

Arbeiten mit Amazon S3

Sie können AWS ParallelCluster den Zugriff auf Amazon S3 über die [S3Access](#) Parameter [HeadNode/Iam/S3Access](#) und [Scheduling/SlurmQueues/- Name/Iam](#) in der AWS ParallelCluster Konfiguration konfigurieren.

Beispiele

Das folgende Beispiel konfiguriert den schreibgeschützten Zugriff auf alle Objekte in *firstbucket/read_only/* und den Lese-/Schreibzugriff auf alle Objekte in *secondbucket/read_and_write/*.

```
...
HeadNode:
  ...
  Iam:
    S3Access:
      - BucketName: firstbucket
        KeyName: read_only/*
        EnableWriteAccess: false
      - BucketName: secondbucket
        KeyName: read_and_write/*
        EnableWriteAccess: true
  ...
```

Im nächsten Beispiel wird der schreibgeschützte Zugriff auf alle Objekte in einem Ordner *read_only/* in einem beliebigen Bucket (*) des Kontos konfiguriert.

```
...
HeadNode:
  ...
  Iam:
    S3Access:
      - BucketName: *
        KeyName: read_only/*
        EnableWriteAccess: false
  ...
```

Das letzte Beispiel konfiguriert den schreibgeschützten Zugriff auf alle Buckets und Objekte im Konto.

```
...
HeadNode:
  ...
  Iam:
    S3Access:
      - BucketName: *
  ...
```

Arbeiten mit Spot-Instances

AWS ParallelCluster verwendet Spot-Instances, wenn Sie `SPOT` in der Cluster-Konfigurationsdatei [SlurmQueues](#) `AwsBatchQueues/CapacityType` oder `CapacityType` auf gesetzt haben. Spot-Instances sind kostengünstiger als On-Demand-Instances, sie können jedoch unterbrochen werden. Es kann hilfreich sein, Benachrichtigungen über Spot-Instance-Unterbrechungen nutzen. Diese stellen zwei Minuten, bevor Amazon EC2 Ihre Spot-Instance beenden muss, eine Warnmeldung bereit. Weitere Informationen finden Sie unter Benachrichtigungen über [Spot-Instance-Unterbrechungen](#) im Amazon EC2 EC2-Benutzerhandbuch für LinInstances. Informationen zur [AwsBatchQueues](#) Funktionsweise mit [Spot-Instances finden Sie im AWS Batch Benutzerhandbuch unter Rechenressourcen](#).

Der AWS ParallelCluster konfigurierte Scheduler weist Rechenressourcen in Warteschlangen mit Spot-Instances Jobs zu, genauso wie er Jobs Rechenressourcen in Warteschlangen mit On-Demand-Instances zuweist.

Wenn Sie Spot-Instances verwenden, muss eine `AWSServiceRoleForEC2Spot` servicegebundene Rolle in Ihrem Konto bereit. Führen Sie den folgenden Befehl aus AWS CLI, um diese Rolle in Ihrem Konto über die zu erstellen:

```
$ aws iam create-service-linked-role --aws-service-name spot.amazonaws.com
```

Weitere Informationen finden Sie unter [Servicegebundene Rolle für Spot-Instance-Unterbrechungen](#) im Amazon EC2 EC2-Benutzerhandbuch für LinInstances.

In den folgenden Abschnitten werden drei Szenarien beschrieben, in denen Spot-Instances bei der Verwendung unterbrochen werden können [SlurmQueues](#).

Szenario 1: Spot-Instance ohne ausgeführte Aufgaben wird unterbrochen

Wenn diese Unterbrechung auftritt, wird AWS ParallelCluster versucht, die Instanz zu ersetzen, wenn die Scheduler-Warteschlange ausstehende Jobs enthält, für die zusätzliche Instanzen erforderlich sind, oder wenn die Anzahl der aktiven Instanzen niedriger als [SlurmQueues/ComputeResources/MinCount](#) ist. Wenn keine neuen Instanzen bereitgestellt werden können, wird eine Anfrage für neue Instanzen regelmäßig wiederholt.

Szenario 2: Spot-Instance mit Einzelknotenaufgaben wird unterbrochen

Der Job schlägt mit einem Statuscode von `fehINODE_FAIL`, und der Job wird in die Warteschlange gestellt (sofern `--no-requeue` er nicht bei der Übermittlung des Jobs angegeben wurde). Wenn der Knoten ein statischer Knoten ist, wird er ersetzt. Wenn der Knoten ein dynamischer Knoten ist, wird der Knoten beendet und zurückgesetzt. Weitere Informationen zu `batch`, einschließlich des `--no-requeue` Parameters, finden Sie [sbatch](#) in der Slurm-Dokumentation.

Szenario 3: Spot-Instance, auf der Aufgaben mit mehreren Knoten ausgeführt werden, wird unterbrochen

Der Job schlägt mit dem Bundesstaatencode `fehINODE_FAIL`, und der Job wird in die Warteschlange gestellt (sofern `--no-requeue` er nicht bei der Einreichung des Jobs angegeben wurde). Wenn der Knoten ein statischer Knoten ist, wird er ersetzt. Wenn der Knoten ein dynamischer Knoten ist, wird der Knoten beendet und zurückgesetzt. Andere Knoten, die die beendeten Jobs ausgeführt haben, wurden möglicherweise anderen ausstehenden Aufträgen zugewiesen oder nach Ablauf der konfigurierten [SlurmSettings/ScaledownIdleTime](#)-Zeit herunterskaliert.

Weitere Informationen über Spot-Instances finden Sie unter [Spot-Instances](#) im Amazon EC2-Benutzerhandbuch für Linux-Instances.

Scheduler werden unterstützt von AWS ParallelCluster

Scheduler werden unterstützt von AWS ParallelCluster

AWS ParallelCluster unterstützt Slurm und AWS Batch Scheduler, die mithilfe der [Scheduler](#) Einstellung eingestellt werden.

Themen

- [Slurm Workload Manager \(slurm\)](#)

- [AWS Batch \(awsbatch\)](#)

Slurm Workload Manager (**slurm**)

Größe und Aktualisierung der Clusterkapazität

Die Kapazität des Clusters wird durch die Anzahl der Rechenknoten definiert, die der Cluster skalieren kann. Rechenknoten werden von EC2-Instances unterstützt, die in der AWS ParallelCluster Konfiguration innerhalb der Rechenressourcen definiert sind. Sie sind in Warteschlangen organisiert (`Scheduling/SlurmQueues/ComputeResources`), (`Scheduling/SlurmQueues`) die Partitionen 1:1 zugeordnet Slurm werden.

Innerhalb einer Rechenressource ist es möglich, die Mindestanzahl von Rechenknoten (Instanzen), die immer im Cluster am Laufen gehalten werden müssen (`MinCount`), und die maximale Anzahl von Instanzen, auf die die Rechenressource skaliert werden kann, zu konfigurieren (`MaxCount3`).

AWS ParallelCluster Startet bei der Clustererstellung oder bei einem Cluster-Update so viele EC2-Instances, wie `MinCount` für jede im Cluster definierte Rechenressource (`Scheduling/SlurmQueues/ComputeResources`) konfiguriert sind. Die Instances, die gestartet werden, um die minimale Anzahl von Knoten für Rechenressourcen im Cluster abzudecken, werden als statische Knoten bezeichnet. Einmal gestartet, sollen statische Knoten im Cluster persistent sein und werden nicht vom System beendet, es sei denn, ein bestimmtes Ereignis oder eine bestimmte Bedingung tritt ein. Zu diesen Ereignissen gehören beispielsweise der Ausfall von EC2-Integritätsprüfungen Slurm oder die Änderung des Status des Slurm-Knotens auf DRAIN oder DOWN.

Die EC2-Instances im Bereich von **1** bis '`MaxCount - MinCount`' (`MaxCount` minus `MinCount`), die bei Bedarf gestartet werden, um die erhöhte Last des Clusters zu bewältigen, werden als dynamische Knoten bezeichnet. Sie sind kurzlebig. Sie werden gestartet, um ausstehende Aufträge zu bearbeiten, und werden beendet, sobald sie für einen `Scheduling/SlurmSettings/ScaledownIdleTime` in der Cluster-Konfiguration festgelegten Zeitraum inaktiv bleiben (Standard: 10 Minuten).

Statische Knoten und dynamische Knoten entsprechen dem folgenden Benennungsschema:

- Statische Knoten `<Queue/Name>-st-<ComputeResource/Name>-<num>` wo `<num>` = `1..ComputeResource/MinCount`
- Dynamische Knoten `<Queue/Name>-dy-<ComputeResource/Name>-<num>` wo `<num>` = `1..(ComputeResource/MaxCount - ComputeResource/MinCount)`

Zum Beispiel bei der folgenden AWS ParallelCluster Konfiguration:

```
Scheduling:
  Scheduler: slurm
  SlurmQueues:
    - Name: queue1
      ComputeResources:
        - Name: c5xlarge
          Instances:
            - InstanceType: c5.xlarge
              MinCount: 100
              MaxCount: 150
```

Die folgenden Knoten werden definiert in Slurm

```
$ sinfo
PARTITION AVAIL  TIMELIMIT  NODES  STATE NODELIST
queue1*    up      infinite   50    idle~ queue1-dy-c5xlarge-[1-50]
queue1*    up      infinite  100    idle queue1-st-c5xlarge-[1-100]
```

Wenn es sich bei einer Rechenressource um statische Rechenknoten handelt `MinCount == MaxCount`, sind alle zugehörigen Rechenknoten statisch und alle Instanzen werden zum Zeitpunkt der Clustererstellung/-aktualisierung gestartet und laufen weiter. Beispielsweise:

```
Scheduling:
  Scheduler: slurm
  SlurmQueues:
    - Name: queue1
      ComputeResources:
        - Name: c5xlarge
          Instances:
            - InstanceType: c5.xlarge
              MinCount: 100
              MaxCount: 100
```

```
$ sinfo
```

PARTITION	AVAIL	TIMELIMIT	NODES	STATE	NODELIST
queue1*	up	infinite	100	idle	queue1-st-c5xlarge-[1-100]

Aktualisierung der Cluster-Kapazität

Die Aktualisierung der Clusterkapazität umfasst das Hinzufügen oder Entfernen von Warteschlangen, Rechenressourcen oder das Ändern MinCount/MaxCount einer Rechenressource. Ab AWS ParallelCluster Version 3.9.0 muss zur Reduzierung der Größe einer Warteschlange die Rechenflotte gestoppt oder auf TERMINATE [QueueUpdateStrategy](#) gesetzt werden, bevor ein Cluster-Update stattfinden kann. In folgenden Fällen ist es nicht erforderlich, die Rechenflotte zu beenden oder auf [QueueUpdateStrategy](#) TERMINATE zu setzen:

- Neue Warteschlangen zu Scheduling hinzufügen/ [SlurmQueues](#)
- Neue Rechenressourcen Scheduling/SlurmQueues/[ComputeResources](#) zu einer Warteschlange hinzufügen
- Erhöhung der [MaxCount](#) Anzahl einer Rechenressource
- Erhöhung MinCount einer Rechenressource und Erhöhung MaxCount derselben Rechenressource um mindestens die gleiche Menge

Überlegungen und Einschränkungen

In diesem Abschnitt sollen alle wichtigen Faktoren, Einschränkungen oder Einschränkungen beschrieben werden, die bei der Größenänderung der Clusterkapazität berücksichtigt werden sollten.

- Beim Entfernen einer Warteschlange aus Scheduling/<https://docs.aws.amazon.com/parallelcluster/latest/ug/Scheduling-v3.html#Scheduling-v3-SlurmQueues> SlurmQueues allen Rechenknoten mit Namen <Queue/Name> - *, sowohl statische als auch dynamische, werden sie aus der Slurm Konfiguration entfernt und die entsprechenden EC2-Instances werden beendet.
- Beim Entfernen einer Rechenressource Scheduling/SlurmQueues/<https://docs.aws.amazon.com/parallelcluster/latest/ug/Scheduling-v3.html#Scheduling-v3-SlurmQueues-ComputeResources> ComputeResources aus einer Warteschlange werden alle Rechenknoten mit Namen <Queue/Name> - * - <ComputeResource/Name> - *, sowohl statische als auch dynamische, aus der Slurm Konfiguration entfernt und die entsprechenden EC2-Instances werden beendet.

Wenn wir den `MinCount` Parameter einer Rechenressource ändern, können wir zwei verschiedene Szenarien unterscheiden: ob gleich gehalten `MaxCount` wird `MinCount` (nur statische Kapazität) und ob größer als `MinCount` (gemischte statische und dynamische Kapazität) `MaxCount` ist.

Die Kapazität ändert sich nur bei statischen Knoten

- Wenn `MinCount == MaxCount` beim Erhöhen `MinCount` (und `MaxCount`) der Cluster konfiguriert wird, indem die Anzahl der statischen Knoten auf den neuen Wert von erhöht wird `MinCount <Queue/Name>-st-<ComputeResource/Name>-<new_MinCount>` und das System weiterhin versucht, EC2-Instances zu starten, um die neue erforderliche statische Kapazität zu erreichen.
- Wenn `MinCount == MaxCount` beim Verringern `MinCount` (und `MaxCount`) der Anzahl `N` der Cluster konfiguriert wird, indem die letzten `N` statischen Knoten `<Queue/Name>-st-<ComputeResource/Name>-<old_MinCount - N>...<old_MinCount>` entfernt werden, beendet das System die entsprechenden EC2-Instances.
- Ausgangszustand `MinCount = MaxCount = 100`

```
$ sinfo
PARTITION AVAIL  TIMELIMIT  NODES  STATE NODELIST
queue1*    up       infinite   100   idle queue1-st-c5xlarge-[1-100]
```

- Update `-30` am `MinCount` und `MaxCount`: `MinCount = MaxCount = 70`

```
$ sinfo
PARTITION AVAIL  TIMELIMIT  NODES  STATE NODELIST
queue1*    up       infinite    70   idle queue1-st-c5xlarge-[1-70]
```

Kapazitätsänderungen bei gemischten Knoten

Wenn bei einer Erhöhung `MinCount` um den Betrag `N` (vorausgesetzt `MinCount < MaxCount`, dass der Wert unverändert bleibt), der Cluster konfiguriert wird, indem die Anzahl der statischen Knoten auf den neuen Wert von `MinCount (old_MinCount + N)`: erhöht wird, `<Queue/Name>-st-<ComputeResource/Name>-<old_MinCount + N>` und

das System versucht weiterhin, EC2-Instances zu starten, um die neue erforderliche statische Kapazität zu erreichen. Um der MaxCount Kapazität der Rechenressource Rechnung zu tragen, wird außerdem die Cluster-Konfiguration aktualisiert, indem die letzten N dynamischen Knoten entfernt werden. `<Queue/Name>-dy-<ComputeResource/Name>-[<MaxCount - old_MinCount - N>...<MaxCount - old_MinCount>]` Das System beendet dann die entsprechenden EC2-Instances.

- Ausgangszustand: MinCount = 100; MaxCount = 150

```
$ sinfo
PARTITION AVAIL  TIMELIMIT  NODES  STATE NODELIST
queue1*    up    infinite   50    idle~ queue1-dy-c5xlarge-[1-50]
queue1*    up    infinite  100    idle queue1-st-c5xlarge-[1-100]
```

- Aktualisieren Sie +30 auf MinCount : MinCount = 130 (MaxCount = 150)

```
$ sinfo
PARTITION AVAIL  TIMELIMIT  NODES  STATE NODELIST
queue1*    up    infinite   20    idle~ queue1-dy-c5xlarge-[1-20]
queue1*    up    infinite  130    idle queue1-st-c5xlarge-[1-130]
```

Wenn `MinCount < MaxCount` MaxCount der Cluster bei Erhöhung MinCount und gleichem Wert N konfiguriert wird, wird der Cluster konfiguriert, indem die Anzahl der statischen Knoten auf den neuen Wert von MinCount (`old_MinCount + N`): erhöht wird `<Queue/Name>-st-<ComputeResource/Name>-<old_MinCount + N>` und das System weiterhin versucht, EC2-Instances zu starten, um die neue erforderliche statische Kapazität zu erreichen. Darüber hinaus werden keine Änderungen an der Anzahl der dynamischen Knoten vorgenommen, um den neuen Anforderungen gerecht zu werden

MaxCount Wert.

- Ausgangszustand: MinCount = 100; MaxCount = 150

```
$ sinfo
```



```

PARTITION AVAIL  TIMELIMIT  NODES  STATE NODELIST
queue1*    up    infinite   50    idle~ queue1-dy-c5xlarge-[1-50]
queue1*    up    infinite  100    idle queue1-st-c5xlarge-[1-100]

```

- Aktualisieren Sie +30 auf MinCount : MinCount = 130 (MaxCount = 180)

```

$ sinfo
PARTITION AVAIL  TIMELIMIT  NODES  STATE NODELIST
queue1*    up    infinite   20    idle~ queue1-dy-c5xlarge-[1-50]
queue1*    up    infinite  130    idle queue1-st-c5xlarge-[1-130]

```

Wenn die MinCount Anzahl N verringert MaxCount wird (vorausgesetzt $\text{MinCount} < \text{MaxCount}$, sie bleibt unverändert), wird der Cluster konfiguriert, indem die letzten N statischen Knoten entfernt werden, `<Queue/Name>-st-<ComputeResource/Name>-[<old_MinCount - N>...<old_MinCount>` und das System beendet die entsprechenden EC2-Instances. Um der MaxCount Kapazität der Rechenressource Rechnung zu tragen, wird außerdem die Cluster-Konfiguration aktualisiert, indem die Anzahl der dynamischen Knoten erhöht wird, um die Lücke zu schließen. `MaxCount - new_MinCount: <Queue/Name>-dy-<ComputeResource/Name>-[1..<MaxCount - new_MinCount>]` In diesem Fall, da es sich um dynamische Knoten handelt, werden keine neuen EC2-Instances gestartet, es sei denn, der Scheduler hat Jobs auf den neuen Knoten ausstehend.

- Ausgangszustand: MinCount = 100; MaxCount = 150

```

$ sinfo
PARTITION AVAIL  TIMELIMIT  NODES  STATE NODELIST
queue1*    up    infinite   50    idle~ queue1-dy-c5xlarge-[1-50]
queue1*    up    infinite  100    idle queue1-st-c5xlarge-[1-100]

```

- Update -30 am MinCount : MinCount = 70 (MaxCount = 120)

```

$ sinfo

```

```

PARTITION AVAIL  TIMELIMIT  NODES  STATE NODELIST
queue1*   up    infinite   80    idle~ queue1-dy-c5xlarge-[1-80]
queue1*   up    infinite   70    idle queue1-st-c5xlarge-[1-70]

```

Wenn $\text{MinCount} < \text{MaxCount}$ der Wert abnimmt MinCount und die MaxCount Anzahl N gleich ist, wird der Cluster konfiguriert, indem die letzten N statischen Knoten entfernt werden, `<Queue/Name>-st-<ComputeResource/Name>-<old_MinCount - N>...<oldMinCount>`] und das System beendet die entsprechenden EC2-Instances.

Außerdem werden keine Änderungen an der Anzahl der dynamischen Knoten vorgenommen, um dem neuen MaxCount Wert Rechnung zu tragen.

- Ausgangszustand: $\text{MinCount} = 100$; $\text{MaxCount} = 150$

```

$ sinfo
PARTITION AVAIL  TIMELIMIT  NODES  STATE NODELIST
queue1*   up    infinite   50    idle~ queue1-dy-c5xlarge-[1-50]
queue1*   up    infinite  100    idle queue1-st-c5xlarge-[1-100]

```

- Update -30 am MinCount : $\text{MinCount} = 70$ ($\text{MaxCount} = 120$)

```

$ sinfo
PARTITION AVAIL  TIMELIMIT  NODES  STATE NODELIST
queue1*   up    infinite   80    idle~ queue1-dy-c5xlarge-[1-50]
queue1*   up    infinite   70    idle queue1-st-c5xlarge-[1-70]

```

Wenn die MaxCount Anzahl N verringert wird (vorausgesetzt $\text{MinCount} < \text{MaxCount}$, sie bleibt unverändert), MinCount wird der Cluster so konfiguriert, dass die letzten N dynamischen Knoten entfernt werden, `<Queue/Name>-dy-<ComputeResource/Name>-<old_MaxCount - N...<oldMaxCount>`] und das System beendet die entsprechenden EC2-Instances, sofern sie ausgeführt wurden. Es sind keine Auswirkungen auf die statischen Knoten zu erwarten.

- Ausgangszustand: MinCount = 100; MaxCount = 150

```
$ sinfo
PARTITION AVAIL  TIMELIMIT  NODES  STATE NODELIST
queue1*   up    infinite   50    idle~ queue1-dy-c5xlarge-[1-50]
queue1*   up    infinite  100    idle  queue1-st-c5xlarge-[1-100]
```

- Update -30 am MaxCount : MinCount = 100 (MaxCount = 120)

```
$ sinfo
PARTITION AVAIL  TIMELIMIT  NODES  STATE NODELIST
queue1*   up    infinite   20    idle~ queue1-dy-c5xlarge-[1-20]
queue1*   up    infinite  100    idle  queue1-st-c5xlarge-[1-100]
```

Auswirkungen auf die Arbeitsplätze

In allen Fällen, in denen Knoten entfernt und EC2-Instances beendet werden, wird ein Sbatch-Job, der auf den entfernten Knoten ausgeführt wird, erneut in die Warteschlange gestellt, es sei denn, es gibt keine anderen Knoten, die die Job-Anforderungen erfüllen. In letzterem Fall schlägt der Job mit dem Status `NODE_FAIL` fehl und verschwindet aus der Warteschlange. In diesem Fall muss er manuell erneut eingereicht werden.

Wenn Sie planen, ein Update zur Änderung der Clustergröße durchzuführen, können Sie verhindern, dass Jobs auf den Knoten ausgeführt werden, die während des geplanten Updates entfernt werden. Dies ist möglich, indem Sie festlegen, dass die Knoten im Rahmen der Wartung entfernt werden. Bitte beachten Sie, dass die Einstellung eines Knotens zur Wartung keine Auswirkungen auf Jobs hat, die eventuell bereits auf dem Knoten ausgeführt werden.

Nehmen wir an, dass Sie mit dem geplanten Update zur Änderung der Clustergröße den Knoten entfernen werden `queue1-st-computeresource-[9-10]`. Sie können eine Slurm Reservierung mit dem folgenden Befehl erstellen

```
sudo -i scontrol create reservation ReservationName=maint_for_update user=root
starttime=now duration=infinite flags=maint,ignore_jobs nodes=queue-st-
computeresource-[9-10]
```

Dadurch wird eine Slurm Reservierung erstellt, die `maint_for_update` auf den Knoten benannt ist `queue-st-computeresource-[9-10]`. Ab dem Zeitpunkt, an dem die Reservierung erstellt wurde, können keine Jobs mehr auf den Knoten ausgeführt werden `queue-st-computeresource-[9-10]`. Bitte beachten Sie, dass die Reservierung nicht verhindert, dass Jobs irgendwann auf den Knoten zugewiesen werden `queue-st-computeresource-[9-10]`.

Wenn die Slurm Reservierung nach der Aktualisierung der Clustergröße nur für Knoten festgelegt wurde, die während der Aktualisierung der Clustergröße entfernt wurden, wird die Wartungsreservierung automatisch gelöscht. Wenn Sie stattdessen eine Slurm Reservierung für Knoten erstellt haben, die nach der Aktualisierung der Clustergröße noch vorhanden sind, möchten wir möglicherweise die Wartungsreservierung auf den Knoten nach der Aktualisierung der Größe entfernen. Verwenden Sie dazu den folgenden Befehl

```
sudo -i scontrol delete ReservationName=maint_for_update
```

[Weitere Informationen zur Slurm Reservierung finden Sie im offiziellen SchedMD-Dokument hier.](#)

Cluster-Aktualisierungsprozess bei Kapazitätsänderungen

Bei einer Änderung der Scheduler-Konfiguration werden während des Cluster-Aktualisierungsvorgangs die folgenden Schritte ausgeführt:

- Stoppen AWS ParallelCluster `clustermgtd` (`supervisorctl stop clustermgtd`)
- Generieren Sie die aktualisierte Slurm Partitionskonfiguration anhand der AWS ParallelCluster Konfiguration
- Neustart `slurmctld` (erfolgt über das Chef-Dienstrezept)
- `slurmctldStatus` überprüfen (`systemctl is-active --quiet slurmctld.service`)
- Konfiguration neu laden Slurm (`scontrol reconfigure`)
- `clustermgtd` (`supervisorctl start clustermgtd`) starten

Informationen zu finden Sie Slurm unter <https://slurm.schedmd.com>. Downloads finden Sie unter <https://github.com/SchedMD/slurm/tags>. Den Quellcode finden Sie unter <https://github.com/SchedMD/slurm>.

AWS ParallelCluster Version (en)	Unterstützte Slurm-Version
3.9.0	23.11.4
3.8.0	23.02,7
3.7.2	23.02,6
3.7.1	23,02,5
3.7.0	23,02,4
3.6.0, 3.6.1	23,02,2
3,5.0, 3,5.1	22.05.8
3,4,0, 3,4,1	22.05.7
3.3.0, 3.3.1	22.05.5
3.1.4, 3.1.5, 3.2.0, 3.2.1	21.08.8-2
3.1.2, 3.1.3	21,08,6
3.1.1	21,08,5
3.0.0	20,11,8

Themen

- [Konfiguration mehrerer Warteschlangen](#)
- [SlurmLeitfaden für den Modus mit mehreren Warteschlangen](#)
- [SlurmCluster-geschützter Modus](#)
- [Slurmschneller Cluster-Failover mit unzureichender Kapazität](#)
- [Slurmspeicherbasierte Terminplanung](#)
- [Zuweisung mehrerer Instanztypen mit Slurm](#)
- [Cluster-Skalierung für dynamische Knoten](#)
- [SlurmAbrechnung mit AWS ParallelCluster](#)

- [SlurmAnpassung der Konfiguration](#)
- [Slurm prolog und epilog](#)
- [Größe und Aktualisierung der Clusterkapazität](#)

Konfiguration mehrerer Warteschlangen

Konfiguration mehrerer Warteschlangen

Mit AWS ParallelCluster Version 3 können Sie mehrere Warteschlangen konfigurieren, indem Sie den Wert [Scheduler](#) auf `slurm` und [SlurmQueues](#) in der Konfigurationsdatei mehrere Warteschlangen für angeben. In diesem Modus existieren verschiedene Instanztypen gleichzeitig in den Rechenknoten, die im [ComputeResources](#) Abschnitt der Konfigurationsdatei angegeben sind. [ComputeResources](#) mit unterschiedlichen Instanztypen werden je nach Bedarf für die nach oben oder unten skaliert. [SlurmQueues](#)

Cluster-Warteschlange und Rechenressourcenkontingente

Ressource	Kontingent
Slurm queues	50 Warteschlangen pro Cluster
Compute resources	50 Rechenressourcen pro Warteschlange 50 Rechenressourcen pro Cluster

Anzahl der Knoten

Jede Rechenressource in [ComputeResources](#) einer Warteschlange muss ein eindeutiges [Name](#), [InstanceTypeMinCount](#), und haben [MaxCount](#). [MinCount](#) und [MaxCount](#) verfügen über Standardwerte, die den Instanzbereich für eine Rechenressource in [ComputeResources](#) einer Warteschlange definieren. Sie können auch Ihre eigenen Werte für [MinCount](#) und [MaxCount](#) angeben. Jede Rechenressource in [ComputeResources](#) besteht aus statischen Knoten, die von 1 bis zum Wert von nummeriert sind, [MinCount](#) und dynamischen Knoten, die vom Wert [MinCount](#) bis zum Wert von nummeriert sind [MaxCount](#).

Beispiel für eine Konfiguration

Im Folgenden finden Sie ein Beispiel für einen [Scheduling-Abschnitt](#) für eine Cluster-Konfigurationsdatei. In dieser Konfiguration gibt es zwei Warteschlangen [ComputeResources](#) mit

dem Namen queue1 und, queue2 und jede der Warteschlangen hat einen bestimmten Wert.

MaxCount

```
Scheduling:
  Scheduler: slurm
  SlurmQueues:
  - Name: queue1
    ComputeResources:
    - InstanceType: c5.xlarge
      MaxCount: 5
      Name: c5xlarge
    - InstanceType: c4.xlarge
      MaxCount: 5
      Name: c4xlarge
  - Name: queue2
    ComputeResources:
    - InstanceType: c5.xlarge
      MaxCount: 5
      Name: c5xlarge
```

Hostnamen

Die Instances, die in die Compute-Flotte gestartet werden, werden dynamisch zugewiesen. Hostnamen werden für jeden Knoten generiert. Standardmäßig AWS ParallelCluster wird das folgende Format des Hostnamens verwendet:

```
$HOSTNAME=$QUEUE-$STATDYN-$COMPUTE_RESOURCE-$NODENUM
```

- \$QUEUE ist der Name der Warteschlange. Wenn der [SlurmQueues](#) Abschnitt beispielsweise einen Eintrag hat, dessen Wert auf „queue-name“ [Name](#) gesetzt ist, dann ist „\$QUEUE“ der Wert „queue-name“.
- \$STATDYN ist st für statische Knoten oder dy für dynamische Knoten.
- \$COMPUTE_RESOURCE ist [Name](#) die [ComputeResources](#) Rechenressource, die diesem Knoten entspricht.
- \$NODENUM ist die Nummer des Knotens. \$NODENUM liegt zwischen eins (1) und dem Wert von [MinCount](#) für statische Knoten und zwischen eins (1) und [MaxCount](#) - [MinCount](#) für dynamische Knoten.

Aus der obigen Beispielkonfigurationsdatei geht hervor, dass ein bestimmter Knoten aus queue1 einer Rechenressource einen Hostnamen c5xlarge hat: queue1-dy-c5xlarge-1.

Sowohl Hostnamen als auch vollqualifizierte Domainnamen (FQDN) werden mithilfe von Amazon Route 53-Hosting-Zonen erstellt. Der FQDN ist `$HOSTNAME.$CLUSTERNAME.pcluster`, wo der Name des `$CLUSTERNAME` Clusters steht.

Beachten Sie, dass dasselbe Format auch für die Slurm-Knotennamen verwendet wird.

Benutzer können wählen, ob sie den standardmäßigen EC2-Hostnamen der Instance verwenden möchten, die den Rechenknoten mit Strom versorgt, anstatt das Standard-Hostnamenformat, das von verwendet wird. AWS ParallelCluster Dies kann erreicht werden, indem der [UseEc2Hostnames](#) Parameter auf `true` gesetzt wird. Slurm-Knotennamen werden jedoch weiterhin das AWS ParallelCluster Standardformat verwenden.

SlurmLeitfaden für den Modus mit mehreren Warteschlangen

Hier erfahren Sie, wie AWS ParallelCluster und wie Sie Warteschlangenknoten (Partitionsknoten) Slurm verwalten und wie Sie die Warteschlangen- und Knotenstatus überwachen können.

Übersicht

Die Skalierungsarchitektur basiert auf dem [Cloud Slurm Scheduling Guide](#) und dem Energiespar-Plugin. Weitere Informationen zum Energiespar-Plugin finden Sie im [SlurmPower Saving Guide](#). In der Architektur sind Ressourcen, die potenziell für einen Cluster verfügbar gemacht werden können, in der Slurm Konfiguration typischerweise als Cloud-Knoten vordefiniert.

Lebenszyklus eines Cloud-Knotens

Während ihres gesamten Lebenszyklus treten Cloud-Knoten in mehrere, wenn nicht sogar alle der folgenden Zustände ein: `POWER_SAVING`, `POWER_UP` (`pow_up`), `ALLOCATED` (`alloc`) und `POWER_DOWN` (`pow_dn`). In einigen Fällen kann ein Cloud-Knoten in den `OFFLINE` Status wechseln. In der folgenden Liste werden verschiedene Aspekte dieser Zustände im Lebenszyklus eines Cloud-Knotens beschrieben.

- Ein Knoten in einem **POWER_SAVING** Status wird mit einem `~` Suffix (zum Beispiel `idle~`) in `sinfo` angezeigt. In diesem Status unterstützen keine EC2-Instances den Knoten. Slurm Kann dem Knoten jedoch weiterhin Jobs zuweisen.
- Ein Knoten, der in einen **POWER_UP** Status übergeht, wird mit einem `#` Suffix (zum Beispiel `idle#`) in `sinfo` angezeigt. Ein Knoten wechselt automatisch in einen `POWER_UP` Status, wenn er einem Knoten in einem Status einen Job Slurm zuweist. `POWER_SAVING`

Alternativ können Sie die Knoten manuell als su Root-Benutzer mit dem folgenden Befehl in den POWER_UP Status überführen:

```
$ scontrol update nodename=nodename state=power_up
```

In dieser Phase ResumeProgram wird der aufgerufen, EC2-Instances werden gestartet und konfiguriert und der Knoten wechselt in den Status. POWER_UP

- Ein Knoten, der derzeit zur Verwendung verfügbar ist, wird ohne Suffix angezeigt (z. B. `idle`) in `sinfo` Nachdem der Knoten eingerichtet wurde und dem Cluster beigetreten ist, steht er für die Ausführung von Jobs zur Verfügung. In dieser Phase ist der Knoten ordnungsgemäß konfiguriert und einsatzbereit.

In der Regel empfehlen wir, dass die Anzahl der EC2-Instances der Anzahl der verfügbaren Knoten entspricht. In den meisten Fällen sind statische Knoten verfügbar, nachdem der Cluster erstellt wurde.

- Ein Knoten, der in einen **POWER_DOWN** Status übergeht, wird mit einem % Suffix (z. B. `idle%`) in `sinfo` angezeigt. Dynamische Knoten wechseln automatisch in den POWER_DOWN Status danach. [ScaledownIdleTime](#) Im Gegensatz dazu werden statische Knoten in den meisten Fällen nicht abgeschaltet. Sie können die Knoten jedoch manuell als su Root-Benutzer mit dem folgenden Befehl in den POWER_DOWN Status versetzen:

```
$ scontrol update nodename=nodename state=down reason="manual draining"
```

In diesem Zustand werden die mit einem Knoten verknüpften Instanzen beendet, und der Knoten wird in den POWER_SAVING Status zurückversetzt und kann danach wieder verwendet [ScaledownIdleTime](#) werden.

Die [ScaledownIdleTime](#) Einstellung wird in der Slurm SuspendTimeout Konfigurationseinstellung gespeichert.

- Ein Knoten, der offline ist, wird mit einem * Suffix (z. B. `down*`) in `sinfo` angezeigt. Ein Knoten geht offline, wenn der Slurm Controller den Knoten nicht kontaktieren kann oder wenn die statischen Knoten deaktiviert sind und die unterstützenden Instanzen beendet werden.

Betrachten Sie die im folgenden `sinfo` Beispiel gezeigten Knotenzustände.

```
$ sinfo
```

PARTITION	AVAIL	TIMELIMIT	NODES	STATE	NODELIST
efa	up	infinite	4	idle~	efa-dy-efacompute1-[1-4]
efa	up	infinite	1	idle	efa-st-efacompute1-1
gpu	up	infinite	1	idle%	gpu-dy-gpucompute1-1
gpu	up	infinite	9	idle~	gpu-dy-gpucompute1-[2-10]
ondemand	up	infinite	2	mix#	ondemand-dy-ondemandcompute1-[1-2]
ondemand	up	infinite	18	idle~	ondemand-dy-ondemandcompute1-[3-10],ondemand-dy-ondemandcompute2-[1-10]
spot*	up	infinite	13	idle~	spot-dy-spotcompute1-[1-10],spot-dy-spotcompute2-[1-3]
spot*	up	infinite	2	idle	spot-st-spotcompute2-[1-2]

Für die `efa-st-efacompute1-1` Knoten `spot-st-spotcompute2-[1-2]` und `ondemand-dy-ondemandcompute1-[1-2]` sind bereits Backing-Instances eingerichtet, sodass sie verwendet werden können. Die `ondemand-dy-ondemandcompute1-[1-2]` Knoten befinden sich im `POWER_UP` Status und sollten innerhalb weniger Minuten verfügbar sein. Der `gpu-dy-gpucompute1-1` Knoten befindet sich im `POWER_DOWN` Status und geht danach in den `POWER_SAVING` Status über [ScaledownIdletime](#) (standardmäßig 10 Minuten).

Alle anderen Knoten befinden sich im `POWER_SAVING` Status, ohne dass sie von EC2-Instances unterstützt werden.

Mit einem verfügbaren Knoten arbeiten

Ein verfügbarer Knoten wird von einer EC2-Instance unterstützt. Standardmäßig kann der Knotenname verwendet werden, um per SSH direkt auf die Instanz zuzugreifen (zum Beispielssh `efa-st-efacompute1-1`). Die private IP-Adresse der Instanz kann mit dem folgenden Befehl abgerufen werden:

```
$ scontrol show nodes nodename
```

Suchen Sie im zurückgegebenen `NodeAddr` Feld nach der IP-Adresse.

Bei Knoten, die nicht verfügbar sind, sollte das `NodeAddr` Feld nicht auf eine laufende EC2-Instance verweisen. Vielmehr sollte es mit dem Knotennamen identisch sein.

Status und Einreichung von Job

In den meisten Fällen werden übermittelte Jobs sofort Knoten im System zugewiesen oder als ausstehend eingestuft, wenn alle Knoten zugewiesen sind.

Wenn die für einen Job zugewiesenen Knoten in einem POWER_SAVING Status enthalten, beginnt der Job mit einem CF oder CONFIGURING. Zu diesem Zeitpunkt wartet der Job darauf, dass die Knoten im POWER_SAVING Status in den Status übergehen POWER_UP und verfügbar werden.

Nachdem alle für einen Job zugewiesenen Knoten verfügbar sind, wechselt der Job in den Status RUNNING (R).

Standardmäßig werden alle Jobs an die Standardwarteschlange (auch Partition in Slurm) übergeben. Dies wird durch ein * Suffix nach dem Warteschlangennamen gekennzeichnet. Sie können mit der Option zum Einreichen von -p Jobs eine Warteschlange auswählen.

Alle Knoten sind mit den folgenden Funktionen konfiguriert, die in Befehlen zur Auftragsübermittlung verwendet werden können:

- Ein Instanztyp (zum Beispiel c5.xlarge)
- Ein Knotentyp (Dies ist entweder dynamic oder static.)

Sie können sich die Funktionen für einen bestimmten Knoten anzeigen lassen, indem Sie den folgenden Befehl verwenden:

```
$ scontrol show nodes nodename
```

Überprüfen Sie im Gegenzug die AvailableFeatures Liste.

Betrachten Sie den Anfangsstatus des Clusters, den Sie anzeigen können, indem Sie den sinfo Befehl ausführen.

```
$ sinfo
PARTITION AVAIL  TIMELIMIT  NODES  STATE NODELIST
efa        up    infinite   4    idle~ efa-dy-efacompute1-[1-4]
efa        up    infinite   1    idle efa-st-efacompute1-1
gpu        up    infinite  10    idle~ gpu-dy-gpucompute1-[1-10]
ondemand  up    infinite  20    idle~ ondemand-dy-ondemandcompute1-
[1-10],ondemand-dy-ondemandcompute2-[1-10]
spot*      up    infinite  13    idle~ spot-dy-spotcompute1-[1-10],spot-dy-
spotcompute2-[1-3]
spot*      up    infinite   2    idle spot-st-spotcompute2-[1-2]
```

Beachten Sie, dass spot dies die Standardwarteschlange ist. Sie wird durch das * Suffix angezeigt.

Sendet einen Job an einen statischen Knoten in der Standardwarteschlange (spot).

```
$ sbatch --wrap "sleep 300" -N 1 -C static
```

Sendet einen Job an einen dynamischen Knoten in der EFA Warteschlange.

```
$ sbatch --wrap "sleep 300" -p efa -C dynamic
```

Sendet einen Job an acht (8) c5.2xlarge Knoten und zwei (2) t2.xlarge Knoten in der ondemand Warteschlange.

```
$ sbatch --wrap "sleep 300" -p ondemand -N 10 -C "[c5.2xlarge*8&t2.xlarge*2]"
```

Senden Sie einen Job an einen GPU-Knoten in der gpu Warteschlange.

```
$ sbatch --wrap "sleep 300" -p gpu -G 1
```

Berücksichtigen Sie den Status der Jobs, die den squeue Befehl verwenden.

```
$ squeue
JOBID PARTITION NAME USER ST TIME NODES NODELIST(REASON)
12 ondemand wrap ubuntu CF 0:36 10 ondemand-dy-ondemandcompute1-
[1-8],ondemand-dy-ondemandcompute2-[1-2]
13 gpu wrap ubuntu CF 0:05 1 gpu-dy-gpucompute1-1
7 spot wrap ubuntu R 2:48 1 spot-st-spotcompute2-1
8 efa wrap ubuntu R 0:39 1 efa-dy-efacompute1-1
```

Die Jobs 7 und 8 (in den efa Warteschlangen spot und) werden bereits ausgeführt (R). Die Jobs 12 und 13 werden noch konfiguriert (CF) und warten wahrscheinlich darauf, dass Instanzen verfügbar werden.

```
# Nodes states corresponds to state of running jobs
$ sinfo
PARTITION AVAIL TIMELIMIT NODES STATE NODELIST
efa up infinite 3 idle~ efa-dy-efacompute1-[2-4]
efa up infinite 1 mix efa-dy-efacompute1-1
efa up infinite 1 idle efa-st-efacompute1-1
gpu up infinite 1 mix~ gpu-dy-gpucompute1-1
gpu up infinite 9 idle~ gpu-dy-gpucompute1-[2-10]
ondemand up infinite 10 mix# ondemand-dy-ondemandcompute1-[1-8],ondemand-
dy-ondemandcompute2-[1-2]
```

ondemand	up	infinite	10	idle~	ondemand-dy-ondemandcompute1-[9-10],ondemand-dy-ondemandcompute2-[3-10]
spot*	up	infinite	13	idle~	spot-dy-spotcompute1-[1-10],spot-dy-spotcompute2-[1-3]
spot*	up	infinite	1	mix	spot-st-spotcompute2-1
spot*	up	infinite	1	idle	spot-st-spotcompute2-2

Status und Funktionen des Knotens

In den meisten Fällen werden Knotenstatus AWS ParallelCluster gemäß den spezifischen Prozessen im Cloud-Knoten-Lebenszyklus, die weiter oben in diesem Thema beschrieben wurden, vollständig verwaltet.

Ersetzt oder beendet jedoch AWS ParallelCluster auch fehlerhafte Knoten in DRAINED Zuständen DOWN und Knoten, die über fehlerhafte Backing-Instances verfügen. Weitere Informationen finden Sie unter [clustermgtd](#).

Status der Partition

AWS ParallelCluster unterstützt die folgenden Partitionsstatus. Eine Slurm Partition ist eine Warteschlange in AWS ParallelCluster.

- UP: Zeigt an, dass sich die Partition in einem aktiven Zustand befindet. Dies ist der Standardstatus einer Partition. In diesem Zustand sind alle Knoten in der Partition aktiv und können verwendet werden.
- INACTIVE: Zeigt an, dass sich die Partition im inaktiven Zustand befindet. In diesem Zustand werden alle Instanzen, die Knoten einer inaktiven Partition unterstützen, beendet. Neue Instanzen werden für Knoten in einer inaktiven Partition nicht gestartet.

pcluster update-compute-fleet

- Stoppen der Rechenflotte — Wenn der folgende Befehl ausgeführt wird, gehen alle Partitionen in den INACTIVE Status über, und AWS ParallelCluster Prozesse behalten die Partitionen im INACTIVE Status bei.

```
$ pcluster update-compute-fleet --cluster-name testSlurm \
  --region eu-west-1 --status STOP_REQUESTED
```

- Starten der Compute-Flotte — Wenn der folgende Befehl ausgeführt wird, wechseln zunächst alle Partitionen in den UP Status. AWS ParallelCluster Prozesse halten die Partition jedoch nicht

in einem UP Zustand. Sie müssen den Partitionsstatus manuell ändern. Alle statischen Knoten sind nach einigen Minuten verfügbar. Beachten Sie, dass das Einstellen einer Partition auf keine dynamische Kapazität erhöht. UP

```
$ pcluster update-compute-fleet --cluster-name testSlurm \  
  --region eu-west-1 --status START_REQUESTED
```

Wenn `update-compute-fleet` es ausgeführt wird, können Sie den Status des Clusters überprüfen, indem Sie den `pcluster describe-compute-fleet` Befehl ausführen und das überprüfenStatus. In der folgenden Liste sind mögliche Status aufgeführt:

- `STOP_REQUESTED`: Die Flottenanforderung „Stop Compute“ wird an den Cluster gesendet.
- `STOPPING`: Der `pcluster` Prozess stoppt derzeit die Compute-Flotte.
- `STOPPED`: Der `pcluster` Prozess hat den Stoppvorgang abgeschlossen, alle Partitionen befinden sich im `INACTIVE` Status und alle Recheninstanzen wurden beendet.
- `START_REQUESTED`: Die Anfrage zum Starten der Compute-Flotte wird an den Cluster gesendet.
- `STARTING`: Der `pcluster` Prozess startet derzeit den Cluster.
- `RUNNING`: Der `pcluster` Prozess hat den Startvorgang abgeschlossen, alle Partitionen befinden sich im `UP` Status und statische Knoten sind nach einigen Minuten verfügbar.
- `PROTECTED`: Dieser Status weist darauf hin, dass bei einigen Partitionen konsistente Bootstrap-Fehler auftreten. Die betroffenen Partitionen sind inaktiv. Bitte untersuchen Sie das Problem und starten Sie dann den Vorgang `update-compute-fleet`, um die Flotte erneut zu aktivieren.

Manuelle Steuerung von Warteschlangen

In einigen Fällen möchten Sie möglicherweise die Knoten oder die Warteschlange (auch Partition genannt `Slurm`) in einem Cluster manuell steuern. Mithilfe des `scontrol` Befehls können Sie Knoten in einem Cluster mithilfe der folgenden gängigen Verfahren verwalten.

- Dynamische Knoten im **POWER_SAVING** Status einschalten

Führen Sie den Befehl als `su` Root-Benutzer aus:

```
$ scontrol update nodename=nodename state=power_up
```

Sie können auch einen `sleep 1` Platzhalter-Job einreichen, der eine bestimmte Anzahl von Knoten anfordert, und sich dann Slurm darauf verlassen, dass die erforderliche Anzahl von Knoten eingeschaltet wird.

- Schalten Sie dynamische Knoten vorher aus [ScaledownIdleTime](#)

Wir empfehlen, dynamische Knoten mit dem DOWN folgenden Befehl auf su Root-Benutzer einzustellen:

```
$ scontrol update nodename=nodename state=down reason="manually draining"
```

AWS ParallelCluster beendet die ausgefallenen dynamischen Knoten automatisch und setzt sie zurück.

Im Allgemeinen wird davon abgeraten, Knoten `POWER_DOWN` direkt mithilfe des Befehls `scontrol update nodename=nodename state=power_down` Das liegt daran, dass der Abschaltvorgang AWS ParallelCluster automatisch abgewickelt wird.

- Deaktivieren Sie eine Warteschlange (Partition) oder stoppen Sie alle statischen Knoten in einer bestimmten Partition

Stellen Sie eine bestimmte Warteschlange mit dem folgenden Befehl `INACTIVE` als su Root-Benutzer ein:

```
$ scontrol update partition=queuename state=inactive
```

Dadurch werden alle Instanzen beendet, die Knoten in der Partition unterstützen.

- Aktiviert eine Warteschlange (Partition)

Stellen Sie mit dem folgenden Befehl `UP` eine bestimmte Warteschlange für einen su Root-Benutzer ein:

```
$ scontrol update partition=queuename state=up
```

Skalierungsverhalten und Anpassungen

Hier ist ein Beispiel für den normalen Skalierungsablauf:

- Der Scheduler empfängt einen Job, für den zwei Knoten erforderlich sind.
- Der Scheduler versetzt zwei Knoten in einen `POWER_UP` Status und ruft `ResumeProgram` mit den Knotennamen auf (zum Beispiel `queue1-dy-spotcompute1-[1-2]`).
- `ResumeProgram` startet zwei EC2-Instances und weist die privaten IP-Adressen und Hostnamen von zu. Wartet darauf `ResumeTimeout` (der Standardzeitraum beträgt 30 Minuten) `queue1-dy-spotcompute1-[1-2]`, bevor die Knoten zurückgesetzt werden.
- Die Instanzen werden konfiguriert und treten dem Cluster bei. Ein Job wird auf Instanzen ausgeführt.
- Der Job wird abgeschlossen und wird nicht mehr ausgeführt.
- Nach Ablauf der konfigurierten `SuspendTime` Zeit (die auf gesetzt ist [ScaledownIdletime](#)), setzt der Scheduler die Instanzen auf den Status `POWER_SAVING`. Der Scheduler wechselt dann `queue1-dy-spotcompute1-[1-2]` auf den `POWER_DOWN` Status und ruft `SuspendProgram` mit den Knotennamen auf.
- `SuspendProgram` wird für zwei Knoten aufgerufen. Knoten bleiben in diesem `POWER_DOWN` Zustand, z. B. indem sie eine Zeit `idle%lang` verbleiben `SuspendTimeout` (der Standardzeitraum beträgt 120 Sekunden (2 Minuten)). Nachdem `clustermgtd` erkannt wurde, dass Knoten heruntergefahren werden, werden die unterstützenden Instances beendet. Anschließend wechselt es in `queue1-dy-spotcompute1-[1-2]` den Ruhezustand und setzt die private IP-Adresse und den Hostnamen zurück, sodass es für future Aufgaben einsatzbereit ist.

Wenn etwas schief geht und eine Instanz für einen bestimmten Knoten aus irgendeinem Grund nicht gestartet werden kann, passiert Folgendes:

- Der Scheduler erhält einen Job, für den zwei Knoten erforderlich sind.
- Der Scheduler versetzt zwei Cloud-Bursting-Knoten in den `POWER_UP` Status und ruft `ResumeProgram` mit den Knotennamen auf (zum Beispiel). `queue1-dy-spotcompute1-[1-2]`
- `ResumeProgram` startet nur eine (1) EC2-Instance und konfiguriert `queue1-dy-spotcompute1-1`, dass eine (1) Instance nicht gestartet werden kann. `queue1-dy-spotcompute1-2`
- `queue1-dy-spotcompute1-1` ist nicht betroffen und geht online, nachdem der Status erreicht wurde. `POWER_UP`

- `queue1-dy-spotcompute1-2` wechselt in den `POWER_DOWN` Status, und der Job wird automatisch in die Warteschlange gestellt, da ein Knotenausfall Slurm erkannt wird.
- `queue1-dy-spotcompute1-2` wird danach verfügbar `SuspendTimeout` (die Standardeinstellung ist 120 Sekunden (2 Minuten)). In der Zwischenzeit wird der Job in die Warteschlange gestellt und kann auf einem anderen Knoten ausgeführt werden.
- Der obige Vorgang wird wiederholt, bis der Job auf einem verfügbaren Knoten ausgeführt werden kann, ohne dass ein Fehler auftritt.

Es gibt zwei Timing-Parameter, die bei Bedarf angepasst werden können:

- **ResumeTimeout** (Die Standardeinstellung ist 30 Minuten): `ResumeTimeout` steuert die Slurm Wartezeit, bis der Knoten in den ausgefallenen Zustand versetzt wird.
 - Eine Verlängerung kann nützlich sein, `ResumeTimeout` wenn der Prozess vor und nach der Installation fast genauso lange dauert.
 - `ResumeTimeout` ist auch die maximale AWS ParallelCluster Wartezeit, bis ein Knoten ersetzt oder zurückgesetzt wird, falls ein Problem auftritt. Rechenknoten beenden sich selbst, wenn während des Starts oder der Einrichtung ein Fehler auftritt. AWS ParallelCluster Prozesse ersetzen einen Knoten, wenn eine beendete Instanz erkannt wird.
- **SuspendTimeout** (Die Standardeinstellung ist 120 Sekunden (2 Minuten)): `SuspendTimeout` steuert, wie schnell Knoten wieder im System platziert werden und wieder einsatzbereit sind.
 - Ein kürzerer Wert `SuspendTimeout` bedeutet, dass Knoten schneller zurückgesetzt werden und häufiger versuchen Slurm können, Instanzen zu starten.
 - Länger `SuspendTimeout` bedeutet, dass ausgefallene Knoten langsamer zurückgesetzt werden. Slurm versucht in der Zwischenzeit, andere Knoten zu verwenden. Wenn `SuspendTimeout` es länger als ein paar Minuten dauert, wird Slurm versucht, alle Knoten im System zu durchlaufen. Eine längere Laufzeit `SuspendTimeout` könnte für große Systeme (über 1.000 Knoten) von Vorteil sein, um die stress zu verringern, Slurm wenn sie versuchen, häufig fehlgeschlagene Jobs in die Warteschlange zu stellen.
 - Beachten Sie, dass sich `SuspendTimeout` dies nicht auf die AWS ParallelCluster Wartezeit bezieht, bis eine Backing-Instance für einen Knoten beendet wird. Backing-Instances für `POWER_DOWN` Knoten werden sofort beendet. Der Terminierungsvorgang ist in der Regel in wenigen Minuten abgeschlossen. Während dieser Zeit verbleibt der Knoten jedoch im `POWER_DOWN` Status und steht dem Scheduler nicht zur Verfügung.

Protokolle für die Architektur

Die folgende Liste enthält die wichtigsten Protokolle. Der mit Amazon Logs verwendete CloudWatch Log-Stream-Name hat das Format `{hostname}.{instance_id}.{logIdentifizier}`, in dem *LogIdentifizier* den Log-Namen folgt.

- ResumeProgram: `/var/log/parallelcluster/slurm_resume.log` (slurm_resume)
- SuspendProgram: `/var/log/parallelcluster/slurm_suspend.log` (slurm_suspend)
- clustermgtd: `/var/log/parallelcluster/clustermgtd.log` (clustermgtd)
- computemgtd: `/var/log/parallelcluster/computemgtd.log` (computemgtd)
- slurmctld: `/var/log/slurmctld.log` (slurmctld)
- slurmd: `/var/log/slurmd.log` (slurmd)

Häufige Probleme und Hinweise zum Debuggen:

Knoten, die nicht gestartet, hochgefahren oder dem Cluster nicht hinzugefügt werden konnten

- Dynamische Knoten:
 - Prüfen Sie im ResumeProgram Protokoll, ob mit dem Knoten aufgerufen ResumeProgram wurde. Falls nicht, überprüfen Sie das slurmctld Protokoll, um festzustellen, ob Slurm versucht wurde, ResumeProgram mit dem Knoten anzurufen. Beachten Sie, dass falsche Berechtigungen dazu führen ResumeProgram können, dass der Vorgang unbemerkt fehlschlägt.
 - Wenn aufgerufen ResumeProgram wird, überprüfen Sie, ob eine Instanz für den Knoten gestartet wurde. Wenn die Instance nicht gestartet wurde, sollte es eine klare Fehlermeldung geben, warum die Instance nicht gestartet werden konnte.
 - Wenn eine Instance gestartet wurde, ist möglicherweise während des Bootstrap-Vorgangs ein Problem aufgetreten. Suchen Sie die entsprechende private IP-Adresse und Instanz-ID aus dem ResumeProgram Protokoll und sehen Sie sich die entsprechenden Bootstrap-Protokolle für die jeweilige Instanz in CloudWatch Logs an.
- Statische Knoten:
 - Prüfen Sie im clustermgtd Protokoll, ob Instanzen für den Knoten gestartet wurden. Wenn Instances nicht gestartet wurden, sollte es eindeutige Fehler darüber geben, warum die Instances nicht gestartet werden konnten.
 - Wenn eine Instance gestartet wurde, liegt ein Problem mit dem Bootstrap-Prozess vor. Suchen Sie die entsprechende private IP und Instanz-ID aus dem clustermgtd Protokoll und sehen

Sie sich die entsprechenden Bootstrap-Protokolle für die jeweilige Instanz unter Logs an CloudWatch .

Knoten, die unerwartet ersetzt oder beendet wurden, und Knotenausfälle

- Knoten wurden unerwartet ersetzt/beendet:
 - In den meisten Fällen `clustermgtd` erledigt es alle Wartungsaktionen für Knoten. Um zu überprüfen, ob ein Knoten `clustermgtd` ersetzt oder beendet wurde, überprüfen Sie das `clustermgtd` Protokoll.
 - Wenn der Knoten `clustermgtd` ersetzt oder beendet wurde, sollte eine Meldung erscheinen, die den Grund für die Aktion angibt. Wenn der Grund mit dem Scheduler zusammenhängt (zum Beispiel der KnotenDOWN), suchen Sie im `slurmctld` Protokoll nach weiteren Einzelheiten. Wenn der Grund mit EC2 zusammenhängt, verwenden Sie Tools wie Amazon CloudWatch oder die AWS EC2-Konsole, CLI oder SDKs, um den Status oder die Protokolle für diese Instance zu überprüfen. Sie können beispielsweise überprüfen, ob es für die Instance geplante Ereignisse gab oder ob die EC2-Integritätsprüfungen nicht bestanden haben.
 - Falls der Knoten `clustermgtd` nicht beendet wurde, überprüfen Sie, ob der Knoten `computemgtd` beendet wurde oder ob EC2 die Instance beendet hat, um eine Spot-Instance zurückzugewinnen.
- Knotenausfälle:
 - In den meisten Fällen werden Jobs automatisch in die Warteschlange gestellt, wenn ein Knoten ausfällt. Schauen Sie im `slurmctld` Protokoll nach, warum ein Job oder ein Knoten ausgefallen ist, und beurteilen Sie die Situation von dort aus.

Fehler beim Ersetzen oder Beenden von Instanzen, Fehler beim Herunterfahren von Knoten

- `clustermgtd`Behandelt im Allgemeinen alle erwarteten Aktionen zum Beenden von Instanzen. Sehen Sie im `clustermgtd` Protokoll nach, warum ein Knoten nicht ersetzt oder beendet werden konnte.
- Wenn dynamische Knoten ausfallen [ScaledownIdleTime](#), schauen Sie im `SuspendProgram` Protokoll nach, ob `slurmctld` Prozesse Aufrufe mit dem spezifischen Knoten als Argument ausgeführt haben. Note führt eigentlich `SuspendProgram` keine bestimmte Aktion aus. Vielmehr protokolliert es nur, wenn es aufgerufen wird. Alle Instanzbeendigungen und `NodeAddr` Resets sind bis abgeschlossen`clustermgtd`. Slurmwechselt von Knoten zu den nachfolgenden KnotenIDLE. `SuspendTimeout`

Andere Probleme:

- AWS ParallelCluster trifft keine Entscheidungen zur Stellenzuweisung oder Skalierung. Es versucht lediglich, Ressourcen gemäß den Anweisungen zu starten, zu beenden und zu Slurm zu verwalten.

Bei Problemen mit der Auftragszuweisung, der Knotenzuweisung und der Skalierungsentscheidung suchen Sie im `slurmctld` Protokoll nach Fehlern.

SlurmCluster-geschützter Modus

Wenn ein Cluster mit aktiviertem Schutzmodus ausgeführt wird, AWS ParallelCluster überwacht und verfolgt er Bootstrap-Fehler beim Starten der Rechenknoten. Auf diese Weise wird festgestellt, ob diese Fehler kontinuierlich auftreten.

Wenn in einer Warteschlange (Partition) Folgendes erkannt wird, wechselt der Cluster in den geschützten Status:

1. Aufeinanderfolgende Bootstrap-Fehler des Rechenknotens treten kontinuierlich auf, ohne dass der Rechenknoten erfolgreich gestartet wird.
2. Die Anzahl der Fehler erreicht einen vordefinierten Schwellenwert.

Sobald der Cluster den geschützten Status erreicht hat, werden Warteschlangen AWS ParallelCluster deaktiviert, deren Fehler den vordefinierten Schwellenwert erreichen oder überschreiten.

SlurmDer clustergeschützte Modus wurde in AWS ParallelCluster Version 3.0.0 hinzugefügt.

Sie können den geschützten Modus verwenden, um den Zeit- und Ressourcenaufwand für den Bootstrap-Fehlerzyklus des Rechenknotens zu reduzieren.

Startmodus-modus-modus-

protected_failure_count

`protected_failure_count` gibt die Anzahl aufeinanderfolgender Fehler in einer Warteschlange (Partition) an, die den Status Cluster-Protected aktivieren.

Die Standardeinstellung `protected_failure_count` ist 10 und der geschützte Modus ist aktiviert.

Wenn `protected_failure_count` der Wert größer als Null ist, ist der geschützte Modus aktiviert.

Wenn dieser `protected_failure_count` Startmodus-Startmodus-Startmodus-Startmodus-Startmodus-Startmodus-Startmodus-Startmodus-Startmodus-Startmodus-Startmodus-Startmodus-Startmodus-Startmodus-Startmodus-Startmodus-

Sie können den `protected_failure_count` Wert ändern, indem Sie den Parameter in der `clustermgtd` Konfigurationsdatei hinzufügen, die sich unter `/etc/parallelcluster/slurm_plugin/parallelcluster_clustermgtd.conf` befindet `HeadNode`.

Sie können diesen Parameter jederzeit aktualisieren und müssen dafür die Rechenflotte nicht anhalten. Wenn ein Start in einer Warteschlange erfolgreich ist, bevor die Anzahl der Fehler erreicht `protected_failure_count`, wird die Anzahl der Fehler auf Null zurückgesetzt.

Überprüfen Sie den Clusterstatus im geschützten Status

Wenn sich ein Cluster im geschützten Status befindet, können Sie den Status der Rechenflotte und den Knotenstatus überprüfen.

Flottenstatus berechnen

Der Status der Rechenflotte befindet sich `PROTECTED` in einem Cluster, der im geschützten Status ausgeführt wird.

```
$ pcluster describe-compute-fleet --cluster-name <cluster-name> --region <region-id>
{
  "status": "PROTECTED",
  "lastStatusUpdateTime": "2022-04-22T00:31:24.000Z"
}
```

Status des Knotens

Um zu erfahren, welche Warteschlangen (Partitionen) Bootstrap-Fehler aufweisen, die den Status „Geschützt“ haben, melden Sie sich beim Cluster an und führen Sie den Befehl aus. `sinfo` Partitionen mit Bootstrap-Fehlern bei oder höher `protected_failure_count` befinden sich im `INACTIVE` Status. Partitionen ohne Bootstrap-Fehler bei oder höher `protected_failure_count` befinden sich im `UP` Status und funktionieren wie erwartet.

`PROTECTED` Der Status wirkt sich nicht auf laufende Jobs aus. Wenn Jobs auf einer Partition mit Bootstrap-Fehlern von mindestens einem Wert ausgeführt werden `protected_failure_count`, wird die Partition auf `gesetzt`, `INACTIVE` nachdem die laufenden Jobs abgeschlossen sind.

Sehen Sie sich die Knotenmodus-Startmodus-Startmodus-Startmodus-Startmodus-Startmodus-Start

```
$ sinfo
PARTITION AVAIL TIMELIMIT NODES STATE NODELIST
queue1*  inact infinite 10  down% queue1-dy-c5xlarge-[1-10]
queue1*  inact infinite 3490 idle~ queue1-dy-c5xlarge-[11-3500]
queue2  up  infinite 10  idle~ queue2-dy-c5xlarge-[1-10]
```

queue1Die Partition ist INACTIVE darauf zurückzuführen, dass 10 aufeinanderfolgende Bootstrap-Fehler des Rechenknotens festgestellt wurden.

Instanzen hinter Knoten queue1-dy-c5xlarge-[1-10] wurden gestartet, konnten dem Cluster jedoch aufgrund eines fehlerhaften Status nicht beitreten.

Der Cluster befindet sich im geschützten Status.

queue2Die Partition ist von den Bootstrap-Fehlern in nicht betroffen. queue1 Es ist im UP Bundesstaat und kann immer noch Jobs ausführen.

So deaktivieren Sie den geschützten Status

Nachdem der Bootstrap-Fehler behoben wurde, können Sie den folgenden Befehl ausführen, um den geschützten Status des Clusters zu beenden.

```
$ pcluster update-compute-fleet --cluster-name <cluster-name> \
  --region <region-id> \
  --status START_REQUESTED
```

Bootstrap-Fehler, die den geschützten Status aktivieren

Bootstrap-Fehler, die den geschützten Status aktivieren, werden in die folgenden drei Typen unterteilt. Um den Typ und das Problem zu identifizieren, können Sie überprüfen, ob Protokolle AWS ParallelCluster generiert wurden. Wenn Protokolle generiert wurden, können Sie sie auf Fehlerdetails überprüfen. Weitere Informationen finden Sie unter [Protokolle abrufen und aufbewahren](#).

1. Bootstrap-Fehler, der dazu führt, dass sich eine Instanz selbst beendet.

Eine Instanz schlägt zu Beginn des Bootstrap-Prozesses fehl, z. B. eine Instanz, die aufgrund von Fehlern im [SlurmQueues\ CustomActions\ OnNodeStart](#) | -Skript automatisch beendet wird. [OnNodeConfigured](#)

Suchen Sie bei dynamischen Knoten nach Fehlern wie folgt:

```
Node bootstrap error: Node ... is in power up state without valid backing instance
```

Suchen Sie bei statischen Knoten im `clustermgtd` log (`/var/log/parallelcluster/clustermgtd`) nach Fehlern, die den folgenden ähneln:

```
Node bootstrap error: Node ... is in power up state without valid backing instance
```

2. Knoten `resume_timeout` oder `node_replacement_timeout` läuft ab.

Eine Instanz kann dem Cluster nicht innerhalb der `resume_timeout` (für dynamische Knoten) oder `node_replacement_timeout` (für statische Knoten) beitreten. Es beendet sich nicht von selbst vor dem Timeout. Beispielsweise ist das Netzwerk für den Cluster nicht richtig eingerichtet und der Knoten wird Slurm nach Ablauf des Timeouts in den DOWN Status versetzt.

Suchen Sie bei dynamischen Knoten nach Fehlern wie folgt:

```
Node bootstrap error: Resume timeout expires for node
```

Suchen Sie bei statischen Knoten im `clustermgtd` log (`/var/log/parallelcluster/clustermgtd`) nach Fehlern, die den folgenden ähneln:

```
Node bootstrap error: Replacement timeout expires for node ... in replacement.
```

3. Knoten bestehen die Integritätsprüfung nicht.

Eine Instance hinter dem Knoten besteht bei einer EC2-Integritätsprüfung oder einer Zustandsprüfung für geplante Ereignisse nicht, und die Knoten werden als Bootstrap-Fehlerknoten behandelt. In diesem Fall wird die Instanz aus einem Grund beendet, der außerhalb der Kontrolle von AWS ParallelCluster liegt.

Suchen Sie im `clustermgtd` Protokoll (`/var/log/parallelcluster/clustermgtd`) nach Fehlern, die den folgenden ähneln:

```
Node bootstrap error: Node %s failed during bootstrap when performing health check.
```

4. Die SlurmRegistrierung von Rechenknoten schlägt fehl.

Die Registrierung des `slurmd` Daemons beim Slurm Kontroll-Daemon (`slurmctld`) schlägt fehl und führt dazu, dass der Status des Rechenknotens in den Status wechselt. `INVALID_REG` Falsch

konfigurierte Slurm Rechenknoten können diesen Fehler verursachen, z. B. bei berechneten Knoten, die mit Spezifikationsfehlern für [CustomSlurmSettings](#) Rechenknoten konfiguriert wurden.

Suchen Sie in der `slurmctld` Protokolldatei (`/var/log/slurmctld.log`) auf dem Hauptknoten oder in der `slurmd` Protokolldatei (`/var/log/slurmd.log`) des ausgefallenen Rechenknotens nach Fehlern, die den folgenden ähneln:

```
Setting node %s to INVALID with reason: ...
```

So debuggen Sie den geschützten Modus

Wenn sich Ihr Cluster im geschützten Status befindet und wenn `clustermgtd` Protokolle von den HeadNode und die `cloud-init-output` Protokolle von problematischen Rechenknoten AWS ParallelCluster generiert wurden, können Sie in den Protokollen nach Fehlerdetails suchen. Weitere Informationen zum Abrufen von Protokollen finden Sie unter [Protokolle abrufen und aufbewahren](#).

`clustermgtdlog` (`/var/log/parallelcluster/clustermgtd`) auf dem Kopfknoten

Protokollmeldungen zeigen, welche Partitionen Bootstrap-Fehler aufweisen, und die entsprechende Anzahl von Bootstrap-Fehlern.

```
[slurm_plugin.clustermgtd:_handle_protected_mode_process] - INFO - Partitions bootstrap failure count: {'queue1': 2}, cluster will be set into protected mode if protected failure count reach threshold.
```

Suchen `clustermgtd` Sie im Protokoll nach `Found the following bootstrap failure nodes` dem Knoten, der nicht gestartet werden konnte.

```
[slurm_plugin.clustermgtd:_handle_protected_mode_process] - WARNING - Found the following bootstrap failure nodes: (x2) ['queue1-st-c5large-1(192.168.110.155)', 'broken-st-c5large-2(192.168.65.215)']
```

Suchen `clustermgtd` Sie im Protokoll `Node bootstrap error` nach dem Grund für den Fehler.

```
[slurm_plugin.clustermgtd:_is_node_bootstrap_failure] - WARNING - Node bootstrap error: Node broken-st-c5large-2(192.168.65.215) is currently in replacement and no backing instance
```


cloud-init-outputlog (/var/log/cloud-init-output.log) auf den Rechenknoten

Nachdem Sie die private IP-Adresse des Bootstrap-Fehlerknotens im `clustermgtd` Protokoll abgerufen haben, können Sie das entsprechende Rechenknotenprotokoll finden, indem Sie sich entweder beim Rechenknoten anmelden oder den Anweisungen [Protokolle abrufen und aufbewahren](#) zum Abrufen von Protokollen folgen. In den meisten Fällen zeigt das `/var/log/cloud-init-output` Protokoll des problematischen Knotens den Schritt, der den Bootstrap-Fehler des Rechenknotens verursacht hat.

Slurmschneller Cluster-Failover mit unzureichender Kapazität

Ab AWS ParallelCluster Version 3.2.0 werden Cluster mit dem standardmäßig aktivierten Failover-Modus für schnelle unzureichende Kapazität ausgeführt. Dadurch wird der Zeitaufwand für den erneuten Versuch, einen Job in die Warteschlange zu stellen, minimiert, wenn EC2-Fehler mit unzureichender Kapazität erkannt werden. Dies ist besonders effektiv, wenn Sie Ihren Cluster mit mehreren Arten von Instanztypen konfigurieren.

EC2 hat unzureichende Kapazitätsausfälle festgestellt:

- `InsufficientInstanceCapacity`
- `InsufficientHostCapacity`
- `InsufficientReservedInstanceCapacity`
- `MaxSpotInstanceCountExceeded`
- `SpotMaxPriceTooLow`: Wird aktiviert, wenn der Spot-Anfragepreis niedriger ist als der erforderliche Mindestpreis für Spot-Anfragen.
- `Unsupported`: Aktiviert mit der Verwendung eines Instanztyps, der in einem bestimmten Instanztyp nicht unterstützt wird AWS-Region.

Wenn im Modus „Schneller Ausfall unzureichender Kapazität“ ein Fehler mit unzureichender Kapazität erkannt wird, wenn ein Job einem [SlurmQueues](#)/zugewiesen wird [compute resource](#), AWS ParallelCluster geht das wie folgt vor:

1. Es setzt die Rechenressource für einen vordefinierten Zeitraum auf den Status Disabled (DOWN).
2. Es wird verwendet `POWER_DOWN_FORCE`, um die ausgefallenen Knotenjobs der Rechenressource abzurechnen und den ausgefallenen Knoten zu suspendieren. Es setzt den ausgefallenen Knoten auf den `POWER_DOWN (!)` Zustand `IDLE` und dann auf `POWERING_DOWN (%)`.
3. Es fordert den Job an eine andere Rechenressource an.

Die statischen und eingeschalteten Knoten der deaktivierten Rechenressource sind nicht betroffen. Jobs können auf diesen Knoten abgeschlossen werden.

Dieser Zyklus wiederholt sich, bis der Job erfolgreich einem oder mehreren Rechenressourcenknoten zugewiesen wurde. Informationen zu Knotenzuständen finden Sie unter [SlurmLeitfaden für den Modus mit mehreren Warteschlangen](#).

Wenn keine Rechenressourcen für die Ausführung des Jobs gefunden werden, wird der Job auf den PENDING Status gesetzt, bis der vordefinierte Zeitraum abgelaufen ist. In diesem Fall können Sie den vordefinierten Zeitraum wie im folgenden Abschnitt beschrieben ändern.

Timeout-Parameter für unzureichende Kapazität

insufficient_capacity_timeout

`insufficient_capacity_timeout` gibt den Zeitraum (in Sekunden) an, für den die Rechenressource im Zustand disabled (down) gehalten wird, wenn ein unzureichender Kapazitätsfehler erkannt wird.

In der Standardeinstellung `insufficient_capacity_timeout` ist aktiviert.

Führen Sie `insufficient_capacity_timeout` die Zeitspanne von 600 Sekunden (10 Minuten) an.

Wenn der `insufficient_capacity_timeout` Wert kleiner oder gleich Null ist, ist der Failure-Over-Modus „Schneller Ausfall bei unzureichender Kapazität“ deaktiviert.

Sie können den `insufficient_capacity_timeout` Wert ändern, indem Sie den Parameter in der `clustermgtd` Konfigurationsdatei hinzufügen, die sich unter `/etc/parallelcluster/slurm_plugin/parallelcluster_clustermgtd.conf` befindet `HeadNode`.

Der Parameter kann jederzeit aktualisiert werden, ohne die Rechenflotte zu stoppen.

Beispiel:

- `insufficient_capacity_timeout=600:`

Wenn ein unzureichender Kapazitätsfehler erkannt wird, wird die Rechenressource auf deaktiviert (DOWN) gesetzt. Nach 10 Minuten wird der ausgefallene Knoten auf den Status `idle~` (POWER_SAVING) gesetzt.

- `insufficient_capacity_timeout=60`:

Wenn ein unzureichender Kapazitätsfehler erkannt wird, befindet sich die Rechenressource in einem deaktivierten (DOWN) Zustand. Nach 1 Minute wird der ausgefallene Knoten auf den `idle` Status gesetzt.

- `insufficient_capacity_timeout=0`:

Der Failure-Over-Modus „Schneller Ausfall bei unzureichender Kapazität“ ist deaktiviert. Die Rechenressource ist nicht deaktiviert.

Note

Zwischen dem Zeitpunkt, an dem Knoten aufgrund unzureichender Kapazitätsfehler ausfallen, und dem Zeitpunkt, zu dem der Cluster-Management-Daemon die Knotenausfälle erkennt, kann es zu einer Verzögerung von bis zu einer Minute kommen. Dies liegt daran, dass der Cluster-Management-Daemon nach Ausfällen mit unzureichender Knotenkapazität sucht und die Rechenressourcen in Intervallen von einer Minute in den `down` Status versetzt.

Schneller Status im Failover-Modus mit unzureichender Kapazität

Wenn sich ein Cluster im Failover-Modus mit schneller unzureichender Kapazität befindet, können Sie seinen Status und seinen Knotenstatus überprüfen.

Knotenzustände

Wenn ein Job an einen dynamischen Knoten mit Rechenressourcen übermittelt wird und ein unzureichender Kapazitätsfehler festgestellt wird, wird der Knoten in den `down` Status mit Begründung versetzt.

```
(Code:InsufficientInstanceCapacity)Failure when resuming nodes.
```

Dann werden ausgeschaltete Knoten (Knoten im `idle` Status) `down` ohne Grund auf gesetzt.

```
(Code:InsufficientInstanceCapacity)Temporarily disabling node due to insufficient capacity.
```

Der Job wird an andere Rechenressourcen in der Warteschlange angefordert.

Die statischen Knoten der Rechenressource und Knoten, dieUP nicht vom schnellen Failover-Modus mit unzureichender Kapazität betroffen sind.

Stellen Sie sich die im folgenden Beispiel veranschaulicht an.

```
$ sinfo
PARTITION AVAIL TIMELIMIT NODES STATE NODELIST
queue1*   up    infinite   30  idle~ queue1-dy-c-1-[1-15],queue1-dy-c-2-[1-15]
queue2    up    infinite   30  idle~ queue2-dy-c-1-[1-15],queue2-dy-c-2-[1-15]
```

Wir senden einen Job an queue1, der einen Knoten benötigt.

```
$ sinfo
PARTITION AVAIL TIMELIMIT NODES STATE NODELIST
queue1*   up    infinite   1   down# queue1-dy-c-1-1
queue1*   up    infinite  15  idle~ queue1-dy-c-2-[1-15]
queue1*   up    infinite  14  down~ queue1-dy-c-1-[2-15]
queue2    up    infinite  30  idle~ queue2-dy-c-1-[1-15],queue2-dy-c-2-[1-15]
```

queue1-dy-c-1-1Der Knoten wird gestartet, um den Job auszuführen. Aufgrund eines Fehlers wegen unzureichender Kapazität konnte jedoch nicht gestartet werden. queue1-dy-c-1-1Der Knoten ist auf gesetztdown. Der ausgeschaltete dynamische Knoten innerhalb der Rechenressource (queue2-dy-c-1) ist auf gesetztdown.

Sie können den Grund für den Knoten mit `scontrol show nodes`.

```
$ scontrol show nodes queue1-dy-c-1-1
NodeName=broken-dy-c-2-1 Arch=x86_64 CoresPerSocket=1
CPUAlloc=0 CPUTot=96 CPULoad=0.00
...
ExtSensorsJoules=n/s ExtSensorsWatts=0 ExtSensorsTemp=n/s
Reason=(Code:InsufficientInstanceCapacity)Failure when resuming nodes
[root@2022-03-10T22:17:50]

$ scontrol show nodes queue1-dy-c-1-2
NodeName=broken-dy-c-2-1 Arch=x86_64 CoresPerSocket=1
CPUAlloc=0 CPUTot=96 CPULoad=0.00
...
ExtSensorsJoules=n/s ExtSensorsWatts=0 ExtSensorsTemp=n/s
Reason=(Code:InsufficientInstanceCapacity)Temporarily disabling node due to
insufficient capacity [root@2022-03-10T22:17:50]
```

Der Job wird in die Warteschlange eines anderen Instanztyps innerhalb der Rechenressourcen der Warteschlange gestellt.

Nachinsufficient_capacity_timeout Ablauf der Gültigkeitsdauer werden die Knoten in der Rechenressource auf denidle~ Status zurückgesetzt.

\$ sinfo

```
PARTITION AVAIL TIMELIMIT NODES STATE NODELIST
queue1*   up    infinite    30  idle~ queue1-dy-c-1-[1-15],queue1-dy-c-2-[1-15]
queue2    up    infinite    30  idle~ queue2-dy-c-1-[1-15],queue2-dy-c-2-[1-15]
```

Nachdem die Knoteninsufficient_capacity_timeout abgelaufen sind und die Knoten in der Rechenressource auf denidle~ Status zurückgesetzt wurden, weist derSlurm Scheduler den Knoten eine niedrigere Priorität zu. Der Scheduler wählt weiterhin Knoten aus anderen Warteschlangenverarbeitungsressourcen mit höheren Gewichtungen aus, sofern nicht einer der folgenden Fälle eintritt:

- Die Einreichungsanforderungen eines Jobs entsprechen der wiederhergestellten Rechenressource.
- Es sind keine anderen Rechenressourcen verfügbar, da sie voll ausgelastet sind.
- slurmctldwird neu gestartet.
- DieAWS ParallelCluster Rechenflotte wird gestoppt und gestartet, um alle Knoten aus- und einzuschalten.

Verwandte Logs

Protokolle, die sich auf Fehler bei unzureichender Kapazität und den schnellen Failover-Modus mit unzureichender Kapazität beziehen, finden Sie imresumeclustermgtd Protokoll und im Protokoll im Hauptknoten.Slurm

Slurm **resume (/var/log/parallelcluster/slurm_resume.log)**

Fehlermeldungen, wenn ein Knoten aufgrund unzureichender Kapazität nicht gestartet werden kann.

```
[slurm_plugin.instance_manager:_launch_ec2_instances] - ERROR - Failed RunInstances
request: dcd0c252-90d4-44a7-9c79-ef740f7ecd87
[slurm_plugin.instance_manager:add_instances_for_nodes] - ERROR - Encountered
exception when launching instances for nodes (x1) ['queue1-dy-c-1-1']: An error
occurred
```

```
(InsufficientInstanceCapacity) when calling the RunInstances operation (reached max
retries: 1): We currently do not have sufficient p4d.24xlarge capacity in the
Availability Zone you requested (us-west-2b). Our system will be working on
provisioning additional capacity. You can currently get p4d.24xlarge capacity by
not
specifying an Availability Zone in your request or choosing us-west-2a, us-west-2c.
```

Slurm `clustermgtd` (`/var/log/parallelcluster/clustermgtd`)

Die Rechenressource c-1 in der Warteschlange 1 ist aufgrund unzureichender Kapazität deaktiviert.

```
[slurm_plugin.clustermgtd:_reset_timeout_expired_compute_resources] - INFO - The
following compute resources are in down state
due to insufficient capacity: {'queue1': {'c-1':
ComputeResourceFailureEvent(timestamp=datetime.datetime(2022, 4, 14, 23, 0, 4,
769380, tzinfo=datetime.timezone.utc),
error_code='InsufficientInstanceCapacity')}}}, compute resources are reset after
insufficient capacity timeout (600 seconds) expired
```

Nach Ablauf des Timeouts für unzureichende Kapazität wird die Rechenressource zurückgesetzt, und die Knoten innerhalb der Rechenressourcen werden auf `gesetztidle~`.

```
[root:_reset_insufficient_capacity_timeout_expired_nodes] - INFO - Reset the
following compute resources because insufficient capacity
timeout expired: {'queue1': ['c-1']}
```

Slurmspeicherbasierte Terminplanung

AWS ParallelCluster Unterstützt ab Version 3.2.0 Slurm speicherbasiertes Scheduling mit dem Cluster-Konfigurationsparameter/. [SlurmSettingsEnableMemoryBasedScheduling](#)

Note

[Ab AWS ParallelCluster Version 3.7.0 EnableMemoryBasedScheduling kann aktiviert werden, wenn Sie mehrere Instanztypen in Instances konfigurieren.](#)

Für die AWS ParallelCluster Versionen 3.2.0 bis 3.6. **x** `EnableMemoryBasedScheduling` kann nicht aktiviert werden, wenn Sie mehrere Instanztypen in [Instances](#) konfigurieren.

⚠ Warning

Wenn Sie mehrere Instance-Typen in einer Slurm Queue-Computing-Ressource mit `EnableMemoryBasedScheduling` aktivierter Option angeben, ist der `RealMemory` Wert die Mindestmenge an Arbeitsspeicher, die allen Instance-Typen zur Verfügung gestellt wird. Dies kann zu erheblichen Mengen an ungenutztem Speicher führen, wenn Sie Instance-Typen mit sehr unterschiedlichen Speicherkapazitäten angeben.

Mit `EnableMemoryBasedScheduling: true` verfolgt der Slurm Scheduler die Menge an Speicher, die jeder Job auf jedem Knoten benötigt. Anschließend verwendet der Slurm Scheduler diese Informationen, um mehrere Jobs auf demselben Rechenknoten zu planen. Die Gesamtmenge an Speicher, die Jobs auf einem Knoten benötigen, darf nicht größer sein als der verfügbare Knotenspeicher. Der Scheduler verhindert, dass ein Job mehr Speicher beansprucht, als beim Absenden des Jobs angefordert wurde.

Mit `EnableMemoryBasedScheduling: false` können Jobs auf einem gemeinsam genutzten Knoten um Arbeitsspeicher konkurrieren und zu Auftragsausfällen und `out-of-memory` Ereignissen führen.

⚠ Warning

Slurm verwendet für seine Bezeichnungen eine Zweierpotenz, z. B. MB oder GB. Lesen Sie diese Beschriftungen jeweils als MiB und GiB.

SlurmKonfiguration und speicherbasiertes Scheduling

Mit Slurm legt `EnableMemoryBasedScheduling: true` die folgenden Slurm Konfigurationsparameter fest:

- [`SelectTypeParameters=CR_CPU_Memory`](#) im `slurm.conf`. Diese Option konfiguriert den Knotenspeicher als verbrauchbare Ressource in Slurm
- [`ConstrainRAMSpace=yes`](#) in der Slurm `cgroup.conf`. Mit dieser Option ist der Speicherzugriff eines Jobs auf die Speichermenge beschränkt, die der Job bei der Übermittlung angefordert hat.

Note

Verschiedene andere Slurm Konfigurationsparameter können sich auf das Verhalten des Slurm Schedulers und des Ressourcenmanagers auswirken, wenn diese beiden Optionen festgelegt sind. Weitere Informationen finden Sie in der [Slurm-Dokumentation](#).

SlurmScheduler und speicherbasiertes Scheduling

EnableMemoryBasedScheduling: false(Standard)

EnableMemoryBasedScheduling ist standardmäßig auf „Falsch“ gesetzt. Wenn der Wert falsch Slurm ist, nimmt der Speicher nicht als Ressource in seinen Planungsalgorithmus auf und verfolgt nicht den Speicher, den Jobs verwenden. Benutzer können die `--mem MEM_PER_NODE` Option angeben, um die Mindestmenge an Speicher pro Knoten festzulegen, die ein Job benötigt. Dadurch wird der Scheduler gezwungen, MEM_PER_NODE bei der Planung des Jobs Knoten mit einem RealMemory Wert von mindestens auszuwählen.

Nehmen wir zum Beispiel an, dass ein Benutzer zwei Jobs mit einreicht. `--mem=5GB` Wenn angeforderte Ressourcen wie CPUs oder GPUs verfügbar sind, können die Jobs gleichzeitig auf einem Knoten mit 8 GiB Arbeitsspeicher ausgeführt werden. Die beiden Jobs sind nicht auf Rechenknoten mit weniger als 5 GiB geplantRealMemory.

Warning

Wenn die speicherbasierte Planung deaktiviert ist, wird Slurm nicht nachverfolgt, wie viel Speicher die Jobs verwenden. Jobs, die auf demselben Knoten ausgeführt werden, konkurrieren möglicherweise um Speicherressourcen und führen dazu, dass der andere Job fehlschlägt.

Wenn die speicherbasierte Planung deaktiviert ist, empfehlen wir Benutzern, die Optionen `--mem-per-cpu` oder `--mem-per-gpu` nicht anzugeben. Diese Optionen können zu einem Verhalten führen, das sich von dem unterscheidet, was in der [Slurm-Dokumentation](#) beschrieben ist.

EnableMemoryBasedScheduling: true

Wenn auf `true` gesetzt `EnableMemoryBasedScheduling` ist, wird die Speichernutzung jedes Jobs Slurm verfolgt und verhindert, dass Jobs mehr Speicher verwenden, als mit den `--mem` Übermittlungsoptionen angefordert wurde.

Im vorherigen Beispiel sendet ein Benutzer zwei Jobs mit `--mem=5GB`. Die Jobs können nicht gleichzeitig auf einem Knoten mit 8 GiB Arbeitsspeicher ausgeführt werden. Das liegt daran, dass die insgesamt benötigte Speichermenge größer ist als der Speicher, der auf dem Knoten verfügbar ist.

Wenn die speicherbasierte Planung aktiviert ist, `--mem-per-cpu` und `--mem-per-gpu` verhalten Sie sich konsistent mit dem, was in der Slurm Dokumentation beschrieben ist. Zum Beispiel wird ein Job mit eingereicht. `--ntasks-per-node=2 -c 1 --mem-per-cpu=2GB` In diesem Fall Slurm weist dem Job insgesamt 4 GiB für jeden Knoten zu.

Warning

Wenn die speicherbasierte Planung aktiviert ist, empfehlen wir Benutzern, bei der Einreichung eines Jobs eine `--mem` Spezifikation anzugeben. Wenn in der Slurm Standardkonfiguration keine Speicheroption enthalten ist (AWS ParallelCluster, oder `--mem-per-gpu`) `--mem--mem-per-cpu`, wird dem Auftrag der gesamte Speicher der zugewiesenen Knoten zugewiesen, auch wenn er nur einen Teil der anderen Ressourcen wie Slurm CPUs oder GPUs anfordert. Dadurch wird die gemeinsame Nutzung von Knoten effektiv verhindert, bis der Job abgeschlossen ist, da kein Speicher für andere Jobs verfügbar ist. Das liegt daran, dass der Arbeitsspeicher pro Knoten für den Job auf den Zeitpunkt Slurm gesetzt wird, zu [DefMemPerNode](#) dem bei der Auftragsübergabe keine Speicherspezifikationen angegeben wurden. Der Standardwert für diesen Parameter ist 0 und gibt unbegrenzten Zugriff auf den Speicher eines Knotens an.

Wenn mehrere Typen von Rechenressourcen mit unterschiedlichen Speichermengen in derselben Warteschlange verfügbar sind, werden einem ohne Speicheroptionen eingereichten Job möglicherweise unterschiedliche Speichermengen auf verschiedenen Knoten zugewiesen. Das hängt davon ab, welche Knoten der Scheduler für den Job zur Verfügung stellt. Benutzer können auf Cluster `DefMemPerNode` - oder Partitionsebene in den Slurm Konfigurationsdateien einen benutzerdefinierten Wert für Optionen wie oder definieren [DefMemPerCPU](#), um dieses Verhalten zu verhindern.

Slurm **RealMemory** und AWS ParallelCluster **SchedulableMemory**

Wird in der AWS ParallelCluster mitgelieferten Slurm Konfiguration als die Menge an Speicher pro Slurm Knoten interpretiert [RealMemory](#), die für Jobs verfügbar ist. Ab Version 3.2.0 wird standardmäßig auf 95 Prozent des Speichers AWS ParallelCluster gesetzt [RealMemory](#), der in [Amazon EC2 EC2-Instanztypen aufgeführt und von der Amazon EC2 EC2-API zurückgegeben wird. DescribeInstanceTypes](#)

Wenn die speicherbasierte Planung deaktiviert ist, filtert der Slurm Scheduler Knoten, wenn Benutzer einen [RealMemory](#) Job mit den angegebenen Werten einreichen. `--mem`

Wenn die speicherbasierte Planung aktiviert ist, interpretiert [RealMemory](#) der Slurm Scheduler die maximale Speichermenge, die für Jobs verfügbar ist, die auf dem Rechenknoten ausgeführt werden.

Die Standardeinstellung ist möglicherweise nicht für alle Instance-Typen optimal:

- Diese Einstellung ist möglicherweise höher als die Speichermenge, auf die Knoten tatsächlich zugreifen können. Dies kann passieren, wenn es sich bei den Rechenknoten um kleine Instance-Typen handelt.
- Diese Einstellung ist möglicherweise niedriger als die Speichermenge, auf die Knoten tatsächlich zugreifen können. Dies kann passieren, wenn es sich bei Rechenknoten um große Instance-Typen handelt, und kann zu einer erheblichen Menge an ungenutztem Speicher führen.

Sie können [SlurmQueues/ComputeResources](#) verwenden [SchedulableMemory](#), um den Wert von [RealMemory](#) configured by AWS ParallelCluster für Compute-Knoten zu optimieren. Um den Standardwert zu überschreiben, definieren Sie einen benutzerdefinierten Wert [SchedulableMemory](#) speziell für Ihre Clusterkonfiguration.

Um den tatsächlich verfügbaren Speicher eines Rechenknotens zu überprüfen, führen Sie den `/opt/slurm/sbin/slurmd -C` Befehl auf dem Knoten aus. Dieser Befehl gibt die Hardwarekonfiguration des Knotens einschließlich des [RealMemory](#) Werts zurück. Weitere Informationen finden Sie unter [slurmd -C](#).

Stellen Sie sicher, dass die Betriebssystemprozesse des Compute-Knotens über ausreichend Arbeitsspeicher verfügen. Beschränken Sie dazu den für Jobs verfügbaren Speicher, indem Sie den [SchedulableMemory](#) Wert auf einen niedrigeren Wert als den vom `slurmd -C` Befehl zurückgegebenen [RealMemory](#) Wert setzen.

Zuweisung mehrerer Instanztypen mit Slurm

Beginnend mit AWS ParallelCluster In Version 3.3.0 können Sie Ihren Cluster so konfigurieren, dass er aus dem Satz definierter Instanztypen einer Rechenressource zuweist. Die Zuweisung kann auf Strategien zur Senkung der Kosten oder der optimalen Kapazität der EC2-Flotte basieren.

Dieser Satz definierter Instance-Typen muss entweder alle dieselbe Anzahl von vCPUs oder, falls Multithreading deaktiviert ist, dieselbe Anzahl von Kernen haben. Darüber hinaus muss dieser Satz von Instance-Typen dieselbe Anzahl von Beschleunigern derselben Hersteller haben. Wenn [Efa/Enabled](#) ist eingestellt auf `true`, für die Instanzen muss EFA unterstützt werden. Weitere Informationen und Anforderungen finden Sie unter [Scheduling/SlurmQueues/AllocationStrategy](#) und [ComputeResources/Instances](#).

Du kannst einstellen [AllocationStrategy](#) zu `lowest-price` oder `capacity-optimized` abhängig von deinem [CapacityType](#) Konfiguration.

In [Instances](#), Sie können eine Reihe von Instance-Typen konfigurieren.

Note

Beginnend mit AWS ParallelCluster Version 3.7.0, `EnableMemoryBasedScheduling` kann aktiviert werden, wenn Sie mehrere Instanztypen konfigurieren in [Instanzen](#).
Für AWS ParallelCluster Versionen 3.2.0 bis 3.6.x, `EnableMemoryBasedScheduling` kann nicht aktiviert werden, wenn Sie mehrere Instanztypen konfigurieren in [Instanzen](#).

Die folgenden Beispiele zeigen, wie Sie Instance-Typen für vCPUs, EFA-Unterstützung und Architektur abfragen können.

Abfrage `InstanceTypes` mit 96 vCPUs und x86_64-Architektur.

```
$ aws ec2 describe-instance-types --region region-id \
  --filters "Name=vcpu-info.default-vcpus,Values=96" "Name=processor-info.supported-
  architecture,Values=x86_64" \
  --query "sort_by(InstanceTypes[*].
  {InstanceType: InstanceType, MemoryMiB: MemoryInfo.SizeInMiB, CurrentGeneration: CurrentGeneration, V
  &InstanceType})" \
  --output table
```

Anfrage `InstanceTypes` mit 64 Kernen, EFA-Unterstützung und arm64-Architektur.

```
$ aws ec2 describe-instance-types --region region-id \
  --filters "Name=vcpu-info.default-cores,Values=64" "Name=processor-
info.supported-architecture,Values=arm64" "Name=network-info.efa-
supported,Values=true" --query "sort_by(InstanceTypes[*].
{InstanceType:InstanceType,MemoryMiB:MemoryInfo.SizeInMiB,CurrentGeneration:CurrentGeneration,V
&InstanceType})" \
  --output table
```

Das nächste Beispiel für eine Cluster-Konfiguration zeigt, wie Sie diese verwenden können InstanceType und AllocationStrategy Eigenschaften.

```
...
Scheduling:
  Scheduler: slurm
  SlurmQueues:
    - Name: queue-1
      CapacityType: ONDEMAND
      AllocationStrategy: lowest-price
      ...
      ComputeResources:
        - Name: computeresource1
          Instances:
            - InstanceType: r6g.2xlarge
            - InstanceType: m6g.2xlarge
            - InstanceType: c6g.2xlarge
          MinCount: 0
          MaxCount: 500
        - Name: computeresource2
          Instances:
            - InstanceType: m6g.12xlarge
            - InstanceType: x2gd.12xlarge
          MinCount: 0
          MaxCount: 500
  ...
```

Cluster-Skalierung für dynamische Knoten

ParallelCluster unterstützt Slurm die Methoden zur dynamischen Skalierung von Clustern mithilfe Slurm des Energiespar-Plugins. Weitere Informationen finden Sie im [Cloud Scheduling Guide](#) und im [SlurmPower Saving Guide](#) in der Slurm Dokumentation.

Ab ParallelCluster Version 3.8.0 wird die Wiederaufnahme auf Jobebene oder die Skalierung auf Jobebene als Standardstrategie für die dynamische Knotenzuweisung ParallelCluster verwendet, um den Cluster zu skalieren: ParallelCluster skaliert den Cluster auf der Grundlage der Anforderungen jedes Jobs, der Anzahl der dem Job zugewiesenen Knoten und der Knoten, die wieder aufgenommen werden müssen. ParallelCluster ruft diese Informationen aus der Umgebungsvariablen `SLURM_RESUME_FILE` ab.

Die Skalierung für dynamische Knoten ist ein zweistufiger Prozess, der den Start der EC2-Instances und die Zuweisung der gestarteten EC2-Instances zu den Slurm-Knoten beinhaltet. Jeder dieser beiden Schritte kann mithilfe einer all-or-nothing oder einer Best-Effort-Logik ausgeführt werden.

Für den Start der EC2-Instances:

- all-or-nothing ruft die EC2-Launch-API auf, wobei das Mindestziel der gesamten Zielkapazität entspricht
- Best-Effort ruft die EC2-API zum Starten auf, wobei das Mindestziel 1 und die Gesamtzielkapazität der angeforderten Kapazität entspricht

Für die Zuweisung der EC2-Instances zu Slurm-Knoten:

- all-or-nothing weist EC2-Instances nur dann Slurm-Knoten zu, wenn es möglich ist, jedem angeforderten Knoten eine EC2-Instance zuzuweisen
- Best-Effort weist EC2-Instances Slurm-Knoten zu, auch wenn nicht alle angeforderten Knoten durch die EC2-Instanzkapazität abgedeckt sind

Aus den möglichen Kombinationen der oben genannten Strategien ergeben sich die Startstrategien. ParallelCluster

Example

<caption>The available ParallelCluster Strategien einführen that can be set into the [ScalingStrategy](#) cluster configuration to be used with Skalierung auf Jobebene are:</caption>

all-or-nothingSkalierung:

Diese Strategie beinhaltet das AWS ParallelCluster Initiieren eines Amazon EC2 EC2-Launch-Instance-API-Aufrufs für jeden Job, der alle Instances erfordert, die für den erfolgreichen Start der angeforderten Rechenknoten erforderlich sind. Dadurch wird sichergestellt, dass der Cluster nur

skaliert wird, wenn die erforderliche Kapazität pro Job verfügbar ist. Dadurch wird vermieden, dass Instances am Ende des Skalierungsprozesses im Leerlauf verbleiben.

Die Strategie verwendet eine all-or-nothingLogik für den Start der EC2-Instances für jeden Job sowie eine all-or-nothingLogik für die Zuweisung der EC2-Instances zu Slurm-Knoten.

Die Strategiegruppen starten Anfragen in Batches, eine für jede angeforderte Rechenressource und jeweils bis zu 500 Knoten. Bei Anfragen, die sich über mehrere Rechenressourcen oder mehr als 500 Knoten erstrecken, werden mehrere ParallelCluster Batches nacheinander verarbeitet.

Der Ausfall eines Batches einer einzelnen Ressource führt zur Kündigung der gesamten zugehörigen ungenutzten Kapazität, wodurch sichergestellt wird, dass am Ende des Skalierungsprozesses keine inaktiven Instanzen übrig bleiben.

Einschränkungen

- Die für die Skalierung benötigte Zeit ist direkt proportional zur Anzahl der Jobs, die pro Ausführung des Slurm-Resume-Programms eingereicht wurden.
- Der Skalierungsvorgang ist durch das RunInstances Ressourcenkontenlimit begrenzt, das standardmäßig auf 1000 Instanzen festgelegt ist. Diese Beschränkung entspricht den AWS EC2-API-Drosselungsrichtlinien. Weitere Informationen finden Sie in der Dokumentation zur [AWS EC2-API-Drosselung](#)
- Wenn Sie einen Job in einer Rechenressource mit einem einzigen Instance-Typ in einer Warteschlange einreichen, die sich über mehrere Availability Zones erstreckt, ist der all-or-nothingEC2-Launch-API-Aufruf nur erfolgreich, wenn die gesamte Kapazität in einer einzigen Availability Zone bereitgestellt werden kann.
- Wenn Sie einen Job in einer Rechenressource mit mehreren Instance-Typen in einer Warteschlange mit einer einzigen Availability Zone einreichen, ist der all-or-nothingEC2-Start-API-Aufruf nur erfolgreich, wenn die gesamte Kapazität von einem einzigen Instance-Typ bereitgestellt werden kann.
- Wenn Sie einen Job in einer Rechenressource mit mehreren Instance-Typen in einer Warteschlange einreichen, die sich über mehrere Availability Zones erstreckt, wird der all-or-nothingEC2-Launch-API-Aufruf nicht unterstützt und stattdessen eine Skalierung nach bestem Aufwand ParallelCluster durchgeführt.

greedy-all-or-nothingSkalierung:

Diese Variante der all-or-nothing Strategie stellt weiterhin sicher, dass der Cluster nur skaliert wird, wenn die erforderliche Kapazität pro Job verfügbar ist, wodurch inaktive Instances am Ende des Skalierungsprozesses vermieden werden. Sie beinhaltet jedoch die ParallelCluster Initiierung eines Amazon EC2 EC2-Launch-Instance-API-Aufrufs, der auf eine Mindestzielkapazität von 1 abzielt und versucht, die Anzahl der gestarteten Knoten bis zur angeforderten Kapazität zu maximieren. Die Strategie verwendet eine Best-Effort-Logik für den Start der EC2-Instances für alle Jobs sowie die all-or-nothing-Logik für die Zuweisung der EC2-Instances zu Slurm-Knoten für jeden Job.

Die Strategie gruppiert Anfragen in Batches, eine für jede angeforderte Rechenressource und jeweils bis zu 500 Knoten. Bei Anfragen, die sich über mehrere Rechenressourcen oder mehr als 500 Knoten erstrecken, verarbeitet ParallelCluster nacheinander mehrere Batches.

Es stellt sicher, dass am Ende des Skalierungsprozesses keine inaktiven Instanzen übrig bleiben, indem der Durchsatz auf Kosten einer vorübergehenden Überskalierung während des Skalierungsprozesses maximiert wird.

Einschränkungen

- Eine vorübergehende Überskalierung ist möglich, was zu zusätzlichen Kosten für Instances führt, die vor Abschluss der Skalierung in den Betriebszustand übergehen.
- Es gilt das gleiche Instance-Limit wie in der all-or-nothing Strategie, abhängig vom Limit für AWS das RunInstances Ressourcenkonto.

Skalierung nach bestem Aufwand:

Bei dieser Strategie wird die EC2-Launch-Instance-API aufgerufen, wobei eine Mindestkapazität von 1 angestrebt wird und versucht wird, die gesamte angeforderte Kapazität zu erreichen, wobei die Instances nach der Ausführung des Skalierungsprozesses im Leerlauf belassen werden müssen, falls nicht die gesamte angeforderte Kapazität verfügbar ist. Die Strategie verwendet eine Best-Effort-Logik für den Start der EC2-Instances für alle Jobs sowie die Best-Effort-Logik für die Zuweisung der Amazon EC2 EC2-Instances zu Slurm-Knoten für jeden Job.

Die Strategie gruppiert Anfragen in Batches, eine für jede angeforderte Rechenressource und jeweils bis zu 500 Knoten. Bei Anfragen, die sich über mehrere Rechenressourcen oder mehr als 500 Knoten erstrecken, werden mehrere ParallelCluster Batches nacheinander verarbeitet.

Diese Strategie ermöglicht eine Skalierung weit über das Standardlimit von 1000 Instanzen bei mehreren Ausführungen von Skalierungsprozessen hinaus, allerdings auf Kosten inaktiver Instanzen für die verschiedenen Skalierungsprozesse.

Einschränkungen

- Mögliche Instances, die am Ende des Skalierungsprozesses im Leerlauf laufen, für den Fall, dass es nicht möglich ist, alle von den Jobs angeforderten Knoten zuzuweisen.

Das folgende Beispiel zeigt, wie sich die Skalierung dynamischer Knoten unter Verwendung der verschiedenen ParallelCluster Startstrategien verhält. Angenommen, Sie haben zwei Jobs eingereicht, bei denen jeweils 20 Knoten angefordert wurden, also insgesamt 40 Knoten desselben Typs, aber es sind nur 30 EC2-Instances verfügbar, um die angeforderte Kapazität auf EC2 abzudecken.

all-or-nothingSkalierung:

- Für den ersten Job wird eine all-or-nothingEC2-Launch-Instance-API aufgerufen, die 20 Instances anfordert. Ein erfolgreicher Aufruf hat zum Start von 20 Instances geführt
- all-or-nothing Die Zuweisung der 20 gestarteten Instanzen zu Slurm-Knoten für den ersten Job ist erfolgreich
- Eine weitere all-or-nothingEC2-Launch-Instance-API wird aufgerufen, die 20 Instances für den zweiten Job anfordert. Der Aufruf ist nicht erfolgreich, da nur Kapazität für weitere 10 Instances vorhanden ist. Derzeit werden keine Instances gestartet

greedy-all-or-nothingSkalierung:

- Es wird eine Best-Effort-EC2-Launch-Instance-API aufgerufen, die 40 Instances anfordert, was der Gesamtkapazität entspricht, die von allen Jobs angefordert wird. Dies führt zum Start von 30 Instances
- Eine all-or-nothingZuweisung von 20 der gestarteten Instances zu Slurm-Knoten für den ersten Job ist erfolgreich
- Eine weitere all-or-nothingZuweisung der verbleibenden gestarteten Instanzen zu Slurm-Knoten für den zweiten Job wird versucht, aber da von den insgesamt 20 vom Job angeforderten Instanzen nur 10 verfügbar sind, ist die Zuweisung nicht erfolgreich
- Die 10 nicht zugewiesenen gestarteten Instances werden beendet

Skalierung nach bestem Aufwand:

- Es wird eine Best-Effort-EC2-Launch-Instance-API aufgerufen, die 40 Instances anfordert, was der Gesamtkapazität entspricht, die von allen Jobs angefordert wird. Dies führt zum Start von 30 Instances.
- Eine Best-Effort-Zuordnung von 20 der gestarteten Instances zu Slurm-Knoten für den ersten Job ist erfolgreich.
- Eine weitere Best-Effort-Zuweisung der verbleibenden 10 gestarteten Instances zu Slurm-Knoten für den zweiten Job ist erfolgreich, auch wenn die angeforderte Gesamtkapazität 20 betrug. Da der Job jedoch die 20 Knoten anforderte und es möglich war, EC2-Instances nur 10 von ihnen zuzuweisen, kann der Job nicht gestartet werden und die Instances laufen im Leerlauf, bis genügend Kapazität gefunden wurde, um die fehlenden 10 Instances bei einem späteren Aufruf des Skalierungsprozesses zu starten, oder der Scheduler plant den Job auf anderen, bereits laufenden Rechenknoten.

SlurmStrategien zur dynamischen Knotenzuweisung in Version 3.7.x

ParallelCluster verwendet zwei Arten von Strategien zur dynamischen Knotenzuweisung, um den Cluster zu skalieren:

- Zuweisung auf der Grundlage verfügbarer angeforderter Knoteninformationen:
 - Wiederaufnahme der Skalierung aller Knoten oder Skalierung der Knotenliste:

ParallelCluster skaliert den Cluster nur auf Slurm der Grundlage der angeforderten Knotenlistenamen, wenn er ausgeführt wird `Slurm. ResumeProgram`. Es weist den Knoten Rechenressourcen nur anhand des Knotennamens zu. Die Liste der Knotennamen kann mehrere Jobs umfassen.

- Lebenslauf auf Jobebene oder Skalierung auf Jobebene:

ParallelCluster skaliert den Cluster auf der Grundlage der Anforderungen der einzelnen Jobs, der aktuellen Anzahl der Knoten, die dem Job zugewiesen sind, und der Knoten, die wieder aufgenommen werden müssen. ParallelCluster ruft diese Informationen aus der `SLURM_RESUME_FILE` Umgebungsvariablen ab.

- Allokation mit einer EC2-Startstrategie:
 - Skalierung nach bestem Aufwand:

ParallelCluster skaliert den Cluster mithilfe eines EC2-Launch-Instance-API-Aufrufs mit einer Mindestzielkapazität von 1, um einige, aber nicht unbedingt alle Instances zu starten, die zur Unterstützung der angeforderten Knoten benötigt werden.

- Eine II-or-nothing Skalierung:

ParallelCluster skaliert den Cluster mithilfe eines EC2-Launch-Instance-API-Aufrufs, der nur erfolgreich ist, wenn alle Instances gestartet werden, die zur Unterstützung der angeforderten Knoten benötigt werden. In diesem Fall ruft er die EC2-Launch-Instance-API auf, wobei die minimale Zielkapazität der angeforderten Gesamtkapazität entspricht.

Standardmäßig ParallelCluster verwendet es die Skalierung der Knotenliste mit einer EC2-Startstrategie nach bestem Wissen und Gewissen, um einige, aber nicht unbedingt alle Instances zu starten, die zur Unterstützung der angeforderten Knoten benötigt werden. Es versucht, so viel Kapazität wie möglich bereitzustellen, um die eingereichte Arbeitslast zu bedienen.

Ab ParallelCluster Version 3.7.0 wird die Skalierung auf Jobebene mit einer all-or-nothing-EC2-Startstrategie für Jobs ParallelCluster verwendet, die im exklusiven Modus eingereicht wurden. Wenn Sie einen Job im exklusiven Modus einreichen, hat der Job exklusiven Zugriff auf die ihm zugewiesenen Knoten. Weitere Informationen finden Sie in der Slurm Dokumentation unter [EXCLUSIVE](#).

Um einen Job im exklusiven Modus einzureichen:

- Übergeben Sie die Exklusivkennzeichnung, wenn Sie einen Slurm Job an den Cluster senden. z. B.
`sbatch ... --exclusive.`

ODER

- Senden Sie einen Job an eine Cluster-Warteschlange, die mit der [JobExclusiveAllocation](#) Einstellung auf konfiguriert wurde `true`.

Wenn Sie einen Job im exklusiven Modus einreichen:

- ParallelCluster führt derzeit Batches von Startanfragen mit bis zu 500 Knoten durch. Wenn ein Job mehr als 500 Knoten anfordert, ParallelCluster stellt er eine all-or-nothing-Startanforderung für jeden Satz von 500 Knoten und eine zusätzliche Startanforderung für die restlichen Knoten.
- Wenn sich die Knotenzuweisung auf eine einzelne Rechenressource ParallelCluster bezieht, wird eine all-or-nothing-Startanforderung für jeden Satz von 500 Knoten und eine zusätzliche

Startanforderung für die restlichen Knoten gestellt. Wenn eine Startanfrage fehlschlägt, wird die ungenutzte Kapazität, die durch alle Startanfragen geschaffen wurde, ParallelCluster beendet.

- Wenn sich die Knotenzuweisung über mehrere Rechenressourcen erstreckt, ParallelCluster muss für jede Rechenressource eine all-or-nothing-Startanforderung gestellt werden. Diese Anfragen werden ebenfalls gebündelt. Wenn eine Startanfrage für eine der Rechenressourcen fehlschlägt, wird die ungenutzte Kapazität ParallelCluster beendet, die durch alle Startanfragen für Rechenressourcen geschaffen wurde.

Skalierung auf Jobebene mit bekannten Einschränkungen der all-or-nothing-Startstrategie:

- Wenn Sie einen Job in einer Rechenressource mit einem einzigen Instance-Typ in einer Warteschlange einreichen, die sich über mehrere Availability Zones erstreckt, ist der all-or-nothing-EC2-Start-API-Aufruf nur erfolgreich, wenn die gesamte Kapazität in einer einzigen Availability Zone bereitgestellt werden kann.
- Wenn Sie einen Job in einer Rechenressource mit mehreren Instance-Typen in einer Warteschlange mit einer einzigen Availability Zone einreichen, ist der all-or-nothing-EC2-Start-API-Aufruf nur erfolgreich, wenn die gesamte Kapazität von einem einzigen Instance-Typ bereitgestellt werden kann.
- Wenn Sie einen Job in einer Rechenressource mit mehreren Instance-Typen in einer Warteschlange einreichen, die sich über mehrere Availability Zones erstreckt, wird der all-or-nothing-EC2-Launch-API-Aufruf nicht unterstützt und stattdessen eine Skalierung nach bestem Aufwand ParallelCluster durchgeführt.

Slurm-Strategien zur dynamischen Knotenzuweisung in Version 3.6.x und früheren Versionen

AWS ParallelCluster verwendet nur eine Art von Strategie für die dynamische Knotenzuweisung, um den Cluster zu skalieren:

- Zuweisung auf der Grundlage verfügbarer angeforderter Knoteninformationen:
 - Wiederaufnahme aller Knoten oder Skalierung der Knotenliste: ParallelCluster skaliert den Cluster nur auf Slurm der Grundlage der angeforderten Knotenlistenamen, wenn er ausgeführt wird. Slurm-Resume-Programme Rechenressourcen werden Knoten nur anhand des Knotennamens zugewiesen. Die Liste der Knotennamen kann mehrere Jobs umfassen.
- Zuteilung mit einer EC2-Startstrategie:
 - Skalierung nach bestem Aufwand: ParallelCluster skaliert den Cluster mithilfe eines EC2-Launch-Instance-API-Aufrufs mit einer Mindestzielkapazität von 1, um einige, aber nicht

unbedingt alle Instances zu starten, die zur Unterstützung der angeforderten Knoten benötigt werden.

ParallelCluster verwendet die Skalierung der Knotenliste mit einer EC2-Startstrategie nach bestem Aufwand, um einige, aber nicht unbedingt alle Instances zu starten, die zur Unterstützung der angeforderten Knoten benötigt werden. Es wird versucht, so viel Kapazität wie möglich bereitzustellen, um die eingereichte Arbeitslast zu bedienen.

Einschränkungen

- Mögliche Instances, die am Ende des Skalierungsprozesses im Leerlauf laufen, für den Fall, dass es nicht möglich ist, alle von den Jobs angeforderten Knoten zuzuweisen.

SlurmAbrechnung mit AWS ParallelCluster

AWS ParallelCluster unterstützt ab Version 3.3.0 die Slurm Abrechnung mit dem Cluster-Konfigurationsparameter [SlurmSettings/Database](#).

Mit Slurm Accounting können Sie eine externe Buchhaltungsdatenbank integrieren, um Folgendes zu tun:

- Verwalten Sie Cluster-Benutzer oder Benutzergruppen und andere Entitäten. Mit dieser Funktion können Sie erweiterte Funktionen wie die Durchsetzung Slurm von Ressourcenbeschränkungen, Fairshare und QoS nutzen.
- Sammeln und speichern Sie Jobdaten, z. B. den Benutzer, der den Job ausgeführt hat, die Dauer des Jobs und die verwendeten Ressourcen. Sie können die gespeicherten Daten mit dem `sacct` Hilfsprogramm anzeigen.

Note

AWS ParallelCluster unterstützt die Slurm Abrechnung [Slurmunterstützter MySQL-Datenbankserver](#).

Arbeiten mit Slurm Buchhaltung in AWS ParallelCluster

Bevor Sie die Slurm Kontoführung konfigurieren, müssen Sie über einen vorhandenen externen Datenbankserver und eine Datenbank verfügen, die `mysql` das Protokoll verwendet.

Um die Slurm Kontoführung mit zu konfigurieren AWS ParallelCluster, müssen Sie Folgendes definieren:

- Der URI für den externen Datenbankserver in [Database//Uri](#). Der Server muss existieren und vom Hauptknoten aus erreichbar sein.
- Anmeldeinformationen für den Zugriff auf die externe Datenbank, die in [Database/PasswordSecretArn](#) und [Database/](#)definiert sind [UserName](#). AWS ParallelCluster verwendet diese Informationen, um die Kontoführung auf der Slurm Ebene und den `slurmdbd` Dienst auf dem Hauptknoten zu konfigurieren. `slurmdbd` ist der Daemon, der die Kommunikation zwischen dem Cluster und dem Datenbankserver verwaltet.

Eine schrittweise Anleitung finden Sie unter [Einen Cluster mit Slurm Accounting erstellen](#).

Note

AWS ParallelCluster führt einen grundlegenden Bootstrap der Slurm Accounting-Datenbank durch, indem der Standard-Clusterbenutzer in der Slurm Datenbank als Datenbankadministrator festgelegt wird. AWS ParallelCluster fügt der Accounting-Datenbank keinen weiteren Benutzer hinzu. Der Kunde ist für die Verwaltung der Buchhaltungseinheiten in der Slurm Datenbank verantwortlich.

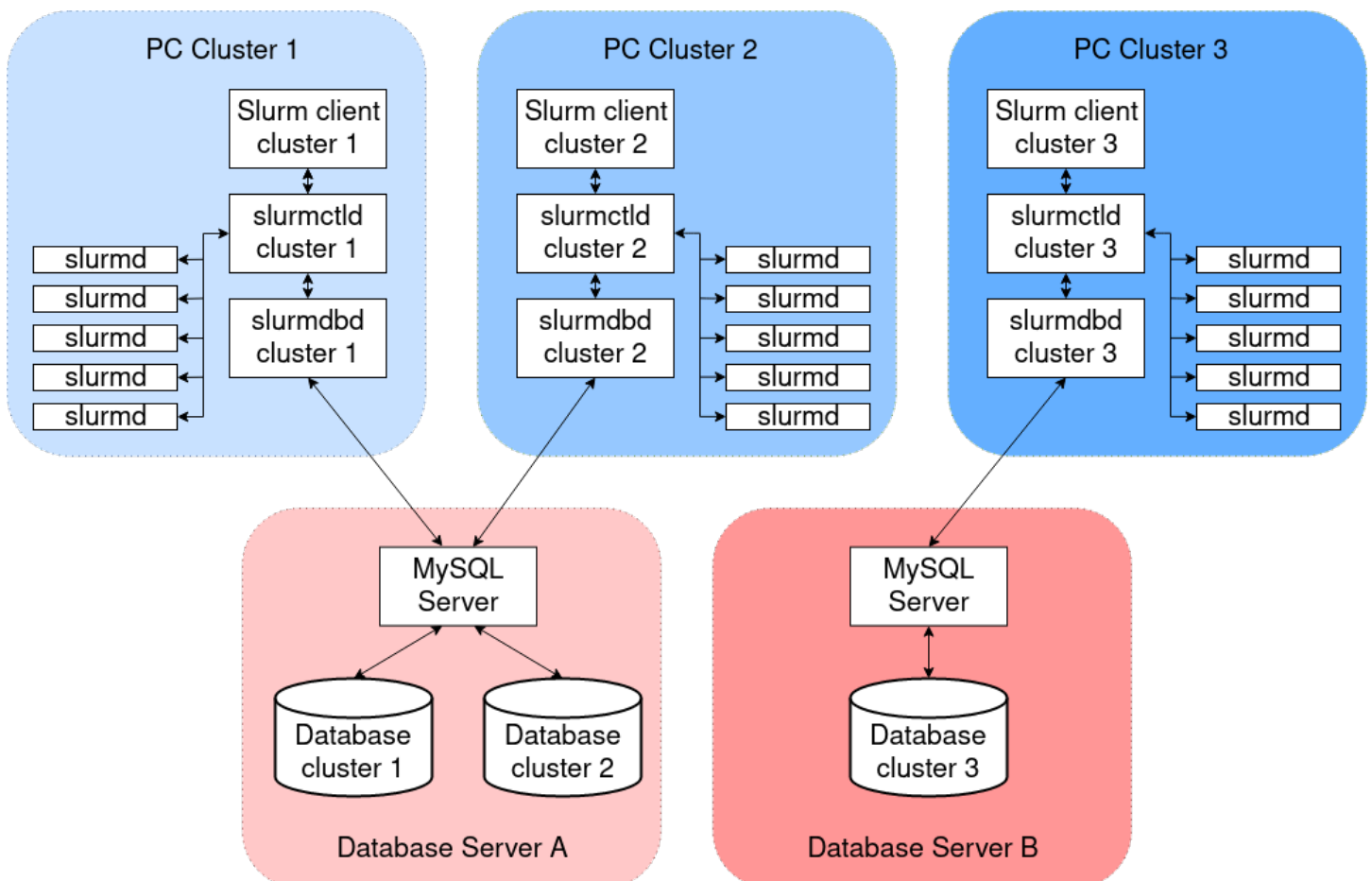
AWS ParallelCluster konfiguriert `slurmdbd`, um sicherzustellen, dass ein Cluster über eine eigene Slurm Datenbank auf dem Datenbankserver verfügt. Derselbe Datenbankserver kann in mehreren Clustern verwendet werden, aber jeder Cluster hat seine eigene separate Datenbank. AWS ParallelCluster verwendet den Clusternamen, um den Namen für die Datenbank im `StorageLoc` Parameter der `slurmdbd` Konfigurationsdatei zu definieren. Stellen Sie sich die folgende Situation vor. Eine Datenbank, die auf dem Datenbankserver vorhanden ist, enthält einen Clusternamen, der keinem aktiven Clusternamen zugeordnet ist. In diesem Fall können Sie einen neuen Cluster mit diesem Clusternamen erstellen, der dieser Datenbank zugeordnet werden soll. Slurm verwendet die Datenbank für den neuen Cluster erneut.

Warning

- Es wird nicht empfohlen, mehr als einen Cluster einzurichten, um dieselbe Datenbank gleichzeitig zu verwenden. Dies kann zu Leistungsproblemen oder sogar zu Datenbank-Deadlock-Situationen führen.

- Wenn Slurm Accounting auf dem Hauptknoten eines Clusters aktiviert ist, empfehlen wir die Verwendung eines Instance-Typs mit einer leistungsstarken CPU, mehr Arbeitsspeicher und höherer Netzwerkbandbreite. SlurmAccounting kann den Hauptknoten des Clusters zusätzlich belasten.

In der aktuellen Architektur der AWS ParallelCluster Slurm Accounting-Funktion hat jeder Cluster seine eigene Instanz des `slurmdbd` Daemons, wie in der folgenden Abbildung mit Beispielkonfigurationen dargestellt.



Wenn Sie Ihrer Slurm Cluster-Umgebung benutzerdefinierte Multi-Cluster- oder Verbundfunktionen hinzufügen, müssen alle Cluster auf dieselbe `slurmdbd` Instanz verweisen. Für diese Alternative empfehlen wir, dass Sie die AWS ParallelCluster Slurm Kontoführung auf einem Cluster aktivieren und die anderen Cluster manuell so konfigurieren, `slurmdbd` dass sie eine Verbindung zu den Clustern herstellen, die auf dem ersten Cluster gehostet werden.

Wenn Sie AWS ParallelCluster Versionen vor Version 3.3.0 verwenden, lesen Sie die alternative Methode zur Implementierung der Slurm Buchhaltung, die in diesem [HPC-Blogbeitrag](#) beschrieben wird.

SlurmÜberlegungen zur Rechnungslegung

Datenbank und Cluster auf verschiedenen VPCs

Um die Slurm Kontoführung zu aktivieren, wird ein Datenbankserver benötigt, der als Backend für die Lese- und Schreiboperationen dient, die der `slurmdbd` Daemon ausführt. Bevor der Cluster erstellt oder aktualisiert wird, um die Slurm Kontoführung zu aktivieren, muss der Hauptknoten in der Lage sein, den Datenbankserver zu erreichen.

Wenn Sie den Datenbankserver auf einer anderen VPC als der, die der Cluster verwendet, bereitstellen müssen, sollten Sie Folgendes berücksichtigen:

- Um die Kommunikation zwischen der `slurmdbd` Clusterseite und dem Datenbankserver zu ermöglichen, müssen Sie die Konnektivität zwischen den beiden VPCs einrichten. Weitere Informationen finden Sie unter [VPC Peering](#) im Amazon Virtual Private Cloud Cloud-Benutzerhandbuch.
- Sie müssen die Sicherheitsgruppe erstellen, die Sie dem Hauptknoten auf der VPC des Clusters zuordnen möchten. Nachdem die beiden VPCs miteinander verbunden wurden, ist eine Querverknüpfung zwischen den Sicherheitsgruppen auf der Datenbankseite und der Clusterseite verfügbar. Weitere Informationen finden Sie unter [Sicherheitsgruppenregeln](#) im Amazon Virtual Private Cloud Cloud-Benutzerhandbuch.

Konfiguration der TLS-Verschlüsselung zwischen `slurmdbd` und dem Datenbankserver

`slurmdbd` stellt mit der standardmäßigen Slurm Kontoführungskonfiguration, die dies AWS ParallelCluster vorsieht, eine TLS-verschlüsselte Verbindung zum Datenbankserver her, sofern der Server die TLS-Verschlüsselung unterstützt. AWS Datenbankdienste wie Amazon RDS Amazon Aurora unterstützen standardmäßig TLS-Verschlüsselung.

Sie können sichere Verbindungen auf der Serverseite verlangen, indem Sie den `require_secure_transport` Parameter auf dem Datenbankserver festlegen. Dies ist in der bereitgestellten CloudFormation Vorlage konfiguriert.

Gemäß den bewährten Sicherheitsmethoden empfehlen wir, dass Sie auch die Überprüfung der Serveridentität auf dem `slurmdbd` Client aktivieren. Konfigurieren Sie dazu den

[StorageParameters](#) in `slurmdbd.conf`. Laden Sie das CA-Zertifikat des Servers auf den Hauptknoten des Clusters hoch. Stellen Sie als Nächstes die [SSL_CA-Option](#) `StorageParameters` in `slurmdbd.conf` auf den Pfad des Server-CA-Zertifikats auf dem Hauptknoten ein. Dadurch wird nebenbei die Überprüfung der Serveridentität aktiviert. `slurmdbd` Nachdem Sie diese Änderungen vorgenommen haben, starten Sie den `slurmdbd` Dienst neu, um die Verbindung zum Datenbankserver mit aktivierter Identitätsprüfung wiederherzustellen.

Aktualisierung der Datenbankanmeldedaten

Um die Werte für [Database/UserName](#) oder zu aktualisieren [PasswordSecretArn](#), müssen Sie zuerst die Rechenflotte beenden. Angenommen, der geheime Wert, der im AWS Secrets Manager Secret gespeichert ist, wird geändert und sein ARN wird nicht geändert. In diesem Fall aktualisiert der Cluster das Datenbankkennwort nicht automatisch auf den neuen Wert. Um den Cluster für den neuen geheimen Wert zu aktualisieren, führen Sie den folgenden Befehl vom Hauptknoten aus aus.

```
$ sudo /opt/parallelcluster/scripts/slurm/update_slurm_database_password.sh
```

Warning

Um den Verlust von Buchhaltungsdaten zu vermeiden, empfehlen wir, das Datenbankkennwort nur zu ändern, wenn die Rechenflotte gestoppt ist.

Datenbank-Überwachung

Wir empfehlen, dass Sie die Überwachungsfunktionen der AWS Datenbankdienste aktivieren. Weitere Informationen finden Sie in der Dokumentation zur [Amazon RDS-Überwachung](#) oder [Amazon Aurora Aurora-Überwachung](#).

SlurmAnpassung der Konfiguration

Beginnend mit `AWS ParallelClusterVersion 3.6.0`, Sie können die `anpassenslurm.conf` SlurmKonfiguration in einem `AWS ParallelClusterCluster-Konfiguration`.

In der Clusterkonfiguration können Sie `anpassenslurmKonfigurationsparameter` mithilfe der folgenden Cluster-Konfigurationseinstellungen:

- `AnpassenslurmParameter` für den gesamten Cluster, indem Sie entweder [SlurmSettings/CustomSlurmSettings](#) oder

der [CustomSlurmSettingsIncludeFile](#) Parameter. AWS ParallelCluster schlägt fehl, wenn Sie beide angeben.

- Anpassen Slurm Parameter für eine Warteschlange mithilfe von [SlurmQueues/CustomSlurmSettings](#) (zugeordnet zu Slurm Partitionen).
- Personalisieren Slurm Parameter für eine Rechenressource mithilfe von [SlurmQueues/ComputeResources/CustomSlurmSettings](#) (zugeordnet zu Slurm Knoten).

Slurm Einschränkungen bei der Anpassung der Konfiguration und Überlegungen bei der Verwendung AWS ParallelCluster

- Für [CustomSlurmSettings](#) und [CustomSlurmSettingsIncludeFile](#) Einstellungen, die Sie nur spezifizieren und aktualisieren können `slurm.conf` Parameter, die in der enthalten sind [SlurmAusführung](#) das wird unterstützt von AWS ParallelCluster Version, die Sie verwenden, um einen Cluster zu konfigurieren.
- Wenn Sie benutzerdefiniert angeben Slurm Konfigurationen in einer [CustomSlurmSettings](#) Parameter, AWS ParallelCluster führt Validierungsprüfungen durch und verhindert das Setzen oder Aktualisieren Slurm Konfigurationsparameter, die in Konflikt stehen mit AWS ParallelCluster Logik. Die Slurm Konfigurationsparameter, von denen bekannt ist, dass sie in Konflikt stehen AWS ParallelCluster sind in Verweigerungslisten aufgeführt. Die Ablehnungslisten können sich in Zukunft ändern AWS ParallelCluster Versionen, falls andere Slurm Funktionen werden hinzugefügt. Weitere Informationen finden Sie unter [Auf der Denim-Liste Slurm Konfigurationsparameter für CustomSlurmSettings](#).
- AWS ParallelCluster prüft nur, ob ein Parameter in einer Sperrliste enthalten ist. AWS ParallelCluster validiert Ihre benutzerdefinierte Einstellung nicht Slurm Syntax oder Semantik von Konfigurationsparametern. Sie sind dafür verantwortlich, Ihren Benutzerdefiniert zu validieren Slurm Konfigurationsparameter. Ungültiger Benutzerdefiniert Slurm Konfigurationsparameter können folgende Ursachen haben Slurm Daemon-Fehler, die zu Fehlern bei der Clustererstellung und -aktualisierung führen können.
- Wenn Sie Benutzerdefiniert angeben Slurm Konfigurationen in [CustomSlurmSettingsIncludeFile](#), AWS ParallelCluster führt keine Validierung durch.
- Sie können aktualisieren [CustomSlurmSettings](#) und [CustomSlurmSettingsIncludeFile](#) ohne die Rechenflotte anzuhalten und zu starten. In diesem Fall AWS ParallelCluster startet das `slurmctl daemon` und führt den `control reconfigure` Befehl.

Einige Slurm-Konfigurationsparameter erfordern möglicherweise unterschiedliche Operationen, bevor eine Änderung im gesamten Cluster registriert wird. Beispielsweise können sie einen Neustart aller Daemons im Cluster erfordern. Sie sind dafür verantwortlich zu überprüfen, ob AWS ParallelCluster-Operationen reichen aus, um Ihre benutzerdefinierte Version zu verbreiten. Slurm-Einstellungen der Konfigurationsparameter bei Updates. Wenn du das findest, AWS ParallelCluster-Operationen reichen nicht aus, es liegt in Ihrer Verantwortung, die zusätzlichen Maßnahmen zu ergreifen, die erforderlich sind, um die aktualisierten Einstellungen zu verbreiten, wie in der [Slurm-Dokumentation](#).

Auf der Denim-Liste Slurm-Konfigurationsparameter für Custom Slurm Settings

In den folgenden Tabellen sind die Parameter mit den AWS ParallelCluster-Versionen, die ihre Verwendung verweigern, beginnend mit Version 3.6.0. Custom Slurm Settings wird nicht unterstützt für AWS ParallelCluster-Versionen vor Version 3.6.0.

Parameter auf Clusterebene, die auf der Denim-Liste stehen:

Slurm Parameter	Auf Verweigerliste gesetzt in AWS ParallelCluster-Versionen
CommunicationParameters	3.6.0
Epilog	3.6.0
GresTypes	3.6.0
LaunchParameters	3.6.0
Prolog	3.6.0
ReconfigFlags	3.6.0
ResumeFailProgram	3.6.0
ResumeProgram	3.6.0
ResumeTimeout	3.6.0
SlurmctlHost	3.6.0

Slurm Parameter	Auf Verweigerliste gesetzt inAWS ParallelClusterVersionen
SlurmctdLogFile	3.6.0
SlurmctdParameters	3.6.0
SlurmdLogFile	3.6.0
SlurmUser	3.6.0
SuspendExcNodes	3.6.0
SuspendProgram	3.6.0
SuspendTime	3.6.0
TaskPlugin	3.6.0
TreeWidth	3.6.0

Parameter auf Clusterebene, die auf der Negativliste stehen, wenn [gebürtigSlurmIntegration der Buchhaltung](#) ist in der Cluster-Konfiguration konfiguriert:

Slurm Parameter	Auf Verweigerliste gesetzt inAWS ParallelClusterVersionen
AccountingStorageType	3.6.0
AccountingStorageHost	3.6.0
AccountingStoragePort	3.6.0
AccountingStorageUser	3.6.0
JobAcctGatherType	3.6.0

Parameter auf der Warteschlangenebene (Partitionsebene) auf der Sperrliste für Warteschlangen, die von verwaltet werdenAWS ParallelCluster:

Slurm Parameter	Auf Verweigerliste gesetzt inAWS ParallelClusterVersionen
Knoten	3.6.0
PartitionName	3.6.0
ResumeTimeout	3.6.0
Status	3.6.0
SuspendTime	3.6.0

Parameter auf der Ebene der Rechenressourcen (Knoten), die auf der Negativliste stehen, für Rechenressourcen, die von verwaltet werdenAWS ParallelCluster:

Slurm Parameter	Auf Verweigerliste gesetzt inAWS ParallelClusterVersion und spätere Versionen
CPUs	3.6.0
Features	3.6.0
Gres	3.6.0
NodeAddr	3.6.0
NodeHostname	3.6.0
NodeName	3.6.0
Gewicht	3.7.0

Slurm **prolog** und **epilog**


Ab AWS ParallelCluster Version 3.6.0 AWS ParallelCluster umfasst die Slurm Konfiguration, mit der bereitgestellt wird, die Epilog Konfigurationsparameter Prolog und:

```
# PROLOG AND EPILOG
Prolog=/opt/slurm/etc/scripts/prolog.d/*
Epilog=/opt/slurm/etc/scripts/epilog.d/*
SchedulerParameters=nohold_on_prolog_fail
BatchStartTimeout=180
```

Weitere Informationen finden Sie im [Prolog- und Epilog-Handbuch](#) in der Slurm Dokumentation.


AWS ParallelCluster beinhaltet die folgenden Prolog- und Epilog-Skripts:

- `90_plcluster_health_check_manager`(im Prolog Ordner)
- `90_pcluster_noop`(im Epilog Ordner)

 Note

Die Epilog Ordner Prolog und müssen mindestens eine Datei enthalten.

Sie können Ihre eigenen benutzerdefinierten prolog oder epilog Skripts verwenden, indem Sie sie zu den entsprechenden Epilog Ordnern Prolog und Ordnern hinzufügen.

 Warning

Slurmführt jedes Skript in den Ordnern in umgekehrter alphabetischer Reihenfolge aus.

Die Laufzeit der prolog und epilog -Skripts wirkt sich auf die Zeit aus, die für die Ausführung eines Jobs benötigt wird. Aktualisieren Sie die BatchStartTimeout Konfigurationseinstellung, wenn Sie mehrere oder lang laufende prolog Skripts ausführen. Die Voreinstellung ist ein Zeitraum von 3 Minuten festgelegt.

Wenn Sie benutzerdefinierte prolog und epilog Skripts verwenden, suchen Sie die Skripts in den entsprechenden Epilog Ordnern Prolog und. Wir empfehlen, `90_plcluster_health_check_manager` das Standardeinstellungen beizubehalten. Weitere Informationen finden Sie unter [SlurmAnpassung der Konfiguration](#).

Größe und Aktualisierung der Clusterkapazität

Die Kapazität des Clusters wird durch die Anzahl der Rechenknoten definiert, die der Cluster skalieren kann. Rechenknoten werden von EC2-Instances unterstützt, die in der AWS ParallelCluster Konfiguration innerhalb der Rechenressourcen definiert sind. Sie sind in Warteschlangen organisiert (`Scheduling/SlurmQueues/ComputeResources`), (`Scheduling/SlurmQueues`) die Partitionen 1:1 zugeordnet Slurm werden.

Innerhalb einer Rechenressource ist es möglich, die Mindestanzahl von Rechenknoten (Instanzen) zu konfigurieren, die immer im Cluster laufen müssen (`MinCount`), und die maximale Anzahl von Instanzen, auf die die Rechenressource skaliert werden kann (`MaxCount3`).

AWS ParallelCluster Startet bei der Clustererstellung oder bei einem Cluster-Update so viele EC2-Instances, wie `MinCount` für jede im Cluster definierte Rechenressource (`Scheduling/SlurmQueues/ComputeResources`) konfiguriert sind. Die Instances, die gestartet werden, um die minimale Anzahl von Knoten für Rechenressourcen im Cluster abzudecken, werden als statische Knoten bezeichnet. Einmal gestartet, sollen statische Knoten im Cluster persistent sein und werden nicht vom System beendet, es sei denn, ein bestimmtes Ereignis oder eine bestimmte Bedingung tritt ein. Zu diesen Ereignissen gehören beispielsweise der Ausfall von EC2-Integritätsprüfungen Slurm oder die Änderung des Status des Slurm-Knotens auf DRAIN oder DOWN.

Die EC2-Instances im Bereich von **1** bis **'MaxCount - MinCount'** (**MaxCount** minus **MinCount**), die bei Bedarf gestartet werden, um die erhöhte Last des Clusters zu bewältigen, werden als dynamische Knoten bezeichnet. Sie sind kurzlebig. Sie werden gestartet, um ausstehende Aufträge zu bearbeiten, und werden beendet, sobald sie für einen `Scheduling/SlurmSettings/ScaledownIdleTime` in der Cluster-Konfiguration festgelegten Zeitraum inaktiv bleiben (Standard: 10 Minuten).

Statische Knoten und dynamische Knoten entsprechen dem folgenden Benennungsschema:

- Statische Knoten `<Queue/Name>-st-<ComputeResource/Name>-<num>` wo `<num>` = `1..ComputeResource/MinCount`
- Dynamische Knoten `<Queue/Name>-dy-<ComputeResource/Name>-<num>` wo `<num>` = `1..(ComputeResource/MaxCount - ComputeResource/MinCount)`

Zum Beispiel bei der folgenden AWS ParallelCluster Konfiguration:

```
Scheduling:
```

```
Scheduler: slurm
SlurmQueues:
  - Name: queue1
    ComputeResources:
      - Name: c5xlarge
        Instances:
          - InstanceType: c5.xlarge
            MinCount: 100
            MaxCount: 150
```

Die folgenden Knoten werden definiert in Slurm

```
$ sinfo
PARTITION AVAIL  TIMELIMIT  NODES  STATE NODELIST
queue1*   up      infinite   50    idle~ queue1-dy-c5xlarge-[1-50]
queue1*   up      infinite  100    idle queue1-st-c5xlarge-[1-100]
```

Wenn es sich bei einer Rechenressource um statische Rechenknoten handelt `MinCount == MaxCount`, sind alle zugehörigen Rechenknoten statisch und alle Instanzen werden bei der Clustererstellung/Aktualisierung gestartet und laufen weiter. Beispielsweise:

```
Scheduling:
Scheduler: slurm
SlurmQueues:
  - Name: queue1
    ComputeResources:
      - Name: c5xlarge
        Instances:
          - InstanceType: c5.xlarge
            MinCount: 100
            MaxCount: 100
```

```
$ sinfo
PARTITION AVAIL  TIMELIMIT  NODES  STATE NODELIST
queue1*   up      infinite   100    idle queue1-st-c5xlarge-[1-100]
```

Aktualisierung der Cluster-Kapazität

Die Aktualisierung der Clusterkapazität umfasst das Hinzufügen oder Entfernen von Warteschlangen, Rechenressourcen oder das Ändern MinCount/MaxCount einer Rechenressource. Ab AWS ParallelCluster Version 3.9.0 muss zur Reduzierung der Größe einer Warteschlange die Rechenflotte gestoppt oder auf TERMINATE [QueueUpdateStrategy](#) gesetzt werden, bevor ein Cluster-Update stattfinden kann. In folgenden Fällen ist es nicht erforderlich, die Rechenflotte zu beenden oder auf [QueueUpdateStrategy](#) TERMINATE zu setzen:

- Neue Warteschlangen zu Scheduling hinzufügen/ [SlurmQueues](#)
- Neue Rechenressourcen Scheduling/SlurmQueues/[ComputeResources](#) zu einer Warteschlange hinzufügen
- Erhöhung der [MaxCount](#) Anzahl einer Rechenressource
- Erhöhung MinCount einer Rechenressource und Erhöhung MaxCount derselben Rechenressource um mindestens die gleiche Menge

Überlegungen und Einschränkungen

In diesem Abschnitt sollen alle wichtigen Faktoren, Einschränkungen oder Einschränkungen beschrieben werden, die bei der Größenänderung der Clusterkapazität berücksichtigt werden sollten.

- Beim Entfernen einer Warteschlange aus Scheduling/<https://docs.aws.amazon.com/parallelcluster/latest/ug/Scheduling-v3.html#Scheduling-v3-SlurmQueues> SlurmQueues allen Rechenknoten mit Namen <Queue/Name> - *, sowohl statische als auch dynamische, werden sie aus der Slurm Konfiguration entfernt und die entsprechenden EC2-Instances werden beendet.
- Beim Entfernen einer Rechenressource Scheduling/SlurmQueues/<https://docs.aws.amazon.com/parallelcluster/latest/ug/Scheduling-v3.html#Scheduling-v3-SlurmQueues-ComputeResources> ComputeResources aus einer Warteschlange werden alle Rechenknoten mit Namen <Queue/Name> - * - <ComputeResource/Name> - *, sowohl statische als auch dynamische, aus der Slurm Konfiguration entfernt und die entsprechenden EC2-Instances werden beendet.

Wenn wir den MinCount Parameter einer Rechenressource ändern, können wir zwei verschiedene Szenarien unterscheiden: ob gleich gehalten MaxCount wird MinCount (nur statische Kapazität) und ob größer als MinCount (gemischte statische und dynamische Kapazität) MaxCount ist.

Die Kapazität ändert sich nur bei statischen Knoten

- Wenn `MinCount == MaxCount` beim Erhöhen `MinCount` (und `MaxCount`) der Cluster konfiguriert wird, indem die Anzahl der statischen Knoten auf den neuen Wert von erhöht wird `MinCount <Queue/Name>-st-<ComputeResource/Name>-<new_MinCount>` und das System weiterhin versucht, EC2-Instances zu starten, um die neue erforderliche statische Kapazität zu erreichen.
- Wenn `MinCount == MaxCount` beim Verringern `MinCount` (und `MaxCount`) der Anzahl `N` der Cluster konfiguriert wird, indem die letzten `N` statischen Knoten `<Queue/Name>-st-<ComputeResource/Name>-<old_MinCount - N>...<old_MinCount>` entfernt werden, beendet das System die entsprechenden EC2-Instances.
- Ausgangszustand `MinCount = MaxCount = 100`

```
$ sinfo
PARTITION AVAIL  TIMELIMIT  NODES  STATE NODELIST
queue1*    up        infinite   100    idle queue1-st-c5xlarge-[1-100]
```

- Update `-30` am `MinCount` und `MaxCount`: `MinCount = MaxCount = 70`

```
$ sinfo
PARTITION AVAIL  TIMELIMIT  NODES  STATE NODELIST
queue1*    up        infinite    70    idle queue1-st-c5xlarge-[1-70]
```

Kapazitätsänderungen bei gemischten Knoten

Wenn bei einer Erhöhung `MinCount` um den Betrag `N` (vorausgesetzt `MinCount < MaxCount`, dass der Wert unverändert bleibt), der Cluster konfiguriert wird, indem die Anzahl der statischen Knoten auf den neuen Wert von `MinCount (old_MinCount + N)`: erhöht wird, `<Queue/Name>-st-<ComputeResource/Name>-<old_MinCount + N>` und das System versucht weiterhin, EC2-Instances zu starten, um die neue erforderliche statische Kapazität zu erreichen. Um der `MaxCount` Kapazität der Rechenressource Rechnung zu tragen, wird außerdem die Cluster-Konfiguration aktualisiert, indem die letzten `N` dynamischen Knoten entfernt werden. `<Queue/Name>-dy-<ComputeResource/Name>-[<MaxCount - old_MinCount -`

N>...<MaxCount - old_MinCount>] Das System beendet dann die entsprechenden EC2-Instances.

- Ausgangszustand: MinCount = 100; MaxCount = 150

```
$ sinfo
PARTITION AVAIL  TIMELIMIT  NODES  STATE NODELIST
queue1*    up    infinite   50    idle~ queue1-dy-c5xlarge-[1-50]
queue1*    up    infinite  100    idle queue1-st-c5xlarge-[1-100]
```

- Aktualisieren Sie +30 auf MinCount : MinCount = 130 (MaxCount = 150)

```
$ sinfo
PARTITION AVAIL  TIMELIMIT  NODES  STATE NODELIST
queue1*    up    infinite   20    idle~ queue1-dy-c5xlarge-[1-20]
queue1*    up    infinite  130    idle queue1-st-c5xlarge-[1-130]
```

Wenn $\text{MinCount} < \text{MaxCount}$ MaxCount der Cluster bei Erhöhung MinCount und gleichem Wert N konfiguriert wird, wird der Cluster konfiguriert, indem die Anzahl der statischen Knoten auf den neuen Wert von MinCount ($\text{old_MinCount} + N$): erhöht wird <Queue/Name>-st-<ComputeResource/Name>-<old_MinCount + N> und das System weiterhin versucht, EC2-Instances zu starten, um die neue erforderliche statische Kapazität zu erreichen. Außerdem werden keine Änderungen an der Anzahl der dynamischen Knoten vorgenommen, um den neuen Anforderungen gerecht zu werden

MaxCount Wert.

- Ausgangszustand: MinCount = 100; MaxCount = 150

```
$ sinfo
PARTITION AVAIL  TIMELIMIT  NODES  STATE NODELIST
queue1*    up    infinite   50    idle~ queue1-dy-c5xlarge-[1-50]
queue1*    up    infinite  100    idle queue1-st-c5xlarge-[1-100]
```

- Aktualisieren Sie +30 auf MinCount : MinCount = 130 (MaxCount = 180)

```
$ sinfo
PARTITION AVAIL  TIMELIMIT  NODES  STATE NODELIST
queue1*    up    infinite   20    idle~ queue1-dy-c5xlarge-[1-50]
queue1*    up    infinite  130    idle queue1-st-c5xlarge-[1-130]
```

Wenn die MinCount Anzahl N verringert MaxCount wird (vorausgesetzt $\text{MinCount} < \text{MaxCount}$, sie bleibt unverändert), wird der Cluster konfiguriert, indem die letzten N statischen Knoten entfernt werden, `<Queue/Name>-st-<ComputeResource/Name>-[<old_MinCount - N>...<old_MinCount>` und das System beendet die entsprechenden EC2-Instances. Um der MaxCount Kapazität der Rechenressource Rechnung zu tragen, wird außerdem die Cluster-Konfiguration aktualisiert, indem die Anzahl der dynamischen Knoten erhöht wird, um die Lücke zu schließen. Da es sich um dynamische Knoten handelt, werden `MaxCount - new_MinCount`: `<Queue/Name>-dy-<ComputeResource/Name>-[1..<MaxCount - new_MinCount>]` in diesem Fall keine neuen EC2-Instances gestartet, es sei denn, der Scheduler hat Jobs auf den neuen Knoten ausstehend.

- Ausgangszustand: MinCount = 100; MaxCount = 150

```
$ sinfo
PARTITION AVAIL  TIMELIMIT  NODES  STATE NODELIST
queue1*    up    infinite   50    idle~ queue1-dy-c5xlarge-[1-50]
queue1*    up    infinite  100    idle queue1-st-c5xlarge-[1-100]
```

- Update -30 am MinCount : MinCount = 70 (MaxCount = 120)

```
$ sinfo
PARTITION AVAIL  TIMELIMIT  NODES  STATE NODELIST
queue1*    up    infinite   80    idle~ queue1-dy-c5xlarge-[1-80]
queue1*    up    infinite   70    idle queue1-st-c5xlarge-[1-70]
```

Wenn $\text{MinCount} < \text{MaxCount}$ der Wert abnimmt MinCount und die MaxCount Anzahl N gleich ist, wird der Cluster konfiguriert, indem die letzten N statischen Knoten entfernt werden, `<Queue/Name>-st-<ComputeResource/Name>-<old_MinCount - N>...<oldMinCount>`] und das System beendet die entsprechenden EC2-Instances.

Außerdem werden keine Änderungen an der Anzahl der dynamischen Knoten vorgenommen, um dem neuen MaxCount Wert Rechnung zu tragen.

- Ausgangszustand: $\text{MinCount} = 100$; $\text{MaxCount} = 150$

```
$ sinfo
PARTITION AVAIL  TIMELIMIT  NODES  STATE NODELIST
queue1*   up    infinite   50    idle~ queue1-dy-c5xlarge-[1-50]
queue1*   up    infinite  100    idle queue1-st-c5xlarge-[1-100]
```

- Update -30 am MinCount : $\text{MinCount} = 70$ ($\text{MaxCount} = 120$)

```
$ sinfo
PARTITION AVAIL  TIMELIMIT  NODES  STATE NODELIST
queue1*   up    infinite   80    idle~ queue1-dy-c5xlarge-[1-50]
queue1*   up    infinite   70    idle queue1-st-c5xlarge-[1-70]
```

Wenn die MaxCount Anzahl N verringert wird (vorausgesetzt $\text{MinCount} < \text{MaxCount}$, sie bleibt unverändert), MinCount wird der Cluster so konfiguriert, dass die letzten N dynamischen Knoten entfernt werden, `<Queue/Name>-dy-<ComputeResource/Name>-<old_MaxCount - N...<oldMaxCount>`] und das System beendet die entsprechenden EC2-Instances, sofern sie ausgeführt wurden. Es sind keine Auswirkungen auf die statischen Knoten zu erwarten.

- Ausgangszustand: $\text{MinCount} = 100$; $\text{MaxCount} = 150$

```
$ sinfo
```

```

PARTITION AVAIL  TIMELIMIT  NODES  STATE NODELIST
queue1*    up    infinite   50    idle~ queue1-dy-c5xlarge-[1-50]
queue1*    up    infinite  100    idle queue1-st-c5xlarge-[1-100]

```

- Update -30 am MaxCount : MinCount = 100 (MaxCount = 120)

```

$ sinfo
PARTITION AVAIL  TIMELIMIT  NODES  STATE NODELIST
queue1*    up    infinite   20    idle~ queue1-dy-c5xlarge-[1-20]
queue1*    up    infinite  100    idle queue1-st-c5xlarge-[1-100]

```

Auswirkungen auf die Arbeitsplätze

In allen Fällen, in denen Knoten entfernt und EC2-Instances beendet werden, wird ein Sbatch-Job, der auf den entfernten Knoten ausgeführt wird, erneut in die Warteschlange gestellt, es sei denn, es gibt keine anderen Knoten, die die Job-Anforderungen erfüllen. In letzterem Fall schlägt der Job mit dem Status `NODE_FAIL` fehl und verschwindet aus der Warteschlange. In diesem Fall muss er manuell erneut eingereicht werden.

Wenn Sie planen, ein Update zur Änderung der Clustergröße durchzuführen, können Sie verhindern, dass Jobs auf den Knoten ausgeführt werden, die während des geplanten Updates entfernt werden. Dies ist möglich, indem Sie festlegen, dass die Knoten im Rahmen der Wartung entfernt werden. Bitte beachten Sie, dass die Einstellung eines Knotens zur Wartung keine Auswirkungen auf Jobs hat, die eventuell bereits auf dem Knoten ausgeführt werden.

Nehmen wir an, dass Sie mit dem geplanten Update zur Änderung der Clustergröße den Knoten entfernen werden `queueu-st-computeresource-[9-10]`. Sie können eine Slurm Reservierung mit dem folgenden Befehl erstellen

```

sudo -i scontrol create reservation ReservationName=maint_for_update user=root
starttime=now duration=infinite flags=maint,ignore_jobs nodes=queueu-st-
computeresource-[9-10]

```

Dadurch wird eine Slurm Reservierung erstellt, die `maint_for_update` auf den Knoten benannt ist `queueu-st-computeresource-[9-10]`. Ab dem Zeitpunkt, an dem die Reservierung erstellt

wurde, können keine Jobs mehr auf den Knoten ausgeführt werden `queue-st-computeresource-[9-10]`. Bitte beachten Sie, dass die Reservierung nicht verhindert, dass Jobs irgendwann auf den Knoten zugewiesen werden `queue-st-computeresource-[9-10]`.

Wenn die Slurm Reservierung nach der Aktualisierung der Clustergröße nur für Knoten festgelegt wurde, die während der Aktualisierung der Clustergröße entfernt wurden, wird die Wartungsreservierung automatisch gelöscht. Wenn Sie stattdessen eine Slurm Reservierung für Knoten erstellt haben, die nach der Aktualisierung der Clustergröße noch vorhanden sind, möchten wir möglicherweise die Wartungsreservierung auf den Knoten nach der Aktualisierung der Größe entfernen. Verwenden Sie dazu den folgenden Befehl

```
sudo -i scontrol delete ReservationName=maint_for_update
```

[Weitere Informationen zur Slurm Reservierung finden Sie im offiziellen SchedMD-Dokument hier.](#)

Cluster-Aktualisierungsprozess bei Kapazitätsänderungen

Bei einer Änderung der Scheduler-Konfiguration werden während des Cluster-Aktualisierungsvorgangs die folgenden Schritte ausgeführt:

- Stoppen AWS ParallelCluster `clustermgtd` (`supervisorctl stop clustermgtd`)
- Generieren Sie die aktualisierte Slurm Partitionskonfiguration anhand der AWS ParallelCluster Konfiguration
- Neustart `slurmctld` (erfolgt über das Chef-Dienstrezept)
- Überprüfen Sie `slurmctld` den Status (`systemctl is-active --quiet slurmctld.service`)
- Konfiguration neu laden Slurm (`scontrol reconfigure`)
- `clustermgtd` (`supervisorctl start clustermgtd`) starten

AWS Batch (**awsbatch**)

Weitere Informationen zu AWS Batch finden Sie unter [AWS Batch](#). Die Dokumentation finden Sie im [AWS Batch Benutzerhandbuch](#).

AWS ParallelCluster CLI-Befehle für AWS Batch

Wenn Sie den `awsbatch` Scheduler verwenden, AWS Batch werden die AWS ParallelCluster CLI-Befehle für automatisch im AWS ParallelCluster Hauptknoten installiert. Die CLI verwendet AWS Batch API-Operationen und erlaubt die folgenden Operationen:

- Übermitteln und Verwalten von Aufgaben.
- Überwachen von Aufgaben, Warteschlangen und Hosts.
- Spiegeln herkömmlicher Scheduler-Befehle.

⚠ Important

AWS ParallelClusterunterstützt keine GPU-Jobs fürAWS Batch. Weitere Informationen finden Sie unter [GPU-Jobs](#).

Diese CLI wird als separates Paket verteilt. Weitere Informationen finden Sie unter [Scheduler-Unterstützung](#).

Themen

- [awsbsub](#)
- [awsbstat](#)
- [awsbout](#)
- [awsbkill](#)
- [awsbqueues](#)
- [awsbhosts](#)

awsbsub

Sendet Jobs an die Auftragswarteschlange des Clusters.

```
awsbsub [-h] [-jn JOB_NAME] [-c CLUSTER] [-cf] [-w WORKING_DIR]  
        [-pw PARENT_WORKING_DIR] [-if INPUT_FILE] [-p VCPUS] [-m MEMORY]  
        [-e ENV] [-eb ENV_DENYLIST] [-r RETRY_ATTEMPTS] [-t TIMEOUT]  
        [-n NODES] [-a ARRAY_SIZE] [-d DEPENDS_ON]  
        [command] [arguments [arguments ...]]
```

⚠ Important

AWS ParallelCluster unterstützt keine GPU-Jobs für AWS Batch. Weitere Informationen finden Sie unter [GPU-Aufträge](#).

Positionale Argumente***command***

Sendet den Job (der angegebene Befehl muss auf den Recheninstanzen verfügbar sein) oder den zu übertragenden Dateinamen. Weitere Informationen finden Sie auch unter `--command-file`.

arguments

(Optional) Gibt Argumente für den Befehl oder die Befehlsdatei an.

Benannte Argumente**`-jn JOB_NAME, --job-name JOB_NAME`**

Benennt die Aufgabe. Das erste Zeichen muss ein Buchstabe oder eine Zahl sein. Der Auftragsname kann Buchstaben (Groß- und Kleinbuchstaben), Ziffern, Bindestriche und Unterstriche enthalten und bis zu 128 Zeichen lang sein.

`-c CLUSTER, --cluster CLUSTER`

Gibt den zu verwendenden Cluster an.

`-cf, --command-file`

Zeigt an, dass der Befehl eine Datei ist, die an Datenverarbeitungs-Instances übertragen werden soll.

Standard: False

`-w WORKING_DIR, --working-dir WORKING_DIR`

Gibt den Ordner an, der als Arbeitsverzeichnis der Aufgabe verwendet werden soll. Wenn kein Arbeitsverzeichnis angegeben ist, wird der Job im `job-<AWS_BATCH_JOB_ID>` Unterordner des Home-Verzeichnisses des Benutzers ausgeführt. Sie können entweder diesen Parameter oder den Parameter `--parent-working-dir` verwenden.

-pw *PARENT_WORKING_DIR*, --parent-working-dir *PARENT_WORKING_DIR*

Gibt den übergeordneten Ordner des Arbeitsverzeichnisses des Jobs an. Wenn kein übergeordnetes Arbeitsverzeichnis angegeben ist, wird standardmäßig das Basisverzeichnis des Benutzers verwendet. Ein Unterordner mit dem Namen `job-<AWS_BATCH_JOB_ID>` wird im übergeordneten Arbeitsverzeichnis erstellt. Sie können entweder diesen Parameter oder den Parameter `--working-dir` verwenden.

-if *INPUT_FILE*, --input-file *INPUT_FILE*

Gibt die an die Datenverarbeitungs-Instances zu übertragende Datei im Arbeitsverzeichnis der Aufgabe an. Sie können mehrere Eingabedateiparameter angeben.

-p *VCPUS*, --vcpus *VCPUS*

Gibt die Anzahl von vCPUs an, die für den Container reserviert werden sollen. Wenn es zusammen mit `used-nodes` verwendet wird, identifiziert es die Anzahl der vCPUs für jeden Knoten.

Standard: 1

-m *MEMORY*, --memory *MEMORY*

Gibt die harte Grenze des Arbeitsspeichers (in MiB) für die Aufgabe an. Wenn Ihr Auftrag versucht, das hier angegebene Limit zu überschreiten, wird der Auftrag beendet.

Standard: 128

-e *ENV*, --env *ENV*

Gibt eine durch Komma getrennte Liste von Umgebungsvariablenamen zum Exportieren in die Aufgabenumgebung an. Zum Exportieren aller Umgebungsvariablen geben Sie „all“ an. Beachten Sie, dass eine Liste von „allen“ Umgebungsvariablen nicht die im `env-blacklist` Parameter aufgelisteten Variablen oder Variablen, die mit dem `AWS_*` Präfix `PCLUSTER_*` oder beginnen, enthält.

-eb *ENV_DENYLIST*, --env-blacklist *ENV_DENYLIST*

Gibt eine durch Komma getrennte Liste von Umgebungsvariablenamen an, die nicht in die Aufgabenumgebung exportiert werden sollen. Standardmäßig werden HOME, PWD, USER, PATH, LD_LIBRARY_PATH, TERM und TERMCAP nicht exportiert.

-r *RETRY_ATTEMPTS*, --retry-attempts *RETRY_ATTEMPTS*

Gibt die Zahl der Wiederholungsversuche an, um einen Auftrag in den `RUNNABLE` Status zu verschieben. Sie können zwischen einem und zehn Versuche angeben. Wenn der Wert der

Versuche größer als 1 ist, wird der Job erneut versucht, falls er fehlschlägt, bis er für die angegebene Anzahl von Malen in einen `RUNNABLE` Status übergegangen ist.


Standard: 1

-t *TIMEOUT*, --timeout *TIMEOUT*

Gibt die Zeitdauer in Sekunden (gemessen ab dem `startedAt` Zeitstempel des Auftrags) an, nach der Ihr Auftrag AWS Batch beendet wird, wenn er noch nicht abgeschlossen ist. Der Timeout-Wert muss mindestens 60 Sekunden betragen.

-n *NODES*, --nodes *NODES*

Gibt die Anzahl der Knoten an, die für die Aufgabe zu reservieren sind. Geben Sie einen Wert für diesen Parameter an, um die parallel Übermittlung mehrerer Knoten zu ermöglichen.

 Note

Wenn der `CapacityType` Parameter `SchedulerAwsBatchQueues` auf `SPOT` gesetzt ist, werden parallel Jobs mit mehreren Knoten nicht unterstützt. Darüber hinaus muss Ihr Konto eine `AWSServiceRoleForEC2Spot` servicegebundene Rolle enthalten. Sie können diese Rolle mit dem folgenden AWS CLI Befehl erstellen:

```
$ aws iam create-service-linked-role --aws-service-name spot.amazonaws.com
```

Weitere Informationen finden Sie unter [Serviceverknüpfte Rolle für Spot-Anfragen](#) im Amazon EC2 EC2-Benutzerhandbuch für Linux-Anfragen.

-a *ARRAY_SIZE*, --array-size *ARRAY_SIZE*

Zeigt die Größe des Arrays an. Sie können einen Wert zwischen 2 und 10.000 auswählen. Wenn Sie Array-Eigenschaften für eine Aufgabe angeben, wird sie zu einer Array-Aufgabe.

-d *DEPENDS_ON*, --depends-on *DEPENDS_ON*

Gibt eine durch Strichpunkte getrennte Liste von Abhängigkeiten für eine Aufgabe an. Eine Aufgabe kann von maximal 20 Aufgaben abhängen. Sie können eine `SEQUENTIAL`-Typabhängigkeit angeben, ohne eine Auftrags-ID für Array-Aufträge anzugeben. Eine sequenzielle Abhängigkeit ermöglicht jeder untergeordneten Array-Aufgabe, sequentiell abgeschlossen zu werden (beginnend mit Index 0). Sie können auch eine Abhängigkeit vom Typ „`N_TO_N`“ mit einer Aufgaben-ID für Array-Aufgaben angeben. Eine Abhängigkeit vom

Typ `N_TO_N` bedeutet, dass jeder untergeordnete Index dieser Aufgabe warten muss, bis der entsprechende untergeordnete Index jeder Abhängigkeit abgeschlossen ist. Die Syntax für diesen Parameter lautet „`jobID=<string>, type=<string>;...`“.

awsbstat

Zeigt die Aufgaben in der Aufgabenwarteschlange des Clusters.

```
awsbstat [-h] [-c CLUSTER] [-s STATUS] [-e] [-d] [job_ids [job_ids ...]]
```

Positionale Argumente

job_ids

Gibt die durch Leerzeichen getrennte Liste von Aufgaben-IDs an, die in der Ausgabe angezeigt werden soll. Wenn die Aufgabe ein Aufgaben-Array ist, werden alle untergeordneten Aufgaben angezeigt. Wenn eine einzelne Aufgabe angefordert wird, wird sie in einer detaillierten Version angezeigt.

Benannte Argumente

-c *CLUSTER*, --cluster *CLUSTER*

Gibt den Cluster an, der verwendet werden soll.

-s *STATUS*, --status *STATUS*

Gibt eine durch Komma getrennte Liste der Job-Status an, die berücksichtigt werden sollen. Der standardmäßige Aufgabenstatus lautet „active“. Akzeptierte Werte sind: SUBMITTED, PENDING, RUNNABLE, STARTING, RUNNING, SUCCEEDED, FAILED und ALL.

Standard: „SUBMITTED,PENDING,RUNNABLE,STARTING,RUNNING“

-e, --expand-children

Erweitert Aufgaben mit untergeordneten Elementen (Array und Multi-Knoten parallel).

Standard: False

-d, --details

Zeigt Aufgabendetails.

Standard: False

awsbout

Zeigt die Ausgabe einer bestimmten Aufgabe an.

```
awsbout [-h] [-c CLUSTER] [-hd HEAD] [-t TAIL] [-s] [-sp STREAM_PERIOD] job_id
```

Positionale Argumente

job_id

Gibt die Aufgaben-ID an.

Benannte Argumente

-c *CLUSTER*, --cluster *CLUSTER*

Gibt den Cluster an, der verwendet werden soll.

-hd *HEAD*, --head *HEAD*

Ruft die ersten *HEAD*-Zeilen der Aufgabenausgabe ab.

-t *TAIL*, --tail *TAIL*

Ruft die letzten <tail>-Zeilen der Aufgabenausgabe ab.

-s, --stream

Ruft die Aufgabenausgabe ab und wartet dann darauf, dass weitere Ausgaben erstellt werden. Dieses Argument kann zusammen mit „-tail“ verwendet werden, um von den aktuellen <tail>-Zeilen der Aufgabenausgabe zu beginnen.

Standard: False

-sp *STREAM_PERIOD*, --stream-period *STREAM_PERIOD*

Legt den Streaming-Zeitraum fest.

Standard: 5

awsbkill

Bricht im Cluster übermittelte Aufgaben ab oder beendet sie.

```
awsbkill [-h] [-c CLUSTER] [-r REASON] job_ids [job_ids ... ]
```

Positionale Argumente

job_ids

Gibt die durch Leerzeichen getrennte Liste von Aufgaben-IDs an, die abgebrochen oder beendet werden sollen.

Benannte Argumente

-c *CLUSTER*, --cluster *CLUSTER*

Gibt den Namen des Clusters an, der verwendet werden soll.

-r *REASON*, --reason *REASON*

Gibt die Nachricht an, die einer Aufgabe angefügt werden soll und den Grund für ihren Abbruch angibt.

Standard: „Terminated by the user“

awsbqueues

Zeigt die Aufgabenwarteschlange an, die mit dem Cluster verknüpft ist.

```
awsbqueues [-h] [-c CLUSTER] [-d] [job_queues [job_queues ... ]]
```

Positionale Argumente

job_queues

Gibt die durch Leerzeichen getrennte Liste von Warteschlangennamen an, die angezeigt werden sollen. Wenn eine einzelne Warteschlange angefordert wird, wird sie in einer detaillierten Version angezeigt.

Benannte Argumente

-c *CLUSTER*, --cluster *CLUSTER*

Gibt den Namen des Clusters an, der verwendet werden soll.

-d, --details

Gibt an, ob die Details der Warteschlangen angezeigt werden sollen.

Standard: False

awsbhosts

Zeigt die Hosts, die zur Datenverarbeitungsumgebung des Clusters gehören.

```
awsbhosts [-h] [-c CLUSTER] [-d] [instance_ids [instance_ids ... ]]
```

Positionale Argumente

instance_ids

Gibt eine durch Leerzeichen getrennte Liste von Instance-IDs an. Wenn eine einzelne Instance angefordert wird, wird sie in einer detaillierten Version angezeigt.

Benannte Argumente

-c *CLUSTER*, --cluster *CLUSTER*

Gibt den Namen des Clusters an, der verwendet werden soll.

-d, --details

Gibt an, ob die Details des Hosts angezeigt werden sollen.

Standard: False

Gemeinsamer Speicher

AWS ParallelCluster [unterstützt entweder die Verwendung von Amazon EBS, FSx for ONTAP und FSx for OpenZFS Shared Storage Volumes, Amazon EFS und FSx for Lustre Shared Storage File](#)

[Systems oder File Caches](#). Wir empfehlen Ihnen, sich an die Grundpfeiler [Zuverlässigkeit eines AWS gut konzipierten Frameworks zu halten und Ihre Volumes](#) und Dateisysteme zu sichern.

Wählen Sie ein Speichersystem, das die I/O-Anforderungen Ihrer HPC-Anwendung erfüllt. Sie können jedes Dateisystem auf der Grundlage Ihres spezifischen Anwendungsfalls optimieren. Weitere Informationen finden Sie in der [Übersicht über die Speicheroptionen](#).

Amazon EBS-Volumes werden an den Hauptknoten angehängt und über NFS für Rechenknoten freigegeben. Diese Option kann kosteneffektiv sein, die Leistung hängt jedoch bei steigendem Speicherbedarf von den Ressourcen des Hauptknotens ab. Dies kann zu einem Engpass werden, wenn dem Cluster mehr Rechenknoten hinzugefügt werden und der Durchsatzbedarf steigt.

Amazon EFS-Dateisysteme werden skaliert, wenn sich die Speicheranforderungen ändern. Sie können diese Dateisysteme für eine Vielzahl von Anwendungsfällen konfigurieren. Verwenden Sie Amazon EFS-Dateisysteme, um parallelisierte und latenzempfindliche Anwendungen auf Ihrem Cluster auszuführen.

FSx for Lustre-Dateisysteme können riesige Datensätze mit einem Durchsatz von bis zu Hunderten von Gigabyte pro Sekunde, Millionen von IOPS und Latenzen unter einer Millisekunde verarbeiten. Verwenden Sie FSx for Lustre-Dateisysteme für anspruchsvolle Hochleistungsrechnerumgebungen.

In der [SharedStorage Abschnitt](#) können Sie entweder externen oder AWS ParallelCluster verwalteten Speicher definieren:

- Externer Speicher bezieht sich auf ein vorhandenes Volume oder Dateisystem, das Sie verwalten. AWS ParallelCluster erstellt oder löscht diesen Speicher nicht.
- Verwalteter Speicher bezieht sich auf ein Volume oder Dateisystem, das AWS ParallelCluster erstellt wurde und gelöscht werden kann.

Externer Speicher

Sie können so konfigurieren AWS ParallelCluster, dass bei der Erstellung oder Aktualisierung des Clusters externer Speicher an den Cluster angehängt wird. Ebenso können Sie es so konfigurieren, dass der externe Speicher vom Cluster getrennt wird, wenn der Cluster gelöscht oder aktualisiert wird. Ihre Daten bleiben erhalten und Sie können sie für langfristigen, dauerhaften, gemeinsam genutzten Speicher außerhalb des Cluster-Lebenszyklus verwenden.

Note

In Versionen AWS ParallelCluster vor 3.8 ist das Mounten von extern verwalteten Dateisystemen nicht möglich. /home Ab Version 3.8 AWS ParallelCluster können Sie es /home als Einhängpunkt für ein externes verwaltetes Dateisystem verwenden. Sie können ein extern verwaltetes Dateisystem einhängen, /home indem Sie /home als Wert für den [MountDir](#)Parameter unter dem angeben. [SharedStorage Abschnitt](#)

Amazon File Cache ist nicht für die Verwendung als /home Systemverzeichnis geeignet und wird daher derzeit nicht für das Mounten unterstützt/home.

Wenn Sie unter der [SharedStorageType](#)Konfigurationsoption ein /home Verzeichnis angeben, wird [SharedStorage Abschnitt](#) das überschrieben, was bedeutet, dass stattdessen die Einstellungen unter verwendet [SharedStorage Abschnitt](#) werden.

Beim Mounten eines externen Dateisystems in das /home Verzeichnis wird der /home Inhalt des Kopfknotens in das externe Dateisystem AWS ParallelCluster kopiert, ohne bestehende Dateien auf dem externen Speicher zu überschreiben. Dies beinhaltet die Übertragung des SSH-Schlüssels des Clusters für den Standardbenutzer, falls er im externen Dateisystem nicht vorhanden ist. Weitere Informationen finden Sie unter. [AWS ParallelClusterÜberlegungen zu gemeinsam genutztem Speicher](#)

AWS ParallelClusterverwalteter Speicher

AWS ParallelClusterDer verwaltete Speicher ist in der Konfiguration standardmäßig vom Lebenszyklus des Clusters abhängig. Der `SharedStorage DeletionPolicy` Konfigurationsparameter ist `Delete` standardmäßig auf eingestellt.

Standardmäßig werden ein AWS ParallelCluster verwaltetes Dateisystem oder ein verwaltetes Volume und die zugehörigen Daten gelöscht, wenn eine der folgenden Bedingungen zutrifft.

- Sie löschen den Cluster.
- Sie ändern die Konfiguration des verwalteten gemeinsam genutzten `SpeichersName`.
- Sie entfernen den verwalteten gemeinsamen Speicher aus der Konfiguration.

Stellen Sie `DeletionPolicy` auf ein `Retain`, um Ihr verwaltetes gemeinsam genutztes Dateisystem oder Volume und Daten beizubehalten. Wir empfehlen Ihnen, Ihre Daten regelmäßig zu sichern, um Datenverlust zu vermeiden. Sie können [AWS Backup](#) damit Backups für all Ihre Speicheroptionen zentral verwalten.

Sie können die Lebenszyklusabhängigkeit mit den Konfigurationseinstellungen entfernen. Weitere Informationen finden Sie unter [Konvertiert AWS ParallelCluster verwalteten Speicher in externen Speicher](#).

Informationen zu gemeinsam genutzten Speicherkontingenten finden Sie unter [Kontingente für gemeinsam genutzten Speicher](#).

Weitere Informationen zu gemeinsam genutztem Speicher und dem Wechsel zu neuen AWS ParallelCluster Versionen finden Sie unter [Bewährte Methoden: Verschieben eines Clusters auf eine neue AWS ParallelCluster Minor- oder Patch-Version](#).

Sie können so konfigurieren AWS ParallelCluster, dass bei der Erstellung oder Aktualisierung des Clusters externer Speicher an den Cluster angehängt wird. Ebenso können Sie es so konfigurieren, dass der externe Speicher vom Cluster getrennt wird, wenn der Cluster gelöscht oder aktualisiert wird. Ihre Daten bleiben erhalten und Sie können sie für langfristige, permanente, gemeinsam genutzte Speicherlösungen verwenden, die unabhängig vom Cluster-Lebenszyklus sind.

Standardmäßig ist verwalteter Speicher vom Lebenszyklus des Clusters abhängig. Sie können die Abhängigkeit mit den unter beschriebenen Konfigurationseinstellungen entfernen [Konvertiert AWS ParallelCluster verwalteten Speicher in externen Speicher](#).

Mit spezifischen Einstellungen können Sie jede der unterstützten Speicherlösungen für Ihre Anwendungsfälle optimieren.

Informationen zu gemeinsam genutzten Speicherkontingenten finden Sie unter [Kontingente für gemeinsam genutzten Speicher](#).

Weitere Informationen zu gemeinsam genutztem Speicher und dem Wechsel zu neuen AWS ParallelCluster Versionen finden Sie unter [Bewährte Methoden: Verschieben eines Clusters auf eine neue AWS ParallelCluster Minor- oder Patch-Version](#).

Themen

- [Konfigurieren Sie gemeinsam genutzten Speicher](#)
- [Arbeiten mit gemeinsam genutztem Speicher in AWS ParallelCluster](#)
- [Kontingente für gemeinsam genutzten Speicher](#)

Konfigurieren Sie gemeinsam genutzten Speicher

Erfahren Sie mehr über die Konfigurationseinstellungen, mit denen Sie gemeinsam genutzten Speicher für Ihren Cluster definieren können.

Themen

- [Amazon Elastic Block Store](#)
- [Amazon Elastic File System](#)
- [Amazon FSx für Lustre](#)
- [Konfiguration von FSx für ONTAP, FSx für OpenZFS und gemeinsam genutzten Datei-Cache-Speicher](#)

Amazon Elastic Block Store

Um ein vorhandenes externes Amazon EBS-Volume für dauerhaften Langzeitspeicher zu verwenden, der unabhängig vom Cluster-Lebenszyklus ist, geben Sie [EbsSettings/VolumeId](#) an.

Wenn Sie dies nicht angeben [VolumeId](#), wird standardmäßig ab dem [EbsSettings](#)Zeitpunkt, an dem Ihr Cluster AWS ParallelCluster erstellt wird, ein verwaltetes EBS-Volume erstellt. AWS ParallelCluster löscht auch das Volume und die Daten, wenn der Cluster gelöscht oder das Volume aus der Cluster-Konfiguration entfernt wird.

Bei einem AWS ParallelCluster verwalteten EBS-Volume können Sie [EbsSettings](#) verwenden, [DeletionPolicy](#) um das Volume AWS ParallelCluster anzuweisen DeleteRetain, oder Snapshot das Volume, wenn entweder der Cluster gelöscht oder wenn das Volume aus der Cluster-Konfiguration entfernt wird. [DeletionPolicy](#) ist standardmäßig auf Delete festgelegt.

Warning

Für AWS ParallelCluster verwalteten gemeinsamen Speicher [DeletionPolicy](#) ist Delete standardmäßig auf eingestellt.

Das bedeutet, dass ein verwaltetes Volume und seine Daten gelöscht werden, wenn eine der folgenden Bedingungen zutrifft:

- Sie löschen den Cluster.
- Sie ändern die Konfiguration des verwalteten gemeinsam genutzten Speichers [SharedStorage/Name](#).

- Sie entfernen den verwalteten gemeinsamen Speicher aus der Konfiguration.

Wir empfehlen Ihnen, Ihren gemeinsam genutzten Speicher regelmäßig mit Snapshots zu sichern, um Datenverlust zu vermeiden. Weitere Informationen zu Amazon EBS-Snapshots finden Sie unter [Amazon EBS-Snapshots](#) im Amazon Elastic Compute Cloud-Benutzerhandbuch für Linux-Instances. Informationen zur Verwaltung von Datensicherungen finden AWS-Services Sie unter [AWSBackup](#) im AWS BackupEntwicklerhandbuch.

Amazon Elastic File System

Um ein vorhandenes externes Amazon EFS-Dateisystem für dauerhaften Langzeitspeicher außerhalb des Cluster-Lebenszyklus zu verwenden, geben Sie [EfsSettings/FileSystemId](#) an. Standardmäßig wird ab dem [EfsSettings](#)Zeitpunkt der Clustererstellung ein verwaltetes Amazon EFS-Dateisystem erstellt. AWS ParallelCluster AWS ParallelClusterlöscht auch das Dateisystem und die Daten, wenn der Cluster gelöscht wird oder wenn das Dateisystem aus der Cluster-Konfiguration entfernt wird.

Bei einem AWS ParallelCluster verwalteten Amazon EFS-Dateisystem können Sie [EfsSettings/](#) verwenden, [DeletionPolicy](#) AWS ParallelClusterum Delete anzugeben, Retain ob der Cluster gelöscht wird oder wann das Dateisystem aus der Cluster-Konfiguration entfernt wird. [DeletionPolicy](#) ist standardmäßig auf Delete festgelegt.

Warning

Für AWS ParallelCluster verwalteten gemeinsamen Speicher [DeletionPolicy](#) ist Delete standardmäßig auf eingestellt.

Das bedeutet, dass ein verwaltetes Dateisystem und seine Daten gelöscht werden, wenn eine der folgenden Bedingungen zutrifft:

- Sie löschen den Cluster.
- Sie ändern die Konfiguration des verwalteten gemeinsam genutzten Speichers [SharedStorage/Name](#).
- Sie entfernen den verwalteten gemeinsamen Speicher aus der Konfiguration.

Wir empfehlen Ihnen, Ihren gemeinsam genutzten Speicher regelmäßig zu sichern, um Datenverlust zu vermeiden. Weitere Informationen zum Sichern einzelner Amazon EFS-

Volumes finden Sie unter [Sichern Ihrer Amazon EFS-Dateisysteme](#) im Amazon Elastic File System-Benutzerhandbuch. Informationen zur Verwaltung von Datensicherungen finden AWS-Services Sie unter [AWSBackup](#) im AWS BackupEntwicklerhandbuch.

Amazon FSx für Lustre

Um ein vorhandenes externes FSx for Lustre-Dateisystem für dauerhaften Langzeitspeicher außerhalb des Cluster-Lebenszyklus zu verwenden, geben Sie [FsxLustreSettings/an.FileSystemId](#)

Wenn Sie [FsxLustreSettings/](#)nicht angeben [FileSystemId](#), wird standardmäßig ab dem [FsxLustreSettings](#)Zeitpunkt der AWS ParallelCluster Clustererstellung ein verwaltetes FSx for Lustre-Dateisystem erstellt. AWS ParallelCluster löscht auch das Dateisystem und die Daten, wenn der Cluster gelöscht wird oder wenn das Dateisystem aus der Clusterkonfiguration entfernt wird.

Bei einem AWS ParallelCluster verwalteten FSx for Lustre-Dateisystem können Sie das [FsxLustreSettings/](#)verwenden, [DeletionPolicy](#)um das Dateisystem AWS ParallelCluster Delete anzuweisen, wenn entweder der Cluster gelöscht oder wenn das Dateisystem aus der Cluster-Konfiguration entfernt wird. Retain DeletionPolicy ist standardmäßig auf Delete festgelegt.

Warning

Für AWS ParallelCluster verwalteten gemeinsamen Speicher DeletionPolicy ist standardmäßig auf Delete eingestellt.

Das bedeutet, dass ein verwaltetes Dateisystem und seine Daten gelöscht werden, wenn eine der folgenden Bedingungen zutrifft:

- Sie löschen den Cluster.
- Sie ändern die Konfiguration des verwalteten gemeinsam genutzten Speichers [SharedStorage/Name](#).
- Sie entfernen den verwalteten gemeinsamen Speicher aus der Konfiguration.

Wir empfehlen Ihnen, Ihren gemeinsam genutzten Speicher regelmäßig zu sichern, um Datenverlust zu vermeiden. Sie können Backups in Ihrem Cluster mit [SharedStorage/FsxLustreSettings/AutomaticBackupRetentionDays](#)und definieren [DailyAutomaticBackupStartTime](#). Informationen zur Verwaltung

von Datensicherungen finden AWS-Services Sie unter [AWSBackup](#) im AWS BackupEntwicklerhandbuch.

Konfiguration von FSx für ONTAP, FSx für OpenZFS und gemeinsam genutzten Datei-Cache-Speicher

Für FSx for ONTAP, FSx for OpenZFS und File Cache können Sie `/,` und [FsxOntapSettingsFileCacheSettings](#) verwenden `VolumeId`, `FileCacheId` und das [FsxOpenZfsSettingsVolumeId](#) Mounten eines externen vorhandenen Volumes oder Datei-Caches für Ihren Cluster anzugeben.

AWS ParallelClusterverwalteter gemeinsam genutzter Speicher wird für FSx for ONTAP, FSx for OpenZFS und File Cache nicht unterstützt.

Arbeiten mit gemeinsam genutztem Speicher in AWS ParallelCluster

Erfahren Sie mehr über die Arbeit mit AWS ParallelCluster und gemeinsam genutztem Speicher.

Themen

- [AWS ParallelClusterÜberlegungen zu gemeinsam genutztem Speicher](#)
- [Konvertiert AWS ParallelCluster verwalteten Speicher in externen Speicher](#)

AWS ParallelClusterÜberlegungen zu gemeinsam genutztem Speicher

Beachten Sie Folgendes, wenn Sie mit gemeinsam genutztem Speicher in arbeitenAWS ParallelCluster.

- Sichern Sie Ihre Dateisystemdaten mit [AWS Backup](#) oder einer anderen Methode, um Backups für alle Ihre Speichersysteme zu verwalten.
- Um gemeinsam genutzten Speicher hinzuzufügen, fügen Sie Ihrer Konfigurationsdatei einen Bereich für gemeinsam genutzten Speicher hinzu und erstellen oder aktualisieren den Cluster.
- Um gemeinsam genutzten Speicher zu entfernen, entfernen Sie den Bereich für gemeinsam genutzten Speicher aus Ihrer Konfigurationsdatei und aktualisieren den Cluster.
- Um den vorhandenen AWS ParallelCluster verwalteten gemeinsam genutzten Speicher durch neuen verwalteten Speicher zu ersetzen, ändern Sie den Wert für [SharedStorage/Name](#) und aktualisieren Sie den Cluster.

⚠ Warning

Standardmäßig werden der vorhandene AWS ParallelCluster verwaltete Speicher und die vorhandenen Daten gelöscht, wenn Sie das Cluster-Update mit einem neuen Name Parameter durchführen. Wenn Sie die vorhandenen verwalteten gemeinsam genutzten Speicherdaten ändern Name und beibehalten müssen, stellen Sie sicher, dass Sie entweder den `DeletionPolicy` Wert auf festlegen `Retain` oder die Daten sichern, bevor Sie den Cluster aktualisieren.

- Wenn Sie die AWS ParallelCluster verwalteten Speicherdaten nicht sichern `Delete`, werden Ihre Daten gelöscht, wenn entweder Ihr Cluster gelöscht wird oder wenn Ihr verwalteter Speicher aus der Clusterkonfiguration entfernt und der Cluster aktualisiert wird. `DeletionPolicy`
- Wenn Sie die AWS ParallelCluster verwalteten Speicherdaten nicht sichern, obwohl dies der `DeletionPolicy` Fall ist `Retain`, wird Ihr Dateisystem getrennt, bevor der Cluster gelöscht wird, und kann als externes Dateisystem wieder an einen anderen Cluster angehängt werden. Ihre Daten bleiben erhalten.
- Wenn AWS ParallelCluster verwalteter Speicher aus der Cluster-Konfiguration entfernt `DeletionPolicy` wird und ist `Retain`, kann er als externes Dateisystem wieder an den Cluster angehängt werden, wobei Ihre Clusterdaten erhalten bleiben.
- Ab AWS ParallelCluster Version 3.4.0 können Sie die Sicherheit für Amazon EFS-Dateisystem-Mounts erhöhen, indem Sie [SharedStorageEfsSettings//EncryptionInTransit](#) und [IamAuthorization](#) Einstellungen konfigurieren.
- Beim Mounten eines externen Dateisystems in das Verzeichnis `/home` wird der Inhalt des Verzeichnisses `/home` des Hauptknotens in das externe Dateisystem AWS ParallelCluster kopiert. Es kopiert bestehende Daten im Verzeichnis `/home`, ohne bestehende Dateien oder Verzeichnisse auf dem externen Speicher zu überschreiben. Dazu gehört auch der SSH-Schlüssel des Clusters für den Standardbenutzer, falls er nicht bereits im externen Dateisystem vorhanden ist. Folglich haben alle anderen Cluster, die dasselbe externe Dateisystem in ihrem jeweiligen `/home`-Verzeichnis mounten, auch denselben SSH-Schlüssel für ihren Standardbenutzer des Clusters.
- In einer Multi-Cluster-Umgebung, in der dasselbe externe Dateisystem in die `/home`-Verzeichnisse von Clustern gemountet wird, werden SSH-Schlüssel, die Zugriff auf die Rechenknoten gewähren, die auf dem Hauptknoten von erstellt wurden, nur einmal generiert AWS ParallelCluster, wenn der erste Cluster das externe Dateisystem in `/home` einhängt. Alle anderen Cluster verwenden denselben SSH-Schlüssel. Daher kann jeder, der den SSH-Schlüssel für den Standardbenutzer

dieser gemeinsam genutzten Cluster besitzt, auf jeden Cluster zugreifen. Alle Rechenknoten ermöglichen Verbindungen unter Verwendung des ursprünglich generierten Schlüssels.

Konvertiert AWS ParallelCluster verwalteten Speicher in externen Speicher

Erfahren Sie, wie Sie AWS ParallelCluster verwalteten Speicher in externen Speicher konvertieren.

Die Verfahren basieren auf dem folgenden Beispielausschnitt einer Konfigurationsdatei.

```
...
- MountDir: /fsx
  Name: fsx
  StorageType: FsxLustre
  FsxLustreSettings:
    StorageCapacity: 1200
    DeletionPolicy: Delete
...
```

Konvertiert AWS ParallelCluster verwalteten Speicher in externen Speicher

1. Stellen Sie `Retain` in der `DeletionPolicy` Cluster-Konfigurationsdatei den Wert auf ein.

```
...
- MountDir: /fsx
  Name: fsx
  StorageType: FsxLustre
  FsxLustreSettings:
    StorageCapacity: 1200
    DeletionPolicy: Retain
...
```

2. Führen Sie den folgenden Befehl aus, um die `DeletionPolicy` Änderung festzulegen.

```
pcluster update-cluster -n cluster-name -c cluster-config.yaml
```

3. Entfernen Sie den `SharedStorage` Abschnitt aus der Cluster-Konfigurationsdatei.

```
...
...
```

4. Führen Sie den folgenden Befehl aus `SharedStorage`, um die Datei verwaltete Datei in eine externe Datei umzuwandeln `SharedStorage` und sie vom Cluster zu trennen.

```
pcluster update-cluster -n cluster-name -c cluster-config.yaml
```

5. Ihr gemeinsam genutzter Speicher ist jetzt extern und vom Cluster getrennt.
6. Gehen Sie wie folgt vor, um Ihr externes Dateisystem an den ursprünglichen Cluster oder einen anderen Cluster anzuhängen.
 - a. Holen Sie sich die FSx for Lustre-Dateisystem-ID.

- i. Um den folgenden Befehl zu verwenden, AWS CLI führen Sie den folgenden Befehl aus und suchen Sie das Dateisystem mit einem Namen, der den Namen Ihres ursprünglichen Clusters enthält, und notieren Sie sich die Dateisystem-ID.

```
aws fsx describe-file-systems
```

- ii. Um den zu verwenden AWS Management Console, melden Sie sich an und navigieren Sie zu <https://console.aws.amazon.com/fsx/>. Suchen Sie in der Liste der Dateisysteme nach dem Dateisystem mit einem Namen, der den Namen Ihres ursprünglichen Clusters enthält, und notieren Sie sich die Dateisystem-ID.
 - b. Aktualisieren Sie die Regeln für die Dateisystem-Sicherheitsgruppe, um den Zugriff auf und von den Dateisystem- und Cluster-Subnetzen zu ermöglichen. Sie finden den Namen und die ID der Dateisystem-Sicherheitsgruppe in der Amazon FSx-Konsole.

Fügen Sie der Dateisystem-Sicherheitsgruppe Regeln hinzu, die eingehenden und ausgehenden TCP-Verkehr vom und zum Hauptknoten und den IP-CIDR-Bereichen oder -Präfixen des Rechenknotens zulassen. Geben Sie die TCP-Ports 988, 1021, 1022 und 1023 für den eingehenden und ausgehenden TCP-Verkehr an.

Weitere Informationen finden Sie unter [SharedStorageFsxLustreSettings//FileSystemId](#) und [Erstellen, Konfigurieren und Löschen von Sicherheitsgruppen für Amazon EC2](#) im AWS Command Line Interface Benutzerhandbuch für Version 2.

- c. Fügen Sie den `SharedStorage` Abschnitt zur Cluster-Konfiguration hinzu.

```
...  
- MountDir: /fsx  
  Name: fsx-external
```



```
StorageType: FsxLustre
FsxLustreSettings:
  FileSystemId: fs-02e5b4b4abd62d51c
...
```

- d. Führen Sie den folgenden Befehl aus, um den externen gemeinsam genutzten Speicher zum Cluster hinzuzufügen.

```
pcluster update-cluster -n cluster-name -c cluster-config.yaml
```

Kontingente für gemeinsam genutzten Speicher

Konfigurieren Sie SharedStorage den Cluster so, dass vorhandener gemeinsam genutzter Dateispeicher bereitgestellt und ein neuer gemeinsam genutzter Dateispeicher auf der Grundlage der in der folgenden Tabelle aufgeführten Kontingente erstellt wird.

Die bereitgestellten Dateispeicherkontingente für jeden Cluster

Speichertyp für gemeinsam genutzte Dateien	AWS ParallelCluster verwalteter Speicher	Externer Speicher	Kontingent netto insgesamt
Amazon EBS	5	5	5
ÜBERFALLEN	1	0	1
Amazon EFS	1	20	21
Amazon FSx †	1 FSx for Lustre	20	21

Note

Diese Tabelle der Kontingente wurde in AWS ParallelCluster Version 3.2.0 hinzugefügt.

† unterstützt AWS ParallelCluster nur das Mounten vorhandener Amazon FSx für NetApp ONTAP-, Amazon FSx für OpenZFS- und File Cache-Systeme. Es unterstützt nicht die Erstellung neuer FSx für ONTAP-, FSx für OpenZFS- und File Cache-Systeme.

Note

Wenn Sie es AWS Batch als Scheduler verwenden, ist FSx for Lustre nur auf dem Cluster-Hauptknoten verfügbar.

Datei-Caches unterstützen keine Scheduler. AWS Batch

AWS ParallelCluster Ressourcen und Tagging

Mit AWS ParallelCluster können Sie Tags erstellen, um Ihre zu verfolgen und zu verwalten AWS ParallelCluster Ressourcen. Sie definieren die gewünschten Tags AWS CloudFormation um Ressourcen zu erstellen und an alle Clusterressourcen in der [Tags Abschnitt](#) der Cluster-Konfigurationsdatei. Sie können auch Tags verwenden, die AWS ParallelCluster automatisch generiert, um Ihre Ressourcen zu verfolgen und zu verwalten.

Wenn Sie einen Cluster erstellen, werden der Cluster und seine Ressourcen mit dem gekennzeichnet AWS ParallelCluster und AWS In diesem Abschnitt definierte System-Tags.

AWS ParallelCluster wendet Tags auf die Cluster-Instances, Volumes und Ressourcen an. Um den Cluster-Stack zu identifizieren, AWS CloudFormation gilt AWS System-Tags für die Cluster-Instanzen. Um die Cluster-EC2-Startvorlagen zu identifizieren, wendet EC2 System-Tags auf die Instances an. Sie können diese Tags verwenden, um Ihre anzuzeigen und zu verwalten AWS ParallelCluster Ressourcen.

Sie können nicht ändern AWS System-Tags. Um Auswirkungen auf zu vermeiden AWS ParallelCluster Funktionalität, nicht ändern AWS ParallelCluster Schlagworte.

Das Folgende ist ein Beispiel für eine AWS System-Tag für ein AWS ParallelCluster Ressource. Sie können sie nicht ändern.

```
"aws:cloudformation:stack-name"="clustername"
```

Das Folgende ist ein Beispiel für eine AWS ParallelCluster Tag, der auf eine Ressource angewendet wurde. Ändern Sie sie nicht.

```
"parallelcluster:cluster-name"="clustername"
```

Sie können sich diese Tags im EC2-Bereich der AWS Management Console.

Anzeigen von Tags

1. Navigieren Sie in der EC2-Konsole unter <https://console.aws.amazon.com/ec2/>.
2. Um alle Cluster-Tags anzuzeigen, wählen Sie **Schlagworte** im Navigationsbereich.
3. Um Cluster-Tags nach Instance anzuzeigen, wählen Sie **Instanzen** im Navigationsbereich.
4. Wählen Sie eine Clusterinstanz aus.
5. Wähle die **Schlagworte verwalten** Öffnen Sie die Instanzdetails und sehen Sie sich die Tags an.
6. Wählen Sie die **Speicher** Tab in den Instanzdetails.
7. Wählen Sie den **Volumen-ID**.
8. In **Bänder**, wählen Sie die **Lautstärke**.
9. Wählen Sie die **Schlagworte** Klicken Sie auf die **Volumendetails** und sehen Sie sich die Tags an.

AWS ParallelCluster Instanz-Tags für den Hauptknoten

Schlüssel	Tag-Wert
parallelcluster:cluster-name	<i>clustername</i>
Name	HeadNode
aws:ec2launchtemplate:id	<i>lt-1234567890abcdef0</i>
aws:ec2launchtemplate:version	<i>1</i>
parallelcluster:node-type	HeadNode
aws:cloudformation:stack-name	<i>clustername</i>
aws:cloudformation:logical-id	HeadNode
aws:cloudformation:stack-id	arn:aws:cloudformation: <i>region-id</i> : <i>ACCOUNTID</i> :stack/ <i>clustername</i> / <i>1234abcd-12ab-12ab-12ab-1234567890abcdef0</i>
parallelcluster:version	<i>3.7.0</i>

AWS ParallelClusterRoot-Volume-Tags für den Hauptknoten

Tag-Schlüssel	Tag-Wert
<code>parallelcluster:cluster-name</code>	<i>clustername</i>
<code>parallelcluster:node-type</code>	HeadNode
<code>parallelcluster:version</code>	<i>3.7.0</i>

AWS ParallelClusterInstanz-Tags für Knoten berechnen

Schlüssel	Tag-Wert
<code>parallelcluster:cluster-name</code>	<i>clustername</i>
<code>parallelcluster:compute-resource-name</code>	<i>compute-resource-name</i>
<code>aws:ec2launchtemplate:id</code>	<i>lt-1234567890abcdef0</i>
<code>aws:ec2launchtemplate:version</code>	<i>1</i>
<code>parallelcluster:node-type</code>	Compute
<code>parallelcluster:queue-name</code>	<i>queue-name</i>
<code>parallelcluster:version</code>	<i>3.7.0</i>

AWS ParallelClusterRoot-Volume-Tags für Knoten berechnen

Tag-Schlüssel	Tag-Wert
<code>parallelcluster:cluster-name</code>	<i>clustername</i>
<code>parallelcluster:compute-resource-name</code>	<i>compute-resource-name</i>
<code>parallelcluster:node-type</code>	Compute

den Zeitpunkt der Aufrufe ermitteln. Weitere Informationen finden Sie im [AWS CloudTrail-Benutzerhandbuch](#).

- Amazon EventBridge ist ein Serverless-Ereignisbus-Service, über den Sie Ihre Anwendungen einfach mit Daten aus einer Vielzahl von Quellen verbinden können. EventBridge stellt einen Stream von Echtzeitdaten aus Ihren eigenen Anwendungen, Software-as-a-Service-Anwendungen und AWS-Services und leitet diese Daten dann an Ziele wie Lambda weiter. Auf diese Weise können Sie Ereignisse überwachen, die in Services auftreten, und ereignisgesteuerte Architekturen erstellen. Weitere Informationen finden im [EventBridgeAmazon-Benutzerhandbuch](#).

Themen

- [Integration mit Amazon CloudWatch Logs](#)
- [CloudWatchAmazon-Dashboard](#)
- [CloudWatchAmazon-Alarme für Cluster-Metriken](#)
- [AWS ParallelClusterkonfigurierte Protokollrotation](#)
- [pclusterCLI-Protokolle](#)
- [Ausgabeprotokolle der EC2-Konsole](#)
- [AWS ParallelClusterUI- und AWS ParallelCluster Laufzeitprotokolle abrufen](#)
- [Protokolle abrufen und aufbewahren](#)

Integration mit Amazon CloudWatch Logs

Weitere Informationen zu CloudWatch Protokollen finden Sie im [Amazon CloudWatch Logs-Benutzerhandbuch](#). Informationen zur Konfiguration der CloudWatch Logs-Integration finden Sie im [Monitoring](#) Abschnitt. Informationen zum Anhängen benutzerdefinierter Protokolle an die CloudWatch Konfiguration mithilfe von `append-config` finden Sie unter [Multiple CloudWatch Agent-Konfigurationsdateien](#) im CloudWatchAmazon-Benutzerhandbuch.

Amazon CloudWatch Logs Clusterprotokolle

Für jeden Cluster wird eine Protokollgruppe mit einem Namen erstellt `/aws/parallelcluster/cluster-name-<timestamp>` (z. B. `/aws/parallelcluster/testCluster-202202050215`). Jedes Protokoll (oder jede Gruppe von Protokollen, wenn der Pfad ein enthält*) auf jedem Knoten hat einen Log-Stream mit dem Namen `{hostname}.{instance_id}.{logIdentifier}`. (Zum

Beispielip-172-31-10-46.i-02587cf29cc3048f3.nodewatcher.) Die Protokolldaten werden CloudWatch vom [CloudWatchAgenten](#) gesendet, der wie root auf allen Cluster-Instances ausgeführt wird.

Bei der Erstellung des Clusters wird ein CloudWatch Amazon-Dashboard erstellt. Dieses Dashboard bietet Ihnen die Möglichkeit, die in CloudWatch Logs gespeicherten Protokolle zu überprüfen. Weitere Informationen finden Sie unter [CloudWatchAmazon-Dashboard](#).

Diese Liste enthält den *LogIdentifier* und den Pfad für die Log-Streams, die für Plattformen, Scheduler und Knoten verfügbar sind.

Log-Streams sind für Plattformen, Scheduler und Knoten verfügbar

Plattformen	Schedulers	Knoten	Protokollstreams
amazon Centos roter Hut ubuntu	als Batch Slurm	HeadNc	DCV-Authentifikator: /var/log/parallelcluster/parallelcluster_dcv_authenticator.log dcv-ext-authenticator: /var/log/parallelcluster/parallelcluster_dcv_connect.log dcv-agent: /var/log/dcv/agent.*.log dcv-x-Sitzung: /var/log/dcv/dcv-xsession.*.log DCV-Server: /var/log/dcv/server.log dcv-session-launcher: /var/log/dcv/sessionlauncher.log Xdc: /var/log/dcv/Xdcv.*.log cfn-init: /var/log/cfn-init.log Chef-Kunde: /var/log/chef-client.log
amazon Centos roter Hut	als Batch Slurm	Comput eet HeadNc	Cloud-Init: /var/log/cloud-init.log beaufsichtigt: /var/log/supervisord.log

Plattformen	Schedulers	Knoten	Protokollstreams
ubuntu			
amazon Centos roter Hut ubuntu	Slurm	Computeet	cloud-init-output: /var/log/cloud-init-output.log computemgtd: /var/log/parallelcluster/computemgtd verleumdet: /var/log/slurmd.log slurm_prolog_epilog: /var/log/parallelcluster/slurm_prolog_epilog.log
amazon Centos roter Hut ubuntu	Slurm	HeadNode	sssd: /var/log/sssds/sssds.log sssd_domain_default: /var/log/sssds/sssds_default.log pam_ssh_schlüsselgenerator: /var/log/parallelcluster/pam_ssh_key_generator.log clusterstatusmgtd: /var/log/parallelcluster/clusterstatusmgtd clustermgtd: /var/log/parallelcluster/clustermgtd Ausgabe der Computerkonsole: /var/log/parallelcluster/compute_console_output slurm_resume: /var/log/parallelcluster/slurm_resume.log slurm_suspend: /var/log/parallelcluster/slurm_suspend.log slurmctld: /var/log/slurmctld.log slurm_fleet_status_manager: /var/log/parallelcluster/slurm_fleet_status_manager.log

Plattformen	Schedulers	Knoten	Protokollstreams
amazon Centos roter Hut	als Batch Slurm	Comput eet HeadNc	Systemmeldungen: <code>/var/log/messages</code>
ubuntu	als Batch Slurm	Comput eet HeadNc	Syslog: <code>/var/log/syslog</code>

Jobs in Clustern, die verwenden, AWS Batch speichern die Ausgabe von Aufträgen, die den Status, oder erreicht haben `RUNNING`, `SUCCEEDED`, `FAILED` in CloudWatch Protokollen. Die Protokollgruppe ist `/aws/batch/job`, und das Format des Log-Stream-Namens ist `jobDefinitionName/default/ecs_task_id`. Standardmäßig sind diese Protokolle so eingestellt, dass sie nicht ablaufen, aber Sie können den Aufbewahrungszeitraum ändern. Weitere Informationen finden Sie im Benutzerhandbuch zu Amazon [CloudWatchLogs unter Ändern der Protokoll-Datenaufbewahrung in CloudWatch Logs](#).

Amazon CloudWatch Logs erstellt Image-Logs

Für jedes benutzerdefinierte Build-Image wird eine Protokollgruppe mit einem Namen erstellt `/aws/imagebuilder/ParallelClusterImage-<image-id>`. Ein eindeutiger Log-Stream mit dem Namen `{pcluster-version}/1` enthält die Ausgabe des Build-Image-Prozesses.

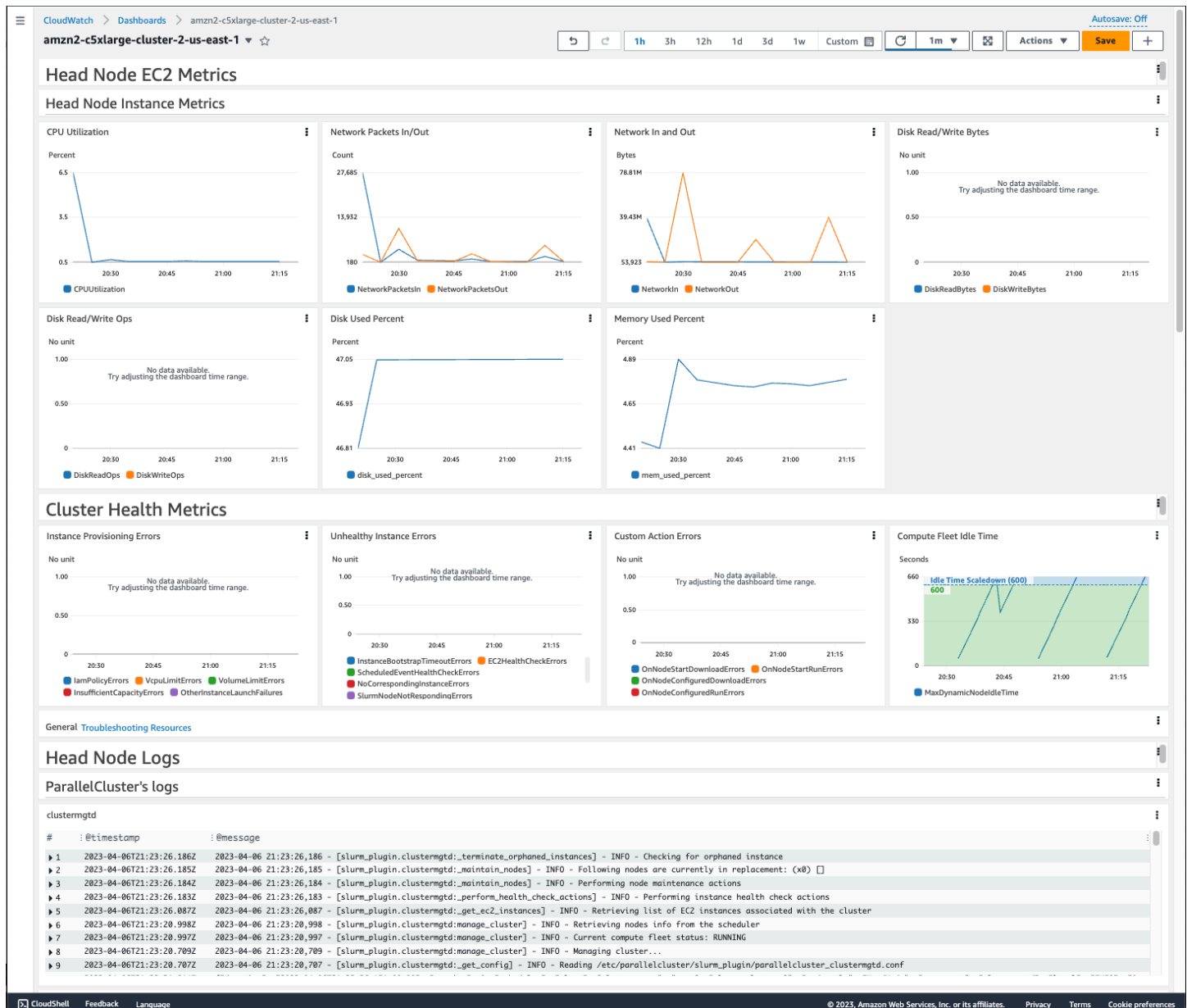
Sie können mit den `pcluster` Image-Befehlen auf die Protokolle zugreifen. Weitere Informationen finden Sie unter [AWS ParallelCluster AMI-Anpassung](#).

CloudWatchAmazon-Dashboard

Ein CloudWatch Amazon-Dashboard wird erstellt, wenn ein Cluster erstellt wird. Dies macht es einfacher, die Knoten in Ihrem Cluster zu überwachen und die in Amazon Logs gespeicherten CloudWatch Protokolle einzusehen. Der Name des Dashboards ist `ClusterName-Region`. `ClusterName` ist der Name Ihres Clusters und `Region` ist der Cluster, in AWS-Region dem sich der Cluster befindet. Sie können in der Konsole oder durch

Öffnen auf das Dashboard zugreifen <https://console.aws.amazon.com/cloudwatch/home?region=Region#dashboards:name=ClusterName-Region>.

Die folgende Abbildung zeigt ein CloudWatch Beispiel für einen Cluster.



Metriken der Head Node-Instance

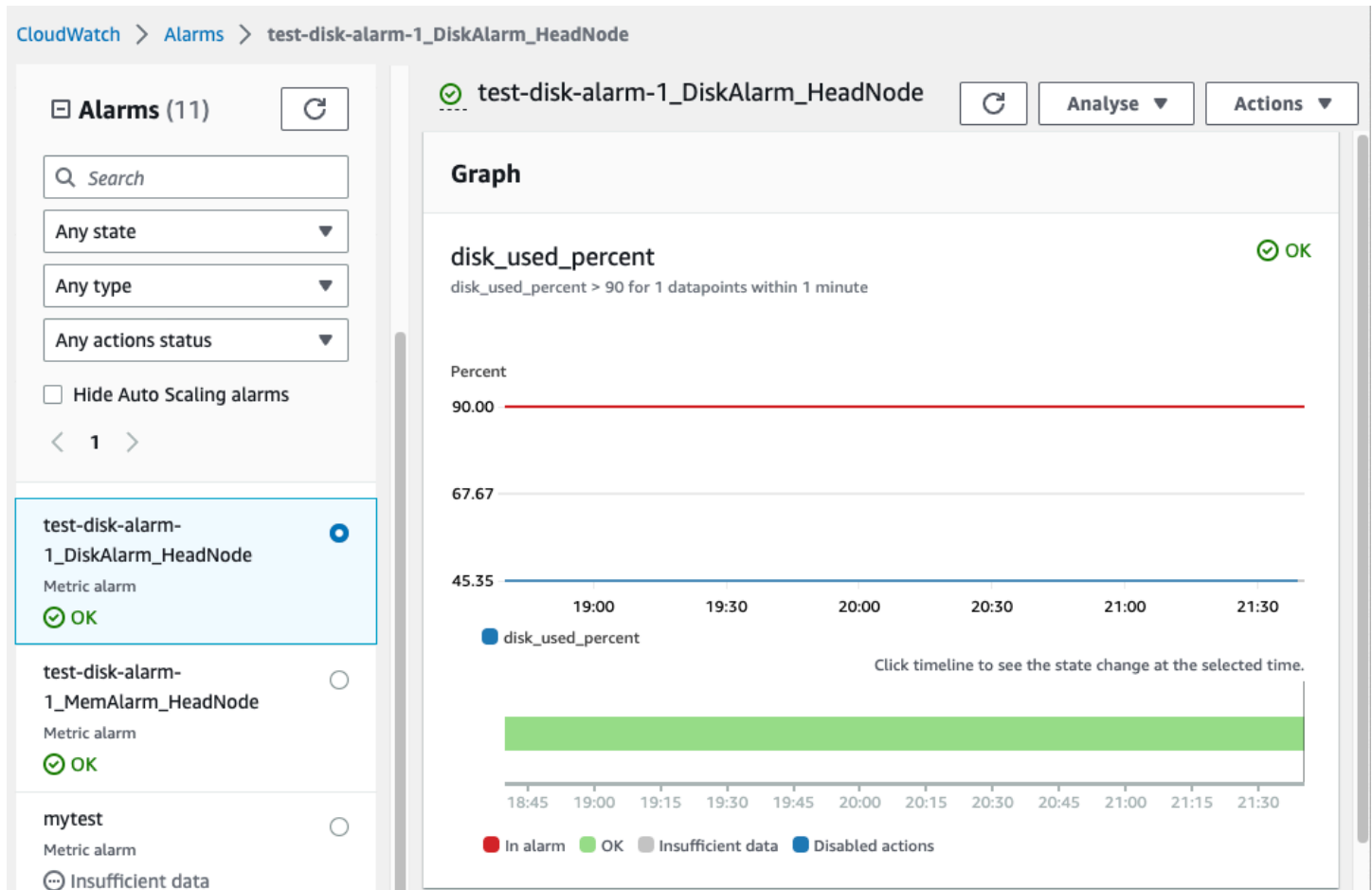
Im ersten Abschnitt des Dashboards werden Diagramme der EC2-Metriken des Hauptknotens angezeigt.

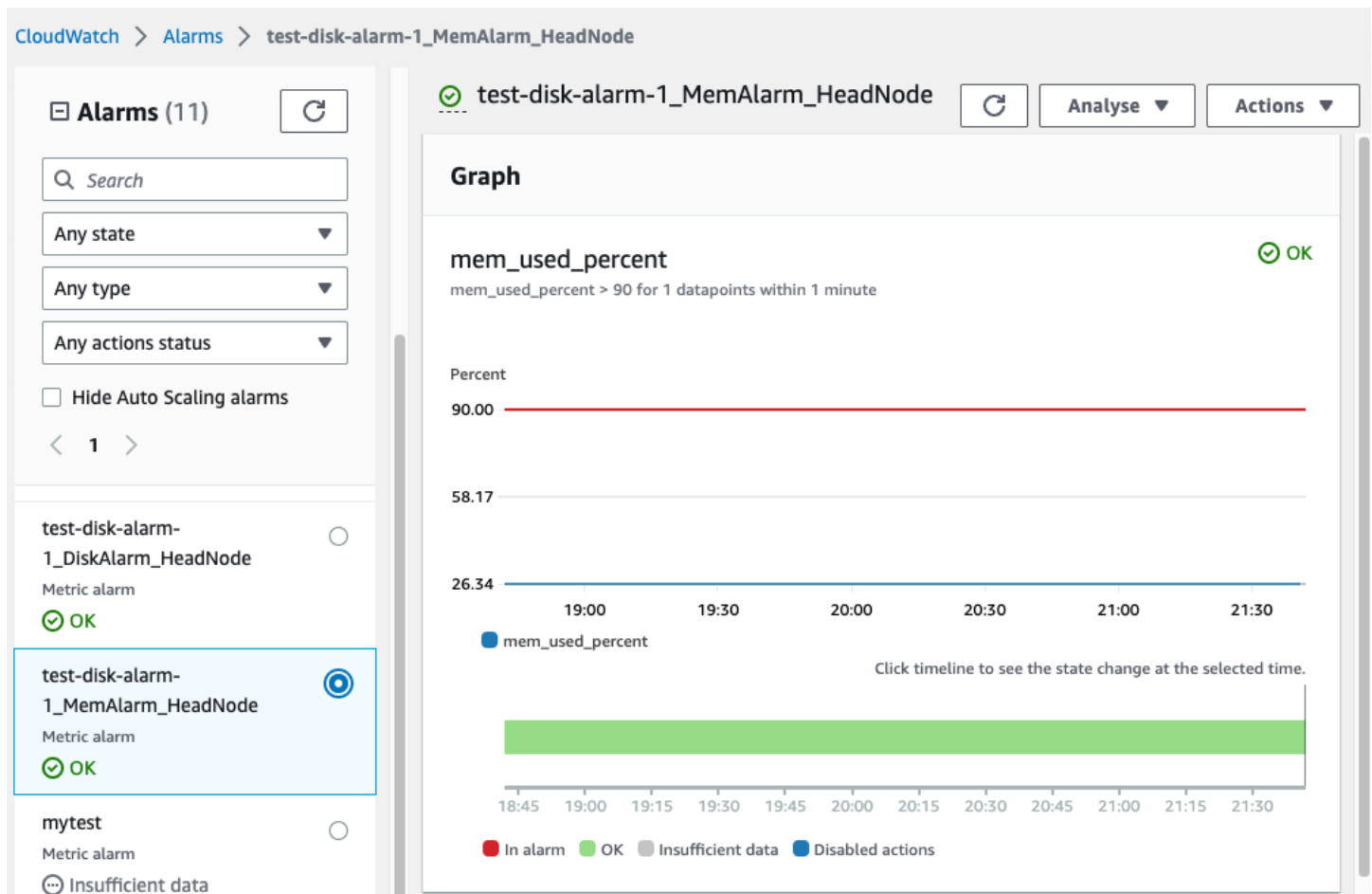
Wenn Ihr Cluster über gemeinsam genutzten Speicher verfügt, werden im nächsten Abschnitt die Metriken für gemeinsam genutzten Speicher angezeigt.

- *cluster-name*_MemAlarm_HeadNode

cluster-name des Clusters.

Greifen Sie in der CloudWatch Konsole auf die Alarme zu, indem Sie im Navigationsbereich Alarme auswählen. Die folgenden Bilder zeigen den Festplattenauslastungsalarm und den Speicherauslastungsalarm für einen Cluster.





Der Festplattenauslastungsalarm befindet sich in dem ALARM Zustand, in dem der Prozentsatz der Festplattenauslastung für einen Datenpunkt innerhalb eines Zeitraums von 1 Minute über 90% liegt.

Der Speicherauslastungsalarm befindet sich in dem ALARM Zustand, in dem der Prozentsatz der Speichernutzung für einen Datenpunkt innerhalb eines Zeitraums von 1 Minute mehr als 90% beträgt.

Note

AWS ParallelCluster konfiguriert standardmäßig keine Alarmaktionen. Informationen zum Einrichten von Alarmaktionen, wie z. B. dem Senden von Benachrichtigungen, finden Sie unter [Alarmaktionen](#). Weitere Informationen zu CloudWatch Amazon-Alarmen finden Sie unter [Verwenden von CloudWatch Amazon-Alarmen](#) im CloudWatch Amazon-Benutzerhandbuch.

Wenn Sie diese CloudWatch Amazon-Alarme nicht erstellen möchten, deaktivieren Sie sie, indem Sie `false` in der [Dashboards](#) Cluster-Konfiguration [MonitoringCloudWatch](#)///[Enabled](#) auf

setzen. Dadurch wird auch die Erstellung des CloudWatch Amazon-Dashboards deaktiviert. Weitere Informationen finden Sie unter [CloudWatchAmazon-Dashboard](#).

Note

Wenn Sie die Erstellung des CloudWatch Amazon-Dashboards deaktivieren, deaktivieren Sie auch Amazon CloudWatch `disk_used_percent` und `memory_used_percent` Alarme für Ihren Cluster.

AWS ParallelClusterkonfigurierte Protokollrotation

Die Konfigurationen für die Protokollrotation befinden sich in `/etc/logrotate.d/parallelcluster*_log_rotation` Dateien. Wenn ein konfiguriertes Protokoll rotiert, wird der aktuelle Protokollinhalt in einer einzigen Sicherung beibehalten, und das geleerte Protokoll nimmt die Protokollierung wieder auf.

Für jedes konfigurierte Protokoll wird nur ein Backup verwaltet.

AWS ParallelCluster konfiguriert ein schnell wachsendes Protokoll so, dass es rotiert, wenn es eine Größe von 50 MB erreicht. Schnell wachsende Logs stehen im Zusammenhang mit Skalierung und Slurm, einschließlich `/var/log/parallelcluster/clustermgtd`, `/var/log/parallelcluster/slurm_resume.log`, und `/var/log/slurmctld.log`.

AWS ParallelCluster konfiguriert ein langsam wachsendes Protokoll so, dass es rotiert, wenn es eine Größe von 10 MB erreicht.

Sie können frühere Protokolle anzeigen, die für die in der Clusterkonfiguration definierte Anzahl von Tagen aufbewahrt werden [Logs/CloudWatch/RetentionInDays](#) Einstellung mit CloudFormation Protokollierung aktiviert. Prüfen Sie die `RetentionInDays` Einstellungen, um zu sehen, ob die Anzahl der Tage für Ihren Anwendungsfall erhöht werden muss.

AWS ParallelCluster konfiguriert und rotiert die folgenden Protokolle:

Logs des Hauptknotens

```
/var/log/cloud-init.log
/var/log/supervisord.log
/var/log/cfn-init.log
/var/log/chef-client.log
```

```
/var/log/dcv/server.log
/var/log/dcv/sessionlauncher.log
/var/log/dcv/agent.*.log
/var/log/dcv/dcv-xsession.*.log
/var/log/dcv/Xdcv.*.log
/var/log/parallelcluster/pam_ssh_key_generator.log
/var/log/parallelcluster/clustermgtd
/var/log/parallelcluster/clusterstatusmgtd
/var/log/parallelcluster/slurm_fleet_status_manager.log
/var/log/parallelcluster/slurm_resume.log
/var/log/parallelcluster/slurm_suspend.log
/var/log/slurmctld.log
/var/log/slurmdbd.log
/var/log/parallelcluster/compute_console_output.log
```

Knotenprotokolle berechnen

```
/var/log/cloud-init.log
/var/log/supervisord.log
/var/log/cloud-init-output.log
/var/log/parallelcluster/computemgtd
/var/log/slurmd.log
```

Logs für Login-Knoten

```
/var/log/cloud-init.log
/var/log/cloud-init.log
/var/log/cloud-init-output.log
/var/log/supervisord.log
/var/log/parallelcluster/pam_ssh_key_generator.log
```

pclusterCLI-Protokolle

Die pcluster CLI schreibt Protokolle Ihrer Befehle in `pcluster.log.# /home/user/.parallelcluster/` Dateien.

Für jeden Befehl enthalten die Protokolle im Allgemeinen den Befehl mit Eingaben, eine Kopie der CLI-API-Version, die für die Ausführung des Befehls verwendet wurde, die Antwort sowie sowohl Informationen als auch Fehlermeldungen. Für einen Create- und Build-Befehl enthalten die Protokolle auch die Konfigurationsdatei, Validierungsvorgänge für die Konfigurationsdatei, die CloudFormation Vorlage und die Stack-Befehle.

Sie können diese Protokolle verwenden, um Fehler, Eingaben, Versionen und `pcluster` CLI-Befehle zu überprüfen. Sie können auch als Aufzeichnung dienen, wann Befehle erteilt wurden.

Ausgabeprotokolle der EC2-Konsole

Wenn AWS ParallelCluster festgestellt wird, dass eine statische Rechenknoteninstanz unerwartet beendet wird, versucht sie nach Ablauf einer bestimmten Zeit, die EC2-Konsolenausgabe von der beendeten Knoteninstanz abzurufen. Wenn der Rechenknoten nicht mit Amazon kommunizieren kann, können auf diese Weise nützliche Informationen zur Fehlerbehebung CloudWatch, warum der Knoten beendet wurde, trotzdem aus der Konsolenausgabe abgerufen werden. Diese Konsolenausgabe wird im `/var/log/parallelcluster/compute_console_output` Protokoll auf dem Hauptknoten aufgezeichnet. Weitere Informationen zur EC2-Konsolenausgabe finden Sie unter [Instance Console-Ausgabe der Instance Console](#) im Amazon EC2 EC2-Benutzerhandbuch für Linux-Instances.

Ruft standardmäßig AWS ParallelCluster nur die Konsolenausgabe von einer Beispieluntermenge terminierter Knoten ab. Dadurch wird verhindert, dass der Cluster-Kopfknoten aufgrund einer großen Anzahl von Terminierungen mit mehreren Konsolenausgabeanforderungen überlastet wird. Wartet standardmäßig 5 Minuten zwischen der Terminierungserkennung und dem Abrufen der Konsolenausgabe, damit EC2 Zeit hat, die endgültige Konsolenausgabe von den Knoten abzurufen.

AWS ParallelCluster

Sie können die Parameterwerte für Stichprobengröße und Wartezeit in der `/etc/parallelcluster/slurm_plugin/parallelcluster_clustermgtd.conf` Datei auf dem Hauptknoten bearbeiten.

Diese Funktion wurde in AWS ParallelCluster Version 3.5.0 hinzugefügt.

Ausgangsparameter der EC2-Konsole

Sie können die Werte der folgenden Ausgabeparameter der EC2-Konsole in der `/etc/parallelcluster/slurm_plugin/parallelcluster_clustermgtd.conf` Datei auf dem Hauptknoten bearbeiten.

`compute_console_logging_enabled`

Um die Erfassung von Konsolenausgabeprotokollen zu deaktivieren, setzen Sie `compute_console_logging_enabled` auf `false`. Der Standardwert ist `true`.

Sie können diesen Parameter jederzeit aktualisieren, ohne die Rechenflotte, ohne die Rechenflotte stoppt.

compute_console_logging_max_sample_size

`compute_console_logging_max_sample_size` legt die maximale Anzahl von Rechenknoten fest, von denen die Konsolenausgaben jedes Mal AWS ParallelCluster erfasst werden, wenn ein unerwarteter Abbruch erkannt wird. Wenn dieser Wert kleiner als 1 ist, wird die Konsolenausgabe von allen terminierten Knoten AWS ParallelCluster abgerufen. Der Standardwert ist 1.

Sie können diesen Parameter jederzeit aktualisieren, ohne die Rechenflotte, ohne die Rechenflotte stoppt.

compute_console_wait_time

`compute_console_wait_time` legt die Zeit in Sekunden fest, die AWS ParallelCluster zwischen der Erkennung eines Knotenausfalls und dem Abrufen der Konsolenausgabe von diesem Knoten vergeht. Sie können die Wartezeit verlängern, wenn Sie feststellen, dass EC2 mehr Zeit benötigt, um die endgültige Ausgabe vom terminierten Knoten abzurufen. Der Standardwert liegt bei 300 Sekunden (5 Minuten).

Sie können diesen Parameter jederzeit aktualisieren, ohne die Rechenflotte, ohne die Rechenflotte stoppt.

AWS ParallelCluster UI- und AWS ParallelCluster Laufzeitprotokolle abrufen

Erfahren Sie, wie Sie die AWS ParallelCluster Benutzeroberfläche und die AWS ParallelCluster Laufzeitprotokolle zur Fehlerbehebung abrufen. Suchen Sie zunächst nach den entsprechenden AWS ParallelCluster UI- und AWS ParallelCluster Stacknamen. Verwenden Sie den Stacknamen, um die Installationsprotokollgruppen zu finden. Exportieren Sie zum Abschluss die Protokolle. Diese Protokolle sind spezifisch für die AWS ParallelCluster Laufzeit. Informationen zu Clusterprotokollen finden Sie unter [Protokolle abrufen und aufbewahren](#).

Voraussetzungen

- Der AWS CLI ist installiert.
- Sie haben Anmeldeinformationen, um AWS CLI Befehle auf dem auszuführen AWS-Konto, auf dem die AWS ParallelCluster Benutzeroberfläche aktiviert ist.
- Sie können auf die CloudWatch Amazon-Konsole zugreifen AWS-Konto, auf der sich die AWS ParallelCluster Benutzeroberfläche befindet.

Schritt 1: Suchen Sie die Stacknamen für die entsprechenden Stacks

Im folgenden Beispiel ersetzen Sie den rot markierten Text und entsprechend dem folgenden Beispiel zusammen.

Listen Sie die Stacks auf, indem Sie den AWS-Region Ort verwenden, an dem Sie die AWS ParallelCluster Benutzeroberfläche installiert haben:

```
$ aws cloudformation list-stacks --region aws-region-id
```

Notieren Sie sich die Stacknamen für die folgenden Stacks:

- Der Name des Stacks, der die AWS ParallelCluster Benutzeroberfläche in Ihrem Konto bereitgestellt hat. Sie haben diesen Namen eingegeben, als Sie die AWS ParallelCluster Benutzeroberfläche installiert haben, `pcluster-ui` z. B.
- Der AWS ParallelCluster Stapel, dem der von Ihnen eingegebene Stapelname als Präfix vorangestellt ist, `pcluster-ui-ParallelClusterApi-ABCD1234EFGH` z. B.

Schritt 2: Suchen Sie die Protokollgruppen

Führen Sie die Tabellen und entsprechend dem AWS ParallelCluster folgenden Beispiel zusammen zusammen, wie im folgenden Beispiel dargestellt:

```
$ aws cloudformation describe-stack-resources \  
  --region aws-region-id \  
  --stack-name pcluster-ui \  
  --query "StackResources[?ResourceType == 'AWS::Logs::LogGroup' &&  
(LogicalResourceId == 'ApiGatewayAccessLog' || LogicalResourceId ==  
'ParallelClusterUILambdaLogGroup')].PhysicalResourceId" \  
  --output text
```

Führen Sie die Tabellen und entsprechend dem AWS ParallelCluster folgenden Beispiel zusammen zusammen, wie im folgenden Beispiel dargestellt:

```
$ aws cloudformation describe-stack-resources \  
  --region aws-region-id \  
  --stack-name pcluster-ui-ParallelClusterApi-ABCD1234EFGH \  
  --query "StackResources[?ResourceType == 'AWS::Logs::LogGroup' && LogicalResourceId  
== 'ParallelClusterFunctionLogGroup'].PhysicalResourceId" \  
  --output text
```

Notieren Sie die Tabellen und entsprechend dem nächsten Schritt zusammen.

Schritt 3:

Gehen Sie wie folgt vor, um die Protokolle zu sammeln und zu exportieren:

1. Melden Sie sich bei der AWS Management Console an und navigieren Sie dann zur [CloudWatch Amazon-Konsole](#), auf der AWS-Konto die AWS ParallelCluster Benutzeroberfläche aktiviert ist.
2. Wählen Sie im Navigationsbereich Logs, Logs Insights aus.
3. Wählen Sie alle im vorherigen Schritt aus.
4. Wählen Sie einen Zeitraum, z. B. 12 Stunden.
5. Führen Sie die folgende Abfrage aus:

```
$ fields @timestamp, @message
| sort @timestamp desc
| limit 10000
```

6. Wählen Sie Ergebnisse exportieren, Tabelle herunterladen (JSON).

Protokolle abrufen und aufbewahren

AWS ParallelCluster erstellt EC2-Metriken für HeadNode und Compute-Instances und Speicher. Sie können die Metriken in der CloudWatch Konsole „Benutzerdefinierte Dashboards“ einsehen. AWS ParallelCluster erstellt auch CloudWatch Cluster-Logstreams in Protokollgruppen. Sie können diese Protokolle in den benutzerdefinierten Dashboards oder Protokollgruppen der CloudWatch Konsole einsehen. Im Abschnitt Clusterkonfiguration [überwachen](#) wird beschrieben, wie Sie die CloudWatch Clusterprotokolle und das Dashboard ändern können. Weitere Informationen erhalten Sie unter [Integration mit Amazon CloudWatch Logs](#) und [CloudWatch Amazon-Dashboard](#).

Protokolle sind eine nützliche Ressource zur Behebung von Problemen. Wenn Sie beispielsweise einen ausgefallenen Cluster löschen möchten, kann es nützlich sein, zunächst ein Archiv der Clusterprotokolle zu erstellen. Folgen Sie den Schritten unter [archiarchiarchiarchi](#), um ein Archiv zu erstellen.

Themen

- [Clusterprotokolle sind nicht verfügbar in CloudWatch](#)
- [archiarchiarchiarchi](#)

- [Konservierte Protokolle](#)
- [Protokolle beendeter Knoten](#)

Clusterprotokolle sind nicht verfügbar in CloudWatch

Wenn Clusterprotokolle in nicht verfügbar sind CloudWatch, stellen Sie sicher, dass Sie die AWS ParallelCluster CloudWatch Protokollkonfiguration nicht überschrieben haben, wenn Sie der Konfiguration benutzerdefinierte Protokolle hinzufügen.

Um benutzerdefinierte Protokolle zur CloudWatch Konfiguration hinzuzufügen, stellen Sie sicher, dass Sie sie an die Konfiguration anhängen, anstatt sie abzurufen und zu überschreiben. Weitere Informationen zu `fetch-config` und `append-config` finden Sie im CloudWatch Benutzerhandbuch unter [Konfigurationsdateien für mehrere CloudWatch Agenten](#).

Um die AWS ParallelCluster CloudWatch Protokollkonfiguration wiederherzustellen, können Sie die folgenden Befehle in einem AWS ParallelCluster Knoten ausführen:

```
$ PLATFORM="$(ohai platform | jq -r ".[]")"
LOG_GROUP_NAME="$(cat /etc/chef/dna.json | jq -r ".cluster.log_group_name")"
SCHEDULER="$(cat /etc/chef/dna.json | jq -r ".cluster.scheduler")"
NODE_ROLE="$(cat /etc/chef/dna.json | jq -r ".cluster.node_type")"
CONFIG_DATA_PATH="/usr/local/etc/cloudwatch_agent_config.json"
/opt/parallelcluster/pyenv/versions/cookbook_virtualenv/bin/python /usr/local/bin/
write_cloudwatch_agent_json.py --platform $PLATFORM --config $CONFIG_DATA_PATH --log-
group $LOG_GROUP_NAME --scheduler $SCHEDULER --node-role $NODE_ROLE
/opt/aws/amazon-cloudwatch-agent/bin/amazon-cloudwatch-agent-ctl -a fetch-config -m ec2
-c file:/opt/aws/amazon-cloudwatch-agent/etc/amazon-cloudwatch-agent.json -s
```

archiarchiarchiarchi

Sie können die Protokolle in Amazon S3 oder in einer lokalen Datei archivieren (abhängig vom `--output-file` Parameter).

Note

Fügen Sie der Amazon-S3-Bucket-Richtlinie CloudWatch hinzu. Weitere Informationen finden Sie unter [Festlegen von Berechtigungen für einen Amazon S3 S3-Bucket](#) im CloudWatchLogs-Benutzerhandbuch.

```
$ pcluster export-cluster-logs --cluster-name mycluster --region eu-west-1 \  
  --bucket bucketname --bucket-prefix logs  
{  
  "url": "https://bucketname.s3.eu-west-1.amazonaws.com/export-log/mycluster-  
logs-202109071136.tar.gz?..."  
}  
  
# use the --output-file parameter to save the logs locally  
$ pcluster export-cluster-logs --cluster-name mycluster --region eu-west-1 \  
  --bucket bucketname --bucket-prefix logs --output-file /tmp/archive.tar.gz  
{  
  "path": "/tmp/archive.tar.gz"  
}
```

Das Archiv enthält die Amazon CloudWatch Logs-Streams und AWS CloudFormation Stack-Ereignisse vom Hauptknoten und den Rechenknoten der letzten 14 Tage, sofern nicht ausdrücklich in der Konfiguration oder in den Parametern für den `export-cluster-logs` Befehl angegeben. Die Zeit, die es dauert, bis der Befehl abgeschlossen ist, hängt von der Anzahl der Knoten im Cluster und der Anzahl der in CloudWatch Logs verfügbaren Protokollstreams ab. Weitere Informationen über die verfügbaren [Integration mit Amazon CloudWatch Logs](#) archiarchiarchiarchiarchiarchiarchiarchiarchiarchiarchiarchiarchiarchiarchiarchiarchi

Konservierte Protokolle

Ab Version 3.0.0 werden CloudWatch Protokolle AWS ParallelCluster standardmäßig beibehalten, wenn ein Cluster gelöscht wird. Wenn Sie einen Cluster löschen und seine Protokolle beibehalten möchten, stellen Sie sicher, dass [Monitoring/Logs/CloudWatch/Delete](#) in der Clusterkonfiguration [DeletionPolicy](#) nicht auf `gesetzt` ist. Andernfalls ändern Sie den Wert für dieses Feld in `Retain` und führen Sie den `pcluster update-cluster` Befehl aus. Führen Sie dann den `pcluster delete-cluster --cluster-name <cluster_name>` Vorgang aus, um den Cluster zu löschen, behalten Sie jedoch die Protokollgruppe bei, die in Amazon gespeichert ist CloudWatch.

Protokolle beendeter Knoten

Wenn ein statischer Rechenknoten unerwartet beendet wird und CloudWatch keine Protokolle für ihn vorliegen, überprüfen Sie, ob die Konsolenausgabe für diesen Rechenknoten auf dem Hauptknoten im Protokoll aufgezeichnet AWS ParallelCluster wurde. `/var/log/parallelcluster/compute_console_output` Weitere Informationen finden Sie unter [Wichtige Protokolle für das Debuggen](#).

1. Fügen Sie einen benutzerdefinierten Ressourcenanbieter-Stack hinzu, der Eigentümer ist und von diesem gehostet wird AWS ParallelCluster.
2. Verweisen Sie in Ihrer CloudFormation Vorlage auf den Provider-Stack als benutzerdefinierte Ressource.

Der Stack für benutzerdefinierte Ressourcenanbieter verarbeitet und beantwortet CloudFormation Anfragen. Wenn Sie beispielsweise Ihren CloudFormation Stack bereitstellen, konfigurieren und erstellen Sie auch einen Cluster. Um einen Cluster zu aktualisieren, aktualisieren Sie Ihren CloudFormation Stack. Sie löschen einen Cluster, wenn Sie Ihren Stack löschen. Weitere Informationen zu CloudFormation benutzerdefinierten Ressourcen finden Sie unter [Benutzerdefinierte Ressourcen](#) im AWS CloudFormation Benutzerhandbuch.

Warning

CloudFormation erkennt keine Drift bei benutzerdefinierten Ressourcen. Wird nur verwendet CloudFormation , um die Clusterkonfiguration zu aktualisieren und einen Cluster zu löschen. Sie können die [pcluster](#) CLI oder die verwenden [AWS ParallelClusterUI](#), um den Status des Clusters zu überwachen oder die Rechenflotte zu aktualisieren, aber Sie dürfen sie nicht verwenden, um die Clusterkonfiguration zu aktualisieren oder den Cluster zu löschen.

Note

Wir empfehlen, dass Sie Ihrem Stack [einen Kündigungsschutz](#) hinzufügen, um ein versehentliches Entfernen zu verhindern.

Provider-Stack, gehostet von AWS ParallelCluster

Der benutzerdefinierte Ressourcenanbieter-Stack ist wie im folgenden CloudFormation Vorlagenausschnitt dargestellt formatiert:

```
PclusterClusterProvider:
  Type: AWS::CloudFormation::Stack
  Properties:
    Parameters:
      CustomLambdaRole: # (Optional) RoleARN to override default
```

```
AdditionalIamPolicies: # (Optional) comma-separated list of IAM policies to add
TemplateURL: !Sub
- https://${AWS::Region}-aws-parallelcluster.s3.${AWS::Region}.${AWS::URLSuffix}/
parallelcluster/${Version}/templates/custom_resource/cluster.yaml
- { Version: 3.7.0 }
```

Eigenschaften:

Parameter:

CustomLambdaRole (optional):

Eine benutzerdefinierte Rolle mit Berechtigungen zum Ausführen der AWS Lambda , die den Cluster erstellt und verwaltet. Standardmäßig verwendet die Rolle dieselben Richtlinien, die standardmäßig in der [AWS ParallelCluster Dokumentation](#) definiert sind.

AdditionalIamPolicies (optional):

Eine durch Kommas getrennte Liste zusätzlicher Amazon Resource Names (ARNs) für IAM-Richtlinien, die der von Lambda verwendeten Rolle hinzugefügt werden sollen. Dies wird nur verwendet, wenn a CustomLambdaRole nicht angegeben ist, und kann leer gelassen werden.

Wenn Sie zusätzliche Richtlinien für den Hauptknoten, die Rechenknoten oder für den Zugriff auf einen Amazon S3 S3-Bucket benötigen, fügen Sie sie der AdditionalIamPolicy Eigenschaft CustomLambdaRole or hinzu.

Weitere Informationen zu den Standardrichtlinien finden Sie unter [AWS Identity and Access Management Berechtigungen in AWS ParallelCluster](#).

TemplateURL (erforderlich):

Die URL der AWS ParallelCluster benutzerdefinierten Ressourcendatei.

Ausgaben:

ServiceToken:

Ein Wert, der als benutzerdefinierte ServiceToken Ressourceneigenschaft verwendet werden kann. Eine benutzerdefinierte Ressource ServiceToken gibt an, wohin Anfragen AWS CloudFormation gesendet werden. Dies ist eine erforderliche Eingabe für eine Clusterressource, die Sie in Ihre AWS CloudFormation Vorlage aufnehmen.

LogGroupArn:

Der ARN, bei dem CloudWatch LogGroup sich die zugrunde liegende Ressource anmeldet.

LambdaLayerArn:

Der ARN der Lambda-Schicht, die für den laufenden AWS ParallelCluster Betrieb verwendet wird.

Cluster-Ressource

Die CloudFormation Clusterressource ist wie im folgenden CloudFormation Vorlagenausschnitt dargestellt formatiert:

```
PclusterCluster:
  Type: Custom::PclusterCluster
  Properties:
    ServiceToken: !GetAtt [ PclusterClusterProvider , Outputs.ServiceToken ]
    ClusterName: !Sub 'c-${AWS::StackName}' # Must be different from StackName
    ClusterConfiguration:
      # Your Cluster Configuration
```

Eigenschaften:

ServiceToken:

Die Ausgabe des AWS ParallelCluster Provider-Stacks. ServiceToken

ClusterName:

Der Name des Clusters, der erstellt und verwaltet werden soll. Der Name darf nicht mit dem Namen des CloudFormation Stacks übereinstimmen. Der Name kann nicht geändert werden, nachdem der Cluster erstellt wurde.

ClusterConfiguration:

Die YAML-Datei für die Cluster-Konfiguration, wie unter beschrieben [Cluster-Konfigurationsdatei](#). Sie können jedoch die üblichen CloudFormation Konstrukte verwenden, z. B. [interne](#) Funktionen.

DeletionPolicy:

Definiert, ob der Cluster gelöscht werden soll, wenn der Root-Stack gelöscht wird. Der Standardwert ist Delete.

Beibehalten:

Behalten Sie den Cluster bei, wenn die benutzerdefinierte Ressource gelöscht wird.

Note

Damit der beibehaltene Cluster weiterhin funktioniert, muss für clusterabhängige Ressourcen, wie Speicher und Netzwerke, eine Löschroutine festgelegt werden, die beibehalten wird.

Löschen:

Löschen Sie den Cluster, wenn die benutzerdefinierte Ressource gelöscht wird.

Fn::GetAttRückgabewerte:

Die `Fn::GetAtt` systeminterne Funktion gibt einen Wert für ein bestimmtes Attribut eines Typs zurück. Weitere Informationen zur Verwendung der `Fn::GetAtt intrinsic` Funktion finden Sie unter [Fn::GetAtt](#)

ClusterProperties:

Die Werte aus der [pcluster describe-cluster](#) Operation.

Bestätigungsnachrichten:

Eine Zeichenfolge, die alle Bestätigungsmeldungen enthält, die während des letzten Erstellungs- oder Aktualisierungsvorgangs aufgetreten sind.

logGroupName:

Der Name der Protokollgruppe, die für die Protokollierung von Lambda-Clustervorgängen verwendet wird. Die Protokollereignisse werden 90 Tage lang aufbewahrt, und die Protokollgruppe wird nach dem Löschen des Clusters aufbewahrt.

Beispiel**Fn::GetAtt**:

```
# Provide the public IP address of the head node as an output of a stack
Outputs:
  HeadNodeIp:
    Description: The public IP address of the head node
```

```
Value: !GetAtt [ PclusterCluster, headNode.publicIpAddress ]
```

Beispiel: Einfache, vollständige CloudFormation Vorlage mit einer AWS ParallelCluster benutzerdefinierten Ressource:

```
AWSTemplateFormatVersion: '2010-09-09'
Description: > AWS ParallelCluster CloudFormation Template

Parameters:
  HeadNodeSubnet:
    Description: Subnet where the HeadNode will run
    Type: AWS::EC2::Subnet::Id

  ComputeSubnet:
    Description: Subnet where the Compute Nodes will run
    Type: AWS::EC2::Subnet::Id

  KeyName:
    Description: KeyPair to login to the head node
    Type: AWS::EC2::KeyPair::KeyName

Resources:
  PclusterClusterProvider:
    Type: AWS::CloudFormation::Stack
    Properties:
      TemplateURL: !Sub
        - https://${AWS::Region}-aws-parallelcluster.s3.${AWS::Region}.
          ${AWS::URLSuffix}/parallelcluster/${Version}/templates/custom_resource/cluster.yaml
        - { Version: 3.7.0 }

  PclusterCluster:
    Type: Custom::PclusterCluster
    Properties:
      ServiceToken: !GetAtt [ PclusterClusterProvider , Outputs.ServiceToken ]
      ClusterName: !Sub 'c-${AWS::StackName}'
      ClusterConfiguration:
        Image:
          Os: alinux2
        HeadNode:
          InstanceType: t2.medium
          Networking:
            SubnetId: !Ref HeadNodeSubnet
        Ssh:
```

```
    KeyName: !Ref KeyName
  Scheduling:
    Scheduler: slurm
    SlurmQueues:
      - Name: queue0
        ComputeResources:
          - Name: queue0-cr0
            InstanceType: t2.micro
        Networking:
          SubnetIds:
            - !Ref ComputeSubnet

Outputs:
  HeadNodeIp:
    Description: The Public IP address of the HeadNode
    Value: !GetAtt [ PclusterCluster, headNode.publicIpAddress ]
  ValidationMessages:
    Description: Any warnings from cluster create or update operations.
    Value: !GetAtt PclusterCluster.validationMessages
```

Weitere Informationen zur Verwendung der CloudFormation AWS ParallelCluster benutzerdefinierten Ressource finden Sie unter [Einen Cluster erstellen mit AWS CloudFormation](#).

Cluster-Operationen

Wenn eine benutzerdefinierte Clusterressource zu einem CloudFormation Stack hinzugefügt wird, CloudFormation kann sie die folgenden Cluster-Operationen ausführen:

- CloudFormation erstellt einen Cluster in einem neuen separaten Stapel, wenn ein Stack bereitgestellt wird, der die AWS ParallelCluster benutzerdefinierte Ressource enthält.
- Wenn Sie die im Stack definierte Clusterkonfiguration gemäß den Richtlinien für CloudFormation Konfigurationsupdates aktualisieren, wird der Cluster aktualisiert. Der AWS ParallelCluster benutzerdefinierte Ressourcenanbieter stoppt die Rechenflotte nicht, bevor er den Cluster aktualisiert hat. Wir empfehlen, dass Sie die [QueueUpdateStrategy](#) Einstellung für Cluster-Updates verwenden. Auf diese Weise können Sie vermeiden, dass vor und nach Updates explizite `pcluster update-compute-fleet` Aufrufe getätigt werden, wenn Sie die AWS ParallelCluster benutzerdefinierte Ressource verwenden.
- Wenn Sie den Stack löschen, wird der Cluster gelöscht.

Fehlerbehebung bei Stacks, die die AWS ParallelCluster benutzerdefinierte Ressource enthalten

CloudFormation stellt mit einer AWS ParallelCluster benutzerdefinierten Ressource einen Cluster aus einem neuen, separaten Stack bereit. Sie können die Clustererstellung überwachen, indem Sie die folgenden Schritte ausführen:

1. Navigieren Sie zu CloudFormation AWS Management Console und wählen Sie im Navigationsbereich Stacks aus.
2. Wählen Sie den Stack mit dem Namen aus, den Sie für den Clusternamen definiert haben.
3. Wenn der Stack-Status lautet `ROLLBACK_COMPLETE`, ist bei der Clustererstellung ein Fehler aufgetreten.
4. Wählen Sie Stack-Details und dann die Registerkarte Ereignisse aus.
5. Suchen Sie nach Ereignissen auf der logischen ID nach dem Namen, den Sie für den Clusternamen definiert haben. Es hat eine `StatusReason`, die einen Grund für ein Problem angibt.
6. Sie können auch das Dropdownmenü Stapel und dann Gelöscht auswählen, um die Liste der gelöschten Stapel zu sehen. Wählen Sie den Stack mit dem Clusternamen aus und sehen Sie sich Ereignisse an, um weitere Informationen zu erhalten.
7. Um die Ausgabe des benutzerdefinierten Ressourcenanbieters anzuzeigen, der den Cluster verwaltet, wählen Sie den Stack mit der Beschreibung „Benutzerdefinierte AWS ParallelCluster Clusterressource“ aus. Wählen Sie die Registerkarte Ressourcen, suchen Sie die Ressource mit der logischen ID `PclusterCfnFunctionLogGroup` und folgen Sie dem angegebenen Link. Sehen Sie sich die Log-Streams an, die die Lambda-Debug-Ausgabe zeigen.
8. Informationen zur Fehlerbehebung im Cluster finden Sie unter [Fehlerbehebung für AWS ParallelCluster](#)

Elastic Fabric Adapter

Elastic Fabric Adapter (EFA) ist ein Netzwerkgerät, das Betriebssystem-Bypass-Funktionen für die Netzwerkkommunikation mit geringer Latenz mit anderen Instances im selben Subnetz besitzt. EFA wird mithilfe von Libfabric verfügbar gemacht und kann von Anwendungen verwendet werden, die die Messaging Passing Interface (MPI) verwenden.

Um EFA mit AWS ParallelCluster einem Slurm Scheduler zu verwenden, setzen Sie [SlurmQueues/ComputeResources/Efa/Enabled](#) auf `true`.

Eine Liste der EC2-Instances, die EFA unterstützen, finden Sie unter [Unterstützte Instanztypen](#) im Amazon EC2 EC2-Benutzerhandbuch für Linux-Instances.

Wir empfehlen, dass Sie Ihre EFA-fähigen Instances in einer Placement-Gruppe ausführen. Auf diese Weise werden die Instances in einer Weitere Informationen zur Konfiguration von Platzierungsgruppen mit AWS ParallelCluster finden Sie unter [SlurmQueues/Networking/PlacementGroup](#).

Weitere Informationen finden Sie unter [Elastic Fabric Adapter](#) im Amazon EC2 EC2-Benutzerhandbuch für Linux-Instances und [Skalierung von HPC-Workloads mit Elastic Fabric-Adapter sowie AWS ParallelCluster](#) im AWS Open Source-Blog.

Note

Elastic Fabric Adapter (EFA) Weitere Informationen finden Sie unter [Scheduling SlurmQueues//Networking/SubnetIds](#).

Note

Standardmäßig aktivieren Ubuntu Distributionen den Schutz `ptrace` (Process Trace). `ptrace` Der Schutz ist deaktiviert, damit `libfabric` ordnungsgemäß funktioniert. Weitere Informationen finden Sie unter [Deaktivieren des PTrace-Schutzes](#) im Amazon EC2 EC2-Benutzerhandbuch für Linux-Instances.

Intel MPI aktivieren

Intel MPI ist verfügbar auf der AWS ParallelCluster AMIs für `linux2`, `centos7`, `rhel8`, `ubuntu2204`, und `ubuntu2004` Werte für die [Image/Os](#) Einstellung.

Note

Um Intel MPI verwenden zu können, müssen Sie die Bedingungen der [Vereinfachte Softwarelizenz von Intel](#).

Standardmäßig wird Open MPI auf dem Pfad platziert. Um Intel MPI anstelle von Open MPI zu aktivieren, müssen Sie zuerst das Intel MPI-Modul laden. Anschließend müssen Sie die neueste Version installieren, indem Sie `module load intelmpi`. Der genaue Name des Moduls ändert sich mit jedem Update. Führen Sie `module avail` aus, um anzuzeigen, welche Module verfügbar sind. Die Ausgabe sieht wie folgt aus.

```
$ module avail
-----/usr/share/Modules/modulefiles
-----
dot                modules
libfabric-aws/1.16.0~amzn3.0  null
module-git         openmpi/4.1.4
module-info        use.own

-----/opt/intel/mpi/2021.6.0/modulefiles
-----
intelmpi
```

Führen Sie `module load modulename` aus, um ein Modul zu laden. Sie können dies dem Skript hinzufügen, mit dem `mpirun` ausgeführt wird.

```
$ module load intelmpi
```

Führen Sie `module list` aus, um anzuzeigen, welche Module geladen werden.

```
$ module list
Currently Loaded Modulefiles:
 1) intelmpi
```

Führen Sie `mpirun --version` aus, um zu überprüfen, ob Intel MPI aktiviert ist.

```
$ mpirun --version
Intel(R) MPI Library for Linux* OS, Version 2021.6 Build 20220227 (id: 28877f3f32)
Copyright 2003-2022, Intel Corporation.
```

Nachdem das Intel MPI-Modul geladen wurde, werden mehrere Pfade geändert, damit die Intel MPI-Tools verwendet werden können. Um den Code auszuführen, der mit den Intel MPI-Tools kompiliert wurde, laden Sie zuerst das Intel MPI-Modul.

Note

Intel MPI ist nicht kompatibel mit AWS Graviton-basierte Instanzen.

Note

Vorher AWS ParallelCluster Version 2.5.0, Intel MPI war nicht verfügbar auf der AWS ParallelCluster AMIs in den Regionen China (Peking) und China (Ningxia).

AWS ParallelCluster-API

Was ist AWS ParallelCluster API?

AWS ParallelCluster API ist eine serverlose Anwendung, die, sobald sie für Sie bereitgestellt ist AWS-Konto, über eine API programmgesteuerten Zugriff auf AWS ParallelCluster Funktionen bietet.

AWS ParallelCluster Die API wird als eigenständige [AWS CloudFormation](#) Vorlage verteilt, die einen [Amazon API Gateway Gateway-Endpunkt](#) enthält, der AWS ParallelCluster Funktionen bereitstellt, und eine [AWS Lambda](#) Funktion, die sich um die Verarbeitung der aufgerufenen Funktionen kümmert.

Die folgende Abbildung zeigt ein übergeordnetes Architekturdiagramm der AWS ParallelCluster API-Infrastruktur.

AWS ParallelCluster API-Dokumentation

Die OpenAPI-Spezifikationsdatei, die die AWS ParallelCluster API beschreibt, kann heruntergeladen werden von:

```
https://<REGION>-aws-parallelcluster.s3.<REGION>.amazonaws.com/  
parallelcluster/<VERSION>/api/ParallelCluster.openapi.yaml
```

[Ausgehend von der OpenAPI-Spezifikationsdatei können Sie die Dokumentation für die AWS ParallelCluster API generieren, indem Sie eines der vielen verfügbaren Tools wie Swagger UI oder Redoc verwenden.](#)

Wie stellt man die API bereit AWS ParallelCluster

Um die AWS ParallelCluster API bereitzustellen, müssen Sie Administrator von sein AWS-Konto.

Die für die Bereitstellung der API verwendete Vorlage ist unter der folgenden URL verfügbar:

```
https://<REGION>-aws-parallelcluster.s3.<REGION>.amazonaws.com/  
parallelcluster/<VERSION>/api/parallelcluster-api.yaml
```

wo **<REGION>** ist der AWS-Region Ort, an dem die API bereitgestellt werden muss, und **<VERSION>** ist die AWS ParallelCluster Version (z. B. 3.7.0).

AWS Lambda verarbeitet die von der API aufgerufenen Funktionen mithilfe einer Lambda-Layer-Schnittstelle mit dem. [AWS ParallelCluster Python-Bibliothek-API](#)

Warning

Jeder Benutzer in den AWS-Konto, der privilegierten Zugriff AWS Lambda auf die Amazon API Gateway Gateway-Service hat, erbt automatisch die Berechtigungen zur Verwaltung von AWS ParallelCluster API-Ressourcen.

Bereitstellen mit AWS CLI

Konfigurieren Sie die AWS Anmeldeinformationen für die Verwendung mit der CLI, falls Sie dies noch nicht getan haben.

```
$ aws configure
```

Führen Sie die folgenden Befehle aus, um die API bereitzustellen:

```
$ REGION=<region>  
$ API_STACK_NAME=<stack-name> # This can be any name  
$ VERSION=3.7.0  
$ aws cloudformation create-stack \  
  --region ${REGION} \  
  --stack-name ${API_STACK_NAME} \  
  --template-url https://${REGION}-aws-parallelcluster.s3.${REGION}.amazonaws.com/  
parallelcluster/${VERSION}/api/parallelcluster-api.yaml \  
  --capabilities CAPABILITY_NAMED_IAM CAPABILITY_AUTO_EXPAND
```

```
$ aws cloudformation wait stack-create-complete --stack-name ${API_STACK_NAME} --region  
${REGION}
```

Passen Sie Ihre Bereitstellung an

Sie können die API-Bereitstellung anpassen, indem Sie die in der Vorlage bereitgestellten AWS CloudFormation Parameter verwenden. Um den Wert eines Parameters bei der Bereitstellung über die CLI zu konfigurieren, kann die folgende Option verwendet werden: `--parameters ParameterKey=KeyName,ParameterValue=Value`.

Die folgenden Parameter sind optional:

- **Region** — Verwenden Sie den Region Parameter, um anzugeben, ob die API alle Ressourcen AWS-Regionen (Standard) oder einzelne Ressourcen kontrollieren kann AWS-Region. Stellen Sie diesen Wert auf AWS-Region die API ein, für die die API bereitgestellt wird, um den Zugriff einzuschränken.
- **ParallelClusterFunctionRole**— Dadurch wird die IAM-Rolle außer Kraft gesetzt, die der AWS Lambda Funktion zugewiesen wird, die Funktionen implementiert AWS ParallelCluster. Der Parameter akzeptiert den ARN einer IAM-Rolle. Eine solche Rolle muss so konfiguriert werden, dass sie AWS Lambda als IAM-Principal verwendet wird.
- **CustomDomainName, CustomDomainCertificate, CustomDomainHostedZoneId** - Verwenden Sie diese Parameter, um eine benutzerdefinierte Domain für den Amazon API Gateway Gateway-Endpunkt festzulegen. **CustomDomainName** ist der Name der zu verwendenden Domain, **CustomDomainCertificate** der ARN eines AWS verwalteten Zertifikats für diesen Domainnamen und **CustomDomainHostedZoneId** die ID der von [Amazon Route 53](#) gehosteten Zone, in der Sie Datensätze erstellen möchten.


Warning

Sie können benutzerdefinierte Domain-Einstellungen konfigurieren, um eine Mindestversion von Transport Layer Security (TLS) für die API durchzusetzen. Weitere Informationen finden Sie unter [Auswahl einer TLS-Mindestversion für eine benutzerdefinierte Domain in API Gateway](#).

- **EnableIamAdminAccess**— Standardmäßig ist die AWS Lambda Funktion, die AWS ParallelCluster API-Operationen verarbeitet, mit einer IAM-Rolle konfiguriert, die jeglichen privilegierten IAM-Zugriff verhindert (`EnableIamAdminAccess=false`). Dadurch kann die API keine Operationen verarbeiten, die die Erstellung von IAM-Rollen oder -Richtlinien erfordern. Aus diesem Grund ist die

Erstellung von Clustern oder benutzerdefinierten Images nur dann erfolgreich, wenn IAM-Rollen als Eingabe im Rahmen der Ressourcenkonfiguration bereitgestellt werden.

Wenn auf `true` die AWS ParallelCluster API gesetzt `EnableIamAdminAccess` ist, werden Berechtigungen zur Verwaltung der Erstellung von IAM-Rollen erteilt, die für die Bereitstellung von Clustern oder die Generierung benutzerdefinierter AMIs erforderlich sind.

 Warning


Wenn dieser Wert auf `true` gesetzt wird, werden IAM-Administratorrechte für die AWS Lambda AWS ParallelCluster Funktionsverarbeitungsvorgänge gewährt.

Weitere Informationen zu [AWS ParallelCluster Benutzerbeispielrichtlinien für die Verwaltung von IAM-Ressourcen](#) den Funktionen, die bei Aktivierung dieses Modus freigeschaltet werden können, finden Sie unter.

- `PermissionsBoundaryPolicy`— Dieser optionale Parameter akzeptiert einen vorhandenen IAM-Richtlinien-ARN, der als Berechtigungsgrenze für alle von der PC-API-Infrastruktur erstellten IAM-Rollen und als Bedingung für die administrativen IAM-Berechtigungen festgelegt wird, sodass nur Rollen mit dieser Richtlinie von der PC-API erstellt werden können.

Weitere Informationen zu den [Modus `PermissionsBoundary`](#) Einschränkungen, die dieser Modus mit sich bringt, finden Sie unter.

- `CreateApiUserRole`— Standardmäßig umfasst die Bereitstellung der AWS ParallelCluster API die Erstellung einer IAM-Rolle, die als einzige Rolle festgelegt ist, die berechtigt ist, die API aufzurufen. Der Amazon API Gateway Gateway-Endpunkt ist mit einer ressourcenbasierten Richtlinie konfiguriert, sodass nur dem erstellten Benutzer die Aufrufberechtigung erteilt wird. Um dies zu ändern, richten Sie API-Zugriff ein `CreateApiUserRole=false` und gewähren Sie diesen dann ausgewählten IAM-Benutzern. Weitere Informationen finden Sie unter [Zugriffskontrolle für das Aufrufen einer API](#) im Amazon API Gateway Developer Guide.

 Warning

Wenn der `CreateApiUserRole=true` Zugriff auf den API-Endpunkt nicht durch die Amazon API Gateway Gateway-Ressourcenrichtlinien eingeschränkt ist, können alle IAM-Rollen mit uneingeschränkten Rechten auf Funktionen `execute-api:Invoke` zugreifen.

AWS ParallelCluster Weitere Informationen finden Sie unter [Steuern des Zugriffs auf eine API mit API-Gateway-Richtlinien](#) im API Gateway Developer Guide.

⚠ Warning

Der `ParallelClusterApiUserRole` hat die Berechtigung, alle AWS ParallelCluster API-Operationen aufzurufen. Informationen zum Einschränken des Zugriffs auf eine Teilmenge von API-Ressourcen finden Sie unter [Steuern, wer eine API-Gateway-API-Methode mit IAM-Richtlinien aufrufen kann](#) im API Gateway Developer Guide.

- `IAM RoleAndPolicyPrefix` — Dieser optionale Parameter akzeptiert eine Zeichenfolge mit maximal 10 Zeichen, die als Präfix sowohl für IAM-Rollen als auch für Richtlinien verwendet wird, die als Teil der PC-API-Infrastruktur erstellt wurden.

Aktualisierung der API

Auf eine neuere AWS ParallelCluster Version aktualisieren

Option 1: Entfernen Sie die vorhandene API, indem Sie den entsprechenden AWS CloudFormation Stack löschen und die neue API wie oben gezeigt bereitstellen.

Option 2: Aktualisieren Sie die bestehende API, indem Sie die folgenden Befehle ausführen:

```
$ REGION=<region>
$ API_STACK_NAME=<stack-name> # This needs to correspond to the existing API stack
name
$ VERSION=3.7.0
$ aws cloudformation update-stack \
  --region ${REGION} \
  --stack-name ${API_STACK_NAME} \
  --template-url https://${REGION}-aws-parallelcluster.s3.${REGION}.amazonaws.com/
parallelcluster/${VERSION}/api/parallelcluster-api.yaml \
  --capabilities CAPABILITY_NAMED_IAM CAPABILITY_AUTO_EXPAND
$ aws cloudformation wait stack-update-complete --stack-name ${API_STACK_NAME} --region
${REGION}
```

API aufrufen AWS ParallelCluster

Der AWS ParallelCluster Amazon API Gateway Gateway-Endpunkt ist mit einem [AWS_IAMAutorisierungstyp](#) konfiguriert und erfordert, dass alle Anfragen mit gültigen IAM-Anmeldeinformationen signiert sind ([API-Referenz: HTTP-Anfragen stellen](#)).

Bei der Bereitstellung mit Standardeinstellungen werden API-Aufrufberechtigungen nur dem standardmäßigen IAM-Benutzer gewährt, der mit der API erstellt wurde.

Um den ARN des Standard-IAM-Benutzers abzurufen, führen Sie folgenden Befehl aus:

```
$ REGION=<region>
$ API_STACK_NAME=<stack-name>
$ aws cloudformation describe-stacks --region ${REGION} --stack-name ${API_STACK_NAME}
  --query "Stacks[0].Outputs[?OutputKey=='ParallelClusterApiUserRole'].OutputValue" --
output text
```

Führen Sie den [AssumeRoleSTS-Befehl](#) aus, um temporäre Anmeldeinformationen für den Standard-IAM-Benutzer zu erhalten.

Sie können den AWS ParallelCluster API-Endpunkt abrufen, indem Sie den folgenden Befehl ausführen:

```
$ REGION=<region>
$ API_STACK_NAME=<stack-name>
$ aws cloudformation describe-stacks --region ${REGION} --stack-name ${API_STACK_NAME}
  --query "Stacks[0].Outputs[?OutputKey=='ParallelClusterApiInvokeUrl'].OutputValue" --
output text
```

Die AWS ParallelCluster API kann von jedem HTTP-Client aufgerufen werden, der den OpenAPI-Spezifikationen entspricht, die Sie hier finden:

```
https://<REGION>-aws-parallelcluster.s3.<REGION>.amazonaws.com/
parallelcluster/<VERSION>/api/ParallelCluster.openapi.yaml
```

[Anfragen müssen mit SigV4 signiert sein, wie hier dokumentiert.](#)

Derzeit bieten wir keine offizielle API-Client-Implementierung an. [API-Clients können jedoch mithilfe von OpenAPI Generator einfach aus dem OpenAPI-Modell generiert werden.](#) Sobald der Client generiert ist, muss die SigV4-Signatur hinzugefügt werden, sofern sie nicht standardmäßig bereitgestellt wird.

Eine Referenzimplementierung für einen Python-API-Client finden Sie im [AWS ParallelClusterRepository](#). Weitere Informationen zur Verwendung des Python-API-Clients finden Sie im [Verwenden der AWS ParallelCluster-API](#) Tutorial.

Um erweiterte Zugriffskontrollmechanismen wie Amazon Cognito oder Lambda Authorizers zu implementieren oder die API mit AWS WAF unseren API-Schlüsseln weiter zu schützen, folgen Sie der [Amazon](#) API Gateway Gateway-Dokumentation.

Warning

Ein IAM-Benutzer, der berechtigt ist, die AWS ParallelCluster API aufzurufen, kann indirekt alle AWS Ressourcen kontrollieren, die in der verwaltet werden. AWS ParallelCluster AWS-Konto Dazu gehört auch die Erstellung von AWS Ressourcen, die der Benutzer aufgrund von Einschränkungen in der Benutzer-IAM-Richtlinie nicht direkt kontrollieren kann. Beispielsweise kann die Erstellung eines AWS ParallelCluster Clusters, abhängig von seiner Konfiguration, die Bereitstellung von Amazon EC2-Instances, Amazon Route 53, Amazon Elastic File System-Dateisystemen, Amazon FSx-Dateisystemen, IAM-Rollen und Ressourcen von anderen AWS-Services Benutzern umfassen, über AWS ParallelCluster die der Benutzer möglicherweise keine direkte Kontrolle hat.

Warning

Bei der Erstellung eines Clusters mit den in der Konfiguration `AdditionalIamPolicies` angegebenen Werten müssen die zusätzlichen Richtlinien einem der folgenden Muster entsprechen:

```
- !Sub arn:${AWS::Partition}:iam:${AWS::AccountId}:policy/parallelcluster*
- !Sub arn:${AWS::Partition}:iam:${AWS::AccountId}:policy/parallelcluster/*
- !Sub arn:${AWS::Partition}:iam:aws:policy/CloudWatchAgentServerPolicy
- !Sub arn:${AWS::Partition}:iam:aws:policy/AmazonSSMManagedInstanceCore
- !Sub arn:${AWS::Partition}:iam:aws:policy/AWSBatchFullAccess
- !Sub arn:${AWS::Partition}:iam:aws:policy/AmazonS3ReadOnlyAccess
- !Sub arn:${AWS::Partition}:iam:aws:policy/service-role/AWSBatchServiceRole
- !Sub arn:${AWS::Partition}:iam:aws:policy/service-role/
AmazonEC2ContainerServiceforEC2Role
- !Sub arn:${AWS::Partition}:iam:aws:policy/service-role/
AmazonECSTaskExecutionRolePolicy
- !Sub arn:${AWS::Partition}:iam:aws:policy/service-role/
AmazonEC2SpotFleetTaggingRole
```

```
- !Sub arn:${AWS::Partition}:iam::aws:policy/EC2InstanceProfileForImageBuilder
- !Sub arn:${AWS::Partition}:iam::aws:policy/service-role/
  AWSLambdaBasicExecutionRole
```

Wenn Sie weitere zusätzliche Richtlinien benötigen, können Sie eine der folgenden Aktionen ausführen:

- Bearbeiten Sie das `DefaultParallelClusterIamAdminPolicy` in:

```
https://<REGION>-aws-parallelcluster.s3.<REGION>.amazonaws.com/
parallelcluster/<VERSION>/api/parallelcluster-api.yaml
```

Fügen Sie die Richtlinie im `ArnLike/iam:PolicyARN` Abschnitt hinzu.

- Lassen Sie die Angabe von Richtlinien für `AdditionalIamPolicies` in der Konfigurationsdatei aus und fügen Sie der im Cluster erstellten AWS ParallelCluster Instanzrolle manuell Richtlinien hinzu.

Zugriff auf die API-Protokolle und -Metriken

API-Protokolle werden CloudWatch mit einer Aufbewahrungsfrist von 30 Tagen auf Amazon veröffentlicht. Führen Sie den folgenden Befehl aus, um den mit einer API-Bereitstellung verknüpften LogGroup Namen abzurufen:

```
$ REGION=<region>
$ API_STACK_NAME=<stack-name>
$ aws cloudformation describe-stacks --region ${REGION} --
  stack-name ${API_STACK_NAME} --query "Stacks[0].Outputs[?
  OutputKey=='ParallelClusterLambdaLogGroup'].OutputValue" --output text
```

Lambda-Metriken, Logs und [AWS X-Ray](#) Trace-Logs können auch über die Lambda-Konsole abgerufen werden. Führen Sie den folgenden Befehl aus, um den ARN der Lambda-Funktion abzurufen, die einer API-Bereitstellung zugeordnet ist:

```
$ REGION=<region>
$ API_STACK_NAME=<stack-name>
$ aws cloudformation describe-stacks --region ${REGION} --stack-name ${API_STACK_NAME}
  --query "Stacks[0].Outputs[?OutputKey=='ParallelClusterLambdaArn'].OutputValue" --
  output text
```

Stellen Sie mit NICE DCV eine Connect zum Hauptknoten her

NICE DCV ist eine Remote-Visualisierungstechnologie, mit der Benutzer eine sichere Verbindung zu grafikintensiven 3D-Anwendungen herstellen können, die auf einem Remote-Hochleistungsserver gehostet werden. Weitere Informationen finden Sie unter [NICE DCV](#).

Die NICE DCV-Software wird automatisch auf dem Hauptknoten installiert und kann mithilfe des [Dcv](#) Abschnitts in der [HeadNode](#)Konfiguration aktiviert werden.

```
HeadNode:
  Dcv:
    Enabled: true
```

Auf diese Weise wird der Hauptknoten auf den [Speicherordner des DCV-Servers gesetzt](#). AWS ParallelCluster/home/<DEFAULT_AMI_USER> Weitere Informationen zu den NICE-DCV-Konfigurationsparametern finden Sie unter [HeadNode/Dcv](#). Verwenden Sie den [pcluster dcv-connect](#) Befehl, um eine Verbindung zur NICE-DCV-Sitzung herzustellen.

NICE DCV HTTPS-Zertifikat

NICE DCV generiert automatisch ein selbstsigniertes Zertifikat, um den Datenverkehr zwischen dem NICE DCV-Client und dem NICE DCV-Server zu sichern.

Um das standardmäßige selbstsignierte NICE-DCV-Zertifikat durch ein anderes Zertifikat zu ersetzen, stellen Sie zuerst eine Verbindung zum Hauptknoten her. Kopieren Sie dann das Zertifikat und den Schlüssel in den /etc/dcv-Ordner, bevor Sie den [pcluster dcv-connect](#)-Befehl ausführen.

Weitere Informationen finden Sie unter [Ändern des TLS-Zertifikats](#) im NICE DCV-Administratorhandbuch.

Lizenzierung NICE DCV

Der NICE DCV-Server benötigt keinen Lizenzserver, wenn er auf Amazon EC2 EC2-Instances ausgeführt wird. Der NICE DCV-Server muss jedoch regelmäßig eine Verbindung zu einem Amazon S3 S3-Bucket herstellen, um festzustellen, ob eine gültige Lizenz verfügbar ist.

AWS ParallelCluster fügt die erforderlichen Berechtigungen automatisch zur Hauptknoten-IAM-Richtlinie hinzu. Verwenden Sie bei der Verwendung einer benutzerdefinierten IAM-Instanzrichtlinie die in [NICE DCV auf Amazon EC2](#) im NICE DCV-Administratorhandbuch beschriebenen Berechtigungen.

Tipps zur Fehlerbehebung finden Sie unter [Behebung von Problemen in NICE DCV](#).

Verwenden von `pcluster update-cluster`

[pcluster update-cluster](#) analysiert in AWS ParallelCluster 3.x die Einstellungen, die zum Erstellen des aktuellen Clusters verwendet wurden, und die Einstellungen in der Konfigurationsdatei auf Probleme. Wenn Probleme entdeckt werden, werden sie gemeldet, und es werden die Schritte zur Behebung der Probleme angezeigt. Wenn beispielsweise die Rechenleistung geändert [InstanceType](#) wird, muss die Rechenflotte gestoppt werden, bevor ein Update durchgeführt werden kann. Dieses Problem wird gemeldet, wenn es entdeckt wird. Wenn keine Blockierungsprobleme festgestellt werden, wird der Aktualisierungsvorgang gestartet und die Änderungen werden gemeldet.

Sie können den verwenden `pcluster update-cluster --dryrun` option, um die Änderungen vor ihrer Ausführung zu sehen. Weitere Informationen finden Sie unter [Beispiele für pcluster update-cluster](#).

Anleitungen zur Fehlerbehebung finden Sie unter [Fehlerbehebung für AWS ParallelCluster](#).

Richtlinie aktualisieren: Definitionen

Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.

Nach dem Ändern dieser Einstellung kann der Cluster mit aktualisiert werden [pcluster update-cluster](#).

Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.

Nach dem Ändern dieser Einstellung kann der Cluster nicht aktualisiert werden. Sie müssen die Einstellungen für den ursprünglichen Cluster rückgängig machen und einen neuen Cluster mit den aktualisierten Einstellungen erstellen. Sie können den ursprünglichen Cluster zu einem späteren Zeitpunkt löschen. Um den neuen Cluster zu erstellen, verwenden Sie [pcluster create-cluster](#). Um den ursprünglichen Cluster zu löschen, verwenden Sie [pcluster delete-cluster](#).

Aktualisierungsrichtlinie: Diese Einstellung wird während eines Updates nicht analysiert.

Diese Einstellungen können geändert und der Cluster aktualisiert werden mit [pcluster update-cluster](#).

Aktualisierungsrichtlinie: Die Rechenflotte muss gestoppt werden, damit diese Einstellung für ein Update geändert werden kann.

Diese Einstellungen können nicht geändert werden, solange die Rechenflotte existiert. Entweder muss die Änderung rückgängig gemacht werden, oder die Rechenflotte muss gestoppt (verwendet [pcluster update-compute-fleet](#)) werden. Nachdem die Compute-Flotte gestoppt wurde, können Sie den Cluster ([pcluster update-cluster](#)) aktualisieren, um die Änderungen zu aktivieren. Wenn Sie beispielsweise einen Slurm Scheduler mit [SlurmQueues//ComputeResourcesName/MinCount](#) > 0 verwenden, wird eine Rechenflotte gestartet.

Aktualisierungsrichtlinie: Die Rechenflotte und die Anmeldeknoten müssen gestoppt werden, damit diese Einstellung für ein Update geändert werden kann.

Diese Einstellungen können nicht geändert werden, solange die Rechenflotte existiert oder wenn die Anmeldeknoten verwendet werden. Entweder muss die Änderung rückgängig gemacht werden oder die Rechenflotte und die Anmeldeknoten müssen gestoppt werden (die Compute-Flotte kann nicht mehr verwendet werden [pcluster update-compute-fleet](#)). Nachdem die Rechenflotte und die Anmeldeknoten gestoppt wurden, können Sie den Cluster ([pcluster update-cluster](#)) aktualisieren, um die Änderungen zu aktivieren.

Aktualisierungsrichtlinie: Diese Einstellung kann während eines Updates nicht verringert werden.


Diese Einstellungen können geändert, aber nicht verringert werden. Wenn diese Einstellungen verringert werden müssen, müssen Sie die Einstellungen für den ursprünglichen Cluster rückgängig machen und einen neuen Cluster mit den aktualisierten Einstellungen erstellen. Sie können den ursprünglichen Cluster zu einem späteren Zeitpunkt löschen. Um den neuen Cluster zu erstellen, verwenden Sie [pcluster create-cluster](#). Um den ursprünglichen Cluster zu löschen, verwenden Sie [pcluster delete-cluster](#).

Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig. Wenn Sie das Update erzwingen, wird der neue Wert ignoriert und der alte Wert verwendet.

Nach dem Ändern dieser Einstellung kann der Cluster nicht aktualisiert werden. Sie müssen die Einstellungen für den ursprünglichen Cluster rückgängig machen und einen neuen Cluster mit den aktualisierten Einstellungen erstellen. Sie können den ursprünglichen Cluster zu einem späteren Zeitpunkt löschen. Um den neuen Cluster zu erstellen, verwenden Sie [pcluster create-cluster](#). Um den ursprünglichen Cluster zu löschen, verwenden Sie [pcluster delete-cluster](#).

Richtlinie aktualisieren: Die Rechenflotte muss gestoppt oder [QueueUpdateStrategy](#) eingerichtet sein, damit diese Einstellung für ein Update geändert werden kann.

Diese Einstellungen können geändert werden. Entweder muss die Rechenflotte gestoppt (verwendet [pcluster update-compute-fleet](#)) oder sie [QueueUpdateStrategy](#) muss eingerichtet werden. Nachdem die Rechenflotte gestoppt oder [QueueUpdateStrategy](#) eingerichtet wurde, können Sie den Cluster ([pcluster update-cluster](#)) aktualisieren, um die Änderungen zu aktivieren.

 Note

Diese Aktualisierungsrichtlinie wird ab AWS ParallelCluster Version 3.2.0 unterstützt.

Richtlinie aktualisieren: Bei dieser Einstellung für Listenwerte kann während eines Updates ein neuer Wert hinzugefügt werden, oder die Rechenflotte muss gestoppt werden, wenn ein vorhandener Wert entfernt wird.

Ein neuer Wert für diese Einstellungen kann während eines Updates hinzugefügt werden. Nachdem der Liste ein neuer Wert hinzugefügt wurde, kann der Cluster mit ([pcluster update-cluster](#)) aktualisiert werden.

Um einen vorhandenen Wert aus der Liste zu entfernen, muss die Rechenflotte gestoppt werden (mit [pcluster update-compute-fleet](#)).

Wenn Sie beispielsweise einen Slurm Scheduler verwenden und [Instances/ einen neuen Instanztyp](#) hinzufügen, können Sie den Cluster aktualisieren `InstanceType`, ohne die Rechenflotte zu stoppen. [Um einen vorhandenen Instanztyp aus Instances/ zu entfernen InstanceType, muss die Rechenflotte zuerst gestoppt werden \(mithilfe von pcluster\). update-compute-fleet](#)

 Note

Diese Update-Richtlinie wird ab Version 3.2.0 unterstützt. AWS ParallelCluster

Aktualisierungsrichtlinie: Um die Größe einer Warteschlange zu reduzieren, [QueueUpdateStrategy](#) muss die Rechenflotte gestoppt oder auf TERMINATE gesetzt werden, damit diese Einstellung für ein Update geändert werden kann.

Diese Einstellungen können geändert werden, aber wenn die Änderung die Größe der Warteschlange verringern würde, muss die Rechenflotte gestoppt (mit `pcluster update-compute-fleet`) oder auf [QueueUpdateStrategy](#) TERMINATE gesetzt werden. Nachdem die Compute-Flotte gestoppt oder auf [QueueUpdateStrategy](#) TERMINATE gesetzt wurde, können Sie den Cluster aktualisieren (`pcluster update-cluster`), um die Änderungen zu aktivieren.

Wenn bei der Größenänderung der Kapazität des Clusters die Einstellung TERMINATE festgelegt wird, werden nur die Knoten beendet, die sich am Ende der Knotenliste befinden, und alle anderen Knoten derselben Partition bleiben unberührt.

Wenn beispielsweise die Anfangskapazität des Clusters `MinCount = 5` und `istMaxCount = 10`, sind es auch die Knoten `st-[1-5]`; `dy-[1-5]`. Wenn Sie die Größe des Clusters auf `MinCount = 3` und `ändernMaxCount = 5`, wird die neue Clusterkapazität aus den Knoten zusammengesetzt `st-[1-3]`; `dy-[1-2]`, die bei der Aktualisierung nicht verändert werden. Nur die Knoten `st-[4-5]`; `dy-[3-5]` werden während des Updates beendet.

Die folgenden Änderungen werden unterstützt und erfordern weder das Stoppen der Rechenflotte noch die [QueueUpdateStrategy](#) Einstellung TERMINATE:

- Eine neue [SlurmQueue](#) wird hinzugefügt
- Ein neues [ComputeResource](#) wird hinzugefügt
- [MaxCount](#) ist erhöht
- [MinCount](#) ist erhöht und [MaxCount](#) wird um mindestens den gleichen Betrag erhöht

Hinweis: Diese Aktualisierungsrichtlinie wird ab AWS ParallelCluster Version 3.9.0 unterstützt.

Richtlinie aktualisieren: Für diese Einstellung mit Listenwerten muss die Rechenflotte gestoppt oder [QueueUpdateStrategy](#) so eingestellt werden, dass sie einen neuen Wert hinzufügt. Die Rechenflotte muss gestoppt werden, wenn ein vorhandener Wert entfernt wird.

Ein neuer Wert für diese Einstellungen kann während eines Updates hinzugefügt werden. Entweder muss die Rechenflotte gestoppt (verwendet `pcluster update-compute-fleet`) oder sie [QueueUpdateStrategy](#) muss eingerichtet werden. Nachdem die Rechenflotte gestoppt oder [QueueUpdateStrategy](#) eingerichtet wurde, können Sie den Cluster (`pcluster update-cluster`) aktualisieren, um die Änderungen zu aktivieren.

Um einen vorhandenen Wert aus der Liste zu entfernen, muss die Rechenflotte gestoppt (verwendet `pcluster update-compute-fleet`) werden.

Note

Diese Update-Richtlinie wird ab AWS ParallelCluster Version 3.3.0 unterstützt.

Aktualisierungsrichtlinie: Alle Rechenknoten müssen gestoppt werden, damit eine verwaltete Platzierungsgruppe gelöscht werden kann. Die Rechenflotte muss gestoppt oder [QueueUpdateStrategy](#) eingerichtet sein, damit diese Einstellung für ein Update geändert werden kann.

Die Rechenflotte muss gestoppt (verwendet `pcluster update-compute-fleet`) werden, um eine verwaltete Platzierungsgruppe zu entfernen. Wenn Sie ein Cluster-Update ausführen, um eine verwaltete Platzierungsgruppe zu entfernen, bevor Sie die Rechenflotte beenden, wird eine ungültige Konfigurationsmeldung zurückgegeben und das Update wird nicht fortgesetzt. Durch das Stoppen der Rechenflotte wird garantiert, dass keine Instances ausgeführt werden.

Beispiele für `pcluster update-cluster`

Diese Einstellungen können geändert werden, aber wenn die Änderung die Größe der Warteschlange verringern würde, muss die Rechenflotte gestoppt (mit `pcluster update-compute-fleet`) oder auf [QueueUpdateStrategy](#) TERMINATE gesetzt werden. Nachdem die Compute-Flotte gestoppt oder auf [QueueUpdateStrategy](#) TERMINATE gesetzt wurde, können Sie den Cluster aktualisieren (`pcluster update-cluster`), [um die Änderungen zu aktivieren](#).

- Dieses Beispiel zeigt ein Update mit einigen zulässigen Änderungen und das Update wird direkt gestartet.

```
$ pcluster update-cluster --cluster-name cluster_name --cluster-config
~/.parallelcluster/test_cluster --region us-east-1
{
  "cluster": {
    "clusterName": cluster_name,
    "cloudformationStackStatus": "UPDATE_IN_PROGRESS",
    "cloudformationStackArn": stack_arn,
    "region": "us-east-1",
    "version": "3.7.0",
```

```

    "clusterStatus": "UPDATE_IN_PROGRESS"
  },
  "changeSet": [
    {
      "parameter": "HeadNode.Networking.AdditionalSecurityGroups",
      "requestedValue": [
        "sg-0cd61884c4ad11234"
      ],
      "currentValue": [
        "sg-0cd61884c4ad16341"
      ]
    }
  ]
}

```

- Dieses Beispiel zeigt ein Testrun-Update mit einigen erlaubten Änderungen. Dryrun ist nützlich, um den Änderungssatz zu melden, ohne das Update zu starten.

```

$ pcluster update-cluster --cluster-name cluster_name --cluster-config
~/.parallelcluster/test_cluster --region us-east-1 --dryrun true
{
  "message": "Request would have succeeded, but DryRun flag is set.",
  "changeSet": [
    {
      "parameter": "HeadNode.Networking.AdditionalSecurityGroups",
      "requestedValue": [
        "sg-0cd61884c4ad11234"
      ],
      "currentValue": [
        "sg-0cd61884c4ad16341"
      ]
    }
  ]
}

```

- Dieses Beispiel zeigt ein Update mit einigen Änderungen, die das Update blockieren.

```

$ pcluster update-cluster --cluster-name cluster_name --cluster-config
~/.parallelcluster/test_cluster --region us-east-1
{
  "message": "Update failure",
  "updateValidationErrors": [
    {

```

```

    "parameter": "HeadNode.Ssh.KeyName",
    "requestedValue": "mykey_2",
    "message": "Update actions are not currently supported for the 'KeyName'
parameter. Restore 'KeyName' value to 'jenkinsjun'. If you need this change, please
consider creating a new cluster instead of updating the existing one.",
    "currentValue": "mykey_1"
  },
  {
    "parameter": "Scheduling.SlurmQueues[queue1].ComputeResources[queue1-
t2micro].InstanceType",
    "requestedValue": "c4.xlarge",
    "message": "All compute nodes must be stopped. Stop the compute fleet with the
pcluster update-compute-fleet command",
    "currentValue": "t2.micro"
  },
  {
    "parameter": "SharedStorage[ebs1].MountDir",
    "requestedValue": "/my/very/very/long/shared_dir",
    "message": "Update actions are not currently supported for the 'MountDir'
parameter. Restore 'MountDir' value to '/shared'. If you need this change, please
consider creating a new cluster instead of updating the existing one.",
    "currentValue": "/shared"
  }
],
"changeSet": [
  {
    "parameter": "HeadNode.Networking.AdditionalSecurityGroups",
    "requestedValue": [
      "sg-0cd61884c4ad11234"
    ],
    "currentValue": [
      "sg-0cd61884c4ad16341"
    ]
  },
  {
    "parameter": "HeadNode.Ssh.KeyName",
    "requestedValue": "mykey_2",
    "currentValue": "mykey_1"
  },
  {
    "parameter": "Scheduling.SlurmQueues[queue1].ComputeResources[queue1-
t2micro].InstanceType",
    "requestedValue": "c4.xlarge",
    "currentValue": "t2.micro"
  }
]

```

```
    },  
    {  
      "parameter": "SharedStorage[ebs1].MountDir",  
      "requestedValue": "/my/very/very/long/shared_dir",  
      "currentValue": "/shared"  
    }  
  ]  
}
```

AWS ParallelCluster AMI-Anpassung

Es gibt Szenarien, in denen die Erstellung eines benutzerdefinierten AMI für erforderlich ist. In diesem Abschnitt wird beschrieben, was beim Erstellen eines benutzerdefinierten AWS ParallelCluster AMI zu beachten ist.

Sie können ein benutzerdefiniertes AWS ParallelCluster AMI mit einer der folgenden Methoden erstellen:

1. Erstellen Sie eine [Build-Image-Konfigurationsdatei](#) und verwenden Sie dann die `pcluster` CLI, um das Image mit EC2 Image Builder zu erstellen. Dieser Prozess ist automatisiert, wiederholbar und unterstützt die Überwachung. Weitere Informationen finden Sie in den `pcluster` Bildbefehlen.
2. Erstellen Sie eine Instanz aus einem AWS ParallelCluster AMI, melden Sie sich dann bei ihr an und nehmen Sie manuelle Änderungen vor. Verwenden Sie zuletzt Amazon EC2, um ein neues AMI aus der geänderten Instance zu erstellen. Dieser Vorgang nimmt weniger Zeit in Anspruch. Es ist jedoch nicht automatisiert oder wiederholbar und unterstützt die Verwendung der `pcluster` CLI-Bildüberwachungsbefehle nicht.

Weitere Informationen zu diesen Methoden finden Sie unter [Erstellen eines benutzerdefinierten AWS ParallelCluster-AMIs](#).

AWS ParallelCluster Überlegungen zur AMI-Anpassung

Unabhängig davon, wie Sie Ihr benutzerdefiniertes Image erstellen, empfehlen wir Ihnen, vorläufige Validierungstests durchzuführen und Vorkehrungen zur Überwachung des Status des erstellten Images zu treffen.

Um ein benutzerdefiniertes AMI zu erstellen, erstellen Sie eine [Build-Image-Konfigurationsdatei](#) mit einem `BuildImage`-Abschnitt, den [EC2 Image Builder](#) verwendet,

um Ihr benutzerdefiniertes Image zu erstellen. In `Build` diesem Abschnitt wird angegeben, was Image Builder zum Erstellen des Images benötigt. Dazu gehören das [ParentImage](#) (Basis-Image) und [Components](#). Eine [Image Builder Builder-Komponente](#) definiert eine Abfolge von Schritten, die erforderlich sind, um eine Instanz anzupassen, bevor ein Image erstellt wird, oder um eine Instanz zu testen, die mit dem erstellten Image gestartet wurde. Beispiele für AWS ParallelCluster Komponenten finden Sie unter [Benutzerdefinierte AMIs](#). ImageIn diesem Abschnitt werden die Image-Eigenschaften angegeben.

Wenn Image Builder von `pcluster` aus aufgerufen wird, [build-image](#) um ein benutzerdefiniertes Image zu erstellen, verwendet es die Build-Image-Konfiguration mit dem AWS ParallelCluster Cookbook, um Ihr Image zu booten. AWS ParallelCluster [ParentImage](#) Image Builder lädt Komponenten herunter, führt Build- und Validierungsphasen aus, erstellt das AMI, startet eine Instance aus dem AMI und führt Tests durch. Wenn der Vorgang abgeschlossen ist, erzeugt Image Builder ein neues Image oder eine Stopfnachricht.

Führen Sie Validierungstests für benutzerdefinierte Komponenten durch

Bevor Sie eine Image Builder Builder-Komponente in eine Konfiguration aufnehmen, testen und validieren Sie sie mit einer der folgenden Methoden. Da der Image Builder Builder-Vorgang bis zu 1 Stunde dauern kann, empfehlen wir, die Komponenten vorher zu testen. Dadurch können Sie eine Menge Zeit sparen.

Skriptkoffer

Testen Sie das Skript in einer laufenden Instanz außerhalb des Build-Image-Prozesses und stellen Sie sicher, dass das Skript mit dem Exit-Code 0 beendet wird.

Fall Amazon Resource Name (ARN)

Testen Sie das Komponentendokument in einer laufenden Instance außerhalb des Build-Image-Prozesses. Eine Liste der Anforderungen finden Sie unter [Component Manager](#) im Image Builder Builder-Benutzerhandbuch.

Fügen Sie die Komponente nach erfolgreicher Überprüfung zu Ihrer Build-Image-Konfiguration hinzu

Nachdem Sie sich vergewissert haben, dass die benutzerdefinierte Komponente funktioniert, fügen Sie sie der [Build-Image-Konfigurationsdatei](#) hinzu.

Überwachen Sie den Image Builder Builder-Prozess mit **pcluster** Befehlen, die beim Debuggen helfen

[describe-image](#)

Verwenden Sie diesen Befehl, um den Status des Build-Images zu überwachen.

[list-image-log-streams](#)

Verwenden Sie diesen Befehl, um die IDs der Protokollstreams abzurufen, mit denen Sie Protokollereignisse abrufen können [get-image-log-events](#).

[get-image-log-events](#)

Verwenden Sie diesen Befehl, um den Protokollstream der Build-Image-Prozessereignisse abzurufen.

Mit dem folgenden Befehl können Sie beispielsweise Build-Image-Ereignisse nachverfolgen.

```
$ watch -n 1 'pcluster get-image-log-events -i <image-id> \
  --log-stream-name/1 <pcluster-version> \
  --query "events[*].message" | tail -n 50'
```

[get-image-stack-events](#)

Verwenden Sie diesen Befehl, um Image-Stack-Ereignisse für den Stapel abzurufen, den Image Builder erstellt.

[export-image-logs](#)

Verwenden Sie diesen Befehl, um Image-Logs zu speichern.

Weitere Informationen zu AWS ParallelCluster Logs und Amazon CloudWatch finden Sie unter [Amazon CloudWatch Logs erstellt Image-Logs](#) und [CloudWatchAmazon-Dashboard](#).

Weitere Überlegungen

Neue AWS ParallelCluster Versionen und benutzerdefinierte AMIs

Wenn Sie ein benutzerdefiniertes AMI erstellen und verwenden, müssen Sie die Schritte, die Sie zur Erstellung Ihres benutzerdefinierten AMI verwendet haben, mit jeder neuen AWS ParallelCluster Version wiederholen.

Benutzerdefinierte Bootstrap-Aktionen

Lesen Sie den [Benutzerdefinierte Bootstrap-Aktionen](#) Abschnitt, um festzustellen, ob die Änderungen, die Sie vornehmen möchten, skriptgesteuert werden können und in future AWS ParallelCluster Versionen unterstützt werden können.

Verwenden von benutzerdefinierten AMIs

Sie können benutzerdefinierte AMIs in der Cluster-Konfiguration in den [CustomAmi](#) Abschnitten [Image/CustomAmi](#) und [SchedulingSlurmQueues//Name/- Image/](#) angeben.

Informationen zur Fehlerbehebung bei benutzerdefinierten AMI-Validierungswarnungen finden Sie unter [Behebung von Problemen mit benutzerdefinierten AMIs](#).

Erstellen Sie Instances mit ODCR Starten Sie Instances

Mit [On-Demand-Kapazitätsreservierungen](#) (On-Demand-Kapazitätsreservierungen Erstellen Sie Kapazitätsreservierungen unabhängig von den in [Savings Plans](#) oder [regionalen angebotenen Reserved Instances](#)

Sie können targeted On-Demand-Kapazitätsreservierungen (ODCRs) konfigurieren. Offene ODCRs decken alle Instanzen ab, die den ODCR-Attributen entsprechen. Diese Attribute Instance-Typ, Plattform und Availability Zone. Sie müssen in der Clusterkonfiguration explizit gezielte ODCRs definieren. Um festzustellen, ob ein ODCR open oder ist targeted, führen Sie den Befehl AWS CLI EC2 [describe-capacity-reservation](#) aus.

Sie können auch eine ODCR in einer Cluster-Platzierungsgruppe erstellen, die als Cluster Placement [Group On-Demand-Capacity-Reservation \(CPG-ODCR\)](#) bezeichnet wird.

In einer Ressourcengruppe können mehrere ODCRs gruppiert werden. Dies kann in der Cluster-Konfigurationsdatei definiert werden. Weitere Informationen Informationen Sie unter [Was sind Ressourcengruppen](#) im Benutzerhandbuch für Resource Groups und Tags.

Verwenden von ODCR mit AWS ParallelCluster

AWS ParallelCluster unterstützt offene ODCRs. Wenn Sie ein offenes ODC AWS ParallelCluster Instanzen werden automatisch für den Cluster ausgewählt. Sie können eine vorhandene Platzierungsgruppe angeben oder Sie AWS ParallelCluster erstellen lassen.

ODCR in der Clusterkonfiguration

Ab AWS ParallelCluster Version 3.3.0 können Sie ODCRs in der Cluster-Konfigurationsdatei definieren, ohne dass Sie EC2-Run-Instance-Overrides angeben müssen.

Sie beginnen mit der Erstellung von [Kapazitätsreservierungen](#) und [Ressourcengruppen](#), indem Sie die Methoden verwenden, die in der jeweiligen verlinkten Dokumentation beschrieben sind. Sie müssen die AWS CLI Methoden verwenden, um Kapazitätsreservierungsgruppen zu erstellen. Wenn Sie die verwenden AWS Management Console, können Sie nur tagbasierte oder stapelbasierte Ressourcengruppen erstellen. Tag-basierte und stapelbasierte Ressourcengruppen werden von AWS ParallelCluster oder von ihnen nicht unterstützt, AWS CLI wenn Instances mit Kapazitätsreservierungen gestartet werden.

Nachdem die Kapazitätsreservierungen und Ressourcengruppen erstellt wurden, geben Sie sie in [SlurmQueues/CapacityReservationTarget](#) oder [SlurmQueues/ComputeResources](#)/an, [CapacityReservationTarget](#) wie in der folgenden Beispielclusterkonfiguration gezeigt. Ersetzen Sie rot hervorgehobene *Werte* durch Ihre gültigen Werte.

```
Image:
  Os: os
HeadNode:
  InstanceType: head_node_instance
  Networking:
    SubnetId: public_subnet_id
  Ssh:
    KeyName: key_name
Scheduling:
  Scheduler: scheduler
SlurmQueues:
  - Name: queue1
    Networking:
      SubnetIds:
        - private_subnet_id
ComputeResources:
  - Name: cr1
    Instances:
      - InstanceType: instance
    MaxCount: max_queue_size
    MinCount: max_queue_size
    Efa:
      Enabled: true
    CapacityReservationTarget:
```

CapacityReservationResourceGroupArn: *capacity_reservation_arn*

VERALTET/NICHT EMPFOHLEN — Gezieltes ODCR mit EC2-Instance-Overrides

Warning

- Ab AWS ParallelCluster Version 3.3.0 empfehlen wir diese Methode nicht. Dieser Abschnitt dient weiterhin als Referenz für Implementierungen, die frühere Versionen verwenden.
- Diese Methode ist nicht mit der Zuweisung mehrerer Instanztypen mit Slurm kompatibel.

Support für `targeted` ODCRs wurde in AWS ParallelCluster 3.1.1 hinzugefügt. In dieser Version wurde ein Mechanismus eingeführt, der `RunInstances` EC2-Parameter außer Kraft setzt und Informationen über die Reservierung weitergibt, die für jede konfigurierte Rechenressource verwendet werden sollen. AWS ParallelCluster Dieser Mechanismus ist mit `targeted` ODCRs kompatibel. Wenn Sie `targeted` ODCRs verwenden, müssen Sie jedoch die `run-instances` Override-Konfiguration angeben. Gezielte ODCRs müssen im AWS CLI `run-instances` EC2-Befehl explizit definiert werden. Um festzustellen, ob es sich um ein `open` ODCR handelt, `targeted` führen Sie den AWS CLI EC2-Befehl aus. [describe-capacity-reservation](#)

In einer Ressourcengruppe können mehrere ODCRs gruppiert werden. Dies kann beim Override von `Run-Instances` verwendet werden, um mehrere ODCRs gleichzeitig anzuvisieren.

Wenn Sie ein `targeted` ODCR verwenden, können Sie eine Platzierungsgruppe angeben. Sie müssen jedoch auch eine `run-instances` Override-Konfiguration angeben.

Angenommen, Sie haben ein `targeted` ODCR für Sie AWS erstellt oder Sie haben einen bestimmten Satz von `Reserved Instances`. Dann können Sie keine Platzierungsgruppe angeben. Die Regeln, die von konfiguriert werden, AWS können mit der Einstellung für die Platzierungsgruppe in Konflikt stehen. Wenn also eine `Vermittlungsgruppe` für Ihre Bewerbung erforderlich ist, verwenden Sie ein [CPG-ODCR](#). In beiden Fällen müssen Sie auch die `run-instances` Override-Konfiguration angeben.

Wenn Sie ein `CPG-ODCR` verwenden, müssen Sie die `run-instances` Override-Konfiguration und dieselbe Platzierungsgruppe in der Cluster-Konfiguration angeben.

Verwenden von Reserved Instances mit AWS ParallelCluster

Reserved Instances [unterscheiden sich](#) von Capacity Reservations (ODCR). Es gibt [zwei Arten von](#) Reserved Instances. Eine Regional Reserved Instance reserviert keine Kapazität. Ein zonaler Reserved Instance reserviert Kapazität in der angegebenen Availability Zone.

Wenn Sie über Regional Reserved Instances verfügen, gibt es keine Kapazitätsreservierung und es kann zu Fehlern bei unzureichender Kapazität kommen. Wenn Sie zonale Reserved Instances haben, haben Sie eine Kapazitätsreservierung, aber es gibt keine `run-instances` API-Parameter, mit denen Sie sie spezifizieren können.

Reserved Instances werden von jeder AWS ParallelCluster Version unterstützt. Sie müssen nichts angeben AWS ParallelCluster und die Instanzen werden automatisch ausgewählt.

Wenn Sie zonale Reserved Instances verwenden, können Sie potenzielle Fehler bei unzureichender Kapazität vermeiden, indem Sie die Platzierungsgruppenspezifikation in der Clusterkonfiguration weglassen.

VERALTET/NICHT EMPFOHLEN — Verwendung der **RunInstances** Anpassung in AWS ParallelCluster 3 für **targeted** On-Demand-Kapazitätsreservierungen (ODCRs)

Warning

- Ab AWS ParallelCluster Version 3.3.0 empfehlen wir diese Methode nicht. Dieser Abschnitt dient weiterhin als Referenz für Implementierungen, die frühere Versionen verwenden.
- Diese Methode ist nicht mit der Zuweisung mehrerer Instanztypen mit Slurm kompatibel.

Sie können `RunInstances` EC2-Parameter für jede Rechenressource überschreiben, die in einer Cluster-Warteschlange konfiguriert ist. Erstellen Sie dazu die `/opt/slurm/etc/pccluster/run_instances_overrides.json` Datei auf dem Hauptknoten des Clusters mit dem folgenden Codefragmentinhalt:

- `{queue_name}` ist der Name der Warteschlange, auf die Sie Überschreibungen anwenden möchten.
- `{compute_resource_name}` ist die Rechen-Ressource, auf die Sie Überschreibungen anwenden möchten.
- `{overrides}` ist ein beliebiges JSON-Objekt, das eine Liste von `RunInstances` Overrides enthält, die für die spezifische Kombination aus Warteschlange und Instanztyp verwendet werden

können. Die Overrides-Syntax muss denselben Spezifikationen entsprechen, die in einem boto3-Aufruf von [run_instances](#) dokumentiert sind.

```
{
  "${queue_name}": {
    "${compute_resource_name}": {
      ${overrides}
    },
    ...
  },
  ...
}
```

Der folgende JSON konfiguriert beispielsweise die ODCR-Gruppe so, dass `group_arn` sie für `p4d.24xlarge` Instanzen verwendet wird, die in `my-queue` und konfiguriert sind. `my-compute-resource`

```
{
  "my-queue": {
    "my-compute-resource": {
      "CapacityReservationSpecification": {
        "CapacityReservationTarget": {
          "CapacityReservationResourceGroupArn": "group_arn"
        }
      }
    }
  }
}
```


Nachdem diese JSON-Datei generiert wurde, verwenden die AWS ParallelCluster Daemons, die für die Clusterskalierung verantwortlich sind, automatisch die Override-Konfiguration für Instanzstarts. Um zu überprüfen, ob die angegebenen Parameter für die Instanzbereitstellung verwendet werden, sehen Sie sich die folgenden Protokolldateien an:

- `/var/log/parallelcluster/clustermgtd`(für statische Kapazität)
- `/var/log/parallelcluster/slurm_resume.log`(für dynamische Kapazität)

Wenn die Parameter korrekt sind, finden Sie einen Logeintrag, der Folgendes enthält:

```
Found RunInstances parameters override. Launching instances with: <parameters_list>
```


VERALTET/NICHT EMPFOHLEN — Erstellen Sie einen Cluster mit **targeted** On-Demand-Kapazitätsreservierungen (ODCRs)

 Warning

- Ab AWS ParallelCluster Version 3.3.0 empfehlen wir diese Methode nicht. Dieser Abschnitt dient weiterhin als Referenz für Implementierungen, die frühere Versionen verwenden.
- Diese Methode ist nicht kompatibel mit [Zuweisung mehrerer Instanztypen mit Slurm](#).

1. Erstellen Sie eine Ressourcengruppe, um die Kapazität zu gruppieren.

```
$ aws resource-groups create-group --name EC2CRGroup \  
  --configuration '{"Type":"AWS::EC2::CapacityReservationPool"}'  
  '{"Type":"AWS::ResourceGroups::Generic", "Parameters": [{"Name": "allowed-  
resource-types", "Values": ["AWS::EC2::CapacityReservation"]}]}'
```

 Note

Eine Ressourcengruppe unterstützt keine Ressourcen, die von anderen Konten gemeinsam genutzt werden.

Wenn das Ziel-ODCR von einem anderen Konto geteilt wird, müssen Sie keine Ressourcengruppe erstellen. Verwenden Sie in Schritt 3 `CapacityReservationId` anstelle einer Ressourcengruppe.

```
#!/bin/bash  
set -e  
  
# Override run_instance attributes  
cat > /opt/slurm/etc/pcluster/run_instances_overrides.json << EOF  
{  
  "my-queue": {  
    "my-compute-resource": {  
      "CapacityReservationSpecification": {  
        "CapacityReservationTarget": {  
          "CapacityReservationId": "cr-abcdef01234567890"  
        }  
      }  
    }  
  }  
}
```



```

    }
  }
}
EOF

```

Fügen Sie der Ressourcengruppe Kapazitätsreservierungen hinzu. Jedes Mal, wenn Sie ein neues ODCR erstellen, fügen Sie es der Gruppenreservierung hinzu. Ersetzen Sie es **ACCOUNT_ID** durch Ihre Konto-ID, **PLACEHOLDER_CAPACITY_RESERVATION** durch Ihre Kapazitätsreservierungs-ID und **REGION_ID** durch Ihre AWS-Region ID (z. B. us-east-1).

```

$ aws resource-groups group-resources --region REGION_ID --group EC2CRGroup \
  --resource-arns arn:aws:ec2:REGION_ID:ACCOUNT_ID:capacity-
reservation/PLACEHOLDER_CAPACITY_RESERVATION

```

Erstellen Sie auf Ihrem lokalen Computer Erstellen Sie auf Ihrem lokalen Computer Ersetzen Sie **ACCOUNT_ID** durch Ihre Konto-ID und **REGION_ID** durch Ihre AWS-Region ID (z. B. us-east-1).

```

cat > policy.json << EOF
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "RunInstancesInCapacityReservation",
      "Effect": "Allow",
      "Action": "ec2:RunInstances",
      "Resource": [
        "arn:aws:ec2:REGION_ID:ACCOUNT_ID:capacity-reservation/*",
        "arn:aws:resource-groups:REGION_ID:ACCOUNT_ID:group/*"
      ]
    }
  ]
}
EOF

```

- Erstellen Sie die IAM-Richtlinie auf Ihrem Computer AWS-Konto mithilfe der von Ihnen erstellten JSON-Datei.

```
$ aws iam create-policy --policy-name RunInstancesCapacityReservation --policy-document file://policy.json
```

- Erstellen Sie das folgende Post-Install-Skript lokal auf der Instanz und benennen Sie **espostinstall.sh**.

Ersetzen Sie **ACCOUNT_ID** durch Ihre AWS-Konto ID und **REGION_ID** durch Ihre AWS-Region ID (z. B. us-east-1).

```
#!/bin/bash
set -e

# Override run_instance attributes
cat > /opt/slurm/etc/pcluster/run_instances_overrides.json << EOF
{
  "my-queue": {
    "my-compute-resource": {
      "CapacityReservationSpecification": {
        "CapacityReservationTarget": {
          "CapacityReservationResourceGroupArn": "arn:aws:resource-
groups:REGION_ID:ACCOUNT_ID:group/EC2CRGroup"
        }
      }
    }
  }
}
EOF
```

Laden Sie die Datei hoch in einen Amazon S3 Bucket hoch. Ersetzen Sie **S3_NAME_BUCKET** durch Ihren spezifischen S3-Bucket-Namen.

```
$ aws s3 mb s3://S3_NAME_BUCKET
aws s3 cp postinstall.sh s3://S3_NAME_BUCKET/postinstall.sh
```

- Erstellen Sie die lokale Cluster-Konfiguration und ersetzen Sie die Platzhalter durch Ihre eigenen Werte.

```
Region: REGION_ID
Image:
  Os: alinux2
HeadNode:
```

```

InstanceType: c5.2xlarge
Ssh:
  KeyName: YOUR_SSH_KEY
Iam:
  S3Access:
    - BucketName: S3_NAME_BUCKET
  AdditionalIamPolicies:
    - Policy: arn:aws:iam::ACCOUNT_ID:policy/RunInstancesCapacityReservation
## This post-install script is executed after the node is configured.
## It is used to install scripts at boot time and specific configurations
## In the script below we are overriding the calls to RunInstance to force
## the provisioning of our my-queue partition to go through
## the On-Demand Capacity Reservation
CustomActions:
  OnNodeConfigured:
    Script: s3://S3_NAME_BUCKET/postinstall.sh
Networking:
  SubnetId: YOUR_PUBLIC_SUBNET_IN_TARGET_AZ

Scheduling:
  Scheduler: slurm
  SlurmQueues:
    - Name: my-queue
      ComputeResources:
        - MinCount: 0
          MaxCount: 100
          InstanceType: p4d.24xlarge
          Name: my-compute-resource
          Efa:
            Enabled: true
      Networking:
        ## PlacementGroup:
        ##   Enabled: true ## Keep PG disabled if using targeted ODCR
        SubnetIds:
          - YOUR_PRIVATE_SUBNET_IN_TARGET_AZ

```

5. Erstellen Sie den Cluster.

Verwenden Sie den folgenden Befehl Erstellen Ersetzen Sie *cluster-config.yaml* durch den Namen Ihrer Konfigurationsdatei, *cluster-dl* durch Ihren Clusternamen und *REGION_ID* durch Ihre Regions-ID (z. B. us-east-1).

```
$ pcluster create-cluster --cluster-configuration cluster-config.yaml --cluster-name cluster-dl --region REGION_ID
```

Nachdem der Cluster erstellt wurde, wird das Skript nach der Installation im Hauptknoten ausgeführt. Das Skript erstellt die `run_instances_overrides.json` Datei und überschreibt die Aufrufe von, um `RunInstances` zu erzwingen, dass die Bereitstellung der Partition über die On-Demand-Kapazitätsreservierung erfolgt.

Die AWS ParallelCluster Daemons, die für die Clusterskalierung verantwortlich sind, verwenden diese Konfiguration automatisch für neue Instanzen, die gestartet werden. Um zu überprüfen, ob die angegebenen Parameter für die Bereitstellung von Instanzen verwendet werden, können Sie sich die folgenden Protokolldateien ansehen:

- `/var/log/parallelcluster/clustermgtd`(für statische Kapazität - [MinCount](#) > 0)
- `/var/log/parallelcluster/slurm_resume.log`(für dynamische Kapazität)

Wenn die Parameter korrekt sind, finden Sie einen Logeintrag, der Folgendes enthält.

```
Found RunInstances parameters override. Launching instances with: <parameters_list>
```

RunInstances Überschreibungen aktualisieren

Sie können die generierte JSON-Konfiguration jederzeit aktualisieren, ohne die Rechenflotte anzuhalten. Nachdem die Änderungen übernommen wurden, werden alle neuen Instances mit der aktualisierten Konfiguration gestartet. Wenn Sie die aktualisierte Konfiguration auf laufende Knoten anwenden müssen, recyceln Sie die Knoten, indem Sie eine Instanzbeendigung erzwingen und warten Sie, AWS ParallelCluster bis diese Knoten ersetzt sind. Sie können dies tun, indem Sie die Instance von der EC2-Konsole aus beenden oder AWS CLI indem Sie die Slurm-Knoten in den Status oder versetzen. DOWN DRAIN

Verwenden Sie den folgenden Befehl, um den Slurm Knoten auf DOWN oder zu setzen DRAIN.

```
$ scontrol update nodename=my-queue-dy-my-compute-resource-1 state=down  
reason=your_reason  
scontrol update nodename=my-queue-dy-my-compute-resource-1 state=drain  
reason=your_reason
```

AMI-Patching und Austausch von EC2-Instances

AWS ParallelCluster deaktiviert die automatischen Betriebssystemupdates der Clusterinstanz, um sicherzustellen, dass sich alle dynamisch gestarteten Cluster-Rechenknoten konsistent verhalten. Darüber hinaus AWS ParallelCluster wird für jede Version von AWS ParallelCluster und die zugehörige CLI ein bestimmter Satz von AMIs erstellt. Dieser spezifische Satz von AMIs bleibt unverändert und wird nur von der AWS ParallelCluster Version unterstützt, für die sie erstellt wurden. AWS ParallelCluster AMIs für veröffentlichte Versionen werden nicht aktualisiert.

Aufgrund neuer Sicherheitsprobleme möchten Kunden diesen AMIs jedoch möglicherweise Patches hinzufügen und dann ihre Cluster mit dem gepatchten AMI aktualisieren. Dies entspricht dem [Modell der AWS ParallelCluster geteilten Verantwortung](#).

Führen Sie Folgendes aus, um die spezifischen AWS ParallelCluster AMIs anzuzeigen, die von der derzeit verwendeten AWS ParallelCluster CLI-Version unterstützt werden:

```
$ pcluster version
$ pcluster list-official-images
```

Der AWS ParallelCluster Head-Knoten ist eine statische Instanz und Sie können ihn manuell aktualisieren. Der Neustart und Neustart des Headknotens werden ab AWS ParallelCluster Version 3.0.0 vollständig unterstützt.

Wenn Ihre Instances temporäre Instanzspeicher haben, müssen Sie daran denken, Instanzspeicherdaten vor manuellen Aktualisierungen zu speichern. Weitere Informationen finden Sie im Amazon EC2 EC2-Benutzerhandbuch für Linux-Instances [unter HeadNodeLocalStorage//EphemeralVolumeCluster-Konfiguration und Instanztypen mit Instanzspeicher-Volumes](#).

Die Datenverarbeitungsknoten sind kurzlebige Instanzen. Standardmäßig können Sie nur vom Kopfknoten aus auf sie zugreifen. Ab AWS ParallelCluster Version 3.0.0 können Sie das mit Compute-Instances verknüpfte AMI aktualisieren, indem Sie den [CustomAmiParameter Scheduling//SlurmQueuesImage](#) ändern und den [pcluster update-cluster](#) Befehl ausführen, nachdem Sie die Compute-Flotte mit dem folgenden Befehl gestoppt haben [pcluster update-compute-fleet](#):

```
$ pcluster update-compute-fleet-status --status STOP_REQUESTED
```

Es ist möglich, die Erstellung eines aktualisierten benutzerdefinierten AMI für die Rechenknoten zu automatisieren, indem Sie eine der folgenden Methoden verwenden:

- Verwenden Sie den `pcluster build-image` Befehl mit einem aktualisierten [Build/ParentImage](#).
- Führe den Build mit `Build/UpdateOsPackages/Enabled: atrue`.

Aktualisierung oder Ersatz der Head-Knoteninstanz

Unter bestimmten Umständen müssen Sie möglicherweise den Head-Knoten neu starten oder neu starten. Dies ist beispielsweise erforderlich, wenn Sie das Betriebssystem manuell aktualisieren oder wenn eine [AWSInstance geplant](#) ist, die einen Neustart der Head-Node-Instance erfordert.

Wenn die Instance nicht hat, können Sie sie und Sie können sie jederzeit und neu starten. Im Falle einer geplanten Stilllegung wird die gestoppte Instance durch das Starten auf die neue Hardware migriert.

In ähnlicher Weise können Sie eine Instance, die keine Instanzspeicher hat, manuell stoppen und starten. Für diesen Fall und für andere Fälle von Fällen ohne vergängliche Bänder fahren Sie fort [und Starten und Starten und Starten und Starten und Starten](#).

Wenn die Instance über die Instance hat und sie angehalten und die Instance hat sie angehalten und die Instance ist die Instance angehalten und die Instance hat sie angehalten und die Instance hat sie Sie können anhand der Tabelle unter [Instanzspeichervolumen feststellen, ob der für den Hauptknoten verwendete Instanztyp über Instanzspeicher](#) verfügt.

Daten von kurzlebigen Laufwerken speichern

Ab AWS ParallelCluster Version 3.0.0 werden der Neustart und der Neustart des Headknotens für jeden Instanztyp vollständig unterstützt. Wenn Instanzen jedoch ein kurzlebiges Laufwerk haben, gehen deren Daten verloren. Folgen Sie den nächsten Schritten, um Ihre Daten vor einem Neustart oder Neustart des Headknotens beizubehalten.

Um zu überprüfen, ob Sie Daten haben, die aufbewahrt werden müssen, sehen Sie sich den Inhalt im `MountDir` Ordner `EphemeralVolume` an (/scratchstandardmäßig).

Sie können die Daten auf das Root-Volume oder die an den Cluster angeschlossenen Shared-Storage-Systeme wie Amazon FSx, Amazon EFS oder Amazon EBS übertragen. Beachten Sie, dass für die Datenübertragung zum Remote-Speicher zusätzliche Kosten anfallen können.

Fahren Sie nach dem Speichern der Daten fort [und Starten und Starten und Starten und Starten und Starten](#).

und Starten und Starten und Starten und Starten und Starten

1. Stellen Sie sicher, dass im Cluster keine laufenden Jobs vorhanden sind.

Wenn Sie einen Slurm Scheduler verwenden:

- Wenn die `batch --no-requeue` Option nicht angegeben ist, werden laufende Jobs in die Warteschlange gestellt.
- Wenn die `--no-requeue` Option angegeben ist, schlagen die laufenden Jobs fehl.

2. Fordern Sie einen Cluster-Compute-Flottenstopp an:

```
$ pcluster update-compute-fleet --cluster-name cluster-name --status STOP_REQUESTED
{
  "status": "STOP_REQUESTED",
  ...
}
```

3. Warten Sie, bis der Status der Rechenflotte wie folgt lautet STOPPED:

```
$ pcluster update-compute-fleet --cluster-name cluster-name --status STOP_REQUESTED
{
  "status": "STOPPED",
  ...
}
```

4. Für manuelle Updates mit einem Betriebssystemneustart oder Instanzneustart können Sie die AWS Management Console oder verwenden AWS CLI. Das Folgende ist ein Beispiel für die Verwendung von AWS CLI.

```
# Retrieve head node instance id
$ pcluster describe-cluster --cluster-name cluster-name --status STOP_REQUESTED
{
  "headNode": {
    "instanceId": "i-1234567890abcdef0",
    ...
  },
  ...
}
```

```
# stop and start the instance
$ aws ec2 stop-instances --instance-ids 1234567890abcdef0
{
  "StoppingInstances": [
    {
      "CurrentState": {
        "Name": "stopping"
        ...
      },
      "InstanceId": "i-1234567890abcdef0",
      "PreviousState": {
        "Name": "running"
        ...
      }
    }
  ]
}
$ aws ec2 start-instances --instance-ids 1234567890abcdef0
{
  "StartingInstances": [
    {
      "CurrentState": {
        "Name": "pending"
        ...
      },
      "InstanceId": "i-1234567890abcdef0",
      "PreviousState": {
        "Name": "stopped"
        ...
      }
    }
  ]
}
```

5. Starten Sie die Cluster-Rechenflotte:

```
$ pcluster update-compute-fleet --cluster-name cluster-name --status
START_REQUESTED
{
  "status": "START_REQUESTED",
  ...
}
```


Betriebssysteme

AWS ParallelCluster unterstützt Amazon Linux 2, CentOS 7, Ubuntu 22.04, Ubuntu 2004, Red Hat Enterprise Linux 8 (RHEL8), Rocky 8, Red Hat Enterprise Linux 9 (RHEL9), Rocky 9. AWS ParallelCluster bietet vorgefertigte AMIs für ausgewählte Betriebssysteme. Weitere Informationen zu AMIs finden Sie unter [AWS ParallelCluster Image Abschnitt](#)

Überlegungen zum Betriebssystem

Ubuntu 22.04

Ubuntu 22.04 benötigt sicherere Schlüssel für SSH und unterstützt standardmäßig keine RSA-Schlüssel. Bitte generieren Sie einen ed25519-Schlüssel und verwenden Sie ihn für die Clustererstellung.

Ubuntu 2204 kann nicht auf den neuesten Kernel aktualisiert werden, da es für diesen Kernel keinen Fsx-Client gibt.

RHEL 8

RedHat Enterprise Linux 8.7 (rhel8) wird ab Version 3.6.0 hinzugefügt. AWS ParallelCluster Wenn Sie Ihren Cluster für die Verwendung von rhel8 konfigurieren, sind die On-Demand-Kosten für jeden Instance-Typ höher als bei der Konfiguration Ihres Clusters für die Verwendung anderer unterstützter Betriebssysteme.

Weitere Informationen zur Preisgestaltung finden Sie unter [On-Demand-Preise](#) und [Wie wird Red Hat Enterprise Linux auf Amazon EC2 angeboten und wie hoch der Preis ist?](#) .

Felsig 8

AWS ParallelCluster 3.8.0 unterstützt Rocky Linux 8, aber vorgefertigte Rocky Linux 8-AMIs (für x86- und ARM-Architekturen) sind nicht verfügbar. AWS ParallelCluster 3.8.0 unterstützt das Erstellen von Clustern mit Rocky Linux 8 mithilfe von benutzerdefinierten AMIs mithilfe der Eigenschaft [CustomAmi](#). Weitere Informationen zum Erstellen benutzerdefinierter AMIs finden Sie unter [AWS ParallelCluster AMI-Anpassung](#)

Um Ihr benutzerdefiniertes AMI auf der Grundlage eines Rocky Linux 8-AMI zu erstellen, können Sie erwägen, die auf dem AWS [Marketplace](#) verfügbaren [Rocky Linux 8-AMIs](#) zu abonnieren. Informieren Sie sich auf jeden Fall über die Preise und Abonnementkosten für Rocky Linux 8-AMIs auf dem

AWS Marketplace. Alternativ können Sie auch die [offiziellen Rocky Linux 8 AMIs](#) als Ihr Basis-AMI verwenden.

Centos 7

[Gdrcopy](#) hat Centos7 aus ihrer Betriebssystem-Supportmatrix entfernt. Das bedeutet, dass gdrcopy 2.3.1 die neueste Version ist, die dieses Betriebssystem unterstützt. Sie müssen die NVIDIA- und gdrcopy-Versionen für Centos7 anheften, da die neuesten NVIDIA-Open-Source-Treiberversionen (OpenRM, d. h. 535.129.03+) mit dieser Version von gdrcopy nicht kompatibel sind. Ab ParallelCluster 3.8.0 werden unsere offiziellen Centos7-AMIs mit gdrcopy 2.3.1 und dem NVIDIA-Treiber 535.129.03 veröffentlicht.

Rocky 9

AWS ParallelCluster 3.9.0 unterstützt Rocky Linux 9, aber vorgefertigte Rocky Linux 9-AMIs (für x86- und ARM-Architekturen) sind nicht verfügbar. AWS ParallelCluster 3.9.0 unterstützt das Erstellen von Clustern mit Rocky Linux 9 mithilfe von benutzerdefinierten AMIs mithilfe der Eigenschaft.

[CustomAmi](#) Weitere Informationen zum Erstellen benutzerdefinierter AMIs finden Sie unter [AWS ParallelCluster AMI-Anpassung](#). Um Ihr benutzerdefiniertes AMI aus einem Basis-Rocky Linux 9-AMI zu erstellen, können Sie auch die [offiziellen Rocky Linux 9-AMIs](#) als Ihr Basis-AMI verwenden. Der benutzerdefinierte Rocky Linux 9-AMI-Build schlägt möglicherweise fehl, wenn das Basis-AMI nicht über den neuesten Kernel verfügt. Um den Kernel vor dem Erstellen des AMI zu aktualisieren:

- [Starten Sie von hier aus eine Instance mit einer Rocky9-AMI-ID: https://rockylinux.org/cloud-images/](https://rockylinux.org/cloud-images/)
- Rufen Sie die Instanz per SSH auf und führen Sie den folgenden Befehl aus: `sudo yum -y update`
- Erstellen Sie ein Image aus der Instanz, das Sie verwenden möchten ParentImage

Referenz für AWS ParallelCluster

Themen

- [AWS ParallelCluster-CLI-Befehle](#)
- [Konfigurationsdateien](#)
- [AWS ParallelCluster-API-Referenz](#)
- [AWS ParallelClusterPython-Bibliothek-API](#)

AWS ParallelCluster-CLI-Befehle

`pcluster` ist der primäre AWS ParallelCluster CLI-Befehl. Sie verwenden `pcluster`, um HPC-Cluster in der zu starten und zu verwalten AWS Cloud und benutzerdefinierte AMI-Images zu erstellen und zu verwalten.

`pcluster3-config-converter` wird verwendet, um Clusterkonfigurationen im Format AWS ParallelCluster Version 2 in das AWS ParallelCluster Version 3-Format zu konvertieren.

```
pcluster [-h] ( build-image | configure |
    create-cluster | dcv-connect |
    delete-cluster | delete-cluster-instances | delete-image |
    describe-cluster | describe-cluster-instances |
    describe-compute-fleet | describe-image |
    export-cluster-logs | export-image-logs |
    get-cluster-log-events | get-cluster-stack-events |
    get-image-log-events | get-image-stack-events |
    list-cluster-log-streams | list-clusters |
    list-images | list-image-log-streams | list-official-images |
    ssh | update-cluster |
    update-compute-fleet | version ) ...
pcluster3-config-converter [-h] [-t CLUSTER_TEMPLATE]
    [-c CONFIG_FILE]
    [--force-convert]
    [-o OUTPUT_FILE]
```

Themen

- [pcluster](#)
- [pcluster3-config-converter](#)

pcluster

`pcluster` ist der primäre AWS ParallelCluster CLI-Befehl. Sie verwenden `pcluster`, um HPC-Cluster in der zu starten und zu verwalten. AWS Cloud

`pcluster` schreibt Protokolle Ihrer Befehle in `pcluster.log.#` Dateien in `/home/user/.parallelcluster/`. Weitere Informationen finden Sie unter [pcluster CLI-Protokolle](#).

Um sie verwenden zu können `pcluster`, benötigen Sie eine IAM-Rolle mit den [für deren Ausführung erforderlichen Berechtigungen](#).

```
pcluster [-h]
```

Argumente

`pcluster` *command*

Mögliche Wahlmöglichkeiten: [build-image](#)[configure](#)[create-cluster](#)[dcv-connect](#)[delete-cluster](#)[delete-cluster-instances](#)[delete-image](#)[describe-cluster](#)[describe-cluster-instances](#)[describe-compute-fleet](#)[describe-image](#)[export-cluster-log](#)[export-image-log](#)[get-cluster-log-events](#)[get-cluster-stack-events](#)[get-image-log-events](#)[get-image-stack-events](#)[list-clusters](#)[list-cluster-log-streams](#)[list-images](#)[list-image-log-streams](#)[list-official-images](#)[sshupdate-cluster](#)[update-cluster](#)[update-compute-fleet](#)[version](#)

Unterbefehle:

Themen

- [pcluster build-image](#)
- [pcluster configure](#)
- [pcluster create-cluster](#)
- [pcluster dcv-connect](#)
- [pcluster delete-cluster](#)
- [pcluster delete-cluster-instances](#)
- [pcluster delete-image](#)
- [pcluster describe-cluster](#)

- [pcluster describe-cluster-instances](#)
- [pcluster describe-compute-fleet](#)
- [pcluster describe-image](#)
- [pcluster export-cluster-logs](#)
- [pcluster export-image-logs](#)
- [pcluster get-cluster-log-events](#)
- [pcluster get-cluster-stack-events](#)
- [pcluster get-image-log-events](#)
- [pcluster get-image-stack-events](#)
- [pcluster list-clusters](#)
- [pcluster list-cluster-log-streams](#)
- [pcluster list-images](#)
- [pcluster list-image-log-streams](#)
- [pcluster list-official-images](#)
- [pcluster ssh](#)
- [pcluster update-cluster](#)
- [pcluster update-compute-fleet](#)
- [pcluster version](#)

pcluster build-image

Erstellen Sie ein benutzerdefiniertes AWS ParallelCluster Bild in der angegebenen Region.

```
pcluster build-image [-h]
                    --image-configuration IMAGE_CONFIGURATION
                    --image-id IMAGE_ID
                    [--debug]
                    [--dryrun DRYRUN]
                    [--query QUERY]
                    [--region REGION]
                    [--rollback-on-failure ROLLBACK_ON_FAILURE]
                    [--suppress-validators SUPPRESS_VALIDATORS [SUPPRESS_VALIDATORS ...]]
                    [--validation-failure-level {INFO,WARNING,ERROR}]
```

Benannte Argumente

-h, --help

Zeigt den Hilfetext für `pcluster build-image`.

--image-configuration, -c *IMAGE_CONFIGURATION*

Gibt die Image-Konfigurationsdatei als YAML-Dokument an.

--image-id, -i *IMAGE_ID*

Gibt die ID des Images an, das erstellt wird.

--debug

Aktivieren von Debugging-Protokollierung.

--dryrun *DRYRUN*

Wenn `true`, führt der Befehl eine Validierung durch, ohne Ressourcen zu erstellen. Sie können dies verwenden, um die Image-Konfiguration zu überprüfen. (Die Standardeinstellung ist `false`.)

--query *QUERY*

JMESPath-Abfrage, die bei der Ausgabe ausgeführt werden soll.

--region, -r *REGION*

Gibt an, welche verwendet werden AWS-Region soll. Das AWS-Region muss mithilfe der [Region-Einstellung](#) in der Image-Konfigurationsdatei, der `AWS_DEFAULT_REGION` Umgebungsvariablen, der `region` Einstellung im `[default]` Abschnitt der `~/.aws/config` Datei oder des `--region` Parameters angegeben werden.

--rollback-on-failure *ROLLBACK_ON_FAILURE*

Wenn `true`, initiiert bei einem Fehler automatisch ein Rollback des Image-Stacks. (Die Standardeinstellung ist.) `false`

--suppress-validators *SUPPRESS_VALIDATORS* [*SUPPRESS_VALIDATORS ...*]

Identifiziert einen oder mehrere Konfigurationsprüfer, die unterdrückt werden sollen.

Format: (|) ALL type:[A-Za-z0-9]+

--validation-failure-level {INFO,WARNING,ERROR}

Gibt die Mindestvalidierungsstufe an, die dazu führen wird, dass die Erstellung fehlschlägt. (Die Standardeinstellung ist `ERROR`.)

Beispiel mit AWS ParallelCluster Version 3.1.2:

```
$ pcluster build-image --image-configuration image-config.yaml --image-id custom-  
alinux2-image  
{  
  "image": {  
    "imageId": "custom-alinux2-image",  
    "imageBuildStatus": "BUILD_IN_PROGRESS",  
    "cloudformationStackStatus": "CREATE_IN_PROGRESS",  
    "cloudformationStackArn": "arn:aws:cloudformation:us-east-1:123456789012:stack/  
custom-alinux2-image/1234abcd-56ef-78gh-90ij-abcd1234efgh",  
    "region": "us-east-1",  
    "version": "3.1.2"  
  }  
}
```

Warning

`pcluster build-image` verwendet die Standard-VPC. Wenn die Standard-VPC gelöscht wurde, möglicherweise mithilfe von AWS Control Tower oder AWS Landing Zone, muss die Subnetz-ID in der Image-Konfigurationsdatei angegeben werden. Weitere Informationen finden Sie unter [SubnetId](#)

pcluster configure

Startet einen interaktiven Konfigurationsassistenten für AWS ParallelCluster Version 3. Weitere Informationen finden Sie unter [Konfigurieren und erstellen Sie einen Cluster mit der AWS ParallelCluster Befehlszeilenschnittstelle](#).

```
pcluster configure [-h]  
                --config CONFIG  
                [--debug]  
                [--region REGION]
```

Benannte Argumente

-h, --help

Zeigt den Hilfetext für `pcluster configure`.

--config *CONFIG*

Pfad zur Ausgabe der generierten Konfigurationsdatei.

--debug

Aktivieren von Debugging-Protokollierung.

--region, -r *REGION*

Gibt die AWS-Region zu verwendende an. Die Region muss mithilfe der [Region-Einstellung](#) in der Image-Konfigurationsdatei, der `AWS_DEFAULT_REGION` Umgebungsvariablen, der `region` Einstellung im `[default]` Abschnitt der `~/.aws/config` Datei oder des `--region` Parameters angegeben werden.

pcluster create-cluster

Erzeugt einen AWS ParallelCluster Cluster.

```
pcluster create-cluster [-h]
                        --cluster-configuration CLUSTER_CONFIGURATION
                        --cluster-name CLUSTER_NAME
                        [--debug]
                        [--dryrun DRYRUN]
                        [--query QUERY]
                        [--region REGION]
                        [--rollback-on-failure ROLLBACK_ON_FAILURE]
                        [--suppress-validators SUPPRESS_VALIDATORS [SUPPRESS_VALIDATORS ...]]
                        [--validation-failure-level {INFO,WARNING,ERROR}]
```

Benannte Argumente

-h, --help

Zeigt den Hilfetext für `pcluster create-cluster`.

--cluster-configuration, -c *CLUSTER_CONFIGURATION*

Gibt die YAML-Cluster-Konfigurationsdatei an.

--cluster-name, -n *CLUSTER_NAME*

Gibt den Namen des zu erstellenden Clusters an.

Der Name muss mit einem alphabetischen Zeichen beginnen. Der Name kann bis zu 60 Zeichen lang sein. Wenn die Slurm Kontoführung aktiviert ist, kann der Name bis zu 40 Zeichen lang sein.

Gültige Zeichen: a-z, A-Z, 0-9 und - (Bindestrich).

--debug

Aktiviert die Debug-Protokollierung.

--dryrun *DRYRUN*

Wenn `true`, führt der Befehl eine Validierung durch, ohne Ressourcen zu erstellen. Sie können dies verwenden, um die Clusterkonfiguration zu validieren. (Die Standardeinstellung ist `false`.)

--query *QUERY*

Gibt die JMESPath-Abfrage an, die für die Ausgabe ausgeführt werden soll.

--region, -r *REGION*

Gibt an, welche verwendet werden AWS-Region soll. Das AWS-Region muss mithilfe der [Region](#) Einstellung in der Cluster-Konfigurationsdatei, der `AWS_DEFAULT_REGION` Umgebungsvariablen, der `region` Einstellung im `[default]` Abschnitt der `~/.aws/config` Datei oder des `--region` Parameters angegeben werden.

--rollback-on-failure *ROLLBACK_ON_FAILURE*

Wenn `true`, initiiert bei Ausfällen automatisch ein Cluster-Stack-Rollback. (Die Standardeinstellung ist.) `true`

--suppress-validators *SUPPRESS_VALIDATORS* [*SUPPRESS_VALIDATORS* ...]

Identifiziert einen oder mehrere Konfigurationsprüfer, die unterdrückt werden sollen.

Format: (ALL|Typ:) [A-Za-z0-9]+

--validation-failure-level {INFO,WARNING,ERROR}

Gibt die Mindestvalidierungsstufe an, die dazu führen wird, dass die Erstellung fehlschlägt. (Die Standardeinstellung ist `ERROR`.)

Beispiel mit AWS ParallelCluster Version 3.1.4:

```
$ pcluster create-cluster -c cluster-config.yaml -n cluster-v3
{
```

```
"cluster": {
  "clusterName": "cluster-v3",
  "cloudformationStackStatus": "CREATE_IN_PROGRESS",
  "cloudformationStackArn": "arn:aws:cloudformation:us-east-1:123456789012:stack/
cluster-v3/1234abcd-56ef-78gh-90ij-abcd1234efgh",
  "region": "us-east-1",
  "version": "3.1.4",
  "clusterStatus": "CREATE_IN_PROGRESS"
}
```

pcluster dcv-connect

Ermöglicht die Verbindung zum Hauptknoten über eine interaktive Sitzung mithilfe von NICE DCV.

```
pcluster dcv-connect [-h]
                    --cluster-name CLUSTER_NAME
                    [--debug]
                    [--key-path KEY_PATH]
                    [--region REGION]
                    [--show-url]
```

Benannte Argumente

-h, --help

Zeigt den Hilfetext für `pcluster dcv-connect`.

--cluster-name, -n *CLUSTER_NAME*

Gibt den Namen des Clusters an.

--debug

Aktiviert die Debug-Protokollierung.

--key-path *KEY_PATH*

Gibt den Pfad des SSH-Schlüssels an, der für die Verbindung verwendet werden soll.

--region, -r *REGION*

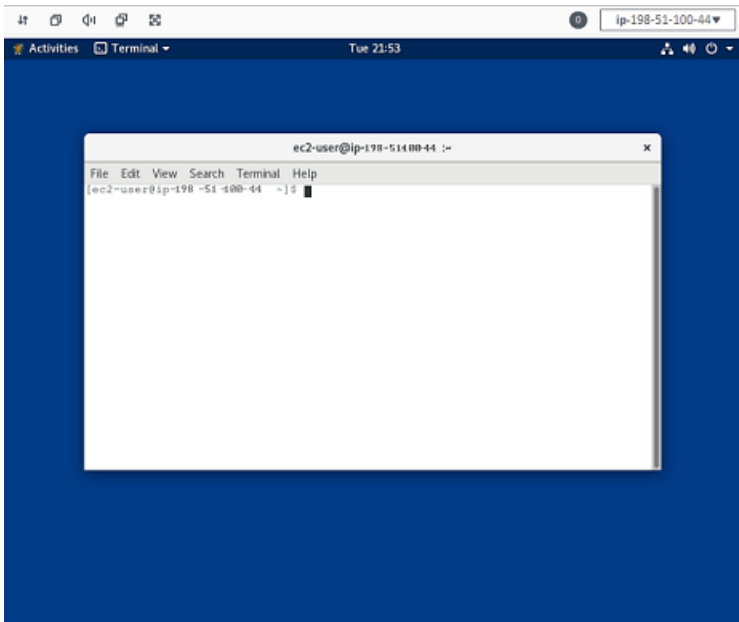
Gibt den AWS-Region zu verwendenden an. Der AWS-Region muss mithilfe der `AWS_DEFAULT_REGION` Umgebungsvariablen, der `region` Einstellung im `[default]` Abschnitt der `~/.aws/config` Datei oder des `--region` Parameters angegeben werden.

--show-url

Druckt die URL, die für die DCV-Verbindung verwendet werden würde, und beendet das Programm.

Beispiel mit AWS ParallelCluster Version 3.1.4:

```
$ pcluster dcv-connect -n cluster-3Dcv -r us-east-1 --key-path /home/user/.ssh/key.pem
```



pcluster delete-cluster

Startet das Löschen eines Clusters.

```
pcluster delete-cluster [-h]
                        --cluster-name CLUSTER_NAME
                        [--debug]
                        [--query QUERY]
                        [--region REGION]
```

Benannte Argumente

-h, --help

Zeigt den Hilfetext für `pcluster delete-cluster`.

--cluster-name, -n *CLUSTER_NAME*

Gibt den Namen des Clusters an.

--debug

Aktiviert die Debug-Protokollierung.

--query *QUERY*

Gibt die JMESPath-Abfrage an, die für die Ausgabe ausgeführt werden soll.

--region, -r *REGION*

Gibt an, welche verwendet werden AWS-Region soll. Die Region muss mithilfe der `AWS_DEFAULT_REGION` Umgebungsvariablen, der `region` Einstellung im `[default]` Abschnitt der `~/.aws/config` Datei oder des `--region` Parameters angegeben werden.

Beispiel mit AWS ParallelCluster Version 3.1.4:

```
$ pcluster delete-cluster -n cluster-v3
{
  "cluster": {
    "clusterName": "cluster-v3",
    "cloudformationStackStatus": "DELETE_IN_PROGRESS",
    "cloudformationStackArn": "arn:aws:cloudformation:us-east-1:123456789012:stack/
cluster-v3/1234abcd-56ef-78gh-90ij-abcd1234efgh",
    "region": "us-east-1",
    "version": "3.1.4",
    "clusterStatus": "DELETE_IN_PROGRESS"
  }
}
```

pcluster delete-cluster-instances

Initiieren Sie die erzwungene Kündigung aller Cluster-Rechenknoten. Dies funktioniert nicht mit AWS Batch Clustern.

```
pcluster delete-cluster-instances [-h]
    --cluster-name CLUSTER_NAME
    [--debug]
    [--force FORCE]
    [--query QUERY]
```

```
[--region REGION]
```

Benannte Argumente

-h, --help

Zeigt den Hilfetext für `pcluster delete-cluster-instances`.

--cluster-name, -n *CLUSTER_NAME*

Gibt den Namen des Clusters an.

--debug

Aktiviert die Debug-Protokollierung.

--force *FORCE*

Wenn `true`, erzwingt das Löschen, indem Validierungsfehler ignoriert werden. (Die Standardeinstellung ist.) `false`

--query *QUERY*

Gibt die JMESPath-Abfrage an, die für die Ausgabe ausgeführt werden soll.

--region, -r *REGION*

Gibt an, welche verwendet werden AWS-Region soll. Der AWS-Region muss mithilfe der `AWS_DEFAULT_REGION` Umgebungsvariablen, der `region` Einstellung im `[default]` Abschnitt der `~/.aws/config` Datei oder des `--region` Parameters angegeben werden.

```
$ pcluster delete-cluster-instances -n cluster-v3
```

pcluster delete-image

Startet das Löschen des benutzerdefinierten AWS ParallelCluster Images.

```
pcluster delete-image [-h]
                      --image-id IMAGE_ID
                      [--debug]
                      [--force FORCE]
                      [--query QUERY]
                      [--region REGION]
```

Benannte Argumente

-h, --help

Zeigt den Hilfetext für `pcluster delete-image`.

--image-id, -i *IMAGE_ID*

Gibt die ID des Bildes an, das gelöscht werden soll.

--debug

Aktiviert die Debug-Protokollierung.

--force *FORCE*

Wenn `true`, erzwingt das Löschen, falls es Instances gibt, die das AMI verwenden, oder wenn das AMI gemeinsam genutzt wird. (Die Standardeinstellung ist `false`.)

--query *QUERY*

Gibt die JMESPath-Abfrage an, die für die Ausgabe ausgeführt werden soll.

--region, -r *REGION*

Gibt an, welche verwendet werden AWS-Region soll. Der AWS-Region muss mithilfe der `AWS_DEFAULT_REGION` Umgebungsvariablen, der `region` Einstellung im `[default]` Abschnitt der `~/.aws/config` Datei oder des `--region` Parameters angegeben werden.

Beispiel mit AWS ParallelCluster Version 3.1.4:

```
$ pcluster delete-image --image-id custom-alinux2-image
{
  "image": {
    "imageId": "custom-alinux2-image",
    "imageBuildStatus": "DELETE_IN_PROGRESS",
    "region": "us-east-1",
    "version": "3.1.4"
  }
}
```

pcluster describe-cluster

Holen Sie sich detaillierte Informationen über einen Cluster.

```
pcluster describe-cluster [-h]
                        --cluster-name CLUSTER_NAME
                        [--debug]
                        [--query QUERY]
                        [--region REGION]
```

Benannte Argumente

-h, --help

Zeigt den Hilfetext für `pcluster describe-cluster`.

--cluster-name, -n *CLUSTER_NAME*

Gibt den Namen des Clusters an.

--debug

Aktiviert die Debug-Protokollierung.

--query *QUERY*

Gibt die JMESPath-Abfrage an, die für die Ausgabe ausgeführt werden soll.

--region, -r *REGION*

Gibt an, welche verwendet werden AWS-Region soll. Der AWS-Region muss mithilfe der `AWS_DEFAULT_REGION` Umgebungsvariablen, der `region` Einstellung im `[default]` Abschnitt der `~/.aws/config` Datei oder des `--region` Parameters angegeben werden.

Beispiele mit AWS ParallelCluster Version 3.1.4:

Beschreiben Sie die Cluster-Details:

```
$ pcluster describe-cluster -n cluster-v3
{
  "creationTime": "2022-07-12T17:19:16.101Z",
  "headNode": {
    "launchTime": "2022-07-12T17:22:21.000Z",
    "instanceId": "i-1234567890abcdef0",
    "publicIpAddress": "198.51.100.44",
    "instanceType": "t2.micro",
    "state": "running",
    "privateIpAddress": "192.0.2.0.196"
```

```

},
"loginNodes": {
  "status": "active",
  "address": "8af2145440569xyz.us-east-1.amazonaws.com",
  "scheme": "internet-facing|internal",
  "healthyNodes": 3,
  "unhealthyNodes": 0
},
"version": "3.1.4",
"clusterConfiguration": {
  "url": "https://parallelcluster-e5ca74255d6c3886-v1-do-not-delete..."
},
"tags": [
  {
    "value": "3.1.4",
    "key": "parallelcluster:version"
  }
],
"cloudFormationStackStatus": "CREATE_COMPLETE",
"clusterName": "cluster-v3",
"computeFleetStatus": "RUNNING",
"cloudFormationStackArn": "arn:aws:cloudformation:us-east-1:123456789012:stack/
cluster-v3/1234abcd-56ef-78gh-90ij-abcd1234efgh",
"lastUpdatedTime": "2022-07-12T17:19:16.101Z",
"region": "us-east-1",
"clusterStatus": "CREATE_COMPLETE"
}

```

Verwenden `Siedescribe-cluster`, um die Cluster-Konfiguration abzurufen:

```

$ curl -o - $(pcluster describe-cluster -n cluster-v3 --query clusterConfiguration.url
| xargs echo)
Region: us-east-1
Image:
  Os: alinux2
HeadNode:
  InstanceType: t2.micro
  Networking:
    SubnetId: subnet-abcdef01234567890
  Ssh:
    KeyName: adpc
  Iam:
    S3Access:

```



```

- BucketName: cluster-v3-bucket
  KeyName: logs
  EnableWriteAccess: true
Scheduling:
  Scheduler: slurm
  SlurmQueues:
- Name: queue1
  ComputeResources:
- Name: t2micro
  InstanceType: t2.micro
  MinCount: 0
  MaxCount: 10
Networking:
  SubnetIds:
- subnet-021345abcdef6789

```

pcluster describe-cluster-instances

Beschreiben Sie die Instanzen in einem Cluster.

```

pcluster describe-cluster-instances [-h]
    --cluster-name CLUSTER_NAME
    [--debug]
    [--next-token NEXT_TOKEN]
    [--node-type {HeadNode,ComputeNode,LoginNode}]
    [--query QUERY]
    [--queue-name QUEUE_NAME]
    [--region REGION]

```

Benannte Argumente

-h, --help

Zeigt den Hilfetext für `pcluster describe-cluster-instances`.

--cluster-name, -n *CLUSTER_NAME*

Gibt den Namen des Clusters an.

--debug

Aktiviert die Debug-Protokollierung.

--next-token *NEXT_TOKEN*

Gibt das Token an, das für paginierte Anfragen verwendet werden soll.

--node-type {*HeadNode,ComputeNode,LoginNode*}

Gibt die aufzulistenden Knotentypen an. Unterstützte Werte sind *HeadNode*, *ComputeNode* und *LoginNode*. Wenn dieser Parameter nicht angegeben ist, werden die *HeadNode* Instanzen *ComputeNode* und *LoginNode* a beschrieben.

--query *QUERY*

Gibt die JMESPath-Abfrage an, die für die Ausgabe ausgeführt werden soll.

--queue-name *QUEUE_NAME*

Gibt den Namen der Warteschlange an, die aufgelistet werden soll. Wenn dieser Parameter nicht angegeben ist, werden Instanzen in allen Warteschlangen beschrieben.

--region, -r *REGION*

Gibt den AWS-Region zu verwendenden an. Der AWS-Region muss mithilfe der *AWS_DEFAULT_REGION* Umgebungsvariablen, der *region* Einstellung im [default] Abschnitt der *~/.aws/config* Datei oder des *--region* Parameters angegeben werden.

Beispiel mit AWS ParallelCluster Version 3.1.4:

```
$ pcluster describe-cluster-instances -n cluster-v3
{
  "instances": [
    {
      "launchTime": "2022-07-12T17:22:21.000Z",
      "instanceId": "i-1234567890abcdef0",
      "publicIpAddress": "198.51.100.44",
      "instanceType": "t2.micro",
      "state": "running",
      "nodeType": "HeadNode",
      "privateIpAddress": "192.0.2.0.196"
    },
    {
      "launchTime": "2022-07-12T17:37:42.000Z",
      "instanceId": "i-021345abcdef6789",
      "queueName": "queue1",
      "publicIpAddress": "198.51.100.44",
      "instanceType": "t2.micro",
```

```
    "state": "pending",
    "nodeType": "ComputeNode",
    "privateIpAddress": "192.0.2.0.196"
  }
]
}
```

pcluster describe-compute-fleet

Beschreiben Sie den Status der Rechenflotte.

```
pcluster describe-compute-fleet [-h]
    --cluster-name CLUSTER_NAME
    [--debug]
    [--query QUERY]
    [--region REGION]
```

Benannte Argumente

-h, --help

Zeigt den Hilfetext für `pcluster describe-compute-fleet`.

--cluster-name, -n *CLUSTER_NAME*

Gibt den Namen des Clusters an.

--debug

Aktiviert die Debug-Protokollierung.

--query *QUERY*

Gibt die JMESPath-Abfrage an, die für die Ausgabe ausgeführt werden soll.

--region, -r *REGION*

Gibt an, welche verwendet werden AWS-Region soll. Der AWS-Region muss mithilfe der `AWS_DEFAULT_REGION` Umgebungsvariablen, der `region` Einstellung im `[default]` Abschnitt der `~/.aws/config` Datei oder des `--region` Parameters angegeben werden.

Beispiel mit AWS ParallelCluster Version 3.1.4:

```
$ pcluster describe-compute-fleet -n pcluster-v3
{
```

```
"status": "RUNNING",
"lastStatusUpdatedTime": "2022-07-12T17:24:26.000Z"
}
```

pcluster describe-image

Holen Sie sich detaillierte Informationen zu einem Bild.

```
pcluster describe-image [-h]
                        --image-id IMAGE_ID
                        [--debug]
                        [--query QUERY]
                        [--region REGION]
```

Benannte Argumente

-h, --help

Zeigt den Hilfetext für `pcluster describe-image`.

--image-id, -i *IMAGE_ID*

Gibt die ID des Bilds an.

--debug

Aktiviert die Debug-Protokollierung.

--query *QUERY*

Gibt die JMESPath-Abfrage an, die für die Ausgabe ausgeführt werden soll.

--region, -r *REGION*

Gibt an, welche verwendet werden AWS-Region soll. Der AWS-Region muss mithilfe der `AWS_DEFAULT_REGION` Umgebungsvariablen, der `region` Einstellung im `[default]` Abschnitt der `~/.aws/config` Datei oder des `--region` Parameters angegeben werden.

Beispiel mit AWS ParallelCluster Version 3.1.2:

```
$ pcluster describe-image --image-id custom-alinux2-image
{
  "imageConfiguration": {
    "url": "https://parallelcluster-1234abcd5678-v1-do-not-delete.../configs/image-
config.yaml"
```

```

},
"imageId": "custom-alinux2-image",
"creationTime": "2022-04-05T20:23:07.000Z"
"imageBuildStatus": "BUILD_COMPLETE",
"region": "us-east-1",
"ec2AmiInfo": {
  "amiName": "custom-alinux2-image 2022-04-05T19-55-22.518Z",
  "amiId": "ami-1234abcd5678efgh",
  "description": "AWS ParallelCluster AMI for alinux2,
kernel-4.14.268-205.500.amzn2.x86_64, lustre-2.10.8-5.amzn2.x86_64,
efa-1.14.2-1.amzn2.x86_64, dcv-2021.3.11591-1.el7.x86_64, slurm-21-08-6-1",
  "state": "AVAILABLE",
"tags": [
  {
    "value": "arn:aws:imagebuilder:us-east-1:123456789012:image/
parallelclusterimage-custom-alinux2-image/3.1.2/1",
    "key": "Ec2ImageBuilderArn"
  },
  {
    "value": "parallelcluster-1234abcd5678efgh-v1-do-not-delete",
    "key": "parallelcluster:s3_bucket"
  },
  {
    "value": "custom-alinux2-image",
    "key": "parallelcluster:image_name"
  },
  {
    "value": "available",
    "key": "parallelcluster:build_status"
  },
  {
    "value": "s3://parallelcluster-1234abcd5678efgh-v1-do-not-delete/
parallelcluster/3.1.2/images/custom-alinux2-image-1234abcd5678efgh/configs/image-
config.yaml",
    "key": "parallelcluster:build_config"
  },
  {
    "value": "EC2 Image Builder",
    "key": "CreatedBy"
  },
  {
    "value": "arn:aws:logs:us-east-1:123456789012:log-group:/aws/imagebuilder/
ParallelClusterImage-custom-alinux2-image",
    "key": "parallelcluster:build_log"
  }
]
}

```

```
  },
  {
    "value": "4.14.268-205.500.amzn2.x86_64",
    "key": "parallelcluster:kernel_version"
  },
  {
    "value": "arn:aws:imagebuilder:us-east-1:444455556666:image/amazon-linux-2-
x86/2022.3.16/1",
    "key": "parallelcluster:parent_image"
  },
  {
    "value": "3.1.2",
    "key": "parallelcluster:version"
  },
  {
    "value": "0.5.14",
    "key": "parallelcluster:munge_version"
  },
  {
    "value": "21-08-6-1",
    "key": "parallelcluster:slurm_version"
  },
  {
    "value": "2021.3.11591-1.el7.x86_64",
    "key": "parallelcluster:dcv_version"
  },
  {
    "value": "alinux2-image",
    "key": "parallelcluster:image_id"
  },
  {
    "value": "3.2.3",
    "key": "parallelcluster:pmix_version"
  },
  {
    "value": "parallelcluster/3.7.0/images/alinux2-image-abcd1234efgh56781234",
    "key": "parallelcluster:s3_image_dir"
  },
  {
    "value": "1.14.2-1.amzn2.x86_64",
    "key": "parallelcluster:efa_version"
  },
  {
    "value": "alinux2",
```

```
    "key": "parallelcluster:os"
  },
  {
    "value": "aws-parallelcluster-cookbook-3.1.2",
    "key": "parallelcluster:bootstrap_file"
  },
  {
    "value": "1.8.23-10.amzn2.1.x86_64",
    "key": "parallelcluster:sudo_version"
  },
  {
    "value": "2.10.8-5.amzn2.x86_64",
    "key": "parallelcluster:lustre_version"
  }
],
"architecture": "x86_64"
},
"version": "3.1.2"
}
```

pcluster export-cluster-logs

Exportieren Sie die Protokolle des Clusters in ein lokales tar.gz Archiv, indem Sie einen Amazon S3 S3-Bucket passieren.

```
pcluster export-cluster-logs [-h]
    --bucket BUCKET_NAME
    --cluster-name CLUSTER_NAME
    [--bucket-prefix BUCKET_PREFIX]
    [--debug]
    [--end-time END_TIME]
    [--filters FILTER [FILTER ...]]
    [--keep-s3-objects KEEP_S3_OBJECTS]
    [--output-file OUTPUT_FILE]
    [--region REGION]
    [--start-time START_TIME]
```

Benannte Argumente

-h, --help

Zeigt den Hilfetext für `pcluster export-cluster-logs`.

--bucket *BUCKET_NAME*

Gibt den Namen des Amazon S3 S3-Buckets an, in den Cluster-Protokolldaten exportiert werden sollen. Er muss sich in derselben Region wie der Cluster befinden.

Note

Sie müssen der Amazon S3 S3-Bucket-Richtlinie Berechtigungen hinzufügen, um CloudWatch Zugriff zu gewähren. Weitere Informationen finden Sie unter [Berechtigungen für einen Amazon S3 S3-Bucket festlegen](#) im CloudWatch Logs-Benutzerhandbuch.

--cluster-name, -n *CLUSTER_NAME*

Gibt den Namen des Clusters an.

--bucket-prefix *BUCKET_PREFIX*

Gibt den Pfad im Amazon S3 S3-Bucket an, in dem exportierte Protokolldaten gespeichert werden sollen.

Standardmäßig lautet das Bucket-Präfix:

```
cluster-name-logs-202209061743.tar.gz
```

202209061743 ist die aktuelle Uhrzeit im %Y%m%d%H%M Format.

--debug

Aktiviert die Debug-Protokollierung.

--end-time *END_TIME*

Gibt das Ende des Zeitbereichs für die Erfassung von Protokollereignissen an, ausgedrückt im ISO 8601-Format (YYYY-MM-DDThh:mm:ssZz. B. 2021-01-01T20:00:00Z '). Ereignisse mit einem Zeitstempel, der diesem Zeitpunkt entspricht oder später liegt, sind nicht enthalten. Zeitelemente (z. B. Minuten und Sekunden) können weggelassen werden. Der Standardwert ist die aktuelle Zeit.

--filters *FILTER [FILTER ...]*

Definiert Filter für das Protokoll. Format:Name=a,Values=1 Name=b,Values=2,3. Unterstützte Filter sind:

private-dns-name

Gibt die Kurzform des privaten DNS-Namens der Instanz an (z. B. `ip-10-0-0-101`).

node-type

Gibt den Knotentyp an, der einzig akzeptierte Wert für diesen Filter ist `HeadNode`.

--keep-s3-objects *KEEP_S3_OBJECTS*

Falls `true`, werden die exportierten Objekte, die nach Amazon S3 exportiert werden, beibehalten. (Die Standardeinstellung ist `false`.)

--output-file *OUTPUT_FILE*

Gibt den Dateipfad an, in dem das Log-Archiv gespeichert werden soll. Wenn dies angegeben ist, werden die Protokolle lokal gespeichert. Andernfalls werden sie mit der in der Ausgabe zurückgegebenen URL auf Amazon S3 hochgeladen. Standardmäßig wird auf Amazon S3 hochgeladen.

--region, -r *REGION*

Gibt an AWS-Region, welche verwendet werden soll. Der AWS-Region muss mithilfe der `AWS_DEFAULT_REGION` Umgebungsvariablen, der `region` Einstellung im `[default]` Abschnitt der `~/.aws/config` Datei oder des `--region` Parameters angegeben werden.

--start-time *START_TIME*

Gibt den Beginn des Zeitbereichs an, ausgedrückt im ISO 8601-Format (zum `YYYY-MM-DDThh:mm:ssZ` Beispiel `2021-01-01T20:00:00Z`). Protokollereignisse mit einem Zeitstempel, der diesem Zeitpunkt oder einem späteren Zeitpunkt entspricht, sind enthalten. Wenn nicht angegeben, ist die Standardeinstellung die Uhrzeit, zu der der Cluster erstellt wurde.

Beispiel mit AWS ParallelCluster Version 3.1.4:

```
$ pcluster export-cluster-logs --bucket cluster-v3-bucket -n cluster-v3
{
  "url": "https://cluster-v3-bucket..."
}
```

pcluster export-image-logs

Exportieren Sie die Protokolle des Image Builder-Stacks in ein lokales `tar.gz` Archiv, indem Sie sie durch einen Amazon S3 S3-Bucket leiten.

```
pcluster export-image-logs [-h]
    --bucket BUCKET
    --image-id IMAGE_ID
    [--bucket-prefix BUCKET_PREFIX]
    [--debug]
    [--end-time END_TIME]
    [--keep-s3-objects KEEP_S3_OBJECTS]
    [--output-file OUTPUT_FILE]
    [--region REGION]
    [--start-time START_TIME]
```

Benannte Argumente

-h, --help

Zeigt den Hilfetext für `pcluster export-image-logs`.

--bucket *BUCKET_NAME*

Gibt den Amazon S3 S3-Bucket-Namen an, in den Image-Build-Logs exportiert werden sollen. Er muss sich in derselben Region wie das Image befinden.

Note

Sie müssen der Amazon S3 S3-Bucket-Richtlinie Berechtigungen hinzufügen, um CloudWatch Zugriff zu gewähren. Weitere Informationen finden Sie unter [Berechtigungen für einen Amazon S3 S3-Bucket festlegen](#) im CloudWatch Logs-Benutzerhandbuch.

--image-id, -i *IMAGE_ID*

Die Bild-ID, deren Protokolle exportiert werden.

--bucket-prefix *BUCKET_PREFIX*

Gibt den Pfad im Amazon S3 S3-Bucket an, in dem exportierte Protokolldaten gespeichert werden sollen.

Standardmäßig lautet das Bucket-Präfix:

```
ami-id-logs-202209061743.tar.gz
```

`202209061743` ist die aktuelle Uhrzeit im `%Y%m%d%H%M` Format.

--debug

Aktiviert die Debug-Protokollierung.

--end-time *END_TIME*

Gibt das Ende des Zeitbereichs für die Erfassung von Protokollereignissen an, ausgedrückt im ISO 8601-Format (YYYY-MM-DDThh:mm:ssZz. B. `2021-01-01T20:00:00Z` '). Ereignisse mit einem Zeitstempel, der diesem Zeitpunkt entspricht oder später liegt, sind nicht enthalten. Zeitelemente (z. B. Minuten und Sekunden) können weggelassen werden. Der Standardwert ist die aktuelle Zeit.

--keep-s3-objects *KEEP_S3_OBJECTS*

Falls `true`, werden die exportierten Objekte, die nach Amazon S3 exportiert werden, beibehalten. (Die Standardeinstellung ist `false`.)

--output-file *OUTPUT_FILE*

Gibt den Dateipfad an, in dem das Log-Archiv gespeichert werden soll. Wenn dies angegeben ist, werden die Protokolle lokal gespeichert. Andernfalls werden sie mit der in der Ausgabe zurückgegebenen URL auf Amazon S3 hochgeladen. Standardmäßig wird auf Amazon S3 hochgeladen.

--region, -r *REGION*

Gibt an AWS-Region, welche verwendet werden soll. Der AWS-Region muss mithilfe der `AWS_DEFAULT_REGION` Umgebungsvariablen, der `region` Einstellung im `[default]` Abschnitt der `~/.aws/config` Datei oder des `--region` Parameters angegeben werden.

--start-time *START_TIME*

Gibt den Beginn des Zeitbereichs an, ausgedrückt im ISO 8601-Format (zum `YYYY-MM-DDThh:mm:ssZ` Beispiel `2021-01-01T20:00:00Z`). Protokollereignisse mit einem Zeitstempel, der diesem Zeitpunkt oder einem späteren Zeitpunkt entspricht, sind enthalten. Wenn nicht angegeben, ist die Standardeinstellung die Uhrzeit, zu der der Cluster erstellt wurde.

Beispiel mit AWS ParallelCluster Version 3.1.4:

```
$ pcluster export-image-logs --bucket image-v3-bucket --image-id ami-1234abcd5678efgh
{
```

```
"url": "https://image-v3-bucket..."
}
```

pcluster get-cluster-log-events

Ruft die Ereignisse ab, die mit einem Protokollstream verknüpft sind.

```
pcluster get-cluster-log-events [-h]
    --cluster-name CLUSTER_NAME
    --log-stream-name LOG_STREAM_NAME
    [--debug]
    [--end-time END_TIME]
    [--limit LIMIT]
    [--next-token NEXT_TOKEN]
    [--query QUERY]
    [--region REGION]
    [--start-from-head START_FROM_HEAD]
    [--start-time START_TIME]
```

Benannte Argumente

-h, --help

Zeigt den Hilfetext für `pcluster get-cluster-log-events`.

--cluster-name, -n *CLUSTER_NAME*

Gibt den Namen des Clusters an.

--log-stream-name *LOG_STREAM_NAME*

Gibt den Namen des Protokollstroms an. Sie können den `list-cluster-log-streams` Befehl verwenden, um einen Protokollstream abzurufen, der einem oder mehreren Ereignissen zugeordnet ist.

--debug

Aktiviert die Debug-Protokollierung.

--end-time *END_TIME*

Gibt das Ende des Zeitbereichs an, ausgedrückt im ISO 8601-Format (zum YYYY-MM-DDThh:mm:ssZ Beispiel 2021-01-01T20:00:00Z). Ereignisse mit einem Zeitstempel, der diesem Zeitpunkt entspricht oder später liegt, sind nicht enthalten.

--limit *LIMIT*

Gibt die maximale Anzahl von zurückgegebenen Protokollereignissen an. Wenn kein Wert angegeben wird, beträgt das Maximum so viele Protokollereignisse, wie in eine Antwortgröße von 1 MB passen, also bis zu 10.000 Protokollereignisse.

--next-token *NEXT_TOKEN*

Gibt das Token an, das für paginierte Anfragen verwendet werden soll.

--query *QUERY*

Gibt die JMESPath-Abfrage an, die für die Ausgabe ausgeführt werden soll.

--region, -r *REGION*

Gibt an, welche verwendet werden AWS-Region soll. Der AWS-Region muss mithilfe der `AWS_DEFAULT_REGION` Umgebungsvariablen, der `region` Einstellung im `[default]` Abschnitt der `~/.aws/config` Datei oder des `--region` Parameters angegeben werden.

--start-from-head *START_FROM_HEAD*

Wenn der Wert `isttrue`, werden die frühesten Protokollereignisse zuerst zurückgegeben. Ist der Wert `gleichfalse`, werden die neuesten Protokollereignisse zuerst zurückgegeben. (Die Standardeinstellung ist `false`.)

--start-time *START_TIME*

Gibt den Beginn des Zeitbereichs an, ausgedrückt im Format ISO 8601 (zum `YYYY-MM-DDThh:mm:ssZ` Beispiel `2021-01-01T20:00:00Z`). Ereignisse mit einem Zeitstempel, der diesem Zeitpunkt oder einem späteren Zeitpunkt entspricht, sind enthalten.

Beispiel mit AWS ParallelCluster Version 3.1.4:

```
$ pcluster get-cluster-log-events \  
  -c cluster-v3 \  
  -r us-east-1 \  
  --log-stream-name ip-198-51-100-44.i-1234567890abcdef0.clustermgtd \  
  --limit 3 \  
{  
  "nextToken": "f/36966906399261933213029082268132291405859205452101451780/s",  
  "prevToken": "b/36966906399239632467830551644990755687586557090595471362/s",  
  "events": [  
    {
```

```

    "message": "2022-07-12 19:16:53,379 - [slurm_plugin.clustermgtd:_maintain_nodes]
- INFO - Performing node maintenance actions",
    "timestamp": "2022-07-12T19:16:53.379Z"
  },
  {
    "message": "2022-07-12 19:16:53,380 - [slurm_plugin.clustermgtd:_maintain_nodes]
- INFO - Following nodes are currently in replacement: (x0) []",
    "timestamp": "2022-07-12T19:16:53.380Z"
  },
  {
    "message": "2022-07-12 19:16:53,380 -
[slurm_plugin.clustermgtd:_terminate_orphaned_instances] - INFO - Checking for
orphaned instance",
    "timestamp": "2022-07-12T19:16:53.380Z"
  }
]
}

```

pcluster get-cluster-stack-events

Ruft die Ereignisse ab, die dem Stack für den angegebenen Cluster zugeordnet sind.

Note

Ab Version 3.6.0 werden verschachtelte Stacks AWS ParallelCluster verwendet, um die Ressourcen zu erstellen, die Warteschlangen und Rechenressourcen zugeordnet sind. Die `GetClusterStackEvents` API und der `pcluster get-cluster-stack-events` Befehl geben nur die Haupt-Stack-Ereignisse des Clusters zurück. Sie können die Cluster-Stack-Ereignisse, einschließlich der Ereignisse im Zusammenhang mit Warteschlangen und Rechenressourcen, in der CloudFormation Konsole anzeigen.

```

pcluster get-cluster-stack-events [-h]
    --cluster-name CLUSTER_NAME
    [--debug]
    [--next-token NEXT_TOKEN]
    [--query QUERY]
    [--region REGION]

```

Benannte Argumente

-h, --help

Zeigt den Hilfetext für `pcluster get-cluster-stack-events`.

--cluster-name, -n *CLUSTER_NAME*

Gibt den Namen des Clusters an.

--debug

Aktiviert die Debug-Protokollierung.

--next-token *NEXT_TOKEN*

Gibt das Token an, das für paginierte Anfragen verwendet werden soll.

--query *QUERY*

Gibt die JMESPath-Abfrage an, die für die Ausgabe ausgeführt werden soll.

--region, -r *REGION*

Gibt an, welche verwendet werden AWS-Region soll. Der AWS-Region muss mithilfe der `AWS_DEFAULT_REGION` Umgebungsvariablen, der `region` Einstellung im `[default]` Abschnitt der `~/.aws/config` Datei oder des `--region` Parameters angegeben werden.

Beispiel mit AWS ParallelCluster Version 3.1.4:

```
$ pcluster get-cluster-stack-events \
  -n cluster-v3 \
  -r us-east-1 \
  --query "events[0]"
{
  "eventId": "1234abcd-56ef-78gh-90ij-abcd1234efgh",
  "physicalResourceId": "arn:aws:cloudformation:us-east-1:123456789012:stack/cluster-
v3/1234abcd-56ef-78gh-90ij-abcd1234efgh",
  "resourceStatus": "CREATE_COMPLETE",
  "stackId": "arn:aws:cloudformation:us-east-1:123456789012:stack/cluster-
v3/1234abcd-56ef-78gh-90ij-abcd1234efgh",
  "stackName": "cluster-v3",
  "logicalResourceId": "cluster-v3",
  "resourceType": "AWS::CloudFormation::Stack",
  "timestamp": "2022-07-12T18:29:12.140Z"
```

```
}
```

pcluster get-image-log-events

Rufen Sie die Ereignisse ab, die mit einem Image-Build verknüpft sind.

```
pcluster get-image-log-events [-h]
    --image-id IMAGE_ID
    --log-stream-name LOG_STREAM_NAME
    [--debug]
    [--end-time END_TIME]
    [--limit LIMIT]
    [--next-token NEXT_TOKEN]
    [--query QUERY]
    [--region REGION]
    [--start-from-head START_FROM_HEAD]
    [--start-time START_TIME]
```

Benannte Argumente

-h, --help

Zeigt den Hilfetext für `pcluster get-image-log-events`.

--image-id, -i *IMAGE_ID*

Gibt die ID des Bildes an.

--log-stream-name *LOG_STREAM_NAME*

Gibt den Namen des Protokollstroms an. Sie können den `list-image-log-streams` Befehl verwenden, um einen Protokollstream abzurufen, der einem oder mehreren Ereignissen zugeordnet ist.

--debug

Aktiviert die Debug-Protokollierung.

--end-time *END_TIME*

Gibt das Ende des Zeitbereichs an, ausgedrückt im ISO 8601-Format (zum `YYYY-MM-DDThh:mm:ssZ` Beispiel `2021-01-01T20:00:00Z`). Ereignisse mit einem Zeitstempel, der diesem Zeitpunkt entspricht oder später liegt, sind nicht enthalten.

--limit *LIMIT*

Gibt die maximale Anzahl von zurückgegebenen Protokollereignissen an. Wenn kein Wert angegeben wird, beträgt das Maximum so viele Protokollereignisse, wie in eine Antwortgröße von 1 MB passen, also bis zu 10.000 Protokollereignisse.

--next-token *NEXT_TOKEN*

Gibt das Token an, das für paginierte Anfragen verwendet werden soll.

--query *QUERY*

Gibt die JMESPath-Abfrage an, die für die Ausgabe ausgeführt werden soll.

--region, -r *REGION*

Gibt an, welche verwendet werden AWS-Region soll. Der AWS-Region muss mithilfe der `AWS_DEFAULT_REGION` Umgebungsvariablen, der `region` Einstellung im `[default]` Abschnitt der `~/.aws/config` Datei oder des `--region` Parameters angegeben werden.

--start-from-head *START_FROM_HEAD*

Wenn der Wert `isttrue`, werden die frühesten Protokollereignisse zuerst zurückgegeben. Ist der Wert `gleichfalse`, werden die neuesten Protokollereignisse zuerst zurückgegeben. (Die Standardeinstellung ist `false`.)

--start-time *START_TIME*

Gibt den Beginn des Zeitbereichs an, ausgedrückt im Format ISO 8601 (zum `YYYY-MM-DDThh:mm:ssZ` Beispiel `2021-01-01T20:00:00Z`). Ereignisse mit einem Zeitstempel, der diesem Zeitpunkt entspricht oder später liegt, sind enthalten.

Beispiel mit AWS ParallelCluster Version 3.1.2:

```
$ pcluster get-image-log-events --image-id custom-alinux2-image --region us-east-1 --  
log-stream-name 3.1.2/1 --limit 3  
{  
  "nextToken": "f/36778317771100849897800729464621464113270312017760944178/s",  
  "prevToken": "b/36778317766952911290874033560295820514557716777648586800/s",  
  "events": [  
    {  
      "message": "ExecuteBash: FINISHED EXECUTION",  
      "timestamp": "2022-04-05T22:13:26.633Z"  
    },  
  ],  
}
```

```
{
  "message": "Document arn:aws:imagebuilder:us-east-1:123456789012:component/parallelclusterimage-test-1234abcd-56ef-78gh-90ij-abcd1234efgh/3.1.2/1",
  "timestamp": "2022-04-05T22:13:26.741Z"
},
{
  "message": "TOE has completed execution successfully",
  "timestamp": "2022-04-05T22:13:26.819Z"
}
]
```

pcluster get-image-stack-events

Ruft die Ereignisse ab, die dem Stack für den angegebenen Image-Build zugeordnet sind.

```
pcluster get-image-stack-events [-h]
    --image-id IMAGE_ID
    [--debug]
    [--next-token NEXT_TOKEN]
    [--query QUERY]
    [--region REGION]
```

Benannte Argumente

-h, --help

Zeigt den Hilfetext für `pcluster get-image-stack-events`.

--image-id, -i *IMAGE_ID*

Gibt die ID des Bilds an.

--debug

Aktiviert die Debug-Protokollierung.

--next-token *NEXT_TOKEN*

Gibt das Token an, das für paginierte Anfragen verwendet werden soll.

--query *QUERY*

Gibt die JMESPath-Abfrage an, die für die Ausgabe ausgeführt werden soll.

--region, -r *REGION*

Gibt an, welche verwendet werden AWS-Region soll. Der AWS-Region muss mithilfe der `AWS_DEFAULT_REGION` Umgebungsvariablen, der `region` Einstellung im `[default]` Abschnitt der `~/.aws/config` Datei oder des `--region` Parameters angegeben werden.

Beispiel mit AWS ParallelCluster Version 3.1.2:

```
$ pcluster get-image-stack-events --image-id custom-alinux2-image --region us-east-1 --
query "events[0]"
{
  "eventId": "ParallelClusterImage-CREATE_IN_PROGRESS-2022-04-05T21:39:24.725Z",
  "physicalResourceId": "arn:aws:imagebuilder:us-east-1:123456789012:image/
parallelclusterimage-custom-alinux2-image/3.1.2/1",
  "resourceStatus": "CREATE_IN_PROGRESS",
  "resourceStatusReason": "Resource creation Initiated",
  "resourceProperties": "{\"InfrastructureConfigurationArn\":
\\\"arn:aws:imagebuilder:us-east-1:123456789012:infrastructure-configuration/
parallelclusterimage-1234abcd-56ef-78gh-90ij-abcd1234efgh\\\", \\\"ImageRecipeArn
\\\": \\\"arn:aws:imagebuilder:us-east-1:123456789012:image-recipe/
parallelclusterimage-custom-alinux2-image/3.1.2\\\", \\\"DistributionConfigurationArn
\\\": \\\"arn:aws:imagebuilder:us-east-1:123456789012:distribution-
configuration/parallelclusterimage-1234abcd-56ef-78gh-90ij-abcd1234efgh\\\",
\\\"EnhancedImageMetadataEnabled\\\": \\\"false\\\", \\\"Tags\\\": {\\\"parallelcluster:image_name\\\":
\\\"custom-alinux2-image\\\", \\\"parallelcluster:image_id\\\": \\\"custom-alinux2-image\\\"}}\",
  "stackId": "arn:aws:cloudformation:us-east-1:123456789012:stack/custom-alinux2-
image/1234abcd-56ef-78gh-90ij-abcd1234efgh",
  "stackName": "custom-alinux2-image",
  "logicalResourceId": "ParallelClusterImage",
  "resourceType": "AWS::ImageBuilder::Image",
  "timestamp": "2022-04-05T21:39:24.725Z"
}
```

pcluster list-clusters

Rufen Sie die Liste der vorhandenen Cluster ab.

```
pcluster list-clusters [-h]
                    [--cluster-status {CREATE_IN_PROGRESS,CREATE_FAILED,CREATE_COMPLETE,
DELETE_IN_PROGRESS,DELETE_FAILED,UPDATE_IN_PROGRESS,
UPDATE_COMPLETE,UPDATE_FAILED}]
                    [{CREATE_IN_PROGRESS,CREATE_FAILED,CREATE_COMPLETE,
```

```
DELETE_IN_PROGRESS,DELETE_FAILED,UPDATE_IN_PROGRESS,
UPDATE_COMPLETE,UPDATE_FAILED} ...]]
```

```
[--debug]
[--next-token NEXT_TOKEN]
[--query QUERY]
[--region REGION]
```

Benannte Argumente

-h, --help

Zeigt den Hilfetext für `pcluster list-clusters`.

```
--cluster-status {CREATE_IN_PROGRESS, CREATE_FAILED, CREATE_COMPLETE,
DELETE_IN_PROGRESS, DELETE_FAILED, UPDATE_IN_PROGRESS, UPDATE_COMPLETE,
UPDATE_FAILED} [{CREATE_IN_PROGRESS, CREATE_FAILED, CREATE_COMPLETE,
DELETE_IN_PROGRESS, DELETE_FAILED, UPDATE_IN_PROGRESS, UPDATE_COMPLETE,
UPDATE_FAILED} ...]
```

Gibt die Liste der Cluster-Status an, nach denen gefiltert werden soll. (Die Standardeinstellung ist.) `all`

--debug

Aktiviert die Debug-Protokollierung.

--next-token *NEXT_TOKEN*

Gibt das Token an, das für paginierte Anfragen verwendet werden soll.

--query *QUERY*

Gibt die JMESPath-Abfrage an, die für die Ausgabe ausgeführt werden soll.

--region, -r *REGION*

Gibt an, welche verwendet werden AWS-Region soll. Der AWS-Region muss mithilfe der `AWS_DEFAULT_REGION` Umgebungsvariablen, der `region` Einstellung im `[default]` Abschnitt der `~/.aws/config` Datei oder des `--region` Parameters angegeben werden.

Beispiel mit AWS ParallelCluster Version 3.1.4:

```
$ pcluster list-clusters
{
  "clusters": [
```

```
{
  "clusterName": "cluster-v3",
  "cloudformationStackStatus": "CREATE_COMPLETE",
  "cloudformationStackArn": "arn:aws:cloudformation:us-east-1:123456789012:stack/
cluster-v3/1234abcd-56ef-78gh-90ij-abcd1234efgh",
  "region": "us-east-1",
  "version": "3.1.4",
  "clusterStatus": "CREATE_COMPLETE"
}
]
```

pcluster list-cluster-log-streams

Ruft die Liste der Log-Streams ab, die einem Cluster zugeordnet sind.

```
pcluster list-cluster-log-streams [-h]
      --cluster-name CLUSTER_NAME
      [--filters FILTERS [FILTERS ...]]
      [--next-token NEXT_TOKEN] [--debug]
      [--query QUERY]
      [--region REGION]
```

Benannte Argumente

-h, --help

Zeigt den Hilfetext für `pcluster list-cluster-log-streams`.

--cluster-name, -n *CLUSTER_NAME*

Gibt den Namen des Clusters an.

--debug

Aktiviert die Debug-Protokollierung.

--filters *FILTERS* [*FILTERS* ...]

Gibt Filter für die Protokolldatenströme an. Format: `Name=a,Values=1` `Name=b,Values=2,3`.
Unterstützte Filter sind:

`private-dns-name`

Gibt die Kurzform des privaten DNS-Namens der Instanz an (z. B. `ip-10-0-0-101`).

node-type

Gibt den Knotentyp an, der einzig akzeptierte Wert für diesen Filter ist `HeadNode`.

--next-token *NEXT_TOKEN*

Gibt das Token an, das für paginierte Anfragen verwendet werden soll.

--query *QUERY*

Gibt die JMESPath-Abfrage an, die für die Ausgabe ausgeführt werden soll.

--region, -r *REGION*

Gibt an, welche verwendet werden AWS-Region soll. Der AWS-Region muss mithilfe der `AWS_DEFAULT_REGION` Umgebungsvariablen, der `region` Einstellung im `[default]` Abschnitt der `~/.aws/config` Datei oder des `--region` Parameters angegeben werden.

Beispiel mit AWS ParallelCluster Version 3.1.4:

```
$ pcluster list-cluster-log-streams \
  -n cluster-v3 \
  -r us-east-1 \
  --query 'logStreams[*].logStreamName'
[
  "ip-172-31-58-205.i-1234567890abcdef0.cfn-init",
  "ip-172-31-58-205.i-1234567890abcdef0.chef-client",
  "ip-172-31-58-205.i-1234567890abcdef0.cloud-init",
  "ip-172-31-58-205.i-1234567890abcdef0.clustermgtd",
  "ip-172-31-58-205.i-1234567890abcdef0.slurmctld",
  "ip-172-31-58-205.i-1234567890abcdef0.supervisord",
  "ip-172-31-58-205.i-1234567890abcdef0.system-messages"
]
```

pcluster list-images

Rufen Sie die Liste der vorhandenen benutzerdefinierten Bilder ab.

```
pcluster list-images [-h]
  --image-status {AVAILABLE,PENDING,FAILED}
  [--debug]
  [--next-token NEXT_TOKEN]
  [--query QUERY]
```

```
[--region REGION]
```

Benannte Argumente

-h, --help

Zeigt den Hilfetext für `pcluster list-images`.

--image-status {AVAILABLE,PENDING,FAILED}

Filtert die zurückgegebenen Bilder nach dem angegebenen Status.

--debug

Aktiviert die Debug-Protokollierung.

--next-token *NEXT_TOKEN*

Gibt das Token an, das für paginierte Anfragen verwendet werden soll.

--query *QUERY*

Gibt die JMESPath-Abfrage an, die für die Ausgabe ausgeführt werden soll.

--region, -r *REGION*

Gibt an, welche verwendet werden AWS-Region soll. Der AWS-Region muss mithilfe der `AWS_DEFAULT_REGION` Umgebungsvariablen, der `region` Einstellung im `[default]` Abschnitt der `~/.aws/config` Datei oder des `--region` Parameters angegeben werden.

Beispiel mit AWS ParallelCluster Version 3.1.2:

```
$ pcluster list-images --image-status AVAILABLE
{
  "images": [
    {
      "imageId": "custom-alinux2-image",
      "imageBuildStatus": "BUILD_COMPLETE",
      "ec2AmiInfo": {
        "amiId": "ami-1234abcd5678efgh"
      },
      "region": "us-east-1",
      "version": "3.1.2"
    }
  ]
}
```

```
}
```

pcluster list-image-log-streams

Ruft die Liste der Protokollstreams ab, die einem Bild zugeordnet sind.

```
pcluster list-image-log-streams [-h]
    --image-id IMAGE_ID
    [--next-token NEXT_TOKEN] [--debug]
    [--query QUERY]
    [--region REGION]
```

Benannte Argumente

-h, --help

Zeigt den Hilfetext für `pcluster list-image-log-streams`.

--image-id, -i *IMAGE_ID*

Gibt die ID des Bilds an.

--debug

Aktiviert die Debug-Protokollierung.

--next-token *NEXT_TOKEN*

Gibt das Token an, das für paginierte Anfragen verwendet werden soll.

--query *QUERY*

Gibt die JMESPath-Abfrage an, die für die Ausgabe ausgeführt werden soll.

--region, -r *REGION*

Gibt an, welche verwendet werden AWS-Region soll. Der AWS-Region muss mithilfe der `AWS_DEFAULT_REGION` Umgebungsvariablen, der `region` Einstellung im `[default]` Abschnitt der `~/.aws/config` Datei oder des `--region` Parameters angegeben werden.

Beispiel mit AWS ParallelCluster Version 3.1.2:

```
$ pcluster list-image-log-streams --image-id custom-alinux2-image --region us-east-1 --
query 'logStreams[*].logStreamName'
[
```



```
"3.0.0/1",  
"3.1.2/1"  
]
```

pcluster list-official-images

Beschreiben Sie offizielle AWS ParallelCluster AMIs.

```
pcluster list-official-images [-h]  
    [--architecture ARCHITECTURE]  
    [--debug]  
    [--os OS]  
    [--query QUERY]  
    [--region REGION]
```

Benannte Argumente

-h, --help

Zeigt den Hilfetext für `pcluster list-official-images`.

--architecture *ARCHITECTURE*

Gibt die Architektur an, die zum Filtern der Ergebnisse verwendet werden soll. Wenn dieser Parameter nicht angegeben wird, werden alle Architekturen zurückgegeben.

--debug

Aktiviert die Debug-Protokollierung.

--os *OS*

Gibt das Betriebssystem an, das zum Filtern der Ergebnisse verwendet werden soll. Wenn dieser Parameter nicht angegeben ist, werden alle Betriebssysteme zurückgegeben.

--query *QUERY*

Gibt die JMESPath-Abfrage an, die für die Ausgabe ausgeführt werden soll.

--region, -r *REGION*

Gibt an, welche verwendet werden AWS-Region soll. Das AWS-Region muss mithilfe der [Region-Einstellung](#) in der Image-Konfigurationsdatei, der `AWS_DEFAULT_REGION` Umgebungsvariablen, der `region` Einstellung im `[default]` Abschnitt der `~/.aws/config` Datei oder des `--region` Parameters angegeben werden.

Beispiel mit AWS ParallelCluster Version 3.1.2:

```
$ pcluster list-official-images
{
  "images": [
    {
      "amiId": "ami-015cfefb4e0d6306b2",
      "os": "ubuntu2004",
      "name": "aws-parallelcluster-3.1.2-ubuntu-2004-lts-hvm-x86_64-202202261505
2022-02-26T15-08-34.759Z",
      "version": "3.1.2",
      "architecture": "x86_64"
    },
    {
      "amiId": "ami-036f23237ce49d25b",
      "os": "ubuntu2204",
      "name": "aws-parallelcluster-3.1.2-ubuntu-1804-lts-hvm-x86_64-202202261505
2022-02-26T15-08-17.558Z",
      "version": "3.1.2",
      "architecture": "x86_64"
    },
    {
      "amiId": "ami-09e5327e694d89ef4",
      "os": "ubuntu2004",
      "name": "aws-parallelcluster-3.1.2-ubuntu-2004-lts-hvm-arm64-202202261505
2022-02-26T15-08-45.736Z",
      "version": "3.1.2",
      "architecture": "arm64"
    },
    {
      "amiId": "ami-0b9b0874c35f626ae",
      "os": "alinux2",
      "name": "aws-parallelcluster-3.1.2-amzn2-hvm-x86_64-202202261505
2022-02-26T15-08-31.311Z",
      "version": "3.1.2",
      "architecture": "x86_64"
    },
    {
      "amiId": "ami-0bf6d01f398f3737e",
      "os": "centos7",
      "name": "aws-parallelcluster-3.1.2-centos7-hvm-x86_64-202202261505
2022-02-26T15-08-25.001Z",
      "version": "3.1.2",
      "architecture": "x86_64"
    }
  ]
}
```

```

    },
    {
      "amiId": "ami-0d0de4f95f56374bc",
      "os": "alinux2",
      "name": "aws-parallelcluster-3.1.2-amzn2-hvm-arm64-202202261505
2022-02-26T15-08-46.088Z",
      "version": "3.1.2",
      "architecture": "arm64"
    },
    {
      "amiId": "ami-0ebf7bc54b8740dc6",
      "os": "ubuntu2204",
      "name": "aws-parallelcluster-3.1.2-ubuntu-1804-lts-hvm-arm64-202202261505
2022-02-26T15-08-45.293Z",
      "version": "3.1.2",
      "architecture": "arm64"
    }
  ]
}

```

pcluster ssh

Führt einen ssh Befehl aus, bei dem der Cluster-Benutzername und die IP-Adresse bereits ausgefüllt sind. Beliebige Argumente werden an das Ende der ssh Befehlszeile angehängt.

```

pcluster ssh [-h]
              --cluster-name CLUSTER_NAME
              [--debug]
              [--dryrun DRYRUN]
              [--region REGION]

```

Benannte Argumente

-h, --help

Zeigt den Hilfetext für `pcluster ssh`.

--cluster-name, -n *CLUSTER_NAME*

Gibt den Namen des Clusters an, mit dem eine Verbindung hergestellt werden soll.

--debug

Aktiviert die Debug-Protokollierung.

--dryrun *DRYRUN*

Wenn `true`, gibt die Befehlszeile aus, die ausgeführt werden würde, und wird beendet. (Die Standardeinstellung ist.) `false`

--region, -r *REGION*

Gibt an, welche verwendet AWS-Region werden soll. Der AWS-Region muss mithilfe der `AWS_DEFAULT_REGION` Umgebungsvariablen, der `region` Einstellung im `[default]` Abschnitt der `~/.aws/config` Datei oder des `--region` Parameters angegeben werden.

Beispiel:

```
$ pcluster ssh --cluster-name mycluster -i ~/.ssh/id_rsa
```

Führt einen `ssh` Befehl aus, bei dem der Benutzername und die IP-Adresse des Clusters vorausgefüllt sind:

```
ssh ec2-user@1.1.1.1 -i ~/.ssh/id_rsa
```

pcluster update-cluster

Aktualisiert einen vorhandenen Cluster so, dass er den Einstellungen einer angegebenen Konfigurationsdatei entspricht.

```
pcluster update-cluster [-h]
                        --cluster-configuration CLUSTER_CONFIGURATION
                        --cluster-name CLUSTER_NAME
                        [--debug]
                        [--dryrun DRYRUN]
                        [--force-update FORCE_UPDATE]
                        [--query QUERY]
                        [--region REGION]
                        [--suppress-validators SUPPRESS_VALIDATORS [SUPPRESS_VALIDATORS ...]]
                        [--validation-failure-level {INFO,WARNING,ERROR}]
```

Benannte Argumente

-h, --help

Zeigt den Hilfetext für `pcluster update-cluster`.

--cluster-configuration, -c *CLUSTER_CONFIGURATION*

Gibt die YAML-Cluster-Konfigurationsdatei an.

--cluster-name, -n *CLUSTER_NAME*

Gibt den Namen des Clusters an.

--debug

Aktiviert die Debug-Protokollierung.

--dryrun *DRYRUN*

Wenn `true`, führt die Validierung durch, ohne den Cluster zu aktualisieren und Ressourcen zu erstellen. Es kann verwendet werden, um die Image-Konfiguration und die Aktualisierungsanforderungen zu überprüfen. (Die Standardeinstellung ist `false`.)

--force-update *FORCE_UPDATE*

Wenn `true`, erzwingt die Aktualisierung, indem die Fehler bei der Aktualisierungsüberprüfung ignoriert werden. (Die Standardeinstellung ist.) `false`

--query *QUERY*

Gibt die JMESPath-Abfrage an, die für die Ausgabe ausgeführt werden soll.

--region, -r *REGION*

Gibt an, welche verwendet werden AWS-Region soll. Das AWS-Region muss mithilfe der [Region](#)Einstellung in der Cluster-Konfigurationsdatei, der `AWS_DEFAULT_REGION` Umgebungsvariablen, der `region` Einstellung im `[default]` Abschnitt der `~/.aws/config` Datei oder des `--region` Parameters angegeben werden.

--suppress-validators *SUPPRESS_VALIDATORS* [*SUPPRESS_VALIDATORS ...*]

Identifiziert einen oder mehrere zu unterdrückende Konfigurationsprüfer.

Format: `(i) ALL type:[A-Za-z0-9]+`

--validation-failure-level *{INFO,WARNING,ERROR}*

Gibt den Grad der bei der Aktualisierung gemeldeten Validierungsfehler an.

Beispiel mit AWS ParallelCluster Version 3.1.4:

```
$ pcluster update-cluster -c cluster-config.yaml -n cluster-v3 -r us-east-1
{
  "cluster": {
    "clusterName": "cluster-v3",
    "cloudformationStackStatus": "UPDATE_IN_PROGRESS",
    "cloudformationStackArn": "arn:aws:cloudformation:us-east-1:123456789012:stack/
cluster-v3/1234abcd-56ef-78gh-90ij-abcd1234efgh",
    "region": "us-east-1",
    "version": "3.1.4",
    "clusterStatus": "UPDATE_IN_PROGRESS"
  },
  "changeSet": [
    {
      "parameter": "HeadNode.Iam.S3Access",
      "requestedValue": {
        "BucketName": "pc-beta-test",
        "KeyName": "output",
        "EnableWriteAccess": false
      }
    },
    {
      "parameter": "HeadNode.Iam.S3Access",
      "currentValue": {
        "BucketName": "pcluster-east-test-bucket",
        "KeyName": "logs",
        "EnableWriteAccess": true
      }
    }
  ]
}
```

pcluster update-compute-fleet

Aktualisiert den Status der Cluster-Rechenflotte.

```
pcluster update-compute-fleet [-h]
    --cluster-name CLUSTER_NAME
    --status {START_REQUESTED,STOP_REQUESTED,ENABLED,DISABLED}

    [--debug]
    [--query QUERY]
    [--region REGION]
```

Benannte Argumente

-h, --help

Zeigt den Hilfetext für `pcluster update-compute-fleet`.

--cluster-name, -n *CLUSTER_NAME*

Gibt den Namen des Clusters an.

--status {*START_REQUESTED, STOP_REQUESTED, ENABLED, DISABLED*}

Gibt den Status an, der auf die Cluster-Rechenflotte angewendet wurde. Die Status A `START_REQUESTED` und B `STOP_REQUESTED` entsprechen dem Slurm-Scheduler, während die Status A `ENABLED` und D dem Scheduler `DISABLED` entsprechen. AWS Batch

--debug

Aktiviert die Debug-Protokollierung.

--query *QUERY*

Gibt die JMESPath-Abfrage an, die für die Ausgabe ausgeführt werden soll.

--region, -r *REGION*

Gibt an, welche verwendet werden AWS-Region soll. Der AWS-Region muss mithilfe der `AWS_DEFAULT_REGION` Umgebungsvariablen, der `region` Einstellung im `[default]` Abschnitt der `~/.aws/config` Datei oder des `--region` Parameters angegeben werden.

Beispiel mit AWS ParallelCluster Version 3.1.4:

```
$ pcluster update-compute-fleet -n cluster-v3 --status STOP_REQUESTED
{
  "status": "STOP_REQUESTED",
  "lastStatusUpdateTime": "2022-07-12T20:19:47.653Z"
}
```

pcluster version

Zeigt die Version von an. AWS ParallelCluster

```
pcluster version [-h] [--debug]
```

Benannte Argumente

-h, --help

Zeigt den Hilfetext für `pcluster version`.

--debug

Aktiviert die Debug-Protokollierung.

Beispiel mit AWS ParallelCluster Version 3.1.4:

```
$ pcluster version
{
  "version": "3.1.4"
}
```

pcluster3-config-converter

Liest eine Konfigurationsdatei der AWS ParallelCluster Version 2 und schreibt eine Konfigurationsdatei der AWS ParallelCluster Version 3.

```
pcluster3-config-converter [-h]
                          [-t CLUSTER_TEMPLATE]
                          [-c CONFIG_FILE]
                          [--force-convert]
                          [-o OUTPUT_FILE]
```

Benannte Argumente

-h, --help

Zeigt den Hilfetext für `pcluster3-config-converter`.

-t *CLUSTER_TEMPLATE*, --cluster-template *CLUSTER_TEMPLATE*

Gibt den [\[cluster\]Abschnitt](#) der Konfigurationsdatei an, der konvertiert werden soll. Wenn nicht angegeben, sucht das Skript nach dem [Cluster-Template-Parameter](#) im [\[global\]Abschnitt](#) oder sucht danach. [cluster default]

-c *CONFIG_FILE*, --config-file *CONFIG_FILE*

Gibt die zu lesende Konfigurationsdatei der AWS ParallelCluster Version 2 an.

--force-convert

Ermöglicht eine Konvertierung, auch wenn eine oder mehrere Einstellungen nicht unterstützt und nicht empfohlen werden.

-o *OUTPUT_FILE*, --output-file *OUTPUT_FILE*

Gibt die Konfigurationsdatei der AWS ParallelCluster Version 3 an, die geschrieben werden soll. Wenn dieser Parameter nicht angegeben ist, wird die Konfiguration nach stdout geschrieben.

Note

Der `pcluster3-config-converter` Befehl wurde in AWS ParallelCluster Version 3.0.1 hinzugefügt.

Konfigurationsdateien

AWS ParallelCluster verwendet YAML 1.1-Dateien für Konfigurationsparameter.

Themen

- [Cluster-Konfigurationsdatei](#)
- [Image-Konfigurationsdateien erstellen](#)

Cluster-Konfigurationsdatei

AWS ParallelCluster Version 3 verwendet separate Konfigurationsdateien, um die Definition der Cluster-Infrastruktur und die Definition von benutzerdefinierten AMIs zu steuern. Alle Konfigurationsdateien verwenden YAML 1.1-Dateien. Detaillierte Informationen zu jeder dieser Konfigurationsdateien sind unten verlinkt. Einige Beispielkonfigurationen finden Sie unter https://github.com/aws/aws-parallelcluster/tree/release-3.0/cli/tests/pcluster/example_configs.

Diese Objekte werden für die Clusterkonfiguration der AWS ParallelCluster Version 3 verwendet.

Themen

- [Eigenschaften der Cluster-Konfigurationsdatei](#)
- [Imds Abschnitt](#)
- [Image Abschnitt](#)

- [HeadNode Abschnitt](#)
- [Scheduling Abschnitt](#)
- [SharedStorage Abschnitt](#)
- [Iam Abschnitt](#)
- [LoginNodes Abschnitt](#)
- [Monitoring Abschnitt](#)
- [Tags Abschnitt](#)
- [AdditionalPackages Abschnitt](#)
- [DirectoryService Abschnitt](#)
- [DeploymentSettings Abschnitt](#)

Eigenschaften der Cluster-Konfigurationsdatei

Region(Fakultativ,String)

Gibt die AWS-Region für den Cluster an. z. B. us-east-2.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

CustomS3Bucket(Fakultativ,String)

Gibt den Namen eines Amazon S3 S3-Buckets an, der in Ihrem erstellt wurde, AWS-Konto um Ressourcen zu speichern, die von Ihren Clustern verwendet werden, z. B. die Cluster-Konfigurationsdatei. AWS ParallelCluster verwaltet in jedem, in dem Sie Cluster erstellen AWS-Region, einen Amazon S3 S3-Bucket. Standardmäßig sind diese Amazon S3 S3-Buckets benannt `parallelcluster-hash-v1-D0-NOT-DELETE`.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)
[Wenn Sie das Update erzwingen, wird der neue Wert ignoriert und der alte Wert verwendet.](#)

AdditionalResources(Fakultativ,String)

Definiert eine zusätzliche AWS CloudFormation Vorlage, die zusammen mit dem Cluster gestartet wird. Diese zusätzliche Vorlage wird für die Erstellung von Ressourcen verwendet, die sich außerhalb des Clusters befinden, aber Teil des Cluster-Lebenszyklus sind.

Der Wert muss eine HTTPS-URL zu einer öffentlichen Vorlage sein, in der alle Parameter angegeben sind.

Es ist kein Standardwert vorhanden.

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

Imds Abschnitt

(Optional) Gibt die Konfiguration des Global Instance Metadata Service (IMDS) an.

```
Imds:  
ImdsSupport: string
```

Imds-Eigenschaften

ImdsSupport(Fakultativ,String)

Gibt an, welche IMDS-Versionen in den Clusterknoten unterstützt werden. Unterstützte Werte sind `v1.0` und `v2.0`. Der Standardwert ist `v2.0`.

Wenn auf gesetzt `ImdsSupport` ist `v1.0`, werden sowohl IMDSv1 als auch IMDSv2 unterstützt.

Wenn auf gesetzt `ImdsSupport` ist `v2.0`, wird nur IMDSv2 unterstützt.

Weitere Informationen finden Sie unter [Verwenden von IMDSv2](#) im EC2-Benutzerhandbuch für Linux-Instances.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

Note

Ab AWS ParallelCluster 3.7.0 ist der `ImdsSupport` Standardwert. `v2.0` Wir empfehlen, dass Sie in Ihren Aufrufen für benutzerdefinierte Aktionen `ImdsSupport` auf IMDSv1 setzen `v2.0` und dieses durch IMDSv2 ersetzen.

Support für [Imds/ImdsSupport](#) wurde mit AWS ParallelCluster Version 3.3.0 hinzugefügt.

Image Abschnitt

(Erforderlich) Definiert das Betriebssystem für den Cluster.

```
Image:
```

`Os`: *string*
`CustomAmi`: *string*

Image-Eigenschaften

`Os`(Erforderlich,String)

Gibt das Betriebssystem an, das für den Cluster verwendet werden soll. Die unterstützten Werte sind `linux2`, `centos7`, `ubuntu2204`, `ubuntu2004`, `rhel8`, `rocky8`, `rhel9`, `rocky9`.

Note

RedHat Enterprise Linux 8.7 (`rhel8`) wird ab AWS ParallelCluster Version 3.6.0 hinzugefügt.

Wenn Sie Ihren Cluster für die Verwendung konfigurieren `rhel`, sind die On-Demand-Kosten für jeden Instance-Typ höher als bei der Konfiguration Ihres Clusters für die Verwendung anderer unterstützter Betriebssysteme. Weitere Informationen zur Preisgestaltung finden Sie unter [On-Demand-Preise](#) und [Wie wird Red Hat Enterprise Linux auf Amazon EC2 angeboten und wie hoch der Preis ist?](#) .


RedHat Enterprise Linux 9 (`rhel9`) wird ab AWS ParallelCluster Version 3.9.0 hinzugefügt.

Mit Ausnahme der in der folgenden Tabelle AWS-Regionen aufgeführten Spezifika, die nicht unterstützt werden. `centos7` Alle anderen AWS kommerziellen Regionen unterstützen alle der folgenden Betriebssysteme.

Partition (AWS-Regionen)	<code>linux2</code>	<code>centos7</code>	<code>ubuntu2204</code> und <code>ubuntu2004</code>	<code>rhel8</code>	<code>rhel9</code>
Kommerziell (Alle AWS-Regionen nicht ausdrücklich erwähnt)	True	True	True	True	True
AWS GovCloud (US-Ost) (<code>us-gov-east-1</code>)	True	False	True	True	True

Partition (AWS-Regionen)	alinux2	centos7	ubuntu2204 und ubuntu2004	rhel8	rhel9
AWS GovCloud (US-West) (us-gov-west-1)	True	False	True	True	True
China (Peking) (cn-north-1)	True	False	True	True	True
China (Ningxia) (cn-northwest-1)	True	False	True	True	True

Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.

 Note

AWS ParallelCluster 3.8.0 unterstützt Rocky Linux 8, aber vorgefertigte Rocky Linux 8-AMIs (für x86- und ARM-Architekturen) sind nicht verfügbar. AWS ParallelCluster 3.8.0 unterstützt die Erstellung von Clustern mit Rocky Linux 8 mithilfe benutzerdefinierter AMIs. Weitere Informationen finden Sie unter [Überlegungen zum Betriebssystem](#) AWS ParallelCluster 3.9.0 unterstützt Rocky Linux 9, aber vorgefertigte Rocky Linux 9-AMIs (für x86- und ARM-Architekturen) sind nicht verfügbar. AWS ParallelCluster 3.9.0 unterstützt die Erstellung von Clustern mit Rocky Linux 9 mithilfe benutzerdefinierter AMIs. Weitere Informationen finden Sie unter [Überlegungen zum Betriebssystem](#).

CustomAmi(Fakultativ), String

Gibt die ID eines benutzerdefinierten AMI an, das für die Head- und Compute-Knoten anstelle des Standard-AMI verwendet werden soll. Weitere Informationen finden Sie unter [AWS ParallelCluster AMI-Anpassung](#).

Wenn das benutzerdefinierte AMI zusätzliche Berechtigungen für seinen Start benötigt, müssen diese Berechtigungen sowohl den Benutzer- als auch den Headnode-Richtlinien hinzugefügt werden.

Wenn einem benutzerdefinierten AMI beispielsweise ein verschlüsselter Snapshot zugeordnet ist, sind die folgenden zusätzlichen Richtlinien sowohl in den Benutzer- als auch in den Headnode-Richtlinien erforderlich:

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "kms:DescribeKey",
        "kms:ReEncrypt*",
        "kms:CreateGrant",
        "kms:Decrypt"
      ],
      "Resource": [
        "arn:aws:kms:<AWS_REGION>:<AWS_ACCOUNT_ID>:key/<AWS_KMS_KEY_ID>"
      ]
    }
  ]
}
```

Um ein benutzerdefiniertes RedHat Enterprise Linux-AMI zu erstellen, müssen Sie das Betriebssystem für die Installation der Pakete konfigurieren, die von den RHUI (AWS) - Repositorys bereitgestellt werden: `rhel-<version>-baseos-rhui-rpms`, `undrhel-<version>-appstream-rhui-rpms`, `codeready-builder-for-rhel-<version>-rhui-rpms`. Darüber hinaus müssen die Repositorys auf dem benutzerdefinierten AMI `kernel-devel` Pakete auf derselben Version wie die laufende Kernelversion enthalten. `Kernel`.

Bekannte Einschränkungen:

- Nur RHEL 8.2 und höhere Versionen unterstützen FSx for Lustre.
- RHEL 8.7 Kernelversion 4.18.0-425.3.1.el8 unterstützt FSx for Lustre nicht.
- Nur RHEL 8.4 und höhere Versionen unterstützen EFA.

Informationen zur Fehlerbehebung bei benutzerdefinierten AMI-Validierungswarnungen finden Sie unter [Behebung von Problemen mit benutzerdefinierten AMIs](#).

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

HeadNode Abschnitt

(Erforderlich) Gibt die Konfiguration für den Hauptknoten an.

```
HeadNode:
  InstanceType: string
  Networking:
    SubnetId: string
    ElasticIp: string/boolean
    SecurityGroups:
      - string
    AdditionalSecurityGroups:
      - string
  Proxy:
    HttpProxyAddress: string
  DisableSimultaneousMultithreading: boolean
  Ssh:
    KeyName: string
    AllowedIps: string
  LocalStorage:
    RootVolume:
      Size: integer
      Encrypted: boolean
      VolumeType: string
      Iops: integer
      Throughput: integer
      DeleteOnTermination: boolean
    EphemeralVolume:
      MountDir: string
  SharedStorageType: string
  Dcv:
    Enabled: boolean
    Port: integer
    AllowedIps: string
  CustomActions:
    OnNodeStart:
      Sequence:
        - Script: string
          Args:
            - string
      Script: string
      Args:
        - string
    OnNodeConfigured:
```

```
Sequence:  
  - Script: string  
    Args:  
      - string  
Script: string  
Args:  
  - string  
OnNodeUpdated:  
  Sequence:  
    - Script: string  
      Args:  
        - string  
  Script: string  
  Args:  
    - string  
Iam:  
  InstanceRole: string  
  InstanceProfile: string  
  S3Access:  
    - BucketName: string  
      EnableWriteAccess: boolean  
      KeyName: string  
  AdditionalIamPolicies:  
    - Policy: string  
Imds:  
  Secured: boolean  
Image:  
  CustomAmi: string
```

HeadNode-Eigenschaften

InstanceType(Erforderlich,String)

Gibt den Instanztyp für den Hauptknoten an.

Gibt den Amazon EC2 EC2-Instance-Typ an, der für den Head-Knoten verwendet wird. Die Architektur des Instance-Typs muss mit der Architektur übereinstimmen, die für die Slurm [InstanceType](#)Einstellung AWS Batch [InstanceType](#)oder verwendet wurde.

Note

AWS ParallelCluster unterstützt die folgenden Instanztypen für die HeadNode Einstellung nicht.

- `hpc6id`

Wenn Sie einen `p4d`-Instance-Typ oder einen anderen Instance-Typ definieren, der über mehrere Netzwerkschnittstellen oder eine Netzwerkschnittstellenkarte verfügt, müssen Sie auf `elasticIp` einstellen, um öffentlichen Zugriff [ElasticIp](#) zu `true` gewähren. AWS Öffentliche IPs können nur Instances zugewiesen werden, die mit einer einzigen Netzwerkschnittstelle gestartet wurden. In diesem Fall empfehlen wir, ein [NAT-Gateway](#) zu verwenden, um öffentlichen Zugriff auf die Cluster-Rechenknoten zu gewähren. Weitere Informationen finden Sie unter [Zuweisen einer öffentlichen IPv4-Adresse beim Instance-Start](#) im Amazon EC2 EC2-Benutzerhandbuch für Linux-Instances.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

`DisableSimultaneousMultithreading`(Optional,) Boolean

Wenn `true`, deaktiviert Hyperthreading auf dem Hauptknoten. Der Standardwert ist `false`.

Nicht alle Instance-Typen können Hyperthreading deaktivieren. Eine Liste der Instance-Typen, die die Deaktivierung von Hyperthreading unterstützen, finden Sie unter [CPU-Kerne und Threads für jeden CPU-Kern pro Instance-Typ](#) im Amazon EC2 EC2-Benutzerhandbuch für Linux-Instances.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

`SharedStorageType`(Optional,) String

Gibt den Speichertyp an, der für intern gemeinsam genutzte Daten verwendet wird. Zu den intern gemeinsam genutzten Daten gehören Daten, die zur Verwaltung des Clusters AWS ParallelCluster verwendet werden, und die Standarddaten, `/home` sofern sie nicht im [SharedStorage Abschnitt](#) Mount-Verzeichnis angegeben sind, zum Mounten eines gemeinsam genutzten Dateisystem-Volumes. Weitere Informationen zu intern gemeinsam genutzten Daten finden Sie unter [AWS ParallelClusterInterne Verzeichnisse](#).

Wenn `Ebs`, was der Standardspeichertyp ist, exportiert der Hauptknoten Teile seines Root-Volumes als gemeinsam genutzte Verzeichnisse für Rechenknoten und Anmeldeknoten, die NFS verwenden.

Falls `Efs`, erstellt Parallelcluster ein EFS-Dateisystem, das für gemeinsam genutzte interne Daten verwendet wird und `/home`

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

Note

Wenn der Cluster horizontal skaliert wird, kann der EBS-Speichertyp zu Leistungsengpässen führen, da der Hauptknoten mithilfe von NFS-Exporten Daten vom Root-Volume mit den Rechenknoten teilt. Mit EFS können Sie NFS-Exporte bei der Skalierung Ihres Clusters vermeiden und damit verbundene Leistungsengpässe vermeiden. Es wird empfohlen, EBS zu wählen, um das maximale Lese-/Schreibpotenzial für kleine Dateien und den Installationsprozess zu nutzen. Wählen Sie EFS für Skalierung.

Networking

(Erforderlich) Definiert die Netzwerkkonfiguration für den Hauptknoten.

Networking:

SubnetId: *string*

ElasticIp: *string/boolean*

SecurityGroups:

- *string*

AdditionalSecurityGroups:

- *string*

Proxy:

HttpProxyAddress: *string*

Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.

Networking-Eigenschaften

SubnetId(Erforderlich,String)

Gibt die ID eines vorhandenen Subnetzes an, in dem der Hauptknoten bereitgestellt werden soll.

Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.

ElasticIp(Optional,String)

Erzeugt oder weist dem Hauptknoten eine Elastic IP-Adresse zu. Unterstützte Werte sind `true`/`false`, oder die ID einer vorhandenen Elastic IP-Adresse. Der Standardwert ist `false`.

Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.

SecurityGroups(Optional,[String])

Liste der Amazon VPC-Sicherheitsgruppen-IDs, die für den Hauptknoten verwendet werden sollen. Diese ersetzen die Sicherheitsgruppen, die AWS ParallelCluster erstellt werden, wenn diese Eigenschaft nicht enthalten ist.

Stellen Sie sicher, dass die Sicherheitsgruppen für Ihre [SharedStorage](#) Systeme richtig konfiguriert sind.

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

AdditionalSecurityGroups(Optional,[String])

Liste zusätzlicher Amazon VPC-Sicherheitsgruppen-IDs, die für den Hauptknoten verwendet werden sollen.

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

Proxy(Fakultativ)

Gibt die Proxyeinstellungen für den Hauptknoten an.

```
Proxy:  
  HttpProxyAddress: string
```

HttpProxyAddress(Optional,String)

Definiert einen HTTP- oder HTTPS-Proxy-Server, in der Regel `https://x.x.x.x:8080`.

Es ist kein Standardwert vorhanden.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

Ssh

(Optional) Definiert die Konfiguration für den SSH-Zugriff auf den Hauptknoten.

```
Ssh:  
  KeyName: string  
  AllowedIps: string
```

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

Ssh-Eigenschaften

KeyName(Optional,String)

Benennt ein vorhandenes Amazon EC2 EC2-Schlüsselpaar, um den SSH-Zugriff auf den Hauptknoten zu ermöglichen.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

AllowedIps(Optional,) String

Gibt den IP-Bereich im CIDR-Format oder eine Präfixlisten-ID für SSH-Verbindungen zum Hauptknoten an. Der Standardwert ist `0.0.0.0/0`.

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

LocalStorage

(Optional) Definiert die lokale Speicherkonfiguration für den Hauptknoten.

```
LocalStorage:  
  RootVolume:  
    Size: integer  
    Encrypted: boolean  
    VolumeType: string  
    Iops: integer  
    Throughput: integer  
    DeleteOnTermination: boolean  
  EphemeralVolume:  
    MountDir: string
```

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

LocalStorage-Eigenschaften

RootVolume(Erforderlich)

Gibt den Root-Volume-Speicher für den Hauptknoten an.

```
RootVolume:  
  Size: integer  
  Encrypted: boolean  
  VolumeType: string
```

```
Iops: integer  
Throughput: integer  
DeleteOnTermination: boolean
```

Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.

Size(Fakultativ,Integer)

Gibt die Größe des Stammvolumens des Kopfknotens in Gibibyte (GiB) an. Die Standardgröße stammt aus dem AMI. Die Verwendung einer anderen Größe erfordert, dass das AMI sie unterstützt `growroot`.

Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.

Encrypted(Fakultativ,Boolean)

Gibt an, ob das Root-Volume verschlüsselt ist. Der Standardwert ist `true`.

Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.

VolumeType(Fakultativ,String)

Gibt den [Amazon EBS-Volumetyp](#) an. Unterstützte Werte sind `gp2gp3`, `io1`, `io2`, `sc1st1`, und `standard`. Der Standardwert ist `gp3`.

Weitere Informationen finden Sie unter [Amazon EBS-Volumen-Typen](#) in der Amazon EC2-Benutzerhandbuch für Linux-Instances.

Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.

Iops(Fakultativ,Integer)

Definiert die Anzahl der IOPS für Volumes `gp3` vom Typ `io1` `io2`, und.

Der Standardwert, die unterstützten Werte und das `volume_size` Verhältnis `volume_iops` zum Verhältnis variieren je nach `VolumeType` und `Size`.

Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.

VolumeType = `io1`

Standard Iops = 100

Unterstützte Werte Iops = 100—64000 †

Maximales Iops Size Verhältnis = 50 IOPS pro GiB. 5000 IOPS erfordern einen Wert `Size` von mindestens 100 GiB.

VolumeType = io2

Standard Iops = 100

Unterstützte Werte Iops = 100—64000 (256000 für io2 Block Express-Volumes) †

Maximales Iops Size Verhältnis = 500 IOPS pro GiB. 5000 IOPS erfordern einen Wert Size von mindestens 10 GiB.

VolumeType = gp3

Standard Iops = 3000

Unterstützte Werte Iops = 3000—16000

Maximales Iops Size Verhältnis = 500 IOPS pro GiB. 5000 IOPS erfordern einen Wert Size von mindestens 10 GiB.

† Maximale IOPS wird nur für [Instances garantiert, die auf dem Nitro-System basieren](#) und mit mehr als 32.000 IOPS ausgestattet sind. Andere Instanzen garantieren bis zu 32.000 IOPS. Ältere io1 Volumes erreichen möglicherweise nicht die volle Leistung, es sei denn, Sie [ändern das](#) Volume. io2 Block Express-Volumes unterstützen Iops Werte bis zu 256000 für R5b Instance-Typen. Weitere Informationen finden Sie unter [io2Block Express-Volumes](#) im Amazon EC2 EC2-Benutzerhandbuch für Linux-Instances.

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

Throughput(Optional,Integer)

Definiert den Durchsatz für gp3 Volumetypen in MiB/s. Diese Einstellung ist nur gültig, wenn sie VolumeType ist. gp3 Der Standardwert ist 125. Unterstützte Werte: 125—1000 MiB/s

Das Verhältnis von Throughput zu Iops darf nicht mehr als 0,25 betragen. Der maximale Durchsatz von 1000 MiB/s setzt voraus, dass die Iops Einstellung mindestens 4000 beträgt.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

DeleteOnTermination(Fakultativ,) Boolean

Gibt an, ob das Root-Volume gelöscht werden soll, wenn der Hauptknoten beendet wird. Der Standardwert ist true.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

EphemeralVolume(Fakultativ)

Gibt Details für jedes Instance-Speicher-Volume an. Weitere Informationen finden Sie unter [Instance-Speicher-Volumes](#) im Amazon EC2 EC2-Benutzerhandbuch für Linux-Instances.

[EphemeralVolume](#):

[MountDir](#): *string*

Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.

MountDir(Optional,String)

Gibt das Mount-Verzeichnis für das Instance-Speicher-Volume an. Der Standardwert ist /scratch.

Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.

Dcv

(Optional) Definiert die Konfigurationseinstellungen für den NICE-DCV-Server, der auf dem Hauptknoten ausgeführt wird.

Weitere Informationen finden Sie unter [Stellen Sie mit NICE DCV eine Connect zum Hauptknoten her](#).

[Dcv](#):

[Enabled](#): *boolean*

[Port](#): *integer*

[AllowedIps](#): *string*

Important

Standardmäßig AWS ParallelCluster ist der NICE-DCV-Port, der eingerichtet wurde, für alle IPv4-Adressen geöffnet. Sie können jedoch nur dann eine Verbindung zu einem NICE-DCV-Port herstellen, wenn Sie die URL für die NICE-DCV-Sitzung haben und innerhalb von 30 Sekunden nach der Rückgabe der URL eine Verbindung zur NICE-DCV-Sitzung herstellen. `pcluster dcv-connect` Verwenden Sie die `AllowedIps` Einstellung, um den Zugriff auf den NICE-DCV-Port mit einem CIDR-formatierten IP-Bereich weiter einzuschränken, und verwenden Sie die `Port` Einstellung, um einen nicht standardmäßigen Port festzulegen.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

Dcv-Eigenschaften

Enabled(Erforderlich,) Boolean

Gibt an, ob NICE DCV auf dem Hauptknoten aktiviert ist. Der Standardwert ist `false`.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

Note

NICE DCV generiert automatisch ein selbstsigniertes Zertifikat, das zur Sicherung des Datenverkehrs zwischen dem NICE-DCV-Client und dem NICE-DCV-Server verwendet wird, der auf dem Hauptknoten ausgeführt wird. Informationen zum Konfigurieren Ihres eigenen Zertifikats finden Sie unter [NICE DCV HTTPS-Zertifikat](#).

Port(Optional,) Integer

Gibt den Port für NICE DCV an. Der Standardwert ist `8443`.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

AllowedIps(Optional, empfohlen,String)

Gibt den CIDR-formatierten IP-Bereich für Verbindungen zu NICE DCV an. Diese Einstellung wird nur beim AWS ParallelCluster Erstellen der Sicherheitsgruppe verwendet. Der Standardwert ist `0.0.0.0/0` und ermöglicht den Zugriff von jeder beliebigen Internetadresse.

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

CustomActions

(Optional) Gibt benutzerdefinierte Skripts an, die auf dem Hauptknoten ausgeführt werden sollen.

```
CustomActions:  
  OnNodeStart:  
    Sequence:  
      - Script: string
```



```

    Args:
      - string
    Script: string
    Args:
      - string
    OnNodeConfigured:
      Sequence:
        - Script: string
          Args:
            - string
          Script: string
          Args:
            - string
      OnNodeUpdated:
        Sequence:
          - Script: string
            Args:
              - string
            Script: string
            Args:
              - string

```

CustomActions-Eigenschaften

OnNodeStart(Fakultativ)

Gibt ein einzelnes Skript oder eine Sequenz von Skripten an, die auf dem Hauptknoten ausgeführt werden sollen, bevor eine Bootstrap-Aktion zur Knotenbereitstellung gestartet wird. Weitere Informationen finden Sie unter [Benutzerdefinierte Bootstrap-Aktionen](#).

Sequence(Optional)

Liste der auszuführenden Skripts. AWS ParallelCluster führt die Skripts in derselben Reihenfolge aus, in der sie in der Konfigurationsdatei aufgeführt sind, beginnend mit der ersten.

Script(Erforderlich,String)

Gibt die zu verwendende Datei an. Der Dateipfad kann mit `https://` oder `s3://` beginnen.

Args(Fakultativ,[String])

Liste der Argumente, die an das Skript übergeben werden sollen.

`Script(Erforderlich,String)`

Gibt die Datei an, die für ein einzelnes Skript verwendet werden soll. Der Dateipfad kann mit `https://` oder `beginners3://`.

`Args(Fakultativ,[String])`

Liste der Argumente, die an das einzelne Skript übergeben werden sollen.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

`OnNodeConfigured(Fakultativ)`

Gibt ein einzelnes Skript oder eine Sequenz von Skripten an, die auf dem Hauptknoten ausgeführt werden sollen, nachdem die Bootstrap-Aktionen des Knotens abgeschlossen sind. Weitere Informationen finden Sie unter [Benutzerdefinierte Bootstrap-Aktionen](#).

`Sequence(Fakultativ)`

Gibt die Liste der auszuführenden Skripten an.

`Script(Erforderlich,String)`

Gibt die zu verwendende Datei an. Der Dateipfad kann mit `https://` oder `beginners3://`.

`Args(Fakultativ,[String])`

Liste der Argumente, die an das Skript übergeben werden sollen.

`Script(Erforderlich,String)`

Gibt die Datei an, die für ein einzelnes Skript verwendet werden soll. Der Dateipfad kann mit `https://` oder `beginners3://`.

`Args(Fakultativ,[String])`

Liste der Argumente, die an das einzelne Skript übergeben werden sollen.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

`OnNodeUpdated(Fakultativ)`

Gibt ein einzelnes Skript oder eine Sequenz von Skripten an, die nach Abschluss der Knotenaktualisierungsaaktionen auf dem Hauptknoten ausgeführt werden sollen. Weitere Informationen finden Sie unter [Benutzerdefinierte Bootstrap-Aktionen](#).

Sequence(Optional)

Gibt die Liste der auszuführenden Skripten an.

`Script(Erforderlich,String)`

Gibt die zu verwendende Datei an. Der Dateipfad kann mit `https://` oder `beginners3://`.

`Args(Fakultativ,[String])`

Liste der Argumente, die an das Skript übergeben werden sollen.

`Script(Erforderlich,String)`

Gibt die Datei an, die für das einzelne Skript verwendet werden soll. Der Dateipfad kann mit `https://` oder `beginners3://`.

`Args(Fakultativ,[String])`

Liste der Argumente, die an das einzelne Skript übergeben werden sollen.

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

Note

`OnNodeUpdated` wird ab AWS ParallelCluster 3.4.0 hinzugefügt.
`Sequence` wird ab AWS ParallelCluster Version 3.6.0 hinzugefügt. Wenn Sie angeben `Sequence`, können Sie mehrere Skripts für eine benutzerdefinierte Aktion auflisten. AWS ParallelCluster unterstützt weiterhin die Konfiguration einer benutzerdefinierten Aktion mit einem einzigen Skript, ohne dies einzuschließen `Sequence`. AWS ParallelCluster unterstützt nicht, sowohl ein einzelnes Skript als auch `Sequence` dieselbe benutzerdefinierte Aktion einzubeziehen.

Iam

(Optional) Gibt entweder eine Instanzrolle oder ein Instanzprofil an, das auf dem Hauptknoten verwendet werden soll, um die Standard-Instanzrolle oder das Instanzprofil für den Cluster zu überschreiben.

`Iam:`

`InstanceRole: string`

```
InstanceProfile: string  
S3Access:  
- BucketName: string  
  EnableWriteAccess: boolean  
  KeyName: string  
AdditionalIamPolicies:  
- Policy: string
```

Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.

Iam-Eigenschaften

InstanceProfile(Optional,String)

Gibt ein Instanzprofil an, um das standardmäßige Instanzprofil des Hauptknotens zu überschreiben. Sie können nicht sowohl InstanceProfile als auch InstanceRole angeben. Das Format ist `arn:Partition:iam::Account:instance-profile/InstanceProfileName`.

Wenn dies angegeben ist, können die AdditionalIamPolicies Einstellungen S3Access und nicht angegeben werden.

Es wird empfohlen, eine oder beide AdditionalIamPolicies Einstellungen für S3Access und anzugeben, da hinzugefügte Funktionen AWS ParallelCluster häufig neue Berechtigungen erfordern.

Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.

InstanceRole(Optional,String)

Gibt eine Instanzrolle an, um die standardmäßige Instanzrolle für den Hauptknoten zu überschreiben. Sie können nicht sowohl InstanceProfile als auch InstanceRole angeben. Das Format ist `arn:Partition:iam::Account:role/RoleName`.

Wenn dies angegeben ist, können die AdditionalIamPolicies Einstellungen S3Access und nicht angegeben werden.

Es wird empfohlen, eine oder beide AdditionalIamPolicies Einstellungen für S3Access und anzugeben, da hinzugefügte Funktionen AWS ParallelCluster häufig neue Berechtigungen erfordern.

Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.

S3Access

S3Access(Fakultativ)

Gibt einen Bucket an. Dies wird verwendet, um Richtlinien zu generieren, um den angegebenen Zugriff auf den Bucket zu gewähren.

Wenn dies angegeben ist, können die InstanceRole Einstellungen InstanceProfile und nicht angegeben werden.

Es wird empfohlen, eine oder beide AdditionalIamPolicies Einstellungen für S3Access und anzugeben, da hinzugefügte Funktionen AWS ParallelCluster häufig neue Berechtigungen erfordern.

S3Access:

- BucketName: *string*
- EnableWriteAccess: *boolean*
- KeyName: *string*

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

BucketName(Erforderlich,String)

Der Name des -Buckets.

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

KeyName(Fakultativ,String)

Der Schlüssel für den Eimer. Der Standardwert lautet "*".

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

EnableWriteAccess(Fakultativ,Boolean)

Gibt an, ob der Schreibzugriff für den Bucket aktiviert ist. Der Standardwert ist `false`.

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

AdditionalIamPolicies

AdditionalIamPolicies(Fakultativ)

Gibt eine Liste von Amazon Resource Names (ARNs) von IAM-Richtlinien für Amazon EC2 an. Diese Liste ist der Root-Rolle angehängt, die für den Hauptknoten verwendet wird, zusätzlich zu den Berechtigungen, die für erforderlich sind. AWS ParallelCluster

Ein IAM-Richtlinienname und sein ARN sind unterschiedlich. Namen können nicht verwendet werden.

Wenn dies angegeben ist, können die InstanceRole Einstellungen InstanceProfile und nicht angegeben werden.

Wir empfehlen die Verwendung, AdditionalIamPolicies da sie zu den erforderlichen Berechtigungen hinzugefügt AdditionalIamPolicies werden und alle erforderlichen Berechtigungen enthalten InstanceRole müssen. AWS ParallelCluster Die erforderlichen Berechtigungen ändern sich häufig von Version zu Version, da Funktionen hinzugefügt werden.

Es ist kein Standardwert vorhanden.

AdditionalIamPolicies:

- Policy: *string*

Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.

Policy(Fakultativ,[String])

Liste der IAM-Richtlinien.

Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.

Imsds

(Optional) Gibt die Eigenschaften für den Instance-Metadatendienst (IMDS) an. Weitere Informationen finden Sie unter [So funktioniert der Instance-Metadaten-Service Version 2](#) im Amazon EC2 EC2-Benutzerhandbuch für Linux-Instances.

Imsds:

Secured: *boolean*

Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.

Imds-Eigenschaften

Secured(Fakultativ,Boolean)

Wenn `true`, beschränkt den Zugriff auf das IMDS des Hauptknotens (und die Anmeldeinformationen des Instanzprofils) auf eine Untergruppe von Superusern.

Wenn `false` jeder Benutzer im Hauptknoten Zugriff auf das IMDS des Hauptknotens hat.

Die folgenden Benutzer haben Zugriff auf das IMDS des Hauptknotens:

- Root-Benutzer
- Cluster-Administratorbenutzer (`pc-cluster-admin` standardmäßig)
- betriebssystemspezifischer Standardbenutzer (unter `ec2-user` Amazon Linux 2 und RedHat unter `ubuntu` Ubuntu 18.04 `centos` unter CentOS 7)

Der Standardwert ist `true`.

Die `default` Benutzer sind dafür verantwortlich, dass ein Cluster über die Berechtigungen verfügt, die er für die Interaktion mit Ressourcen benötigt. AWS Wenn Sie den `default` Benutzer-IMDS-Zugriff deaktivieren, AWS ParallelCluster können Sie die Rechenknoten nicht verwalten und funktionieren nicht mehr. Deaktivieren Sie den `default` Benutzer-IMDS-Zugriff nicht.

Wenn einem Benutzer Zugriff auf das IMDS des Hauptknotens gewährt wird, kann er die im [Instanzprofil des Hauptknotens](#) enthaltenen Berechtigungen verwenden. Sie können diese Berechtigungen beispielsweise verwenden, um EC2-Instances zu starten oder das Passwort für eine AD-Domäne zu lesen, für deren Verwendung der Cluster für die Authentifizierung konfiguriert ist.

Um den IMDS-Zugriff einzuschränken, AWS ParallelCluster verwaltet eine Kette von `iptables` Cluster-Benutzer mit `sudo` Zugriff können den Zugriff auf das IMDS des Hauptknotens für andere einzelne Benutzer, einschließlich `default` Benutzer, selektiv aktivieren oder deaktivieren, indem sie den folgenden Befehl ausführen:

```
$ sudo /opt/parallelcluster/scripts/imds/imds-access.sh --allow <USERNAME>
```

Sie können den Benutzer-IMDS-Zugriff mit der `--deny` Option für diesen Befehl deaktivieren.

Wenn Sie unwissentlich den `default` Benutzer-IMDS-Zugriff deaktivieren, können Sie die Berechtigung mithilfe der Option wiederherstellen. `--allow`

Note

Jede Anpassung von iptables oder ip6tables Regeln kann den Mechanismus beeinträchtigen, der zur Beschränkung des IMDS-Zugriffs auf den Hauptknoten verwendet wird.

Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.

Image

(Optional) Definiert ein benutzerdefiniertes Image für den Hauptknoten.

Image:

CustomAmi: *string*

Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.

Image-Eigenschaften

CustomAmi(Optional,String)

Gibt die ID eines benutzerdefinierten AMI an, das für den Hauptknoten anstelle des Standard-AMI verwendet werden soll. Weitere Informationen finden Sie unter [AWS ParallelCluster AMI-Anpassung](#).

Wenn das benutzerdefinierte AMI zusätzliche Berechtigungen für seinen Start benötigt, müssen diese Berechtigungen sowohl den Benutzer- als auch den Headnode-Richtlinien hinzugefügt werden.

Wenn einem benutzerdefinierten AMI beispielsweise ein verschlüsselter Snapshot zugeordnet ist, sind die folgenden zusätzlichen Richtlinien sowohl in den Benutzer- als auch in den Headnode-Richtlinien erforderlich:

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "kms:DescribeKey",
```



```

        "kms:ReEncrypt*",
        "kms:CreateGrant",
        "kms:Decrypt"
    ],
    "Resource": [
        "arn:aws:kms:<AWS_REGION>:<AWS_ACCOUNT_ID>:key/<AWS_KMS_KEY_ID>"
    ]
}
]
}

```

Informationen zur Fehlerbehebung bei benutzerdefinierten AMI-Validierungswarnungen finden Sie unter [Behebung von Problemen mit benutzerdefinierten AMIs](#).

Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.

Scheduling Abschnitt

(Erforderlich) Definiert den Job Scheduler, der im Cluster verwendet wird, und die Compute-Instances, die der Job Scheduler verwaltet. Sie können entweder den Slurm oder AWS Batch - Scheduler verwenden. Jeder unterstützt einen anderen Satz von Einstellungen und Eigenschaften.

Themen

- [Scheduling-Eigenschaften](#)
- [AwsBatchQueues](#)
- [SlurmQueues](#)
- [SlurmSettings](#)

Scheduling:

Scheduler: slurm

ScalingStrategy: *string*

SlurmSettings:

MungeKeySecretArn: *string*

ScaledownIdleTime: *integer*

QueueUpdateStrategy: *string*

EnableMemoryBasedScheduling: *boolean*

CustomSlurmSettings: *[dict]*

CustomSlurmSettingsIncludeFile: *string*

Database:

Uri: *string*

UserName: *string*

PasswordSecretArn: *string*

DatabaseName: *string*

Dns:

DisableManagedDns: *boolean*

HostedZoneId: *string*

UseEc2Hostnames: *boolean*

SlurmQueues:

- Name: *string*

ComputeSettings:

LocalStorage:

RootVolume:

Size: *integer*

Encrypted: *boolean*

VolumeType: *string*

Iops: *integer*

Throughput: *integer*

EphemeralVolume:

MountDir: *string*

CapacityReservationTarget:

CapacityReservationId: *string*

CapacityReservationResourceGroupArn: *string*

CapacityType: *string*

AllocationStrategy: *string*

JobExclusiveAllocation: *boolean*

CustomSlurmSettings: *dict*

Tags:

- Key: *string*

Value: *string*

HealthChecks:

Gpu:

Enabled: *boolean*

Networking:

SubnetIds:

- *string*

AssignPublicIp: *boolean*

SecurityGroups:

- *string*

AdditionalSecurityGroups:

- *string*

PlacementGroup:

Enabled: *boolean*

Id: *string*

```
Name: string
Proxy:
  HttpProxyAddress: string
ComputeResources:
  - Name: string
    InstanceType: string
    Instances:
      - InstanceType: string
    MinCount: integer
    MaxCount: integer
    DynamicNodePriority: integer
    StaticNodePriority: integer
    SpotPrice: float
    DisableSimultaneousMultithreading: boolean
    SchedulableMemory: integer
    HealthChecks:
      Gpu:
        Enabled: boolean
      Efa:
        Enabled: boolean
        GdrSupport: boolean
    CapacityReservationTarget:
      CapacityReservationId: string
      CapacityReservationResourceGroupArn: string
    Networking:
      PlacementGroup:
        Enabled: boolean
        Name: string
      CustomSlurmSettings: dict
    Tags:
      - Key: string
        Value: string
  CustomActions:
    OnNodeStart:
      Sequence:
        - Script: string
          Args:
            - string
          Script: string
          Args:
            - string
      OnNodeConfigured:
        Sequence:
          - Script: string
```

```
    Args:
      - string
    Script: string
    Args:
      - string
  Iam:
    InstanceProfile: string
    InstanceRole: string
    S3Access:
      - BucketName: string
        EnableWriteAccess: boolean
        KeyName: string
    AdditionalIamPolicies:
      - Policy: string
  Image:
    CustomAmi: string
```

Scheduling:

Scheduler: awsbatch

AwsBatchQueues:

- Name: *string*

CapacityType: *string*

Networking:

SubnetIds:

- *string*

AssignPublicIp: *boolean*

SecurityGroups:

- *string*

AdditionalSecurityGroups:

- *string*

ComputeResources: # this maps to a Batch compute environment (initially we support only 1)

- Name: *string*

InstanceTypes:

- *string*

MinvCpus: *integer*

DesiredvCpus: *integer*

MaxvCpus: *integer*

SpotBidPercentage: *float*

Scheduling-Eigenschaften

Scheduler(Erforderlich,String)

Gibt den Typ des verwendeten Schedulers an. Unterstützte Werte sind `slurm` und `awsbatch`.

Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.

Note

`awsbatch` unterstützt nur das `alinux2` Betriebssystem und die `x86_64` Plattform.

ScalingStrategy(Fakultativ,String)

Ermöglicht es Ihnen, auszuwählen, wie dynamische Slurm-Knoten skaliert werden sollen. Unterstützte Werte sind `all-or-nothing`, `greedy-all-or-nothing` und `best-effort`. Der Standardwert ist `all-or-nothing`.

Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.

Note

Die Skalierungsstrategie gilt nur für Knoten, die von Slurm wieder aufgenommen werden sollen, nicht für Knoten, die irgendwann schon laufen.

- `all-or-nothing` Diese Strategie folgt strikt einer `all-or-nothing-approach`, die darauf abzielt, inaktive Instanzen am Ende des Skalierungsprozesses zu vermeiden. Sie arbeitet auf einer `all-or-nothing` Basis, was bedeutet, dass sie entweder vollständig oder gar nicht skaliert wird. Beachten Sie, dass aufgrund vorübergehend gestarteter Instanzen zusätzliche Kosten anfallen können, wenn Jobs mehr als 500 Knoten erfordern oder sich über mehrere Rechenressourcen erstrecken. Diese Strategie hat den niedrigsten Durchsatz unter den drei möglichen Skalierungsstrategien. Die Skalierungszeit hängt von der Anzahl der Jobs ab, die pro Ausführung des Slurm-Resume-Programms eingereicht wurden. Außerdem können Sie nicht weit über das Standardlimit für RunInstances Ressourcenkonten pro Ausführung hinaus skalieren, das standardmäßig bei 1000 Instanzen liegt. Weitere Einzelheiten finden Sie in der Dokumentation zur Drosselung der [AWS EC2-API](#)
- `greedy-all-or-nothing` Ähnlich wie bei der `all-or-nothing` Strategie zielt sie darauf ab, inaktive Instanzen nach der Skalierung zu vermeiden. Diese Strategie ermöglicht

eine vorübergehende Überskalierung während des Skalierungsprozesses, um einen höheren Durchsatz als bei der all-or-nothing Methode zu erreichen, hat aber auch dasselbe Skalierungslimit von 1000 Instanzen wie beim RunInstances Ressourcenkontolimit.

- **best-effort** Bei dieser Strategie wird ein hoher Durchsatz priorisiert, auch wenn dies bedeutet, dass einige Instanzen am Ende des Skalierungsprozesses möglicherweise inaktiv sind. Es wird versucht, so viele Knoten zuzuweisen, wie von den Jobs angefordert werden, aber es besteht die Möglichkeit, dass nicht die gesamte Anfrage erfüllt wird. Im Gegensatz zu den anderen Strategien können beim Best-Effort-Ansatz mehr Instanzen als das RunInstances Standardlimit akkumuliert werden, allerdings auf Kosten ungenutzter Ressourcen während der Ausführung mehrerer Skalierungsprozesse.

Jede Strategie ist so konzipiert, dass sie unterschiedlichen Skalierungsanforderungen gerecht wird, sodass Sie eine auswählen können, die Ihren spezifischen Anforderungen und Einschränkungen entspricht.

AwsBatchQueues

(Optional) Die AWS Batch Warteschlangeneinstellungen. Es wird nur eine Warteschlange unterstützt. Wenn auf gesetzt [Scheduler](#) ist `awsbatch`, ist dieser Abschnitt erforderlich. Weitere Informationen zum `awsbatch` Scheduler finden Sie unter [Netzwerkconfiguration](#) und [AWS Batch \(awsbatch\)](#).

AwsBatchQueues:

- Name: *string*
- CapacityType: *string*
- Networking:
 - SubnetIds:
 - *string*
 - AssignPublicIp: *boolean*
 - SecurityGroups:
 - *string*
 - AdditionalSecurityGroups:
 - *string*
- ComputeResources: # this maps to a Batch compute environment (initially we support only 1)
 - Name: *string*
 - InstanceTypes:
 - *string*
 - MinvCpus: *integer*
 - DesiredvCpus: *integer*
 - MaxvCpus: *integer*

`SpotBidPercentage`: *float*

Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.

AwsBatchQueues-Eigenschaften

Name(Erforderlich,**String**)

Der Name der AWS Batch Warteschlange.

Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.

CapacityType(Fakultativ,**String**)

Der Typ der Rechenressourcen, die die AWS Batch Warteschlange verwendet. Unterstützte Werte sind ONDEMAND, SPOT oder CAPACITY_BLOCK. Der Standardwert ist ONDEMAND.

Note

Wenn Sie diese Einstellung CapacityType auf festlegen SPOT, muss Ihr Konto eine AWSServiceRoleForEC2Spot dienstbezogene Rolle enthalten. Sie können diese Rolle mit dem folgenden AWS CLI Befehl erstellen.

```
$ aws iam create-service-linked-role --aws-service-name spot.amazonaws.com
```

Weitere Informationen finden Sie unter [Service-verknüpfte Rolle für Spot-Instance-Anfragen](#) im Amazon EC2 EC2-Benutzerhandbuch für Linux-Instances.

Aktualisierungsrichtlinie: Die Rechenflotte muss gestoppt werden, damit diese Einstellung für ein Update geändert werden kann.

Networking

(Erforderlich) Definiert die Netzwerkkonfiguration für die AWS Batch Warteschlange.

Networking:

SubnetIds:

- *string*

AssignPublicIp: *boolean*

SecurityGroups:

- *string*

AdditionalSecurityGroups:

- *string*

Networking-Eigenschaften**SubnetIds**(Erforderlich, **[String]**)

Gibt die ID eines vorhandenen Subnetzes an, in dem die AWS Batch Warteschlange bereitgestellt werden soll. Derzeit wird nur ein Subnetz unterstützt.

[Aktualisierungsrichtlinie: Die Rechenflotte muss gestoppt werden, damit diese Einstellung für ein Update geändert werden kann.](#)

AssignPublicIp(Fakultativ, **String**)

Erzeugt oder weist den Knoten in der AWS Batch Warteschlange eine öffentliche IP-Adresse zu. Unterstützte Werte sind `true` und `false`. Die Standardeinstellung hängt von dem Subnetz ab, das Sie angegeben haben.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

SecurityGroups(Optional, **[String]**)

Liste der Sicherheitsgruppen, die die AWS Batch Warteschlange verwendet. Wenn Sie keine Sicherheitsgruppen angeben, AWS ParallelCluster erstellt neue Sicherheitsgruppen.

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

AdditionalSecurityGroups(Optional, **[String]**)

Liste der Sicherheitsgruppen, die die AWS Batch Warteschlange verwendet.

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

ComputeResources

(Erforderlich) Definiert die ComputeResources Konfiguration für die AWS Batch Warteschlange.

```
ComputeResources: # this maps to a Batch compute environment (initially we support only 1)
- Name: string
  InstanceTypes:
    - string
  MinvCpus: integer
  DesiredvCpus: integer
```


`MaxvCpus`: *integer*
`SpotBidPercentage`: *float*

ComputeResources-Eigenschaften

Name(Erforderlich,**String**)

Der Name der AWS Batch Warteschlangencomputer-Umgebung.

[Aktualisierungsrichtlinie: Die Rechenflotte muss gestoppt werden, damit diese Einstellung für ein Update geändert werden kann.](#)

InstanceTypes(Erforderlich,**[String]**)

Das Array der Instanztypen für die AWS Batch Rechenumgebung. Alle Instanztypen müssen die x86_64 Architektur verwenden.

[Aktualisierungsrichtlinie: Die Rechenflotte muss gestoppt werden, damit diese Einstellung für ein Update geändert werden kann.](#)

MinvCpus(Fakultativ,**Integer**)

Die Mindestanzahl von vCPUs, die eine AWS Batch Rechenumgebung verwenden kann.

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

DesiredVcpus(Optional,) **Integer**

Die gewünschte Anzahl von vCPUs in der AWS Batch Rechenumgebung. AWS Batch passt diesen Wert zwischen `MinvCpus` und `MaxvCpus` basierend auf dem Bedarf in der Auftragswarteschlange an.

[Aktualisierungsrichtlinie: Diese Einstellung wird während eines Updates nicht analysiert.](#)

MaxvCpus(Fakultativ,**Integer**)

Die maximale Anzahl von vCPUs für die AWS Batch Rechenumgebung. Sie können diesen Wert nicht auf einen Wert setzen, der niedriger ist als `DesiredVcpus`.

[Aktualisierungsrichtlinie: Diese Einstellung kann während eines Updates nicht verringert werden.](#)

SpotBidPercentage(Fakultativ,**Float**)

Der maximale Prozentsatz des On-Demand-Preises für den Instance-Typ, den ein EC2-Spot-Instance-Preis erreichen kann, bevor Instances gestartet werden. Der Standardwert ist 100 (100%). Der unterstützte Bereich ist 1 -100.

Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.

SlurmQueues

(Optional) Einstellungen für die Slurm Warteschlange. Wenn auf gesetzt [Scheduleristslurm](#), ist dieser Abschnitt erforderlich.

SlurmQueues:

- Name: *string*
- ComputeSettings:
 - LocalStorage:
 - RootVolume:
 - Size: *integer*
 - Encrypted: *boolean*
 - VolumeType: *string*
 - Iops: *integer*
 - Throughput: *integer*
 - EphemeralVolume:
 - MountDir: *string*
 - CapacityReservationTarget:
 - CapacityReservationId: *string*
 - CapacityReservationResourceGroupArn: *string*
 - CapacityType: *string*
 - AllocationStrategy: *string*
 - JobExclusiveAllocation: *boolean*
 - CustomSlurmSettings: *dict*
- Tags:
 - Key: *string*
 - Value: *string*
- HealthChecks:
 - Gpu:
 - Enabled: *boolean*
- Networking:
 - SubnetIds:
 - *string*
 - AssignPublicIp: *boolean*
 - SecurityGroups:
 - *string*
 - AdditionalSecurityGroups:
 - *string*
 - PlacementGroup:
 - Enabled: *boolean*
 - Id: *string*

```
Name: string
Proxy:
  HttpProxyAddress: string
ComputeResources:
  - Name: string
    InstanceType: string
    Instances:
      - InstanceType: string
    MinCount: integer
    MaxCount: integer
    DynamicNodePriority: integer
    StaticNodePriority: integer
    SpotPrice: float
    DisableSimultaneousMultithreading: boolean
    SchedulableMemory: integer
    HealthChecks:
      Gpu:
        Enabled: boolean
      Efa:
        Enabled: boolean
        GdrSupport: boolean
    CapacityReservationTarget:
      CapacityReservationId: string
      CapacityReservationResourceGroupArn: string
    Networking:
      PlacementGroup:
        Enabled: boolean
        Name: string
      CustomSlurmSettings: dict
    Tags:
      - Key: string
        Value: string
  CustomActions:
    OnNodeStart:
      Sequence:
        - Script: string
          Args:
            - string
          Script: string
          Args:
            - string
      OnNodeConfigured:
        Sequence:
          - Script: string
```

```
  Args:
    - string
  Script: string
  Args:
    - string
  Iam:
    InstanceProfile: string
    InstanceRole: string
    S3Access:
      - BucketName: string
        EnableWriteAccess: boolean
        KeyName: string
    AdditionalIamPolicies:
      - Policy: string
  Image:
    CustomAmi: string
```

Richtlinie aktualisieren: Bei dieser Einstellung für Listenwerte kann während eines Updates ein neuer Wert hinzugefügt werden, oder die Rechenflotte muss gestoppt werden, wenn ein vorhandener Wert entfernt wird.

SlurmQueues-Eigenschaften

Name(Erforderlich,String)

Der Name der Slurm Warteschlange.

Note

Die Clustergröße kann sich während eines Updates ändern. Weitere Informationen finden Sie unter [Größe und Aktualisierung der Clusterkapazität](#)

Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.

CapacityReservationTarget

Note

CapacityReservationTarget wird mit AWS ParallelCluster Version 3.3.0 hinzugefügt.

CapacityReservationTarget:**CapacityReservationId:** *string***CapacityReservationResourceGroupArn:** *string*

Gibt die On-Demand-Kapazitätsreservierung für die Rechenressourcen der Warteschlange an.

CapacityReservationId(Optional,**String**)

Die ID der vorhandenen Kapazitätsreservierung, die für die Rechenressourcen der Warteschlange als Ziel verwendet werden soll. Die ID kann sich auf ein [ODCR](#) oder einen [Kapazitätsblock für ML](#) beziehen.

Die Reservierung muss dieselbe Plattform verwenden, die die Instanz verwendet. Wenn Ihre Instances beispielsweise ausgeführt werden `rhel8`, muss Ihre Kapazitätsreservierung auf der Red Hat Enterprise Linux-Plattform laufen. Weitere Informationen finden Sie unter [Unterstützte Plattformen](#) im Amazon-EC2-Benutzerhandbuch für Linux-Instances.

Note

Wenn Sie diese Einstellung [Instances](#) in die Cluster-Konfiguration einbeziehen, müssen Sie diese `CapacityReservationId` Einstellung für die Warteschlangenebene aus der Konfiguration ausschließen.

CapacityReservationResourceGroupArn(Optional,**String**)

Der Amazon-Ressourcename (ARN) der Ressourcengruppe, die als serviceverknüpfte Gruppe von Kapazitätsreservierungen für die Rechenressourcen der Warteschlange dient. AWS ParallelCluster identifiziert und verwendet die am besten geeignete Kapazitätsreservierung aus der Ressourcengruppe auf der Grundlage der folgenden Bedingungen:

- Wenn in [SlurmQueues/Networking](#) oder [SlurmQueues](#)//aktiviert `PlacementGroup` ist [Networking](#), wird eine Ressourcengruppe AWS ParallelCluster ausgewählt, die auf den Instanztyp abzielt, und `PlacementGroup` für eine Rechenressource, falls die Rechenressource vorhanden ist. [ComputeResources](#)

Das `PlacementGroup` muss auf einen der Instanztypen abzielen, der in definiert ist [ComputeResources](#).

- Wenn es in [SlurmQueues/Networking](#) oder [SlurmQueues//PlacementGroup](#) nicht aktiviert ist [Networking](#), AWS ParallelCluster wird eine Ressourcengruppe ausgewählt, die nur auf den Instanztyp einer Rechenressource abzielt, sofern die Rechenressource existiert. [ComputeResources](#)

Die Ressourcengruppe muss mindestens einen ODCR für jeden Instanztyp haben, der in einer Availability Zone für alle Rechenressourcen und Availability Zones der Warteschlange reserviert ist. Weitere Informationen finden Sie unter [Erstellen Sie Instances mit ODCR Starten Sie Instances](#).

Weitere Informationen zu den Konfigurationsanforderungen für mehrere Subnetze finden Sie unter [Networking/ SubnetIds](#)

Note

In AWS ParallelCluster Version 3.4.0 wurden mehrere Availability Zones hinzugefügt.

[Richtlinie aktualisieren: Die Rechenflotte muss gestoppt oder QueueUpdateStrategieingerichtet sein, damit diese Einstellung für ein Update geändert werden kann.](#)

CapacityType(Fakultativ,String)

Der Typ der Rechenressourcen, die die Slurm Warteschlange verwendet. Unterstützte Werte sind ONDEMAND oder SPOT. Der Standardwert ist ONDEMAND.

Note

Wenn Sie das CapacityType auf setzenSPOT, muss Ihr Konto über eine AWSServiceRoleForEC2Spot dienstbezogene Rolle verfügen. Sie können diese Rolle mit dem folgenden AWS CLI Befehl erstellen.

```
$ aws iam create-service-linked-role --aws-service-name spot.amazonaws.com
```

Weitere Informationen finden Sie unter [Service-verknüpfte Rolle für Spot-Instance-Anfragen](#) im Amazon EC2 EC2-Benutzerhandbuch für Linux-Instances.

[Richtlinie aktualisieren: Die Rechenflotte muss gestoppt oder QueueUpdateStrategieingerichtet sein, damit diese Einstellung für ein Update geändert werden kann.](#)

AllocationStrategy(Fakultativ,) **String**

Geben Sie die Zuweisungsstrategie für alle Rechenressourcen an, die in [definiert sind Instances](#).

Zulässige Werte: `lowest-price` | `capacity-optimized`

Standard: `lowest-price`

lowest-price

- Wenn Sie diese Option verwenden `CapacityType = ONDEMAND`, bestimmt EC2 Fleet anhand des Preises die Reihenfolge und startet zuerst die Instances mit dem niedrigsten Preis.
- Wenn Sie diese Option verwenden `CapacityType = SPOT`, startet EC2 Fleet Instances aus dem Spot-Instance-Pool mit dem niedrigsten Preis, der über verfügbare Kapazität verfügt. Wenn die Kapazität eines Pools knapp wird, bevor Ihre erforderliche Kapazität erreicht ist, erfüllt EC2 Fleet Ihre Anfrage, indem Instances für Sie gestartet werden. Insbesondere startet EC2 Fleet Instances aus dem Spot-Instance-Pool mit dem niedrigsten Preis, der über verfügbare Kapazität verfügt. EC2 Fleet kann Spot-Instances aus mehreren verschiedenen Pools starten.
- Wenn Sie festlegen `CapacityType = CAPACITY_BLOCK`, gibt es keine Zuweisungsstrategien, sodass der `AllocationStrategy` Parameter nicht konfiguriert werden kann.

capacity-optimized

- Wenn Sie festlegen `CapacityType = ONDEMAND`, `capacity-optimized` ist es nicht verfügbar.
- Wenn Sie diese Option festlegen `CapacityType = SPOT`, startet EC2 Fleet Instances aus Spot-Instance-Pools mit optimaler Kapazität für die Anzahl der zu startenden Instances.

[Richtlinie aktualisieren: Die Rechenflotte muss gestoppt oder QueueUpdateStrategieingerichtet sein, damit diese Einstellung für ein Update geändert werden kann.](#)

Note

`AllocationStrategy` wird ab AWS ParallelCluster Version 3.3.0 unterstützt.

`JobExclusiveAllocation`(Fakultativ, `String`)

Wenn auf `gesetzttrue`, ist das Slurm `OverSubscribe` Partitionsflag auf `gesetztEXCLUSIVE`. Bei `OverSubscribe = EXCLUSIVE` haben Jobs in der Partition exklusiven Zugriff auf alle zugewiesenen Knoten. Weitere Informationen finden Sie unter [EXCLUSIVE](#) in der Slurm Dokumentation.

Zulässige Werte: `true` | `false`

Standard: `false`

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

Note

`JobExclusiveAllocation` wird ab AWS ParallelCluster Version 3.7.0 unterstützt.

`CustomSlurmSettings`(Fakultativ, `Dict`)

Definiert die Konfigurationseinstellungen für die benutzerdefinierte Slurm Partition (Warteschlange).

Gibt ein Wörterbuch mit Schlüssel-Wert-Paaren für benutzerdefinierte Slurm Konfigurationsparameter an, die für Warteschlangen (Partitionen) gelten.

Jedes einzelne Schlüssel-Wert-Paar, z. B. `Param1: Value1`, wird separat am Ende der Slurm Partitionskonfigurationszeile im Format hinzugefügt. `Param1=Value1`

Sie können nur Slurm Konfigurationsparameter angeben, die nicht auf der Sperrliste stehen. `CustomSlurmSettings` Hinweise zu Slurm Konfigurationsparametern auf der Sperrliste finden Sie unter. [Auf der Denim-Liste Slurm Konfigurationsparameter für CustomSlurmSettings](#)


AWS ParallelCluster prüft nur, ob ein Parameter auf einer Sperrliste steht. AWS ParallelCluster validiert die Syntax oder Semantik Ihrer benutzerdefinierten Slurm Konfigurationsparameter

nicht. Sie sind für die Validierung Ihrer benutzerdefinierten Slurm Konfigurationsparameter verantwortlich. Ungültige benutzerdefinierte Slurm Konfigurationsparameter können zu Slurm Daemon-Fehlern führen, die zu Fehlern bei der Clustererstellung und -aktualisierung führen können.

Weitere Hinweise zum Angeben von benutzerdefinierten Slurm Konfigurationsparametern mit finden Sie AWS ParallelCluster unter [SlurmAnpassung der Konfiguration](#).

Weitere Informationen zu Slurm Konfigurationsparametern finden Sie in der [Dokumentation unter slurm.conf](#). Slurm

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

 Note

CustomSlurmSettings wird ab Version 3.6.0 unterstützt. AWS ParallelCluster

Tags(Optional, [Zeichenfolge])

Eine Liste von Tag-Schlüssel-Wert-Paaren. [ComputeResource](#) Tags überschreiben doppelte Tags, die in [Tags Abschnitt](#) oder in `SlurmQueues`/angegeben sind. Tags

Key(Optional, **String**)

Der Tag-Schlüssel.

Value(Fakultativ, **String**)

Der Tag-Wert.

[Richtlinie aktualisieren: Die Rechenflotte muss gestoppt oder QueueUpdateStrategieingerichtet sein, damit diese Einstellung für ein Update geändert werden kann.](#)

HealthChecks(Fakultativ)

Geben Sie Integritätsprüfungen für Rechenknoten für alle Rechenressourcen in der Warteschlange an.

Gpu(Fakultativ)

Geben Sie GPU-Zustandsprüfungen für alle Rechenressourcen in einer Warteschlange an.

Note

AWS ParallelCluster unterstützt HealthChecks/nicht Gpu in Knoten, die a1linux2 ARM-Betriebssysteme verwenden. Diese Plattformen unterstützen den [NVIDIA Data Center GPU Manager \(DCGM\)](#) nicht.

Enabled(Fakultativ, **Boolean**)

Gibt an AWS ParallelCluster, ob GPU-Zustandsprüfungen auf Rechenknoten durchgeführt werden. Der Standardwert ist `false`.

GpuVerhalten bei der Integritätsprüfung

- Falls `Gpu`/auf gesetzt `Enabled` ist `true`, werden AWS ParallelCluster GPU-Integritätsprüfungen für Rechenressourcen in der Warteschlange durchgeführt.
- Bei der `Gpu` Integritätsprüfung werden GPU-Integritätsprüfungen für Rechenressourcen durchgeführt, um zu verhindern, dass Jobs auf Knoten mit einer herabgesetzten GPU gesendet werden.
- Wenn ein Rechenknoten eine `Gpu` Zustandsprüfung nicht besteht, ändert sich der Status des Rechenknotens auf `DRAIN`. Neue Jobs werden auf diesem Knoten nicht gestartet. Bestehende Jobs werden bis zum Abschluss ausgeführt. Wenn alle laufenden Jobs abgeschlossen sind, wird der Rechenknoten beendet, wenn es sich um einen dynamischen Knoten handelt, und er wird ersetzt, wenn es sich um einen statischen Knoten handelt.
- Die Dauer der `Gpu` Zustandsprüfung hängt vom ausgewählten Instanztyp, der Anzahl der GPUs in der Instanz und der Anzahl der `Gpu` Integritätsprüfungsziele ab (entspricht der Anzahl der Job-GPU-Ziele). Bei einer Instanz mit 8 GPUs beträgt die typische Dauer weniger als 3 Minuten.
- Wenn die `Gpu` Zustandsprüfung auf einer Instanz ausgeführt wird, die nicht unterstützt wird, wird sie beendet und der Job wird auf dem Rechenknoten ausgeführt. Wenn eine Instanz beispielsweise keine GPU hat oder, wenn eine Instanz über eine GPU verfügt, es sich aber nicht um eine NVIDIA-GPU handelt, wird die Integritätsprüfung beendet und der Job wird auf dem Rechenknoten ausgeführt. Es werden nur NVIDIA-GPUs unterstützt.
- Die `Gpu` Integritätsprüfung verwendet das `dcgmi` Tool, um Integritätsprüfungen an einem Knoten durchzuführen, und umfasst die folgenden Schritte:


Wenn die `Gpu` Zustandsprüfung in einem Knoten beginnt:

1. Es erkennt, ob die `nvidia-fabricmanager` Dienste `nvidia-dcgm` und ausgeführt werden.
2. Wenn diese Dienste nicht ausgeführt werden, werden sie durch die Gpu Integritätsprüfung gestartet.
3. Es erkennt, ob der Persistenzmodus aktiviert ist.
4. Wenn der Persistenzmodus nicht aktiviert ist, wird er durch die Gpu Integritätsprüfung aktiviert.

Am Ende der Zustandsprüfung werden diese Dienste und Ressourcen durch die Gpu Zustandsprüfung in ihren ursprünglichen Zustand zurückversetzt.

- Wenn der Job einem bestimmten Satz von Knoten-GPUs zugewiesen ist, wird die Gpu Integritätsprüfung nur für diesen bestimmten Satz ausgeführt. Andernfalls wird die Gpu Integritätsprüfung auf allen GPUs im Knoten ausgeführt.
- Wenn ein Rechenknoten zwei oder mehr Gpu Integritätsprüfungsanfragen gleichzeitig empfängt, wird nur die erste Zustandsprüfung ausgeführt und die anderen werden übersprungen. Dies ist auch bei Zustandsprüfungen der Fall, die auf Knoten-GPUs abzielen. Sie können in den Protokolldateien nach weiteren Informationen zu dieser Situation suchen.
- Das Protokoll der Integritätsprüfung für einen bestimmten Rechenknoten ist in der `/var/log/parallelcluster/slurm_health_check.log` Datei verfügbar. Die Datei ist in Amazon CloudWatch in der CloudWatch Cluster-Protokollgruppe verfügbar. Dort finden Sie:
 - Einzelheiten zu der Aktion, die im Rahmen der Gpu Integritätsprüfung ausgeführt wurde, einschließlich der Aktivierung und Deaktivierung von Diensten und des Persistenzmodus.
 - Die GPU-Kennung, die serielle ID und die UUID.
 - Die Ausgabe des Integritätschecks.

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

 Note

HealthCheck wird ab AWS ParallelCluster Version 3.6.0 unterstützt.

Networking

(Erforderlich) Definiert die Netzwerkkonfiguration für die Slurm Warteschlange.

Networking:**SubnetIds:**

- *string*

AssignPublicIp: *boolean*

SecurityGroups:

- *string*

AdditionalSecurityGroups:

- *string*

PlacementGroup:

Enabled: *boolean*

Id: *string*

Name: *string*

Proxy:

HttpProxyAddress: *string*

Richtlinie aktualisieren: Die Rechenflotte muss gestoppt oder QueueUpdateStrategieingerichtet sein, damit diese Einstellung für ein Update geändert werden kann.

Networking-Eigenschaften**SubnetIds**(Erforderlich, **[String]**)

Die IDs vorhandener Subnetze, in denen Sie die Slurm Warteschlange bereitstellen.

Wenn Sie Instanztypen in [SlurmQueues/ComputeResources](#)/konfigurieren [InstanceType](#), können Sie nur ein Subnetz definieren.

Wenn Sie Instanztypen in [SlurmQueues/ComputeResources](#)/konfigurieren [Instances](#), können Sie ein einzelnes Subnetz oder mehrere Subnetze definieren.

Wenn Sie mehrere Subnetze verwenden, müssen sich alle für eine Warteschlange definierten Subnetze in derselben VPC befinden, wobei sich jedes Subnetz in einer separaten Availability Zone (AZ) befindet.


Nehmen wir beispielsweise an, Sie definieren Subnetz-1 und Subnetz-2 für Ihre Warteschlange.

subnet-1 und subnet-2 können nicht beide in AZ-1 sein.

subnet-1 kann in AZ-1 sein und subnet-2 kann in AZ-2 sein.

Wenn Sie nur einen Instance-Typ konfigurieren und mehrere Subnetze verwenden möchten, definieren Sie Ihren Instance-Typ in `Instances` und nicht `InstanceType`

Definieren Sie beispielsweise `ComputeResources/Instances/InstanceType=instance.type` statt `ComputeResources/InstanceType=instance.type`.

 Note

Elastic Fabric Adapter (EFA) wird in verschiedenen Availability Zones nicht unterstützt.

Die Verwendung mehrerer Availability Zones kann zu einer Erhöhung der Speichernetzwerklatenz und zu zusätzlichen Kosten für die Datenübertragung zwischen den einzelnen AZ-Datenbanken führen. Dies könnte beispielsweise der Fall sein, wenn eine Instance auf einen Dateispeicher zugreift, der sich in einer anderen AZ befindet. Weitere Informationen finden Sie unter [Datenübertragung innerhalb derselben AWS-Region](#).

Cluster-Updates zur Umstellung von der Verwendung eines einzelnen Subnetzes auf mehrere Subnetze:

- Angenommen, die Subnetzdefinition eines Clusters ist mit einem einzelnen Subnetz und einem AWS ParallelCluster verwalteten FSx for Lustre-Dateisystem definiert. Dann können Sie diesen Cluster nicht direkt mit einer aktualisierten Subnetz-ID-Definition aktualisieren. Um das Cluster-Update durchzuführen, müssen Sie zuerst das verwaltete Dateisystem in ein externes Dateisystem ändern. Weitere Informationen finden Sie unter [Konvertiert AWS ParallelCluster verwalteten Speicher in externen Speicher](#).
- Angenommen, die Subnetzdefinition eines Clusters wird mit einem einzelnen Subnetz und einem externen Amazon EFS-Dateisystem definiert, wenn EFS-Mount-Ziele nicht für alle AZs für die mehreren Subnetze existieren, die hinzugefügt werden sollen. Dann können Sie diesen Cluster nicht direkt mit einer aktualisierten Subnetz-ID-Definition aktualisieren. Um den Cluster zu aktualisieren oder einen Cluster zu erstellen, müssen Sie zunächst alle Mount-Ziele für alle AZs für die definierten mehreren Subnetze erstellen.

Availability Zones und Cluster-Kapazitätsreservierungen, definiert in [CapacityReservationResourceGroupArn](#):

- Sie können keinen Cluster erstellen, wenn es keine Überschneidung zwischen den Instanztypen und Verfügbarkeitszonen, die von der definierten Ressourcengruppe für die Kapazitätsreservierung abgedeckt werden, und den für die Warteschlange definierten Instanztypen und Verfügbarkeitszonen gibt.
- Sie können einen Cluster erstellen, wenn es eine teilweise Überschneidung zwischen den Instanztypen und Verfügbarkeitszonen, die von der definierten Ressourcengruppe für die

Kapazitätsreservierung abgedeckt werden, und den für die Warteschlange definierten Instanztypen und Verfügbarkeitszonen gibt. AWS ParallelCluster sendet in diesem Fall eine Warnmeldung über die teilweise Überlappung.

- Weitere Informationen finden Sie unter [Erstellen Sie Instances mit ODCR Starten Sie Instances](#).

Note

In AWS ParallelCluster Version 3.4.0 wurden mehrere Availability Zones hinzugefügt.

Warning

Diese Warnung gilt für alle AWS ParallelCluster 3.x.y-Versionen vor Version 3.3.1. AWS ParallelCluster Version 3.3.1 ist nicht betroffen, wenn dieser Parameter geändert wird.

Für AWS ParallelCluster 3 Versionen vor Version 3.3.1:

Wenn Sie diesen Parameter ändern und einen Cluster aktualisieren, wird ein neues verwaltetes FSx for Lustre-Dateisystem erstellt und das bestehende verwaltete FSx for Lustre-Dateisystem gelöscht, ohne die vorhandenen Daten beizubehalten. Dies führt zu Datenverlust. Bevor Sie fortfahren, stellen Sie sicher, dass Sie die Daten aus dem vorhandenen FSx for Lustre-Dateisystem sichern, wenn Sie Daten beibehalten möchten. Weitere Informationen finden Sie unter [Arbeiten mit Backups](#) im FSx for Lustre-Benutzerhandbuch.

Wenn ein neuer Subnetzwert hinzugefügt wird, [Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden](#).

Wenn ein Subnetzwert entfernt wird, [Richtlinie aktualisieren: Die Rechenflotte muss gestoppt oder QueueUpdateStrategyeingerichtet sein, damit diese Einstellung für ein Update geändert werden kann](#).

AssignPublicIp(Fakultativ,**String**)

Erzeugt oder weist den Knoten in der Slurm Warteschlange eine öffentliche IP-Adresse zu. Unterstützte Werte sind `true` und `false`. Das von Ihnen angegebene Subnetz bestimmt den Standardwert. Ein Subnetz mit öffentlichen IP-Adressen weist standardmäßig öffentliche IP-Adressen zu.

Wenn Sie einen hpc6id Instanztyp p4d oder einen anderen Instanztyp mit mehreren Netzwerkschnittstellen oder einer Netzwerkschnittstellenkarte definieren, müssen Sie [HeadNode/Networking/ElasticIp](#) auf festlegen, um öffentlichen Zugriff `true` zu gewähren. AWS Öffentliche IPs können nur Instances zugewiesen werden, die mit einer einzigen Netzwerkschnittstelle gestartet wurden. In diesem Fall empfehlen wir, ein [NAT-Gateway](#) zu verwenden, um öffentlichen Zugriff auf die Cluster-Rechenknoten zu gewähren. Stellen Sie in diesem Fall `AssignPublicIp` auf `false`. Weitere Informationen zu IP-Adressen finden Sie unter [Zuweisen einer öffentlichen IPv4-Adresse beim Instance-Start](#) im Amazon EC2 EC2-Benutzerhandbuch für Linux-Instances.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

SecurityGroups(Optional.) **[String]**

Eine Liste von Sicherheitsgruppen, die für die Slurm Warteschlange verwendet werden sollen. Wenn keine Sicherheitsgruppen angegeben sind, AWS ParallelCluster erstellt es Sicherheitsgruppen für Sie.

Stellen Sie sicher, dass die Sicherheitsgruppen für Ihre [SharedStorage](#) Systeme korrekt konfiguriert sind.

Warning

Diese Warnung gilt für alle 3. *x.y* AWS ParallelCluster Versionen vor Version 3.3.0. AWS ParallelCluster Version 3.3.0 ist nicht betroffen, wenn dieser Parameter geändert wird.

Für AWS ParallelCluster 3 Versionen vor Version 3.3.0:

Wenn Sie diesen Parameter ändern und einen Cluster aktualisieren, wird ein neues verwaltetes FSx for Lustre-Dateisystem erstellt und das bestehende verwaltete FSx for Lustre-Dateisystem gelöscht, ohne die vorhandenen Daten beizubehalten. Dies führt zu Datenverlust. Stellen Sie sicher, dass Sie die Daten aus dem vorhandenen FSx for Lustre-Dateisystem sichern, wenn Sie Daten erhalten möchten. Weitere Informationen finden Sie unter [Arbeiten mit Backups](#) im FSx for Lustre-Benutzerhandbuch.

Warning

Wenn Sie [Efa](#) für Ihre Compute-Instances aktivieren, stellen Sie sicher, dass Ihre EFA-fähigen Instances Mitglieder einer Sicherheitsgruppe sind, die den gesamten eingehenden und ausgehenden Datenverkehr für sich selbst zulässt.

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

AdditionalSecurityGroups([**String**]Optional,)

Eine Liste zusätzlicher Sicherheitsgruppen, die für die Slurm Warteschlange verwendet werden sollen.

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

PlacementGroup(Fakultativ)

Gibt die Platzierungsgruppeneinstellungen für die Slurm Warteschlange an.

```
PlacementGroup:  
  Enabled: boolean  
  Id: string  
  Name: string
```

[Aktualisierungsrichtlinie: Alle Rechenknoten müssen gestoppt werden, damit eine verwaltete Platzierungsgruppe gelöscht werden kann. Die Rechenflotte muss gestoppt oder QueueUpdateStrategieingerichtet sein, damit diese Einstellung für ein Update geändert werden kann.](#)

Enabled(Optional,**Boolean**)

Gibt an, ob eine Platzierungsgruppe für die Slurm Warteschlange verwendet wird. Der Standardwert ist `false`.

[Richtlinie aktualisieren: Die Rechenflotte muss gestoppt oder QueueUpdateStrategieingerichtet sein, damit diese Einstellung für ein Update geändert werden kann.](#)

Id(Optional,**String**)

Der Name der Platzierungsgruppe für eine bestehende Cluster-Platzierungsgruppe, die von der Slurm Warteschlange verwendet wird. Stellen Sie sicher, dass Sie den Namen der Platzierungsgruppe und nicht die ID angeben.

[Richtlinie aktualisieren: Die Rechenflotte muss gestoppt oder QueueUpdateStrategieingerichtet sein, damit diese Einstellung für ein Update geändert werden kann.](#)

Name(Fakultativ,String)

Der Name der Platzierungsgruppe für eine bestehende Cluster-Platzierungsgruppe, die von der Slurm Warteschlange verwendet wird. Stellen Sie sicher, dass Sie den Namen der Platzierungsgruppe und nicht die ID angeben.

[Richtlinie aktualisieren: Die Rechenflotte muss gestoppt oder QueueUpdateStrategieingerichtet sein, damit diese Einstellung für ein Update geändert werden kann.](#)

Note

- Wenn `PlacementGroup/auf` gesetzt `Enabled` ist `true`, ohne dass ein `Name` oder `Id` definiert ist, wird jeder Rechenressource ihre eigene verwaltete Platzierungsgruppe zugewiesen, es sei denn, `ComputeResources/Networking/PlacementGroup` ist so definiert, dass es diese Einstellung überschreibt.
- Ab AWS ParallelCluster Version 3.3.0 `Name` wurde `SlurmQueues/Networking/PlacementGroup/als` bevorzugte Alternative zu `SlurmQueues//NetworkingPlacementGroup/Id` hinzugefügt.

`PlacementGroup/Id` und `PlacementGroup/Name` sind gleichwertig. Sie können beide verwenden.

Wenn Sie sowohl `PlacementGroup/als` auch `PlacementGroup/Id` angeben `Name`, AWS ParallelCluster schlägt dies fehl. Sie können nur das eine oder das andere wählen.

Sie müssen Ihren Cluster nicht aktualisieren, um `PlacementGroup/` verwenden zu können `Name`.

Proxy(Fakultativ)

Gibt die Proxyeinstellungen für die Slurm Warteschlange an.

```
Proxy:  
  HttpProxyAddress: string
```

[Richtlinie aktualisieren: Die Rechenflotte muss gestoppt oder QueueUpdateStrategieeingesetzt sein, damit diese Einstellung für ein Update geändert werden kann.](#)

HttpProxyAddress(Optional,**String**)

Definiert einen HTTP- oder HTTPS-Proxyserver für die Slurm Warteschlange. In der Regel ist es `https://x.x.x.x:8080`.

Es gibt keinen Standardwert.

[Richtlinie aktualisieren: Die Rechenflotte muss gestoppt oder QueueUpdateStrategieeingesetzt sein, damit diese Einstellung für ein Update geändert werden kann.](#)

Image

(Optional) Gibt das Bild an, das für die Slurm Warteschlange verwendet werden soll. Um dasselbe AMI für alle Knoten zu verwenden, verwenden Sie die [CustomAmi](#)-Einstellung im [ImageAbschnitt](#).

[Image](#):
[CustomAmi](#): *string*

[Richtlinie aktualisieren: Die Rechenflotte muss gestoppt oder QueueUpdateStrategieeingesetzt sein, damit diese Einstellung für ein Update geändert werden kann.](#)

ImageEigenschaften

CustomAmi(Fakultativ,**String**)

Das AMI, das für die Slurm Warteschlange anstelle der Standard-AMIs verwendet werden soll. Sie können den `pcluster CLI`-Befehl verwenden, um eine Liste der Standard-AMIs anzuzeigen.

Note

Das AMI muss auf demselben Betriebssystem basieren, das vom Hauptknoten verwendet wird.

```
pcluster list-official-images
```

Wenn das benutzerdefinierte AMI zusätzliche Berechtigungen für seinen Start benötigt, müssen Sie diese Berechtigungen zur Head-Node-Richtlinie hinzufügen.

Wenn einem benutzerdefinierten AMI beispielsweise ein verschlüsselter Snapshot zugeordnet ist, sind die folgenden zusätzlichen Richtlinien in den Richtlinien für den Hauptknoten erforderlich.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "kms:DescribeKey",
        "kms:ReEncrypt*",
        "kms:CreateGrant",
        "kms:Decrypt"
      ],
      "Resource": [
        "arn:aws:kms:<AWS_REGION>:<AWS_ACCOUNT_ID>:key/<AWS_KMS_KEY_ID>"
      ]
    }
  ]
}
```

Informationen zur Fehlerbehebung bei benutzerdefinierten AMI-Validierungswarnungen finden Sie unter [Behebung von Problemen mit benutzerdefinierten AMIs](#).

[Richtlinie aktualisieren: Die Rechenflotte muss gestoppt oder QueueUpdateStrategieeingesetzt sein, damit diese Einstellung für ein Update geändert werden kann.](#)

ComputeResources

(Erforderlich) Definiert die ComputeResources Konfiguration für die Slurm Warteschlange.

Note

Die Clustergröße kann sich während eines Updates ändern. Weitere Informationen finden Sie unter [Größe und Aktualisierung der Clusterkapazität](#)

ComputeResources:

- Name: *string*
- InstanceType: *string*
- Instances:
 - InstanceType: *string*
- MinCount: *integer*
- MaxCount: *integer*
- DynamicNodePriority: *integer*
- StaticNodePriority: *integer*
- SpotPrice: *float*
- DisableSimultaneousMultithreading: *boolean*
- SchedulableMemory: *integer*
- HealthChecks:
 - Gpu:
 - Enabled: *boolean*
- Efa:
 - Enabled: *boolean*
 - GdrSupport: *boolean*
- CapacityReservationTarget:
 - CapacityReservationId: *string*
 - CapacityReservationResourceGroupArn: *string*
- Networking:
 - PlacementGroup:
 - Enabled: *boolean*
 - Name: *string*
- CustomSlurmSettings: *dict*
- Tags:
 - Key: *string*
 - Value: *string*

Richtlinie aktualisieren: Bei dieser Einstellung für Listenwerte kann während eines Updates ein neuer Wert hinzugefügt werden, oder die Rechenflotte muss gestoppt werden, wenn ein vorhandener Wert entfernt wird.

ComputeResources-Eigenschaften

Name(Erforderlich, **String**)

Der Name der Slurm Warteschlangencomputer-Umgebung. Der Name kann bis zu 25 Zeichen lang sein.

Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.

InstanceType(Erforderlich,String)

Der Instanztyp, der in dieser Slurm Rechenressource verwendet wird. Alle Instance-Typen in einem Cluster müssen dieselbe Prozessorarchitektur verwenden. Instanzen können entweder die x86_64 arm64 Oder-Architektur verwenden.

Die Clusterkonfiguration muss entweder [Instanzen InstanceType](#) oder definieren. Wenn beide definiert sind, AWS ParallelCluster schlägt dies fehl.

Wenn Sie definieren InstanceType, können Sie nicht mehrere Subnetze definieren. Wenn Sie nur einen Instanztyp konfigurieren und mehrere Subnetze verwenden möchten, definieren Sie Ihren Instanztyp in und Instances nicht in. InstanceType Weitere Informationen finden Sie unter [Networking/SubnetIds](#).

Wenn Sie einen hpc6id Instance-Typ p4d oder einen anderen Instance-Typ mit mehreren Netzwerkschnittstellen oder einer Netzwerkschnittstellenkarte definieren, müssen Sie die Compute-Instances in einem privaten Subnetz starten, wie unter beschrieben. [AWS ParallelCluster mit zwei Subnetzen](#) AWS Öffentliche IPs können nur Instances zugewiesen werden, die mit einer einzigen Netzwerkschnittstelle gestartet werden. Weitere Informationen finden Sie unter [Zuweisen einer öffentlichen IPv4-Adresse beim Instance-Start](#) im Amazon EC2 EC2-Benutzerhandbuch für Linux-Instances.

[Aktualisierungsrichtlinie: Die Rechenflotte muss gestoppt werden, damit diese Einstellung für ein Update geändert werden kann.](#)

Instances(Erforderlich)

Gibt die Liste der Instanztypen für eine Rechenressource an. Informationen zur Angabe der Zuweisungsstrategie für die Liste der Instanztypen finden Sie unter [AllocationStrategy](#).

Die Clusterkonfiguration muss entweder [InstanceType](#) oder definieren [Instances](#). Wenn beide definiert sind, AWS ParallelCluster schlägt dies fehl.

Weitere Informationen finden Sie unter [Zuweisung mehrerer Instanztypen mit Slurm](#).

[Instances](#):

- [InstanceType](#): *string*

Note

Ab AWS ParallelCluster Version 3.7.0 `EnableMemoryBasedScheduling` kann aktiviert werden, wenn Sie mehrere Instanztypen in [Instances](#) konfigurieren.

Für die AWS ParallelCluster Versionen 3.2.0 bis 3.6. **x**

`EnableMemoryBasedScheduling` kann nicht aktiviert werden, wenn Sie mehrere Instanztypen in [Instances](#) konfigurieren.

Richtlinie aktualisieren: Bei dieser Einstellung für Listenwerte kann während eines Updates ein neuer Wert hinzugefügt werden, oder die Rechenflotte muss gestoppt werden, wenn ein vorhandener Wert entfernt wird.

InstanceType(Erforderlich,**String**)

Der Instanztyp, der in dieser Slurm Rechenressource verwendet werden soll. Alle Instanztypen in einem Cluster müssen dieselbe Prozessorarchitektur verwenden, entweder `x86_64` oder `arm64`.

Die unter aufgeführten Instanztypen [Instances](#) müssen Folgendes aufweisen:

- Dieselbe Anzahl von vCPUs oder, falls [DisableSimultaneousMultithreading](#) auf `true` eingestellt, dieselbe Anzahl von Kernen.
- Dieselbe Anzahl von Beschleunigern derselben Hersteller.
- EFA wird unterstützt, falls [Efa](#)/auf [Enabled](#) gesetzt ist. `true`

Die unter aufgelisteten Instance-Typen [Instances](#) können Folgendes haben:

- Unterschiedliche Speichermenge.

In diesem Fall muss der Mindestspeicher als verbrauchbare Slurm Ressource festgelegt werden.

Note

Ab AWS ParallelCluster Version 3.7.0 `EnableMemoryBasedScheduling` kann aktiviert werden, wenn Sie mehrere Instance-Typen in [Instances](#) konfigurieren.

Für die AWS ParallelCluster Versionen 3.2.0 bis 3.6. **x**

`EnableMemoryBasedScheduling` kann nicht aktiviert werden, wenn Sie mehrere Instanztypen in [Instances](#) konfigurieren.

- Verschiedene Netzwerkkarten.

In diesem Fall wird die Anzahl der für die Rechenressource konfigurierten Netzwerkschnittstellen durch den Instanztyp mit der geringsten Anzahl von Netzwerkkarten definiert.

- Unterschiedliche Netzwerkbandbreite.
- Unterschiedliche Größe des Instance-Speichers.

Wenn Sie einen hpc6id Instance-Typ p4d oder einen anderen Instance-Typ mit mehreren Netzwerkschnittstellen oder einer Netzwerkschnittstellenkarte definieren, müssen Sie die Compute-Instances in einem privaten Subnetz starten, wie unter beschrieben. [AWS ParallelCluster mit zwei Subnetzen](#) AWS Öffentliche IPs können nur Instances zugewiesen werden, die mit einer einzigen Netzwerkschnittstelle gestartet wurden. Weitere Informationen finden Sie unter [Zuweisen einer öffentlichen IPv4-Adresse beim Instance-Start](#) im Amazon EC2 EC2-Benutzerhandbuch für Linux-Instances.

[Aktualisierungsrichtlinie: Die Rechenflotte muss gestoppt werden, damit diese Einstellung für ein Update geändert werden kann.](#)

i Note

Instances wird ab AWS ParallelCluster Version 3.3.0 unterstützt.

MinCount(Fakultativ,**Integer**)

Die Mindestanzahl von Instanzen, die die Slurm Rechenressource verwendet. Der Standardwert ist 0.

i Note

Die Clustergröße kann sich während eines Updates ändern. Weitere Informationen finden Sie unter [Größe und Aktualisierung der Clusterkapazität](#)

[Aktualisierungsrichtlinie: Die Rechenflotte muss gestoppt werden, damit diese Einstellung für ein Update geändert werden kann.](#)

MaxCount(Optional,Integer)

Die maximale Anzahl von Instanzen, die die Slurm Rechenressource verwendet. Der Standardwert ist 10.

MaxCountMuss bei der Verwendung CapacityType = CAPACITY_BLOCK gleich oder größer als 0 sein, da alle Instanzen, die Teil der Capacity Block-Reservierung sind, als statische Knoten verwaltet werden. MinCount

Bei der Clustererstellung wartet der Hauptknoten, bis alle statischen Knoten bereit sind, bevor er den Erfolg der Clustererstellung signalisiert. Bei der Verwendung werden die KnotenCapacityType = CAPACITY_BLOCK, die Teil der Rechenressourcen sind, die Kapazitätsblöcken zugeordnet sind, bei dieser Prüfung jedoch nicht berücksichtigt. Der Cluster wird auch dann erstellt, wenn nicht alle konfigurierten Kapazitätsblöcke aktiv sind.

Note

Die Clustergröße kann sich während eines Updates ändern. Weitere Informationen finden Sie unter [Größe und Aktualisierung der Clusterkapazität](#)

DynamicNodePriority(Optional,Integer)

Die Priorität dynamischer Knoten in einer Queue-Rechenressource. Die Priorität ist dem Slurm [Weight](#)Knotenkonfigurationsparameter für die dynamischen Rechenressourcenknoten zugeordnet. Der Standardwert ist 1000.

Slurmpriorisiert zuerst Knoten mit den niedrigsten Weight Werten.

Warning

Die Verwendung vieler verschiedener Weight Werte in einer Slurm Partition (Warteschlange) kann die Geschwindigkeit der Jobplanung in der Warteschlange verlangsamen.

In AWS ParallelCluster Versionen vor Version 3.7.0 wurde sowohl statischen als auch dynamischen Knoten dieselbe Standardgewichtung von 1 zugewiesen. In diesem Fall Slurm könnte aufgrund des Benennungsschemas für statische und dynamische Knoten im Leerlauf dynamische Knoten Vorrang vor inaktiven statischen Knoten eingeräumt werden. Wenn alles andere gleich ist, werden Knoten Slurm alphabetisch nach Namen sortiert.

Note

`DynamicNodePriority` wurde in AWS ParallelCluster Version 3.7.0 hinzugefügt.

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

StaticNodePriority(Fakultativ,**Integer**)

Die Priorität statischer Knoten in einer Queue-Rechenressource. Die Priorität ist dem Slurm [Weight](#) Knotenkonfigurationsparameter für die statischen Rechenressourcenknoten zugeordnet. Der Standardwert ist 1.

Slurmpriorisiert zuerst Knoten mit den niedrigsten Weight Werten.

Warning

Die Verwendung vieler verschiedener Weight Werte in einer Slurm Partition (Warteschlange) kann die Geschwindigkeit der Jobplanung in der Warteschlange verlangsamen.

Note

`StaticNodePriority` wurde in AWS ParallelCluster Version 3.7.0 hinzugefügt.

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

SpotPrice(Fakultativ,**Float**)

Der Höchstpreis, der für eine EC2-Spot-Instance bezahlt wurde, bevor Instances gestartet wurden. Der Standardwert ist der On-Demand-Preis.

[Richtlinie aktualisieren: Die Rechenflotte muss gestoppt oder QueueUpdateStrategieingerichtet sein, damit diese Einstellung für ein Update geändert werden kann.](#)

DisableSimultaneousMultithreading(Optional,**Boolean**)

Wenn `true` Multithreading auf den Knoten in der Slurm Warteschlange deaktiviert ist. Der Standardwert ist `false`.

Nicht alle Instance-Typen können Multithreading deaktivieren. Eine Liste der Instance-Typen, die die Deaktivierung von Multithreading unterstützen, finden Sie unter [CPU-Kerne und Threads für jeden CPU-Kern pro Instance-Typ](#) im Amazon EC2 EC2-Benutzerhandbuch für Linux-Instances.

[Aktualisierungsrichtlinie: Die Rechenflotte muss gestoppt werden, damit diese Einstellung für ein Update geändert werden kann.](#)

SchedulableMemory(Optional,) **Integer**

Die Speichermenge in MiB, die im Slurm Parameter `RealMemory` für die Rechenknoten einer Rechenressource konfiguriert ist. Dieser Wert ist die Obergrenze für den Knotenspeicher, der für Jobs verfügbar ist, wenn [SlurmSettings](#)/aktiviert [EnableMemoryBasedScheduling](#) ist. Der Standardwert ist 95 Prozent des Speichers, der in [Amazon EC2 EC2-Instanztypen aufgeführt und von der Amazon EC2](#) EC2-API zurückgegeben wird. [DescribeInstanceTypes](#) Achten Sie darauf, Werte, die in GiB angegeben sind, in MiB umzurechnen.

Unterstützte Werte: 1-`EC2Memory`

`EC2Memory` ist der Speicher (in MiB), der in [Amazon EC2 EC2-Instanztypen aufgeführt und von der Amazon EC2](#) EC2-API zurückgegeben wird. [DescribeInstanceTypes](#) Achten Sie darauf, Werte, die in GiB angegeben sind, in MiB umzurechnen.

Diese Option ist am relevantesten, wenn [SlurmSettings](#)/aktiviert [EnableMemoryBasedScheduling](#) ist. Weitere Informationen finden Sie unter [Slurmspeicherbasierte Terminplanung](#).

Note

`SchedulableMemory` wird ab AWS ParallelCluster Version 3.2.0 unterstützt. Ab Version 3.2.0 werden `RealMemory` für Slurm Rechenknoten standardmäßig 95 Prozent des Speichers AWS ParallelCluster konfiguriert, der von der Amazon EC2 EC2-API zurückgegeben wird. `DescribeInstanceTypes` Diese Konfiguration ist unabhängig vom Wert von `EnableMemoryBasedScheduling`

[Richtlinie aktualisieren: Die Rechenflotte muss gestoppt oder QueueUpdateStrategieingerichtet sein, damit diese Einstellung für ein Update geändert werden kann.](#)

HealthChecks(Fakultativ)

Geben Sie Integritätsprüfungen für eine Rechenressource an.

Gpu(Fakultativ)

Geben Sie GPU-Zustandsprüfungen für eine Rechenressource an.

Enabled(Fakultativ, Boolean)

Gibt an AWS ParallelCluster, ob GPU-Zustandsprüfungen bei der Berechnung einer Ressource in einer Warteschlange durchgeführt werden. Der Standardwert ist `false`.

Note

AWS ParallelCluster unterstützt HealthChecks/nicht Gpu in Knoten, die `alinux2` ARM-Betriebssysteme verwenden. Diese Plattformen unterstützen den [NVIDIA Data Center GPU Manager \(DCGM\)](#) nicht.

GpuVerhalten bei der Gesundheitsprüfung

- Wenn `Gpu`/auf gesetzt `Enabled` ist `true`, AWS ParallelCluster werden GPU-Integritätsprüfungen für eine Rechenressource durchgeführt.
- Die Gpu Zustandsprüfung führt Integritätsprüfungen für eine Rechenressource durch, um zu verhindern, dass Jobs auf Knoten mit einer herabgesetzten GPU gesendet werden.
- Wenn ein Rechenknoten eine Gpu Zustandsprüfung nicht besteht, ändert sich der Status des Rechenknotens auf `DRAIN`. Neue Jobs werden auf diesem Knoten nicht gestartet. Bestehende Jobs werden bis zum Abschluss ausgeführt. Wenn alle laufenden Jobs abgeschlossen sind, wird der Rechenknoten beendet, wenn es sich um einen dynamischen Knoten handelt, und er wird ersetzt, wenn es sich um einen statischen Knoten handelt.
- Die Dauer der Gpu Zustandsprüfung hängt vom ausgewählten Instanztyp, der Anzahl der GPUs in der Instanz und der Anzahl der Gpu Integritätsprüfungsziele ab (entspricht der Anzahl der Job-GPU-Ziele). Bei einer Instanz mit 8 GPUs beträgt die typische Dauer weniger als 3 Minuten.
- Wenn die Gpu Zustandsprüfung auf einer Instanz ausgeführt wird, die nicht unterstützt wird, wird sie beendet und der Job wird auf dem Rechenknoten ausgeführt. Wenn eine Instanz beispielsweise keine GPU hat oder, wenn eine Instanz über eine GPU verfügt, es sich aber nicht um eine NVIDIA-GPU handelt, wird die Integritätsprüfung beendet und der Job wird auf dem Rechenknoten ausgeführt. Es werden nur NVIDIA-GPUs unterstützt.

- Die Gpu Integritätsprüfung verwendet das `dcgmi` Tool, um Integritätsprüfungen an einem Knoten durchzuführen, und umfasst die folgenden Schritte:

Wenn die Gpu Zustandsprüfung in einem Knoten beginnt:

1. Es erkennt, ob die `nvidia-fabricmanager` Dienste `nvidia-dcgm` und ausgeführt werden.
2. Wenn diese Dienste nicht ausgeführt werden, werden sie durch die Gpu Integritätsprüfung gestartet.
3. Es erkennt, ob der Persistenzmodus aktiviert ist.
4. Wenn der Persistenzmodus nicht aktiviert ist, wird er durch die Gpu Integritätsprüfung aktiviert.

Am Ende der Zustandsprüfung werden diese Dienste und Ressourcen durch die Gpu Zustandsprüfung in ihren ursprünglichen Zustand zurückversetzt.

- Wenn der Job einem bestimmten Satz von Knoten-GPUs zugewiesen ist, wird die Gpu Integritätsprüfung nur für diesen bestimmten Satz ausgeführt. Andernfalls wird die Gpu Integritätsprüfung auf allen GPUs im Knoten ausgeführt.
- Wenn ein Rechenknoten zwei oder mehr Gpu Integritätsprüfungsanfragen gleichzeitig empfängt, wird nur die erste Zustandsprüfung ausgeführt und die anderen werden übersprungen. Dies ist auch bei Zustandsprüfungen der Fall, die auf Knoten-GPUs abzielen. Sie können in den Protokolldateien nach weiteren Informationen zu dieser Situation suchen.
- Das Protokoll der Integritätsprüfung für einen bestimmten Rechenknoten ist in der `/var/log/parallelcluster/slurm_health_check.log` Datei verfügbar. Diese Datei ist in Amazon CloudWatch in der CloudWatch Cluster-Protokollgruppe verfügbar. Dort finden Sie:
 - Einzelheiten zu der Aktion, die durch die Gpu Zustandsprüfung ausgeführt wurde, einschließlich der Aktivierung und Deaktivierung von Diensten und des Persistenzmodus.
 - Die GPU-Kennung, die serielle ID und die UUID.
 - Die Ausgabe des Integritätschecks.

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

 Note

HealthCheckswird ab AWS ParallelCluster Version 3.6.0 unterstützt.

Efa(Fakultativ)

Gibt die Elastic Fabric Adapter (EFA) -Einstellungen für die Knoten in der Slurm Warteschlange an.

Efa:

Enabled: *boolean*

GdrSupport: *boolean*

Richtlinie aktualisieren: Die Rechenflotte muss gestoppt oder QueueUpdateStrategyeingesetzt sein, damit diese Einstellung für ein Update geändert werden kann.

Enabled(Optional,**Boolean**)

Gibt an, dass der Elastic Fabric Adapter (EFA) aktiviert ist. Eine Liste der EC2-Instances, die EFA unterstützen, finden Sie unter [Unterstützte Instance-Typen](#) im Amazon EC2 EC2-Benutzerhandbuch für Linux-Instances. Weitere Informationen finden Sie unter [Elastic Fabric Adapter](#). Wir empfehlen die Verwendung eines Clusters [SlurmQueues/Networking/](#), um die Latenzen zwischen [PlacementGroup](#)den Instances zu minimieren.

Der Standardwert ist `false`.

Note

Elastic Fabric Adapter (EFA) wird in verschiedenen Availability Zones nicht unterstützt. Weitere Informationen finden Sie unter [SubnetIds](#).

Warning

Wenn Sie eine benutzerdefinierte Sicherheitsgruppe in definieren [SecurityGroups](#), stellen Sie sicher, dass Ihre EFA-fähigen Instances Mitglieder einer Sicherheitsgruppe sind, die den gesamten eingehenden und ausgehenden Datenverkehr für sich selbst zulässt.

Richtlinie aktualisieren: Die Rechenflotte muss gestoppt oder QueueUpdateStrategyeingesetzt sein, damit diese Einstellung für ein Update geändert werden kann.

GdrSupport(Optional,) **Boolean**

(Optional) Ab AWS ParallelCluster Version 3.0.2 hat diese Einstellung keine Auswirkung. Die Unterstützung des Elastic Fabric Adapter (EFA) für GPUDirect RDMA (Remote Direct Memory Access) ist immer aktiviert, wenn sie vom Instanztyp für die Slurm Rechenressource und das Betriebssystem unterstützt wird.

Note

AWS ParallelCluster Version 3.0.0 bis 3.0.1: Die Support für GPUDirect RDMA ist für Rechenressourcen aktiviert. Slurm Die Support für GPUDirect RDMA wird von bestimmten Instanztypen (p4d.24xlarge) auf bestimmten Betriebssystemen unterstützt ([Osistalinux2](#), [centos7ubuntu1804](#), oder). [ubuntu2004](#) Der Standardwert ist "false".

[Richtlinie aktualisieren: Die Rechenflotte muss gestoppt oder QueueUpdateStrategieingerichtet sein, damit diese Einstellung für ein Update geändert werden kann.](#)

CapacityReservationTarget

[CapacityReservationTarget](#):

[CapacityReservationId](#): *string*

[CapacityReservationResourceGroupArn](#): *string*

Gibt die On-Demand-Kapazitätsreservierung an, die für die Rechenressource verwendet werden soll.

CapacityReservationId(Fakultativ,**String**)

Die ID der vorhandenen Kapazitätsreservierung, die für die Rechenressourcen der Warteschlange als Ziel verwendet werden soll. Die ID kann sich auf ein [ODCR](#) oder einen [Kapazitätsblock für ML](#) beziehen.

Wenn dieser Parameter auf Rechenressourcenebene angegeben wird, InstanceType ist er optional und wird automatisch aus der Reservierung abgerufen.

CapacityReservationResourceGroupArn(Optional,**String**)

Gibt den Amazon-Ressourcennamen (ARN) der Ressourcengruppe an, die als serviceverknüpfte Gruppe von Kapazitätsreservierungen für die Rechenressource dient. AWS


ParallelCluster identifiziert und verwendet die am besten geeignete Kapazitätsreservierung aus der Gruppe. Die Ressourcengruppe muss mindestens ein ODCR für jeden Instanztyp haben, der für die Rechenressource aufgeführt ist. Weitere Informationen finden Sie unter [Erstellen Sie Instances mit ODCR Starten Sie Instances](#).

- Wenn die Option in [SlurmQueues/Networking](#) oder [SlurmQueues](#)//aktiviert PlacementGroup ist [Networking](#), wird eine Ressourcengruppe AWS ParallelCluster ausgewählt, die auf den Instanztyp abzielt, und PlacementGroup für eine Rechenressource, falls diese existiert. [ComputeResources](#)

Sie PlacementGroup muss auf einen der in definierten Instanztypen abzielen [ComputeResources](#).

- Wenn sie in [SlurmQueues/Networking](#) oder [SlurmQueuesComputeResources](#)/PlacementGroup nicht aktiviert ist [Networking](#), AWS ParallelCluster wird eine Ressourcengruppe ausgewählt, die nur auf den Instanztyp einer Rechenressource abzielt, sofern diese existiert.

[Richtlinie aktualisieren: Die Rechenflotte muss gestoppt oder QueueUpdateStrategieeingesetzt sein, damit diese Einstellung für ein Update geändert werden kann.](#)

 Note

CapacityReservationTarget wird mit AWS ParallelCluster Version 3.3.0 hinzugefügt.

Networking

```
Networking:  
  PlacementGroup:  
    Enabled: boolean  
    Name: string
```

[Aktualisierungsrichtlinie: Alle Rechenknoten müssen gestoppt werden, damit eine verwaltete Platzierungsgruppe gelöscht werden kann. Die Rechenflotte muss gestoppt oder QueueUpdateStrategieeingesetzt sein, damit diese Einstellung für ein Update geändert werden kann.](#)

PlacementGroup(Fakultativ)

Gibt die Platzierungsgruppeneinstellungen für die Rechenressource an.

Enabled(Optional,Boolean)

Gibt an, ob eine Platzierungsgruppe für die Rechenressource verwendet wird.

- Wenn dieser Wert auf `true` gesetzt ist und kein Name definierter Wert angegeben ist, wird dieser Rechenressource unabhängig von der [PlacementGroup](#)Einstellung [SlurmQueues/Networking](#)eine eigene verwaltete Platzierungsgruppe zugewiesen.
- Wenn dieser Wert auf `true` gesetzt ist und ein Name Wert definiert ist, wird dieser Rechenressource unabhängig von den [SlurmQueues/Networking/PlacementGroup](#)-Einstellungen die benannte Platzierungsgruppe zugewiesen.

[Richtlinie aktualisieren: Die Rechenflotte muss gestoppt oder QueueUpdateStrategieingerichtet sein, damit diese Einstellung für ein Update geändert werden kann.](#)

Name(Optional,String)

Der Name der Platzierungsgruppe für eine bestehende Cluster-Platzierungsgruppe, die für die Rechenressource verwendet wird.

[Richtlinie aktualisieren: Die Rechenflotte muss gestoppt oder QueueUpdateStrategieingerichtet sein, damit diese Einstellung für ein Update geändert werden kann.](#)

Note

- Wenn `Enabled` sowohl `PlacementGroup` als auch `Name` nicht gesetzt sind, werden für ihre jeweiligen Werte standardmäßig die [PlacementGroup](#)Einstellungen [SlurmQueues/Networking](#) verwendet.
- `ComputeResources/Networking/PlacementGroup` wird mit AWS ParallelCluster Version 3.3.0 hinzugefügt.

CustomSlurmSettings(Fakultativ,Dict)

(Optional) Definiert die Konfigurationseinstellungen für den benutzerdefinierten Slurm Knoten (Rechenressource).

Gibt ein Wörterbuch mit Schlüssel-Wert-Paaren für benutzerdefinierte Slurm Konfigurationsparameter an, die für Slurm Knoten (Rechenressourcen) gelten.

Jedes einzelne Schlüssel-Wert-Paar, z. B. `Param1: Value1`, wird separat am Ende der Slurm Knotenkonfigurationszeile im Format hinzugefügt. `Param1=Value1`


Sie können nur Slurm Konfigurationsparameter angeben, die nicht auf der Negativliste stehen. CustomSlurmSettings Hinweise zu Slurm Konfigurationsparametern auf der Sperrliste finden Sie unter. [Auf der Denim-Liste Slurm Konfigurationsparameter für CustomSlurmSettings](#)

AWS ParallelCluster prüft nur, ob ein Parameter auf einer Sperrliste steht. AWS ParallelCluster validiert die Syntax oder Semantik Ihrer benutzerdefinierten Slurm Konfigurationsparameter nicht. Sie sind für die Validierung Ihrer benutzerdefinierten Slurm Konfigurationsparameter verantwortlich. Ungültige benutzerdefinierte Slurm Konfigurationsparameter können zu Slurm Daemon-Fehlern führen, die zu Fehlern bei der Clustererstellung und -aktualisierung führen können.

Weitere Hinweise zum Angeben von benutzerdefinierten Slurm Konfigurationsparametern mit finden Sie AWS ParallelCluster unter [SlurmAnpassung der Konfiguration](#).

Weitere Informationen zu Slurm Konfigurationsparametern finden Sie in der [Dokumentation unter slurm.conf](#). Slurm

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

 Note

CustomSlurmSettings wird ab Version 3.6.0 unterstützt. AWS ParallelCluster

Tags(Optional, [Zeichenfolge])

Eine Liste von Tag-Schlüssel-Wert-Paaren. ComputeResourceTags überschreiben doppelte Tags, die in [Tags Abschnitt](#) oder [SlurmQueues](#)/angegeben sind. Tags

Key(Optional, **String**)

Der Tag-Schlüssel.

Value(Fakultativ, **String**)

Der Tag-Wert.

[Richtlinie aktualisieren: Die Rechenflotte muss gestoppt oder QueueUpdateStrategieingerichtet sein, damit diese Einstellung für ein Update geändert werden kann.](#)

ComputeSettings

(Erforderlich) Definiert die ComputeSettings Konfiguration für die Slurm Warteschlange.

ComputeSettings-Eigenschaften

Gibt die Eigenschaften ComputeSettings der Knoten in der Slurm Warteschlange an.

```
ComputeSettings:  
LocalStorage:  
RootVolume:  
  Size: integer  
  Encrypted: boolean  
  VolumeType: string  
  Iops: integer  
  Throughput: integer  
EphemeralVolume:  
  MountDir: string
```

[Richtlinie aktualisieren: Die Rechenflotte muss gestoppt oder QueueUpdateStrategieingerichtet sein, damit diese Einstellung für ein Update geändert werden kann.](#)

LocalStorage(Fakultativ)

Gibt die Eigenschaften LocalStorage der Knoten in der Slurm Warteschlange an.

```
LocalStorage:  
RootVolume:  
  Size: integer  
  Encrypted: boolean  
  VolumeType: string  
  Iops: integer  
  Throughput: integer  
EphemeralVolume:  
  MountDir: string
```

[Richtlinie aktualisieren: Die Rechenflotte muss gestoppt oder QueueUpdateStrategieingerichtet sein, damit diese Einstellung für ein Update geändert werden kann.](#)

RootVolume(Fakultativ)

Gibt die Details des Root-Volumes der Knoten in der Slurm Warteschlange an.

RootVolume:

Size: *integer*

Encrypted: *boolean*

VolumeType: *string*

Iops: *integer*

Throughput: *integer*

[Richtlinie aktualisieren: Die Rechenflotte muss gestoppt oder QueueUpdateStrategieingerichtet sein, damit diese Einstellung für ein Update geändert werden kann.](#)

Size(Fakultativ,**Integer**)

Gibt die Größe des Root-Volumes in Gibibyte (GiB) für die Knoten in der Slurm Warteschlange an. Die Standardgröße stammt aus dem AMI. Die Verwendung einer anderen Größe erfordert, dass das AMI sie unterstütztgrowroot.

[Richtlinie aktualisieren: Die Rechenflotte muss gestoppt oder QueueUpdateStrategieingerichtet sein, damit diese Einstellung für ein Update geändert werden kann.](#)

Encrypted(Fakultativ,**Boolean**)

Falls `true`, ist das Root-Volume der Knoten in der Slurm Warteschlange verschlüsselt. Der Standardwert ist `false`.

[Richtlinie aktualisieren: Die Rechenflotte muss gestoppt oder QueueUpdateStrategieingerichtet sein, damit diese Einstellung für ein Update geändert werden kann.](#)

VolumeType(Fakultativ,**String**)

Gibt den [Amazon EBS-Volumetyp](#) der Knoten in der Slurm Warteschlange an. Unterstützte Werte sind `gp2gp3`, `io1`, `io2`, `sc1st1`, und `standard`. Der Standardwert ist `gp3`.

Weitere Informationen finden Sie unter [Amazon EBS-Volumen-Typen](#) in der Amazon EC2-Benutzerhandbuch für Linux-Instances.

[Richtlinie aktualisieren: Die Rechenflotte muss gestoppt oder QueueUpdateStrategieeingesetzt sein, damit diese Einstellung für ein Update geändert werden kann.](#)

Iops(Fakultativ,**Boolean**)

Definiert die Anzahl der IOPS für Volumes gp3 vom Typ io1io2, und.

Der Standardwert, die unterstützten Werte und das volume_size Verhältnis volume_iops zum Verhältnis variieren je nach VolumeType undSize.

VolumeType = io1

Standard Iops = 100

Unterstützte Werte Iops = 100—64000 †

Maximales volume_iops volume_size Verhältnis = 50 IOPS pro GiB. 5000 IOPS erfordern einen Wert volume_size von mindestens 100 GiB.

VolumeType = io2

Standard Iops = 100

Unterstützte Werte Iops = 100—64000 (256000 für io2 Block Express-Volumes) †

Maximales Iops Size Verhältnis = 500 IOPS pro GiB. 5000 IOPS erfordern einen Wert Size von mindestens 10 GiB.

VolumeType = gp3

Standard Iops = 3000

Unterstützte Werte Iops = 3000—16000 †

Maximales Iops Size Verhältnis = 500 IOPS pro GiB für Volumes mit mehr als 3000 IOPS.

† Maximale IOPS wird nur für [Instances garantiert, die auf dem Nitro-System basieren](#) und auch mit mehr als 32.000 IOPS ausgestattet sind. Andere Instanzen können bis zu 32.000 IOPS haben. Frühere io1 Volumes erreichen möglicherweise nicht die volle Leistung, es sei denn, Sie [ändern das](#) Volume. io2 Block Express-Volumes unterstützen

`volume_iops` Werte bis zu 256000 für R5b Instance-Typen. Weitere Informationen finden Sie unter [io2Block Express-Volumes](#) im Amazon EC2 EC2-Benutzerhandbuch für Linux-Instances.

[Richtlinie aktualisieren: Die Rechenflotte muss gestoppt oder QueueUpdateStrategieeingesetzt sein, damit diese Einstellung für ein Update geändert werden kann.](#)

Throughput(Optional,**Integer**)

Definiert den Durchsatz für gp3 Volumetypen in MiB/s. Diese Einstellung ist nur gültig, wenn sie `VolumeType` ist. gp3 Der Standardwert ist 125. Unterstützte Werte: 125—1000 MiB/s

Das Verhältnis von Throughput zu Iops darf nicht mehr als 0,25 betragen. Der maximale Durchsatz von 1000 MiB/s setzt voraus, dass die Iops Einstellung mindestens 4000 beträgt.

[Richtlinie aktualisieren: Die Rechenflotte muss gestoppt oder QueueUpdateStrategieeingesetzt sein, damit diese Einstellung für ein Update geändert werden kann.](#)

EphemeralVolume(Fakultativ,) **Boolean**

Gibt die Einstellungen für das kurzlebige Volumen an. Das ephemere Volume wird erstellt, indem alle Instance-Speicher-Volumes zu einem einzigen logischen Volume zusammengefasst werden, das mit dem Dateisystem formatiert ist. ext4 Der Standardwert ist `/scratch`. Wenn der Instance-Typ keine Instance-Speicher-Volumes hat, wird kein ephemeres Volume erstellt. Weitere Informationen finden Sie unter [Instance-Speicher-Volumes](#) im Amazon EC2 EC2-Benutzerhandbuch für Linux-Instances.

[EphemeralVolume](#):
[MountDir](#): *string*

[Richtlinie aktualisieren: Die Rechenflotte muss gestoppt oder QueueUpdateStrategieeingesetzt sein, damit diese Einstellung für ein Update geändert werden kann.](#)

MountDir(Optional,**String**)

Das Mount-Verzeichnis für das ephemere Volume für jeden Knoten in der Slurm Warteschlange.

[Richtlinie aktualisieren: Die Rechenflotte muss gestoppt oder QueueUpdateStrategieingerichtet sein, damit diese Einstellung für ein Update geändert werden kann.](#)

CustomActions

(Optional) Gibt benutzerdefinierte Skripts an, die auf den Knoten in der Slurm Warteschlange ausgeführt werden sollen.

```
CustomActions:
  OnNodeStart:
    Sequence:
      - Script: string
        Args:
          - string
    Script: string
    Args:
      - string
  OnNodeConfigured:
    Sequence:
      - Script: string
        Args:
          - string
    Script: string
    Args:
      - string
```

[Richtlinie aktualisieren: Die Rechenflotte muss gestoppt oder QueueUpdateStrategieingerichtet sein, damit diese Einstellung für ein Update geändert werden kann.](#)

CustomActionsEigenschaften

OnNodeStart(Fakultativ, **String**)

Gibt eine Sequenz von Skripten oder ein einzelnes Skript an, das auf den Knoten in der Slurm Warteschlange ausgeführt werden soll, bevor eine Bootstrap-Aktion zur Knotenbereitstellung gestartet wird. AWS ParallelCluster unterstützt nicht, sowohl ein einzelnes Skript als auch Sequence dieselbe benutzerdefinierte Aktion einzubeziehen. Weitere Informationen finden Sie unter [Benutzerdefinierte Bootstrap-Aktionen.](#)

Sequence(Fakultativ)

Liste der auszuführenden Skripts.

[Richtlinie aktualisieren: Die Rechenflotte muss gestoppt oder QueueUpdateStrategieingerichtet sein, damit diese Einstellung für ein Update geändert werden kann.](#)

Script(Erforderlich,**String**)

Die zu verwendende Datei. Der Dateipfad kann mit `https://` oder `beginns3://`.

[Richtlinie aktualisieren: Die Rechenflotte muss gestoppt oder QueueUpdateStrategieingerichtet sein, damit diese Einstellung für ein Update geändert werden kann.](#)

Args(Fakultativ,**[String]**)

Die Liste der Argumente, die an das Skript übergeben werden sollen.

[Richtlinie aktualisieren: Die Rechenflotte muss gestoppt oder QueueUpdateStrategieingerichtet sein, damit diese Einstellung für ein Update geändert werden kann.](#)

Script(Erforderlich,**String**)

Die Datei, die für ein einzelnes Skript verwendet werden soll. Der Dateipfad kann mit `https://` oder `beginns3://`.

[Richtlinie aktualisieren: Die Rechenflotte muss gestoppt oder QueueUpdateStrategieingerichtet sein, damit diese Einstellung für ein Update geändert werden kann.](#)

Args(Fakultativ,**[String]**)

Die Liste der Argumente, die an das einzelne Skript übergeben werden sollen.

[Richtlinie aktualisieren: Die Rechenflotte muss gestoppt oder QueueUpdateStrategieingerichtet sein, damit diese Einstellung für ein Update geändert werden kann.](#)

[Richtlinie aktualisieren: Die Rechenflotte muss gestoppt oder QueueUpdateStrategieingerichtet sein, damit diese Einstellung für ein Update geändert werden kann.](#)

OnNodeConfigured(Fakultativ,**String**)

Gibt eine Sequenz von Skripten oder ein einzelnes Skript an, das auf den Knoten in der Slurm Warteschlange ausgeführt wird, nachdem alle Knoten-Bootstrap-Aktionen abgeschlossen sind. AWS ParallelCluster unterstützt nicht, sowohl ein einzelnes Skript als auch Sequence dieselbe benutzerdefinierte Aktion einzubeziehen. Weitere Informationen finden Sie unter [Benutzerdefinierte Bootstrap-Aktionen](#).

Sequence(Fakultativ)

Liste der auszuführenden Skripts.

[Richtlinie aktualisieren: Die Rechenflotte muss gestoppt oder QueueUpdateStrategieingerichtet sein, damit diese Einstellung für ein Update geändert werden kann.](#)

Script(Erforderlich,**String**)

Die zu verwendende Datei. Der Dateipfad kann mit `https://` oder `beginners3://`.

[Richtlinie aktualisieren: Die Rechenflotte muss gestoppt oder QueueUpdateStrategieingerichtet sein, damit diese Einstellung für ein Update geändert werden kann.](#)

Args(Fakultativ,**[String]**)

Die Liste der Argumente, die an das Skript übergeben werden sollen.

[Richtlinie aktualisieren: Die Rechenflotte muss gestoppt oder QueueUpdateStrategieingerichtet sein, damit diese Einstellung für ein Update geändert werden kann.](#)

Script(Erforderlich,**String**)

Die Datei, die für ein einzelnes Skript verwendet werden soll. Der Dateipfad kann mit `https://` oder `beginners3://`.

[Richtlinie aktualisieren: Die Rechenflotte muss gestoppt oder QueueUpdateStrategieingerichtet sein, damit diese Einstellung für ein Update geändert werden kann.](#)

Args(Fakultativ,**[String]**)

Eine Liste von Argumenten, die an das einzelne Skript übergeben werden sollen.

Richtlinie aktualisieren: Die Rechenflotte muss gestoppt oder QueueUpdateStrategieingerichtet sein, damit diese Einstellung für ein Update geändert werden kann.

Richtlinie aktualisieren: Die Rechenflotte muss gestoppt oder QueueUpdateStrategieingerichtet sein, damit diese Einstellung für ein Update geändert werden kann.

Note

Sequence wird ab AWS ParallelCluster Version 3.6.0 hinzugefügt. Wenn Sie angebenSequence, können Sie mehrere Skripts für eine benutzerdefinierte Aktion auflisten. AWS ParallelCluster unterstützt weiterhin die Konfiguration einer benutzerdefinierten Aktion mit einem einzigen Skript, ohne dies einzuschließenSequence. AWS ParallelCluster unterstützt nicht, sowohl ein einzelnes Skript als auch Sequence dieselbe benutzerdefinierte Aktion einzubeziehen.

Iam

(Optional) Definiert optionale IAM-Einstellungen für die Slurm Warteschlange.

Iam:

S3Access:

- BucketName: *string*
- EnableWriteAccess: *boolean*
- KeyName: *string*

AdditionalIamPolicies:

- Policy: *string*

InstanceProfile: *string*

InstanceRole: *string*

Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.

IamEigenschaften

InstanceProfile(Fakultativ,**String**)

Gibt ein Instanzprofil an, um die Standard-Instanzrolle oder das Instanzprofil für die Slurm Warteschlange zu überschreiben. Sie können nicht sowohl als InstanceProfile

auch angeben `InstanceRole`. Das Format ist `arn:${Partition}:iam::
${Account}:instance-profile/${InstanceProfileName}`.

Wenn dies angegeben ist, können die `AdditionalIamPolicies` Einstellungen `S3Access` und nicht angegeben werden.

Es wird empfohlen, eine oder beide `AdditionalIamPolicies` Einstellungen für `S3Access` und anzugeben, da hinzugefügte Funktionen AWS ParallelCluster häufig neue Berechtigungen erfordern.

[Aktualisierungsrichtlinie: Die Rechenflotte muss gestoppt werden, damit diese Einstellung für ein Update geändert werden kann.](#)

InstanceRole(Optional,**String**)

Gibt eine Instanzrolle an, um die Standard-Instanzrolle oder das Instanzprofil für die Slurm Warteschlange zu überschreiben. Sie können nicht sowohl als `InstanceProfile` auch angeben `InstanceRole`. Das Format ist `arn:${Partition}:iam::
${Account}:role/${RoleName}`.

Wenn dies angegeben ist, können die `AdditionalIamPolicies` Einstellungen `S3Access` und nicht angegeben werden.

Es wird empfohlen, eine oder beide `AdditionalIamPolicies` Einstellungen für `S3Access` und anzugeben, da hinzugefügte Funktionen AWS ParallelCluster häufig neue Berechtigungen erfordern.

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

S3Access(Fakultativ)

Gibt einen Bucket für die Slurm Warteschlange an. Dies wird verwendet, um Richtlinien zu generieren, um den angegebenen Zugriff auf den Bucket in der Slurm Warteschlange zu gewähren.

Wenn dies angegeben ist, können die `InstanceRole` Einstellungen `InstanceProfile` und nicht angegeben werden.

Es wird empfohlen, eine oder beide `AdditionalIamPolicies` Einstellungen für `S3Access` und anzugeben, da hinzugefügte Funktionen AWS ParallelCluster häufig neue Berechtigungen erfordern.

S3Access:

- **BucketName**: *string*
- EnableWriteAccess**: *boolean*
- KeyName**: *string*

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

BucketName(Erforderlich,**String**)

Der Name des -Buckets.

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

KeyName(Fakultativ,**String**)

Der Schlüssel für den Eimer. Der Standardwert ist *.

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

EnableWriteAccess(Fakultativ,**Boolean**)

Gibt an, ob der Schreibzugriff für den Bucket aktiviert ist.

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

AdditionalIamPolicies(Fakultativ)

Gibt eine Liste von Amazon Resource Names (ARNs) von IAM-Richtlinien für Amazon EC2 an. Diese Liste ist zusätzlich zu den Berechtigungen, die für erforderlich sind, an die Root-Rolle angehängt, die für die Slurm Warteschlange verwendet wird. AWS ParallelCluster

Ein IAM-Richtlinienname und sein ARN sind unterschiedlich. Namen können nicht verwendet werden.

Wenn dies angegeben ist, können die InstanceRole Einstellungen InstanceProfile und nicht angegeben werden.

Wir empfehlen die Verwendung, AdditionalIamPolicies da sie zu den erforderlichen Berechtigungen hinzugefügt AdditionalIamPolicies werden und alle erforderlichen Berechtigungen enthalten InstanceRole müssen. AWS ParallelCluster Die erforderlichen Berechtigungen ändern sich häufig von Version zu Version, da Funktionen hinzugefügt werden.

Es gibt keinen Standardwert.

AdditionalIamPolicies:

- **Policy:** *string*

Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.

Policy(Erforderlich, **[String]**)

Liste der IAM-Richtlinien.

Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.

SlurmSettings

(Optional) Definiert die Einstellungen Slurm, die für den gesamten Cluster gelten.

SlurmSettings:

ScaledownIdleTime: *integer*
QueueUpdateStrategy: *string*
EnableMemoryBasedScheduling: *boolean*
CustomSlurmSettings: *[dict]*
CustomSlurmSettingsIncludeFile: *string*
Database:
Uri: *string*
UserName: *string*
PasswordSecretArn: *string*
Dns:
DisableManagedDns: *boolean*
HostedZoneId: *string*
UseEc2Hostnames: *boolean*

SlurmSettingsEigenschaften**ScaledownIdleTime**(Fakultativ, **Integer**)

Definiert den Zeitraum (in Minuten), für den es keinen Job gibt und der Slurm Knoten beendet wird.

Der Standardwert ist 10.

Aktualisierungsrichtlinie: Die Rechenflotte muss gestoppt werden, damit diese Einstellung für ein Update geändert werden kann.

MungeKeySecretArn(Optional,String)

Der Amazon-Ressourcenname (ARN) des AWS Secrets Manager Manager-Geheimnisses im Klartext, das den Base64-kodierten Munge-Schlüssel enthält, der im Slurm-Cluster verwendet werden soll. Dieser Munge-Schlüssel wird verwendet, um RPC-Aufrufe zwischen Slurm-Client-Befehlen und Slurm-Daemons, die als Remote-Server agieren, zu authentifizieren. Wenn MungeKeySecretArn nicht angegeben, AWS ParallelCluster wird ein zufälliger Munge-Schlüssel für den Cluster generiert.

Note

MungeKeySecretArn wird ab AWS ParallelCluster Version 3.8.0 unterstützt.

Warning

Wenn der MungeKeySecretArn neu zu einem vorhandenen Cluster hinzugefügt ParallelCluster wird, wird der vorherige munge Key im Falle eines Rollbacks oder beim späteren Entfernen des nicht wiederhergestellt. MungeKeySecretArn Stattdessen wird ein neuer zufälliger Munge-Schlüssel generiert.

Ob der AWS ParallelCluster Benutzer die Erlaubnis hat, [DescribeSecret](#) auf diese bestimmte geheime Ressource zuzugreifen, MungeKeySecretArn wird überprüft. MungeKeySecretArn ist gültig, wenn:

- Das angegebene Geheimnis ist vorhanden, und
- Das Geheimnis ist Klartext und enthält eine gültige Base64-kodierte Zeichenfolge, und
- Der dekodierte binäre Munge-Schlüssel hat eine Größe zwischen 256 und 8192 Bit.

Wenn die IAM-Richtlinie für den Pcluster-Benutzer nicht einschließt DescribeSecret, MungeKeySecretArn wird sie nicht validiert und es wird eine Warnmeldung angezeigt. Weitere Informationen finden Sie unter [AWS ParallelCluster pclusterGrundlegende Benutzerrichtlinie](#).

Wenn Sie ein Update MungeKeySecretArn durchführen, müssen die Rechenflotte und alle Anmeldeknoten gestoppt werden.

Wenn der geheime Wert im geheimen ARN geändert wird, während der ARN gleich bleibt, wird der Cluster nicht automatisch mit dem neuen Munge-Schlüssel aktualisiert. Um den neuen

Munge-Schlüssel des geheimen ARN zu verwenden, müssen Sie die Compute-Flotte und die Anmeldeknoten stoppen und dann den folgenden Befehl vom Hauptknoten aus ausführen.

```
sudo /opt/parallelcluster/scripts/slurm/update_munge_key.sh
```

Nachdem Sie den Befehl ausgeführt haben, können Sie sowohl die Rechenflotte als auch die Anmeldeknoten wieder aufnehmen: Die neu bereitgestellten Rechen- und Anmeldeknoten werden automatisch mit dem neuen Munge-Schlüssel gestartet.

Um einen Base64-codierten benutzerdefinierten Munge-Schlüssel zu generieren, können Sie das im Lieferumfang der [Munge-Software enthaltene Mungekey-Hilfsprogramm](#) verwenden und es dann mit dem Base64-Hilfsprogramm codieren, das allgemein in Ihrem Betriebssystem verfügbar ist. Alternativ können Sie entweder bash verwenden (bitte setzen Sie den bs-Parameter zwischen 32 und 1024)

```
dd if=/dev/random bs=128 count=1 2>/dev/null | base64 -w 0
```

oder Python wie folgt:

```
import random
import os
import base64

# key length in bytes
key_length=128

base64.b64encode(os.urandom(key_length)).decode("utf-8")
```

Richtlinie aktualisieren: **NEUE AKTUALISIERUNGSRICHTLINIE MIT GESTOPPTEN COMPUTE-FLOTTEN- UND ANMELDEKNOTEN** (fälschlicherweise nicht in 3.7.0 hinzugefügt).

QueueUpdateStrategy(Fakultativ,) **String**

Gibt die Ersatzstrategie für die [SlurmQueues](#) Abschnittsparameter an, für die die folgende Aktualisierungsrichtlinie gilt:

[Aktualisierungsrichtlinie: Die Rechenflotte muss gestoppt oder QueueUpdateStrategy eingerichtet sein, damit diese Einstellung für ein Update geändert werden kann.](#)

Der QueueUpdateStrategy Wert wird nur verwendet, wenn ein Cluster-Aktualisierungsprozess gestartet wird.

Zulässige Werte: COMPUTE_FLEET_STOP | DRAIN | TERMINATE

Standardwert: COMPUTE_FLEET_STOP

DRAIN

Knoten in Warteschlangen mit geänderten Parameterwerten sind auf DRAINING eingestellt. Knoten in diesem Status akzeptieren keine neuen Jobs und laufende Jobs werden bis zum Abschluss fortgesetzt.

Wenn ein Knoten zu `idle` (DRAINED) wird, wird ein Knoten ersetzt, wenn es sich um einen statischen Knoten handelt, und der Knoten wird beendet, wenn der Knoten dynamisch ist. Andere Knoten in anderen Warteschlangen ohne geänderte Parameterwerte sind nicht betroffen.

Die Zeit, die diese Strategie benötigt, um alle Warteschlangenknoten durch geänderte Parameterwerte zu ersetzen, hängt von der laufenden Arbeitslast ab.

COMPUTE_FLEET_STOP

Der Standardwert des `QueueUpdateStrategy` Parameters. Bei dieser Einstellung müssen Sie zum Aktualisieren der Parameter [SlurmQueues](#) im Abschnitt [die Rechenflotte beenden](#), bevor Sie ein Cluster-Update durchführen:

```
$ pcluster update-compute-fleet --status STOP_REQUESTED
```

TERMINATE

In Warteschlangen mit geänderten Parameterwerten werden laufende Jobs beendet und die Knoten werden sofort heruntergefahren.

Statische Knoten werden ersetzt und dynamische Knoten werden beendet.

Andere Knoten in anderen Warteschlangen ohne geänderte Parameterwerte sind nicht betroffen.

[Aktualisierungsrichtlinie: Diese Einstellung wird während eines Updates nicht analysiert.](#)

Note

`QueueUpdateStrategy` wird ab AWS ParallelCluster Version 3.2.0 unterstützt.

EnableMemoryBasedScheduling(Fakultativ, Boolean)

Falls `true` die speicherbasierte Planung in aktiviert ist. Slurm Weitere Informationen finden Sie unter [SlurmQueues//ComputeResources. SchedulableMemory](#)

Der Standardwert ist `false`.

Warning

Die Aktivierung der speicherbasierten Planung wirkt sich auf die Art und Weise aus, wie der Slurm Scheduler Jobs und die Knotenzuweisung verarbeitet.

Weitere Informationen finden Sie unter [Slurmspeicherbasierte Terminplanung](#).

Note

EnableMemoryBasedScheduling wird ab Version 3.2.0 unterstützt. AWS ParallelCluster

Note

[Ab AWS ParallelCluster Version 3.7.0 EnableMemoryBasedScheduling kann aktiviert werden, wenn Sie mehrere Instanztypen in Instances konfigurieren.](#)

Für die AWS ParallelCluster Versionen 3.2.0 bis 3.6. **x**

EnableMemoryBasedScheduling kann nicht aktiviert werden, wenn Sie mehrere Instanztypen in [Instances](#) konfigurieren.

[Aktualisierungsrichtlinie: Die Rechenflotte muss gestoppt werden, damit diese Einstellung für ein Update geändert werden kann.](#)

CustomSlurmSettings(Optional, [Dict])

Definiert die benutzerdefinierten Slurm Einstellungen, die für den gesamten Cluster gelten.

Gibt eine Liste von Slurm Konfigurationswörterbüchern mit Schlüssel-Wert-Paaren an, die an das Ende der generierten Datei angehängt werden. `slurm.conf` AWS ParallelCluster

Jedes Wörterbuch in der Liste wird als separate Zeile angezeigt, die der Konfigurationsdatei hinzugefügt wird. Slurm Sie können entweder einfache oder komplexe Parameter angeben.

Einfache Parameter bestehen aus einem einzigen key pair, wie in den folgenden Beispielen gezeigt:

```
- Param1: 100
- Param2: "SubParam1,SubParam2=SubValue2"
```

In der Slurm Konfiguration gerendertes Beispiel:

```
Param1=100
Param2=SubParam1,SubParam2=SubValue2
```

Komplexe Slurm Konfigurationsparameter bestehen aus mehreren durch Leerzeichen getrennten Schlüsselwertpaaren, wie in den nächsten Beispielen gezeigt:

```
- nodeName: test-nodes[1-10]
  CPUs: 4
  RealMemory: 4196
  ... # other node settings
- NodeSet: test-nodeset
  Nodes: test-nodes[1-10]
  ... # other nodeset settings
- PartitionName: test-partition
  Nodes: test-nodeset
  ... # other partition settings
```

Beispiel, gerendert in der Konfiguration: Slurm

```
nodeName=test-nodes[1-10] CPUs=4 RealMemory=4196 ... # other node settings
NodeSet=test-nodeset Nodes=test-nodes[1-10] ... # other nodeset settings
PartitionName=test-partition Nodes=test-nodeset ... # other partition settings
```

Note

Benutzerdefinierte Slurm Knoten dürfen die `-st-` `-dy-` Oder-Muster nicht in ihren Namen enthalten. Diese Muster sind Knoten vorbehalten, die von verwaltet werden AWS ParallelCluster.

Wenn Sie benutzerdefinierte Slurm Konfigurationsparameter in `angebenCustomSlurmSettings`, dürfen Sie keine benutzerdefinierten Slurm Konfigurationsparameter für `angebenCustomSlurmSettingsIncludeFile`.


Sie können nur Slurm Konfigurationsparameter angeben, die in nicht auf der Sperrliste stehen. `CustomSlurmSettings` Hinweise zu Slurm Konfigurationsparametern auf der Sperrliste finden Sie unter [Auf der Denim-ListeSlurmKonfigurationsparameter fürCustomSlurmSettings](#)

AWS ParallelCluster prüft nur, ob ein Parameter auf einer Sperrliste steht. AWS ParallelCluster validiert die Syntax oder Semantik Ihrer benutzerdefinierten Slurm Konfigurationsparameter nicht. Sie sind für die Validierung Ihrer benutzerdefinierten Slurm Konfigurationsparameter verantwortlich. Ungültige benutzerdefinierte Slurm Konfigurationsparameter können zu Slurm Daemon-Fehlern führen, die zu Fehlern bei der Clustererstellung und -aktualisierung führen können.

Weitere Hinweise zum Angeben von benutzerdefinierten Slurm Konfigurationsparametern mit finden Sie AWS ParallelCluster unter [SlurmAnpassung der Konfiguration](#).

Weitere Informationen zu Slurm Konfigurationsparametern finden Sie in der [Dokumentation unter slurm.conf](#). Slurm

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

 Note

`CustomSlurmSettings` wird ab Version 3.6.0 unterstützt. AWS ParallelCluster

`CustomSlurmSettingsIncludeFile`(Fakultativ, **String)**

Definiert die benutzerdefinierten Slurm Einstellungen, die für den gesamten Cluster gelten.

Gibt die benutzerdefinierte Slurm Datei an, die aus benutzerdefinierten Slurm Konfigurationsparametern besteht, die am Ende der AWS ParallelCluster generierten `slurm.conf` Datei angehängt werden.

Sie müssen den Pfad zur Datei angeben. Der Pfad kann mit `https://` oder `beginns3://`.

Wenn Sie benutzerdefinierte Slurm Konfigurationsparameter für `angebenCustomSlurmSettingsIncludeFile`, dürfen Sie keine benutzerdefinierten Slurm Konfigurationsparameter für `angebenCustomSlurmSettings`.

Note

Benutzerdefinierte Slurm Knoten dürfen keine `-dy-` Oder-Muster in ihren Namen enthalten. `-st-` Diese Muster sind Knoten vorbehalten, die von verwaltet werden AWS ParallelCluster.

Sie können nur Slurm Konfigurationsparameter angeben, die nicht auf der Sperrliste stehen. CustomSlurmSettingsIncludeFile Hinweise zu Slurm Konfigurationsparametern auf der Sperrliste finden Sie unter. [Auf der Denim-ListeSlurmKonfigurationsparameter fürCustomSlurmSettings](#)

AWS ParallelCluster prüft nur, ob ein Parameter auf einer Sperrliste steht. AWS ParallelCluster validiert die Syntax oder Semantik Ihrer benutzerdefinierten Slurm Konfigurationsparameter nicht. Sie sind für die Validierung Ihrer benutzerdefinierten Slurm Konfigurationsparameter verantwortlich. Ungültige benutzerdefinierte Slurm Konfigurationsparameter können zu Slurm Daemon-Fehlern führen, die zu Fehlern bei der Clustererstellung und -aktualisierung führen können.

Weitere Hinweise zum Angeben von benutzerdefinierten Slurm Konfigurationsparametern mit finden Sie AWS ParallelCluster unter [SlurmAnpassung der Konfiguration](#).

Weitere Informationen zu Slurm Konfigurationsparametern finden Sie in der [Dokumentation unter slurm.conf](#). Slurm

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

Note

CustomSlurmSettingswird ab Version 3.6.0 unterstützt. AWS ParallelCluster

Database

(Optional) Definiert die Einstellungen zur Aktivierung von Slurm Accounting auf dem Cluster. Weitere Informationen finden Sie unter [SlurmAbrechnung mit AWS ParallelCluster](#).

Database:

Uri: *string*

```
UserName: string  
PasswordSecretArn: string
```

[Aktualisierungsrichtlinie: Die Rechenflotte muss gestoppt werden, damit diese Einstellung für ein Update geändert werden kann.](#)

Database-Eigenschaften

Uri(Erforderlich,String)

Die Adresse des Datenbankservers, der als Backend für die Slurm Buchhaltung verwendet wird. Dieser URI muss als `host:port` formatiert sein und darf kein Schema enthalten, wie z. `mysql://`. Der Host kann entweder eine IP-Adresse oder ein DNS-Name sein, der vom Hauptknoten aufgelöst werden kann. Wenn kein Port bereitgestellt wird, wird der MySQL Standardport 3306 AWS ParallelCluster verwendet.

AWS ParallelCluster bootet die Slurm Accounting-Datenbank in den Cluster und muss auf die Datenbank zugreifen.

Die Datenbank muss erreichbar sein, bevor Folgendes passiert:

- Ein Cluster wird erstellt.
- SlurmDie Kontoführung wird mit einem Cluster-Update aktiviert.

[Aktualisierungsrichtlinie: Die Rechenflotte muss gestoppt werden, damit diese Einstellung für ein Update geändert werden kann.](#)


UserName(Erforderlich,String)

Die Identität, die Slurm verwendet wird, um eine Verbindung mit der Datenbank herzustellen, Kontoführungsprotokolle zu schreiben und Abfragen durchzuführen. Der Benutzer muss sowohl Lese- als auch Schreibberechtigungen für die Datenbank haben.

[Aktualisierungsrichtlinie: Die Rechenflotte muss gestoppt werden, damit diese Einstellung für ein Update geändert werden kann.](#)

PasswordSecretArn(Erforderlich,String)

Der Amazon-Ressourcenname (ARN) des AWS Secrets Manager Geheimnisses, das das `UserName` Klartext-Passwort enthält. Dieses Passwort wird zusammen mit `UserName` Slurm Accounting zur Authentifizierung auf dem Datenbankserver verwendet.

 Note


Achten Sie beim Erstellen eines Geheimnisses mit der AWS Secrets Manager Konsole darauf, „Andere Art von Geheimnis“ und Klartext auszuwählen und nur den Passworttext in das Geheimnis aufzunehmen.

Weitere Informationen zur Erstellung eines Geheimnisses finden Sie AWS Secrets Manager unter [Create an AWS Secrets Manager Secret](#)

Ob der Benutzer dazu berechtigt `PasswordSecretArn` ist [DescribeSecret](#), wird überprüft. `PasswordSecretArn` ist gültig, wenn das angegebene Geheimnis existiert. Wenn die Benutzer-IAM-Richtlinie dies nicht beinhaltet `DescribeSecret`, `PasswordSecretArn` nicht validiert wird und eine Warnmeldung angezeigt wird. Weitere Informationen finden Sie unter [AWS ParallelCluster pclusterGrundlegende Benutzerrichtlinie](#).

Wenn Sie ein `Update PasswordSecretArn` durchführen, muss die Rechenflotte gestoppt werden. Wenn sich der geheime Wert ändert und der geheime ARN sich nicht ändert, wird der Cluster nicht automatisch mit dem neuen Datenbankkennwort aktualisiert. Um den Cluster für den neuen geheimen Wert zu aktualisieren, müssen Sie den folgenden Befehl vom Hauptknoten aus ausführen, nachdem die Compute-Flotte gestoppt wurde.

```
$ sudo /opt/parallelcluster/scripts/slurm/update_slurm_database_password.sh
```

 Warning

Wir empfehlen, das Datenbankkennwort nur zu ändern, wenn die Rechenflotte gestoppt ist, um den Verlust von Buchhaltungsdaten zu vermeiden.

[Aktualisierungsrichtlinie: Die Rechenflotte muss gestoppt werden, damit diese Einstellung für ein Update geändert werden kann.](#)

DatabaseName(Fakultativ, **String**)

Name der Datenbank auf dem Datenbankserver (definiert durch den Parameter `Uri`), die für Slurm Accounting verwendet werden soll.

Der Name der Datenbank kann Kleinbuchstaben, Zahlen und Unterstriche enthalten. Der Name darf nicht länger als 64 Zeichen sein.

Dieser Parameter ist dem StorageLoc Parameter von [slurmdbd.conf](#) zugeordnet.

Wenn DatabaseName nicht angegeben, ParallelCluster wird der Name des Clusters verwendet, um einen Wert für zu definieren. StorageLoc

Die Aktualisierung von DatabaseName ist zulässig, wobei die folgenden Überlegungen zu beachten sind:

- Wenn eine Datenbank mit einem Namen noch DatabaseName nicht auf dem Datenbankserver existiert, erstellt slurmdbd sie. Es liegt in Ihrer Verantwortung, die neue Datenbank nach Bedarf neu zu konfigurieren (z. B. Hinzufügen der Buchhaltungseinheiten — Cluster, Konten, Benutzer, Assoziationen, QoS usw.).
- Wenn auf dem Datenbankserver DatabaseName bereits eine Datenbank mit einem Namen existiert, verwendet slurmdbd sie für die Slurm-Accounting-Funktionalität.

[Aktualisierungsrichtlinie: Die Rechenflotte muss gestoppt werden, damit diese Einstellung für ein Update geändert werden kann.](#)

Note

Datase wird ab Version 3.3.0 hinzugefügt.

Dns

(Optional) Definiert die EinstellungenSlurm, die für den gesamten Cluster gelten.

Dns:

DisableManagedDns: *boolean*

HostedZoneId: *string*

UseEc2Hostnames: *boolean*


DnsEigenschaften

DisableManagedDns(Fakultativ, **Boolean**)

Fallst `true`, werden die DNS-Einträge für den Cluster nicht erstellt und die Slurm Knotennamen können nicht aufgelöst werden.

AWS ParallelCluster Erstellt standardmäßig eine Route 53-Hosting-Zone, in der Knoten beim Start registriert werden. Der Standardwert ist `false`. Wenn auf gesetzt `DisableManagedDns` ist `true`, wird die Hosting-Zone nicht von erstellt AWS ParallelCluster.

Informationen zur Verwendung dieser Einstellung zur Bereitstellung von Clustern in Subnetzen ohne Internetzugang finden Sie unter [AWS ParallelCluster in einem einzelnen Subnetz ohne Internetzugang ohne Internetzugang ohne Internetzugang ohne Internetzugang..](#)

 **Warning**

Für den ordnungsgemäßen Betrieb des Clusters ist ein System zur Namensauflösung erforderlich. Wenn auf gesetzt `DisableManagedDns` ist `true`, müssen Sie ein System zur Namensauflösung bereitstellen. Um den EC2-Standard-DNS zu verwenden, setzen Sie ihn `UseEc2Hostnames` auf `true`. Alternativ können Sie Ihren eigenen DNS-Resolver konfigurieren und sicherstellen, dass die Knotennamen registriert werden, wenn Instances gestartet werden. Sie können dies beispielsweise tun, indem Sie [CustomActions/OnNodeStart](#) konfigurieren.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

HostedZoneId(Fakultativ, **String**)

Definiert eine benutzerdefinierte Route 53-Hosting-Zonen-ID, die für die DNS-Namensauflösung für den Cluster verwendet wird. Falls angegeben, werden Clusterknoten in der angegebenen Hosting-Zone AWS ParallelCluster registriert und keine verwaltete Hosting-Zone erstellt.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

UseEc2Hostnames(Optional, **Boolean**)

Fall `true`, sind Cluster-Rechenknoten mit dem standardmäßigen EC2-Hostnamen konfiguriert. Der Slurm `NodeHostName` wird ebenfalls mit diesen Informationen aktualisiert. Der Standardwert ist `false`.

Informationen zur Verwendung dieser Einstellung zur Bereitstellung von Clustern in Subnetzen ohne Internetzugang finden Sie unter [AWS ParallelCluster in einem einzelnen Subnetz ohne Internetzugang ohne Internetzugang ohne Internetzugang ohne Internetzugang..](#)

Note

Dieser Hinweis ist ab AWS ParallelCluster Version 3.3.0 nicht relevant.

Für AWS ParallelCluster unterstützte Versionen vor 3.3.0:

Wenn auf gesetzt `UseEc2Hostnames` ist `true`, wird die Slurm-Konfigurationsdatei mit den Skripten AWS ParallelCluster `prolog` und `epilog` gesetzt:

- `prolog` wird ausgeführt, um Knoteninformationen zu `/etc/hosts` den Rechenknoten hinzuzufügen, wenn jeder Job zugewiesen ist.
- `epilog` wird ausgeführt, um Inhalte zu bereinigen, die von geschrieben wurden `prolog`. Um benutzerdefinierte `epilog` Skripts `prolog` oder Skripts hinzuzufügen, fügen Sie sie den jeweiligen `/opt/slurm/etc/pcluster/epilog.d/` Ordnern `/opt/slurm/etc/pcluster/prolog.d/` oder hinzu.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

SharedStorage Abschnitt

(Optional) Die Einstellungen für den gemeinsam genutzten Speicher für den Cluster.

AWS ParallelCluster [unterstützt entweder die Verwendung von Amazon EBS, FSx for ONTAP und FSx for OpenZFS Shared Storage Volumes, Amazon EFS und FSx for Lustre Shared Storage File Systems oder File Caches.](#)

In SharedStorage diesem Abschnitt können Sie entweder externen oder verwalteten Speicher definieren:

- Externer Speicher bezieht sich auf ein vorhandenes Volume oder Dateisystem, das Sie verwalten. AWS ParallelCluster erstellt oder löscht es nicht.
- AWS ParallelCluster Verwalteter Speicher bezieht sich auf ein Volume oder Dateisystem, das AWS ParallelCluster erstellt wurde und gelöscht werden kann.

[Kontingente für gemeinsam genutzten Speicher](#) und weitere Informationen zur Konfiguration Ihres gemeinsam genutzten Speichers finden Sie [Gemeinsamer Speicher](#) unter Verwenden AWS ParallelCluster.

Note

Wenn AWS Batch es als Scheduler verwendet wird, ist FSx for Lustre nur auf dem Cluster-Hauptknoten verfügbar.

SharedStorage:

- MountDir: *string*
Name: *string*
StorageType: Ebs
EbsSettings:
 - VolumeType: *string*
 - Iops: *integer*
 - Size: *integer*
 - Encrypted: *boolean*
 - KmsKeyId: *string*
 - SnapshotId: *string*
 - Throughput: *integer*
 - VolumeId: *string*
 - DeletionPolicy: *string*
 - Raid:
 - Type: *string*
 - NumberOfVolumes: *integer*
- MountDir: *string*
Name: *string*
StorageType: Efs
EfsSettings:
 - Encrypted: *boolean*
 - KmsKeyId: *string*
 - EncryptionInTransit: *boolean*
 - IamAuthorization: *boolean*
 - PerformanceMode: *string*
 - ThroughputMode: *string*
 - ProvisionedThroughput: *integer*
 - FileSystemId: *string*
 - DeletionPolicy: *string*
- MountDir: *string*
Name: *string*
StorageType: FsxLustre
FsxLustreSettings:
 - StorageCapacity: *integer*
 - DeploymentType: *string*

```
ImportedFileChunkSize: integer
DataCompressionType: string
ExportPath: string
ImportPath: string
WeeklyMaintenanceStartTime: string
AutomaticBackupRetentionDays: integer
CopyTagsToBackups: boolean
DailyAutomaticBackupStartTime: string
PerUnitStorageThroughput: integer
BackupId: string
KmsKeyId: string
FileSystemId: string
AutoImportPolicy: string
DriveCacheType: string
StorageType: string
DeletionPolicy: string
DataRepositoryAssociations:
- Name: string
  BatchImportMetaDataOnCreate: boolean
  DataRepositoryPath: string
  FileSystemPath: string
  ImportedFileChunkSize: integer
  AutoExportPolicy: string
  AutoImportPolicy: string
- MountDir: string
  Name: string
  StorageType: FsxOntap
  FsxOntapSettings:
    VolumeId: string
- MountDir: string
  Name: string
  StorageType: FsxOpenZfs
  FsxOpenZfsSettings:
    VolumeId: string
- MountDir: string
  Name: string
  StorageType: FileCache
  FileCacheSettings:
    FileCacheId: string
```

SharedStorageRichtlinien aktualisieren

- Für verwaltetes/externes EBS, verwaltetes EFS und verwaltetes FSx Lustre lautet die Aktualisierungsrichtlinie [Richtlinie aktualisieren: Für diese Einstellung mit Listenwerten muss die Rechenflotte gestoppt oder QueueUpdateStrategyso eingestellt werden, dass sie einen neuen Wert hinzufügt. Die Rechenflotte muss gestoppt werden, wenn ein vorhandener Wert entfernt wird.](#)
- Für externe EFS, FSx Lustre, FSx ONTAP, FSx OpenZfs und File Cache lautet die Aktualisierungsrichtlinie: [Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

SharedStorage-Eigenschaften

MountDir(Erforderlich,) String

Der Pfad, in dem der gemeinsam genutzte Speicher bereitgestellt wird.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

Name(Erforderlich,String)

Der Name des gemeinsam genutzten Speichers. Sie verwenden diesen Namen, wenn Sie die Einstellungen aktualisieren.

Warning

Wenn Sie AWS ParallelCluster verwalteten gemeinsamen Speicher angeben und den Wert für ändernName, werden der vorhandene verwaltete gemeinsame Speicher und die Daten gelöscht und ein neuer verwalteter gemeinsam genutzter Speicher erstellt. Das Ändern des Werts für Name mit einem Cluster-Update entspricht dem Ersetzen des vorhandenen verwalteten gemeinsam genutzten Speichers durch einen neuen. Stellen Sie sicher, dass Sie Ihre Daten sichern, bevor Sie die Änderung vornehmen, Name falls Sie die Daten aus dem vorhandenen gemeinsam genutzten Speicher behalten müssen.

[Richtlinie aktualisieren: Für diese Einstellung mit Listenwerten muss die Rechenflotte gestoppt oder QueueUpdateStrategyso eingestellt werden, dass sie einen neuen Wert hinzufügt. Die Rechenflotte muss gestoppt werden, wenn ein vorhandener Wert entfernt wird.](#)

StorageType(Erforderlich,String)

Der Typ des gemeinsam genutzten Speichers. Unterstützte Werte sind EbsEfs,FsxLustre,FsxOntap, undFsxOpenZfs.

Weitere Informationen finden Sie unter [FsxLustreSettings](#), [FsxOntapSettings](#) und [FsxOpenZfsSettings](#).

Note

Wenn Sie es AWS Batch als Scheduler verwenden, ist FSx for Lustre nur auf dem Cluster-Hauptknoten verfügbar.

Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.

EbsSettings

(Optional) Die Einstellungen für ein Amazon EBS-Volume.

EbsSettings:

```
VolumeType: string
Iops: integer
Size: integer
Encrypted: boolean
KmsKeyId: string
SnapshotId: string
VolumeId: string
Throughput: integer
DeletionPolicy: string
Raid:
  Type: string
  NumberOfVolumes: integer
```

Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.

EbsSettings-Eigenschaften

Wenn der [DeletionPolicy](#)Wert auf gesetzt istDelete, wird ein verwaltetes Volume mit seinen Daten gelöscht, wenn der Cluster gelöscht wird oder wenn das Volume mit einem Cluster-Update entfernt wird.

Weitere Informationen finden Sie [Gemeinsamer Speicher](#) unter Verwenden AWS ParallelCluster.

VolumeType(Optional,String)

Gibt den [Amazon EBS-Volumetyp](#) an. Unterstützte Werte sind gp2gp3,io1,io2, sc1st1, undstandard. Der Standardwert ist gp3.

Weitere Informationen finden Sie unter [Amazon EBS-Volume-Typen](#) in der Amazon EC2-Benutzerhandbuch für Linux-Instances.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

Iops(Fakultativ,Integer)

Definiert die Anzahl der IOPS für Volumes gp3 vom Typ io1io2, und.

Der Standardwert, die unterstützten Werte und das volume_size Verhältnis volume_iops zum Verhältnis variieren je nach VolumeType undSize.

VolumeType = io1

Standard Iops = 100

Unterstützte Werte Iops = 100—64000 †

Maximales volume_iops volume_size Verhältnis = 50 IOPS für jedes GiB. 5000 IOPS erfordern einen Wert volume_size von mindestens 100 GiB.

VolumeType = io2

Standard Iops = 100

Unterstützte Werte Iops = 100—64000 (256000 für io2 Block Express-Volumes) †

Maximales Iops Size Verhältnis = 500 IOPS für jedes GiB. 5000 IOPS erfordern einen Wert Size von mindestens 10 GiB.

VolumeType = gp3

Standard Iops = 3000

Unterstützte Werte Iops = 3000—16000

Maximales Iops Size Verhältnis = 500 IOPS für jedes GiB. 5000 IOPS erfordern einen Wert Size von mindestens 10 GiB.

† Maximale IOPS wird nur für [Instances garantiert, die auf dem Nitro-System basieren](#) und mit mehr als 32.000 IOPS ausgestattet sind. Andere Instanzen garantieren bis zu 32.000 IOPS. Wenn Sie [das Volume nicht ändern](#), erreichen frühere io1 Volumes möglicherweise nicht die volle Leistung. io2 Block Express-Volumes unterstützen `volume_iops` Werte bis zu 256000 für R5b Instance-Typen. Weitere Informationen finden Sie unter [io2Block Express-Volumes](#) im Amazon EC2 EC2-Benutzerhandbuch für Linux-Instances.

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

Size(Optional,Integer)

Gibt die Datenträgergröße in Gibibytes (GiB) an. Der Standardwert ist 35.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

Encrypted(Fakultativ,Boolean)

Gibt an, ob das Volume verschlüsselt ist. Der Standardwert ist `true`.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

KmsKeyId(Fakultativ,String)

Gibt einen benutzerdefinierten AWS KMS Schlüssel an, der für die Verschlüsselung verwendet werden soll. Für diese Einstellung muss die `Encrypted` Einstellung auf `gesetzt` sein `true`.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

SnapshotId(Optional,String)


Gibt die Amazon EBS-Snapshot-ID an, wenn Sie einen Snapshot als Quelle für das Volume verwenden.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

VolumeId(Optional,String)

Gibt die Amazon EBS-Volume-ID an. Wenn dies für eine `EbsSettings` Instance angegeben wird, kann auch nur der `MountDir` Parameter angegeben werden.

Das Volume muss in derselben Availability Zone wie das erstellt werden `HeadNode`.

 Note

In AWS ParallelCluster Version 3.4.0 wurden mehrere Availability Zones hinzugefügt.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

Throughput(Fakultativ,Integer)

Der Durchsatz in MiB/s bis zur Bereitstellung für ein Volumen mit einem Maximum von 1.000 MiB/s.

Diese Einstellung ist nur gültig, wenn sie gültig ist. VolumeType gp3 Der unterstützte Bereich liegt zwischen 125 und 1000, der Standardwert ist 125.

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

DeletionPolicy(Optional,String)

Gibt an, ob das Volume beibehalten, gelöscht oder als Snapshot erstellt werden soll, wenn der Cluster gelöscht oder das Volume entfernt wird. Die unterstützten Werte sind DeleteRetain, und. Snapshot Der Standardwert ist Delete.

Wenn die [DeletionPolicy](#)Einstellung auf gesetzt istDelete, wird ein verwaltetes Volume mit seinen Daten gelöscht, wenn der Cluster gelöscht wird oder wenn das Volume mit einem Cluster-Update entfernt wird.

Weitere Informationen finden Sie unter [Gemeinsamer Speicher](#).

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

Note

DeletionPolicywird ab AWS ParallelCluster Version 3.2.0 unterstützt.

Raid

(Optional) Definiert die Konfiguration eines RAID-Volumes.

Raid:

Type: *string*

NumberOfVolumes: *integer*

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

Raid-Eigenschaften

Type(Erforderlich,String)

Definiert den Typ des RAID-Arrays. Unterstützte Werte sind „0“ (gestreift) und „1“ (gespiegelt).

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

NumberOfVolumes(Optional,) Integer

Definiert die Anzahl der Amazon EBS-Volumes, die zur Erstellung des RAID-Arrays verwendet werden sollen. Der unterstützte Wertebereich liegt zwischen 2 und 5. Der Standardwert (wenn die Raid Einstellung definiert ist) ist 2.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

EfsSettings

(Optional) Die Einstellungen für ein Amazon EFS-Dateisystem.

[EfsSettings:](#)

[Encrypted:](#) *boolean*

[KmsKeyId:](#) *string*

[EncryptionInTransit:](#) *boolean*

[IamAuthorization:](#) *boolean*

[PerformanceMode:](#) *string*

[ThroughputMode:](#) *string*

[ProvisionedThroughput:](#) *integer*

[FileSystemId:](#) *string*

[DeletionPolicy:](#) *string*

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

EfsSettings-Eigenschaften

Bei [DeletionPolicy](#) Einstellung auf Delete wird ein verwaltetes Dateisystem mit seinen Daten gelöscht, wenn der Cluster gelöscht wird oder wenn das Dateisystem mit einem Cluster-Update entfernt wird.

Weitere Informationen finden Sie [Gemeinsamer Speicher](#) unter Verwenden AWS ParallelCluster.

Encrypted(Optional,Boolean)

Gibt an, ob das Amazon EFS-Dateisystem verschlüsselt ist. Der Standardwert ist false.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

`KmsKeyId(Optional,String)`

Gibt einen benutzerdefinierten AWS KMS Schlüssel an, der für die Verschlüsselung verwendet werden soll. Für diese Einstellung muss die `Encrypted` Einstellung auf `gesetzt` sein `true`.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

`EncryptionInTransit(Optional,Boolean)`

Wenn auf `gesetzt` `true`, werden Amazon EFS-Dateisysteme mit Transport Layer Security (TLS) bereitgestellt. Standardmäßig ist dies auf `eingestellt` `false`.

Note

AWS Batch Wird als Scheduler verwendet, wird `EncryptionInTransit` nicht unterstützt.

Note

`EncryptionInTransit` wird ab AWS ParallelCluster Version 3.4.0 hinzugefügt.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

`IamAuthorization(Fakultativ,Boolean)`

Wenn auf `gesetzt` `true`, wird Amazon EFS mithilfe der IAM-Identität des Systems authentifiziert. Standardmäßig ist dies auf `eingestellt` `false`.

Note

Wenn `IamAuthorization` auf `true` festgelegt ist, muss `EncryptionInTransit` auch auf `true` festgelegt werden.

Note

AWS Batch Wird als Scheduler verwendet, wird `IamAuthorization` nicht unterstützt.

Note

IamAuthorization wird ab AWS ParallelCluster Version 3.4.0 hinzugefügt.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

PerformanceMode(Fakultativ,String)

Gibt den Leistungsmodus des Amazon EFS-Dateisystems an. Unterstützte Werte sind `generalPurpose` und `maxIO`. Der Standardwert ist `generalPurpose`. Weitere Informationen finden Sie unter [Leistungsmodi](#) im Benutzerhandbuch zu Amazon Elastic File System.

Wir empfehlen den `generalPurpose`-Leistungsmodus für die meisten Dateisysteme.

Dateisysteme, die den `maxIO`-Leistungsmodus verwenden, können auf einen höheren Gesamtdurchsatz und mehr Vorgänge pro Sekunde skalieren. Bei den meisten Dateioperationen gibt es jedoch einen Kompromiss zwischen etwas höheren Latenzen.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

ThroughputMode(Fakultativ,) String

Gibt den Durchsatzmodus des Amazon EFS-Dateisystems an. Unterstützte Werte sind `bursting` und `provisioned`. Der Standardwert ist `bursting`. Wann verwendet `provisioned` wird, `ProvisionedThroughput` muss angegeben werden.

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

ProvisionedThroughput(Erforderlich, wenn ThroughputMode istprovisioned,Integer)

Definiert den bereitgestellten Durchsatz (in MIB/s) des Amazon EFS-Dateisystems, gemessen in MIB/s. Dies entspricht dem [ProvisionedThroughputInMibps](#) Parameter in der Amazon EFS API-Referenz.

Wenn Sie diesen Parameter verwenden, müssen Sie `ThroughputMode` auf `provisioned` einstellen.

Der unterstützte Bereich ist 1 -1024. Um eine Erhöhung des Limits zu beantragen, wenden Sie sich an AWS Support.

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

FileSystemId(Fakultativ,String)

Definiert die Amazon EFS-Dateisystem-ID für ein vorhandenes Dateisystem.

Wenn der Cluster so konfiguriert ist, dass er sich über mehrere Availability Zones erstreckt, müssen Sie in jeder Availability Zone, die vom Cluster verwendet wird, ein Dateisystem-Mount-Ziel definieren.

Wenn dies angegeben ist, MountDir kann nur angegeben werden. Es EfsSettings kann kein anderer angegeben werden.

Wenn Sie diese Option festlegen, muss Folgendes für die von Ihnen definierten Dateisysteme gelten:

- Die Dateisysteme verfügen über ein vorhandenes Mount-Ziel in jeder Availability Zones des Clusters, wobei eingehender und ausgehender NFS-Verkehr von und aus zulässig ist. HeadNode ComputeNodes [In Scheduling//Networking SlurmQueues/werden mehrere Availability Zones konfiguriert. SubnetIds](#)

Um sicherzustellen, dass Datenverkehr zwischen dem Cluster und dem Dateisystem zulässig ist, können Sie einen der folgenden Schritte ausführen:

- Konfigurieren Sie die Sicherheitsgruppen des Mount-Ziels so, dass der Datenverkehr zum und vom CIDR oder der Präfixliste der Cluster-Subnetze zugelassen wird.

Note

AWS ParallelCluster überprüft, ob die Ports geöffnet sind und ob die CIDR- oder Präfixliste konfiguriert ist. AWS ParallelCluster validiert den Inhalt der CIDR-Block- oder Präfixliste nicht.

- Legen Sie mithilfe von [SecurityGroups](#) und [SlurmQueuesHeadNode/Networking/NetworkingSecurityGroups](#) benutzerdefinierte Sicherheitsgruppen für Clusterknoten fest. Die benutzerdefinierten Sicherheitsgruppen müssen so konfiguriert werden, dass sie den Datenverkehr zwischen dem Cluster und dem Dateisystem zulassen.

Note

Wenn alle Clusterknoten benutzerdefinierte Sicherheitsgruppen verwenden, wird AWS ParallelCluster nur überprüft, ob die Ports geöffnet sind. AWS ParallelCluster überprüft nicht, ob Quelle und Ziel richtig konfiguriert sind.

Warning

EFS OneZone wird nur unterstützt, wenn sich alle Rechenknoten und der Hauptknoten in derselben Availability Zone befinden. EFS OneZone kann nur ein Mount-Ziel haben.

Note

In AWS ParallelCluster Version 3.4.0 wurden mehrere Availability Zones hinzugefügt.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

DeletionPolicy(Fakultativ,String)

Gibt an, ob das Dateisystem beibehalten oder gelöscht werden soll, wenn das Dateisystem aus dem Cluster entfernt oder der Cluster gelöscht wird. Die unterstützten Werte sind `Delete` und `Retain`. Der Standardwert ist `Delete`.

Wenn auf gesetzt `DeletionPolicy` ist `Delete`, wird ein verwaltetes Dateisystem mit seinen Daten gelöscht, wenn der Cluster gelöscht wird oder wenn das Dateisystem mit einem Cluster-Update entfernt wird.

Weitere Informationen finden Sie unter [Gemeinsamer Speicher](#).

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

Note

`DeletionPolicy` wird ab AWS ParallelCluster Version 3.3.0 unterstützt.

FsxLustreSettings

Note

Sie müssen definieren `FsxLustreSettings`, ob für [StorageType](#) angegeben `FsxLustre` ist.

(Optional) Die Einstellungen für ein FSx for Lustre-Dateisystem.

FsxLustreSettings:

```
StorageCapacity: integer
DeploymentType: string
ImportedFileChunkSize: integer
DataCompressionType: string
ExportPath: string
ImportPath: string
WeeklyMaintenanceStartTime: string
AutomaticBackupRetentionDays: integer
CopyTagsToBackups: boolean
DailyAutomaticBackupStartTime: string
PerUnitStorageThroughput: integer
BackupId: string # BackupId cannot coexist with some of the fields
KmsKeyId: string
FileSystemId: string # FileSystemId cannot coexist with other fields
AutoImportPolicy: string
DriveCacheType: string
StorageType: string
DeletionPolicy: string
```

Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.

Note

Wenn AWS Batch es als Scheduler verwendet wird, ist FSx for Lustre nur auf dem Cluster-Hauptknoten verfügbar.

FsxLustreSettings-Eigenschaften

Wenn der [DeletionPolicy](#) Wert auf `Delete` gesetzt ist, wird ein verwaltetes Dateisystem mit seinen Daten gelöscht, wenn der Cluster gelöscht wird oder wenn das Dateisystem mit einem Cluster-Update entfernt wird.

Weitere Informationen finden Sie unter [Gemeinsamer Speicher](#).

StorageCapacity(Erforderlich,Integer)

Legt die Speicherkapazität des FSx for Lustre-Dateisystems in GiB fest. `StorageCapacity` ist erforderlich, wenn Sie ein neues Dateisystem erstellen. Geben Sie nicht an `StorageCapacity`, ob `BackupId` oder angegeben `FileSystemId` ist.

- Gültige Werte für `SCRATCH_2PERSISTENT_1`, und `PERSISTENT_2` Bereitstellungstypen sind 1200 GiB, 2400 GiB und Inkremente von 2400 GiB.
- Für den `SCRATCH_1`-Bereitstellungstyp lauten die gültigen Werte 1200 GiB, 2400 GiB sowie Inkremente von 3600 GiB.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

DeploymentType(Optional,String)

Gibt den Bereitstellungstyp des FSx for Lustre-Dateisystems an. Unterstützte Werte sind `SCRATCH_1`, `SCRATCH_2`, `PERSISTENT_1` und `PERSISTENT_2`. Der Standardwert ist `SCRATCH_2`.

Wählen Sie `SCRATCH_1` `SCRATCH_2` Bereitstellungstypen aus, wenn Sie temporären Speicher und eine kurzfristige Verarbeitung von Daten benötigen. Der `SCRATCH_2` Bereitstellungstyp bietet Verschlüsselung von Daten bei der Übertragung und eine höhere Burst-Durchsatzkapazität als `SCRATCH_1`.


Wählen Sie den `PERSISTENT_1` Bereitstellungstyp für längerfristige Speicherung und für durchsatzorientierte Workloads, die nicht latenzempfindlich sind. `PERSISTENT_1` unterstützt die Verschlüsselung von Daten während der Übertragung. Es ist überall verfügbar, AWS-Regionen wo FSx for Lustre verfügbar ist.

Wählen Sie den `PERSISTENT_2` Bereitstellungstyp für längerfristige Speicherung und für latenzempfindliche Workloads, die ein Höchstmaß an IOPS und Durchsatz erfordern. `PERSISTENT_2` unterstützt SSD-Speicher und bietet mehr `PerUnitStorageThroughput` (bis zu 1000 MB/s/TiB). `PERSISTENT_2` ist in einer begrenzten Anzahl von erhältlich. AWS-Regionen
Weitere Informationen zu Bereitstellungstypen und eine Liste der verfügbaren AWS-Regionen

PERSISTENT_2 Bereitstellungsarten finden Sie unter [Dateisystem-Bereitstellungsoptionen für FSx for Lustre](#) im Amazon FSx for Lustre-Benutzerhandbuch.

Die Verschlüsselung von Daten während der Übertragung wird automatisch aktiviert SCRATCH_2, wenn Sie über Amazon EC2 EC2-Instances PERSISTENT_1, die [diese Funktion](#) unterstützen, auf Dateisysteme vom Typ PERSISTENT_2 Deployment zugreifen.

Die Verschlüsselung von Daten während der Übertragung für SCRATCH_2 PERSISTENT_1, und PERSISTENT_2 Bereitstellungstypen wird unterstützt, wenn auf sie über unterstützte Instance-Typen zugegriffen wird AWS-Regionen. Weitere Informationen finden Sie unter [Verschlüsselung von Daten bei der Übertragung](#) im Amazon FSx for Lustre-Benutzerhandbuch.

 Note


Support für den PERSISTENT_2 Bereitstellungstyp wurde mit AWS ParallelCluster Version 3.2.0 hinzugefügt.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

ImportedFileChunkSize(Fakultativ,Integer)

Für Dateien, die aus einem Datenrepository importiert werden, bestimmt dieser Wert die Stripe-Anzahl und die maximale Datenmenge für jede Datei (in MiB), die auf einer einzelnen physischen Festplatte gespeichert ist. Die maximale Anzahl von Datenträgern, über die eine einzelne Datei als Stripeset zugeordnet werden kann, ist durch die Gesamtzahl der Datenträger begrenzt, aus denen sich das Dateisystem zusammensetzt.

Die Standard-Chunk-Größe ist 1.024 MiB (1 GiB) und kann bis auf 512.000 MiB (500 GiB) steigen. Amazon S3-Objekte haben eine maximale Größe von 5 TB.

 Note

Dieser Parameter wird für Dateisysteme, die den PERSISTENT_2 Bereitstellungstyp verwenden, nicht unterstützt. Anweisungen zur Konfiguration von Datenrepository-Verknüpfungen finden Sie unter [Verknüpfen Ihres Dateisystems mit einem S3-Bucket](#) im Amazon FSx for Lustre-Benutzerhandbuch.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

DataCompressionType(Optional,) String

Legt die Datenkomprimierungskonfiguration für das FSx for Lustre-Dateisystem fest. Der unterstützte Wert ist `LZ4`. `LZ4` gibt an, dass die Datenkomprimierung mit dem LZ4-Algorithmus aktiviert ist. Wenn `DataCompressionType` nicht angegeben, wird die Datenkomprimierung ausgeschaltet, wenn das Dateisystem erstellt wird.

Weitere Informationen finden Sie unter [Lustre-Datenkomprimierung](#).

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

ExportPath(Fakultativ,String)

Der Pfad in Amazon S3, in den das Stammverzeichnis Ihres FSx for Lustre-Dateisystems exportiert wird. Diese Einstellung wird nur unterstützt, wenn der `ImportPath` Parameter angegeben ist. Der Pfad muss denselben Amazon S3 S3-Bucket verwenden, wie unter `ImportPath`. Sie können ein optionales Präfix angeben, an das neue und geänderte Daten aus Ihrem FSx for Lustre-Dateisystem exportiert werden sollen. Wenn kein `ExportPath` Wert angegeben wird, legt FSx for Lustre einen Standardexportpfad fest. `s3://import-bucket/FSxLustre[creation-timestamp]` Der Zeitstempel weist das UTC-Format auf, z. B. `s3://import-bucket/FSxLustre20181105T222312Z`.

Der Amazon S3-Export-Bucket muss derselbe sein, wie der Import-Bucket, der von `ImportPath` angegeben wurde. Wenn Sie nur einen Bucket-Namen angeben, z. B. `s3://import-bucket` erhalten Sie eine 1:1 -Zuordnung von Dateisystemobjekten zu Amazon S3 S3-Bucket-Objekten. Diese Zuordnung bedeutet, dass die Eingabedaten in Amazon S3 beim Export überschrieben werden. Wenn Sie im Exportpfad ein benutzerdefiniertes Präfix angeben, z. B. `s3://import-bucket/[custom-optional-prefix]` exportiert FSx for Lustre den Inhalt Ihres Dateisystems in dieses Exportpräfix im Amazon S3 S3-Bucket.

Note

Dieser Parameter wird für Dateisysteme, die den `PERSISTENT_2` Bereitstellungstyp verwenden, nicht unterstützt. Konfigurieren Sie Datenrepository-Verknüpfungen wie unter [Verknüpfen Ihres Dateisystems mit einem S3-Bucket](#) im Amazon FSx for Lustre-Benutzerhandbuch beschrieben.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

ImportPath(Optional,) String

Der Pfad zum Amazon S3 S3-Bucket (einschließlich des optionalen Präfixes), den Sie als Daten-Repository für Ihr FSx for Lustre-Dateisystem verwenden. Das Stammverzeichnis Ihres FSx for Lustre-Dateisystems wird dem Stammverzeichnis des Amazon S3-Buckets zugeordnet, den Sie auswählen. Ein Beispiel ist `s3://import-bucket/optional-prefix`. Wenn Sie ein Präfix nach dem Amazon S3-Bucket-Namen angeben, werden nur Objektschlüssel mit diesem Präfix in das Dateisystem geladen.

Note

Dieser Parameter wird für Dateisysteme, die den `PERSISTENT_2` Bereitstellungstyp verwenden, nicht unterstützt. Konfigurieren Sie Datenrepository-Verknüpfungen wie unter [Verknüpfen Ihres Dateisystems mit einem S3-Bucket](#) im Amazon FSx for Lustre-Benutzerhandbuch beschrieben.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

WeeklyMaintenanceStartTime(Optional,) String

Die bevorzugte Startzeit für wöchentliche Wartungsarbeiten. Sie hat das "d:HH:MM" Format der Zeitzone UTC+0. Bei diesem Format d handelt es sich um die Wochentagsnummer von 1 bis 7, die mit Montag beginnt und mit Sonntag endet. Für dieses Feld sind Anführungszeichen erforderlich.

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

AutomaticBackupRetentionDays(Fakultativ,Integer)

Die Anzahl der Tage für die Aufbewahrung automatischer Sicherungen. Wenn Sie diesen Wert auf „0“ festlegen, werden automatische Sicherungen deaktiviert. Der unterstützte Bereich liegt zwischen 0 und 90. Der Standardwert ist 0. Diese Einstellung ist nur für die Verwendung mit `PERSISTENT_1` und für `PERSISTENT_2` Bereitstellungstypen gültig. Weitere Informationen finden Sie unter [Arbeiten mit Sicherungen](#) im Amazon FSx for Lustre-Benutzerhandbuch.

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

CopyTagsToBackups(Optional,Boolean)

Falls `true`, kopieren Sie die Tags für das FSx for Lustre-Dateisystem in Backups. Dieser Wert ist standardmäßig `false`. Wenn er auf `true` festgelegt ist, werden alle Tags für das Dateisystem

in alle automatischen und vom Benutzer angeordneten Backups kopiert, in denen der Benutzer keine Tags angibt. Wenn dieser Wert `true` ist und Sie einen oder mehrere Tags angeben, werden nur die angegebenen Tags in die Sicherungen kopiert. Wenn Sie beim Erstellen einer vom Benutzer initiierten Sicherung ein oder mehrere Tags angeben, werden unabhängig von diesem Wert keine Tags aus dem Dateisystem kopiert. Diese Einstellung ist nur für die Verwendung mit `PERSISTENT_1` und für `PERSISTENT_2` Bereitstellungstypen gültig.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

`DailyAutomaticBackupStartTime`(Optional,String)

Eine sich wiederholende tägliche Uhrzeit im HH:MM Format. HH ist die mit Nullen aufgefüllte Stunde des Tages (00-23). MM ist die mit Nullen aufgefüllte Minute der Stunde (00-59). `05:00` gibt beispielsweise täglich 5 Uhr morgens an. Diese Einstellung ist nur für die Verwendung mit `PERSISTENT_1` und für `PERSISTENT_2` Bereitstellungstypen gültig.

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

`PerUnitStorageThroughput`(Erforderlich für **`PERSISTENT_1`** und für **`PERSISTENT_2`** Bereitstellungstypen Integer)

Beschreibt die Menge des Lese- und Schreibdurchsatzes für jedes 1 Tebibyte Speicher in MB/s/TiB. Die Dateisystem-Durchsatzkapazität wird durch Multiplizieren der Dateisystem-Speicherkapazität (TiB) mit dem `PerUnitStorageThroughput` (MB/s/TiB) berechnet. Bei einem Dateisystem mit 2,4 TiB ergibt die Bereitstellung von 50 MB/s/TiB von `PerUnitStorageThroughput` einen Dateisystem-Durchsatz von 120 MB/s. Sie zahlen den Durchsatz, den Sie bereitstellen. Dies entspricht der Eigenschaft. [PerUnitStorageThroughput](#)

Zulässige Werte:

`PERSISTENT_1` SSD-Speicher: 50, 100, 200 MB/s/TiB.

`PERSISTENT_1` Festplattenspeicher: 12, 40 MB/s/TiB.

`PERSISTENT_2` SSD-Speicher: 125, 250, 500, 1000 MB/s/TiB.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

`BackupId`(**String**Fakultativ,)

Gibt die ID der Sicherung an, die für die Wiederherstellung des FSx for Lustre-Dateisystems aus einer vorhandenen Sicherung verwendet werden soll. Wenn die `BackupId` Einstellung angegeben ist, dürfen die `AutoImportPolicy` `PerUnitStorageThroughput` Einstellungen `DeploymentType` `ExportPath`, `KmsKeyId`, `ImportPath`, `ImportedFileChunkSize`, `StorageCapacity`

und nicht angegeben werden. Diese Einstellungen werden aus dem Backup gelesen. Darüber hinaus dürfen die `ImportedFileChunkSize` Einstellungen `AutoImportPolicy` `ExportPath` `ImportPath`,, und nicht angegeben werden. Dies entspricht der [BackupId](#)Eigenschaft.

Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.

`KmsKeyId`(Fakultativ,String)


Die ID der AWS Key Management Service (AWS KMS) -Schlüssel-ID, die verwendet wird, um die Daten des FSx for Lustre-Dateisystems für persistente FSx for Lustre-Dateisysteme im Ruhezustand zu verschlüsseln. Wenn nicht angegeben, wird der verwaltete Schlüssel FSx for Lustre verwendet. Die Dateisysteme `SCRATCH_1` und `SCRATCH_2` FSx for Lustre werden im Ruhezustand immer mit von FSx for Lustre verwalteten Schlüsseln verschlüsselt. Weitere Informationen finden Sie unter [Verschlüsseln](#) in der API-Referenz.AWS Key Management Service

Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.


`FileSystemId`(Fakultativ,String)

Gibt die ID eines vorhandenen FSx for Lustre-Dateisystems an.

Wenn diese Option angegeben ist, werden nur die `FileSystemId` Einstellungen `MountDir` und in `FsxLustreSettings` verwendet. Alle anderen Einstellungen in der `FsxLustreSettings` werden ignoriert.

 Note

Wenn der AWS Batch Scheduler verwendet wird, ist FSx for Lustre nur auf dem Hauptknoten verfügbar.

 Note

Das Dateisystem muss einer Sicherheitsgruppe zugeordnet sein, die eingehenden und ausgehenden TCP-Verkehr über die Ports 988, 1021, 1022 und 1023 zulässt.

Stellen Sie sicher, dass Datenverkehr zwischen dem Cluster und dem Dateisystem zulässig ist, indem Sie einen der folgenden Schritte ausführen:

- Konfigurieren Sie die Sicherheitsgruppen des Dateisystems so, dass der Datenverkehr zum und vom CIDR oder der Präfixliste der Cluster-Subnetze zugelassen wird.

Note

AWS ParallelCluster überprüft, ob die Ports geöffnet sind und ob die CIDR- oder Präfixliste konfiguriert ist. AWS ParallelCluster validiert den Inhalt der CIDR-Block- oder Präfixliste nicht.

- Legen Sie mithilfe von [SecurityGroups](#) und [SlurmQueuesHeadNode/Networking/](#) benutzerdefinierte Sicherheitsgruppen für Clusterknoten fest. [NetworkingSecurityGroups](#) Die benutzerdefinierten Sicherheitsgruppen müssen so konfiguriert werden, dass sie den Datenverkehr zwischen dem Cluster und dem Dateisystem zulassen.

Note

Wenn alle Clusterknoten benutzerdefinierte Sicherheitsgruppen verwenden, wird AWS ParallelCluster nur überprüft, ob die Ports geöffnet sind. AWS ParallelCluster überprüft nicht, ob Quelle und Ziel richtig konfiguriert sind.


[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

AutoImportPolicy(Fakultativ,String)

Wenn Sie Ihr FSx for Lustre-Dateisystem erstellen, werden Ihre vorhandenen Amazon S3 S3-Objekte als Datei- und Verzeichnislisten angezeigt. Verwenden Sie diese Eigenschaft, um auszuwählen, wie FSx for Lustre Ihre Datei- und Verzeichnislisten auf dem neuesten Stand hält, wenn Sie Objekte in Ihrem verknüpften Amazon S3 S3-Bucket hinzufügen oder ändern. AutoImportPolicy kann die folgenden Werte haben:

- NEW- Der automatische Import ist aktiviert. FSx for Lustre importiert automatisch Verzeichnislisten aller neuen Objekte, die dem verknüpften Amazon S3 S3-Bucket hinzugefügt wurden und die derzeit nicht im FSx for Lustre-Dateisystem existieren.
- NEW_CHANGED- Der automatische Import ist aktiviert. FSx for Lustre importiert automatisch Datei- und Verzeichnislisten aller neuen Objekte, die dem Amazon S3 S3-Bucket hinzugefügt wurden, und aller vorhandenen Objekte, die im Amazon S3 S3-Bucket geändert wurden, nachdem Sie diese Option ausgewählt haben.
- NEW_CHANGED_DELETED- Der automatische Import ist aktiviert. FSx for Lustre importiert automatisch Datei- und Verzeichnislisten aller neuen Objekte, die dem Amazon S3 S3-Bucket


hinzugefügt wurden, aller vorhandenen Objekte, die im Amazon S3 S3-Bucket geändert wurden, und aller Objekte, die im Amazon S3 S3-Bucket gelöscht wurden, nachdem Sie diese Option ausgewählt haben.

 Note

Support für NEW_CHANGED_DELETED wurde in AWS ParallelCluster Version 3.1.1 hinzugefügt.

Wenn `AutoImportPolicy` nicht angegeben, ist der automatische Import ausgeschaltet. FSx for Lustre aktualisiert nur Datei- und Verzeichnislisten aus dem verknüpften Amazon S3 S3-Bucket, wenn das Dateisystem erstellt wird. FSx for Lustre aktualisiert keine Datei- und Verzeichnislisten für neue oder geänderte Objekte, nachdem diese Option ausgewählt wurde.

Weitere Informationen finden Sie unter [Automatisches Importieren von Updates aus Ihrem S3-Bucket](#) im Amazon FSx for Lustre-Benutzerhandbuch.

 Note

Dieser Parameter wird für Dateisysteme, die den PERSISTENT_2 Bereitstellungstyp verwenden, nicht unterstützt. Anweisungen zur Konfiguration von Datenrepository-Verknüpfungen finden Sie unter [Verknüpfen Ihres Dateisystems mit einem S3-Bucket](#) im Amazon FSx for Lustre-Benutzerhandbuch.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

`DriveCacheType`(Optional,) String

Gibt an, dass das Dateisystem über einen SSD-Laufwerkscache verfügt. Dies kann nur festgelegt werden `HDD`, wenn die `StorageType` Einstellung auf `gesetzt` ist und die `DeploymentType` Einstellung auf `gesetzt` ist `PERSISTENT_1`. Dies entspricht der `DriveCacheType` Eigenschaft. Weitere Informationen finden Sie unter [FSx for Lustre-Bereitstellungsoptionen](#) im Amazon FSx for Lustre-Benutzerhandbuch.

Der einzige gültige Wert ist `READ`. Um den SSD-Laufwerk-Cache zu deaktivieren, geben Sie die Einstellung nicht an. `DriveCacheType`

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

StorageType(Optional,String)

Legt den Speichertyp für das FSx for Lustre-Dateisystem fest, das Sie erstellen. Gültige Werte sind SSD und HDD.

- Setzen Sie den Wert auf SSD, um Solid-State-Laufwerksspeicher zu verwenden.
- Legt fest, HDD dass Festplattenspeicher verwendet werden soll. HDD wird für PERSISTENT Bereitstellungstypen unterstützt.

Der Standardwert ist SSD. Weitere Informationen finden Sie unter [Speichertypoptionen](#) im Amazon FSx for Windows-Benutzerhandbuch und unter [Mehrfachspeicheroptionen](#) im Amazon FSx for Lustre Benutzerhandbuch.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

DeletionPolicy(Fakultativ,String)

Gibt an, ob das Dateisystem beibehalten oder gelöscht werden soll, wenn das Dateisystem aus dem Cluster entfernt oder der Cluster gelöscht wird. Die unterstützten Werte sind Delete und Retain. Der Standardwert ist Delete.

Wenn auf gesetzt [DeletionPolicy](#) ist Delete, wird ein verwaltetes Dateisystem mit seinen Daten gelöscht, wenn der Cluster gelöscht wird oder wenn das Dateisystem mit einem Cluster-Update entfernt wird.

Weitere Informationen finden Sie unter [Gemeinsamer Speicher](#).

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

Note

DeletionPolicy wird ab AWS ParallelCluster Version 3.3.0 unterstützt.

DataRepositoryAssociations(Fakultativ,String)

Liste der DRAs (bis zu 8 pro Dateisystem)

Jeder Datenrepository-Zuordnung muss ein eindeutiges Amazon-FSx-Dateisystemverzeichnis und ein eindeutiges S3-Bucket oder Präfix zugeordnet sein.

Sie können [ExportPath](#) und [ImportPath](#) nicht gleichzeitig mit DRAs verwenden. FsxLustreSettings

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

Name(Erforderlich,String)

Der Name des DRA. Sie verwenden diesen Namen, wenn Sie die Einstellungen aktualisieren.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

BatchImportMetaDataOnCreate(Fakultativ,Boolean)

Eine boolesche Kennzeichnung, die angibt, ob eine Aufgabe zum Importieren von Metadaten ausgeführt werden soll, nachdem die Zuordnung zum Datenspeicher erstellt wurde. Die Aufgabe wird ausgeführt, wenn diese Kennzeichnung auf `true` gesetzt ist.

Standardwert: `false`

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

DataRepositoryPath(Erforderlich,String)

Der Pfad zum Amazon-S3-Daten-Repository, das mit dem Dateisystem verknüpft werden soll. Der Pfad kann ein S3-Bucket oder ein Präfix im Format `s3://myBucket/myPrefix/` sein. Dieser Pfad gibt an, wohin im S3-Daten-Repository Dateien importiert oder exportiert werden.

Kann sich nicht mit anderen DRAs überschneiden

Pattern: `^[^\u0000\u0085\u2028\u2029\r\n]{3,4357}$`

Minimum: 3

Maximum: 4357

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

FileSystemPath(Erforderlich,String)

Ein Pfad im Dateisystem von Amazon FSx für Lustre, der auf ein übergeordnetes Verzeichnis (z. B. `/ns1/`) oder Unterverzeichnis (z. B. `/ns1/subdir/`) verweist, das `DataRepositoryPath` 1–1 zugeordnet wird. Der führende Schrägstrich im Namen ist erforderlich. Zwei Daten-Repository-Verknüpfungen dürfen keine überlappenden Dateisystempfade haben. Wenn beispielsweise ein Daten-Repository dem Dateisystempfad `/ns1/` zugeordnet ist, können Sie kein anderes Daten-Repository mit dem Dateisystempfad `/ns1/ns2` verknüpfen.

Dieser Pfad gibt an, wohin in Ihrem Dateisystem Dateien exportiert oder importiert werden. Dieses Dateisystemverzeichnis kann nur mit einem Amazon-S3-Bucket verknüpft werden und kein anderer S3-Bucket kann mit dem Verzeichnis verknüpft werden.

Kann sich nicht mit anderen DRAs überschneiden

Note

Wenn Sie als Dateisystempfad nur einen Schrägstrich (/) angeben, können Sie nur ein Daten-Repository mit dem Dateisystem verknüpfen. Sie können nur "/" als Dateisystempfad für das erste Datenrepository angeben, das einem Dateisystem zugeordnet ist.

Pattern: `^[^\u0000\u0085\u2028\u2029\r\n]{1,4096}$`

Minimum: 1

Maximum: 4096

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

ImportedFileChunkSize(Optional,Integer)

Für Dateien, die aus einem Daten-Repository importiert werden, bestimmt dieser Wert die Anzahl der Stripes und die maximale Datenmenge pro Datei (in MiB), die auf einem einzigen physischen Datenträger gespeichert wird. Die maximale Anzahl von Datenträgern, auf die eine einzelne Datei verteilt werden kann, ist durch die Gesamtzahl von Datenträgern begrenzt, aus denen das Dateisystem oder der Cache besteht.

Die Standard-Chunk-Größe ist 1.024 MiB (1 GiB) und kann bis auf 512.000 MiB (500 GiB) steigen. Amazon S3-Objekte haben eine maximale Größe von 5 TB.

Minimum: 1

Maximum: 4096

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

AutoExportPolicy(Fakultativ,Array of strings)

Die Liste kann einen oder mehrere der folgenden Werte enthalten:

- NEW – Neue Dateien und Verzeichnisse werden automatisch in das Daten-Repository exportiert, wenn sie dem Dateisystem hinzugefügt werden.
- CHANGED – Änderungen an Dateien und Verzeichnissen im Dateisystem werden automatisch in das Daten-Repository exportiert.

- DELETED – Dateien und Verzeichnisse werden automatisch auf dem Daten-Repository gelöscht, wenn sie auf dem Dateisystem gelöscht werden.

Sie können eine beliebige Kombination von Ereignistypen für Ihre `AutoExportPolicy` definieren.

Maximum: 3

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

`AutoImportPolicy`(Fakultativ,Array of strings)

Die Liste kann einen oder mehrere der folgenden Werte enthalten:

- NEW – Amazon FSx importiert automatisch Metadaten von Dateien, die dem verknüpften S3-Bucket hinzugefügt wurden und derzeit nicht im FSx-Dateisystem vorhanden sind.
- CHANGED – Amazon FSx aktualisiert Dateimetadaten automatisch und macht vorhandene Dateiinhalte im Dateisystem ungültig, wenn sich Dateien im Daten-Repository ändern.
- DELETED – Amazon FSx löscht automatisch Dateien im Dateisystem, wenn entsprechende Dateien im Daten-Repository gelöscht werden.

Sie können eine beliebige Kombination von Ereignistypen für Ihre `AutoImportPolicy` definieren.

Maximum: 3

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

FsxOntapSettings

Note

Sie müssen definieren `FsxOntapSettings`, ob für angegeben `FsxOntap` ist [StorageType](#).

(Optional) Die Einstellungen für ein FSx for ONTAP-Dateisystem.

`FsxOntapSettings`:

`VolumeId`: *string*

FsxOntapSettings-Eigenschaften

VolumeId(Erforderlich,) String

Gibt die Volume-ID des vorhandenen FSx for ONTAP-Systems an.

Note

- Wenn ein AWS Batch Scheduler verwendet wird, ist FSx for ONTAP nur auf dem Hauptknoten verfügbar.
- Wenn der Bereitstellungstyp FSx for ONTAP lautet `Multi-AZ`, stellen Sie sicher, dass die Routing-Tabelle des Head-Knoten-Subnetzes ordnungsgemäß konfiguriert ist.
- Support für FSx for ONTAP wurde in AWS ParallelCluster Version 3.2.0 hinzugefügt.
- Das Dateisystem muss einer Sicherheitsgruppe zugeordnet sein, die eingehenden und ausgehenden TCP- und UDP-Verkehr über die Ports 111, 635, 2049 und 4046 zulässt.

Stellen Sie sicher, dass Datenverkehr zwischen dem Cluster und dem Dateisystem zulässig ist, indem Sie eine der folgenden Aktionen ausführen:

- Konfigurieren Sie die Sicherheitsgruppen des Dateisystems so, dass der Datenverkehr zum und vom CIDR oder der Präfixliste der Cluster-Subnetze zugelassen wird.

Note

AWS ParallelCluster überprüft, ob die Ports geöffnet sind und ob die CIDR- oder Präfixliste konfiguriert ist. AWS ParallelCluster validiert den Inhalt der CIDR-Block- oder Präfixliste nicht.

- Legen Sie mithilfe von [SecurityGroups](#) und [SlurmQueuesHeadNode/Networking/](#) benutzerdefinierte Sicherheitsgruppen für Clusterknoten fest. [NetworkingSecurityGroups](#) Die benutzerdefinierten Sicherheitsgruppen müssen so konfiguriert werden, dass sie den Datenverkehr zwischen dem Cluster und dem Dateisystem zulassen.

Note

Wenn alle Clusterknoten benutzerdefinierte Sicherheitsgruppen verwenden, wird AWS ParallelCluster nur überprüft, ob die Ports geöffnet sind. AWS ParallelCluster überprüft nicht, ob Quelle und Ziel richtig konfiguriert sind.

Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.

FsxOpenZfsSettings**Note**

Sie müssen definieren `FsxOpenZfsSettings`, ob für angegeben `FsxOpenZfs` ist [StorageType](#).

(Optional) Die Einstellungen für ein FSx for OpenZFS-Dateisystem.

[FsxOpenZfsSettings](#):

[VolumeId](#): *string*

Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.

FsxOpenZfsSettings-Eigenschaften

`VolumeId`(Erforderlich,) `String`

Spezifiziert die Volume-ID des bestehenden FSx for OpenZFS-Systems.

Note

- Wenn ein AWS Batch Scheduler verwendet wird, ist FSx for OpenZFS nur auf dem Hauptknoten verfügbar.
- Support für FSx for OpenZFS wurde in AWS ParallelCluster Version 3.2.0 hinzugefügt.

- Das Dateisystem muss einer Sicherheitsgruppe zugeordnet sein, die eingehenden und ausgehenden TCP- und UDP-Verkehr über die Ports 111, 2049, 20001, 20002 und 20003 zulässt.

Stellen Sie sicher, dass Datenverkehr zwischen dem Cluster und dem Dateisystem zulässig ist, indem Sie einen der folgenden Schritte ausführen:

- Konfigurieren Sie die Sicherheitsgruppen des Dateisystems so, dass der Datenverkehr zum und vom CIDR oder der Präfixliste der Cluster-Subnetze zugelassen wird.

Note

AWS ParallelCluster überprüft, ob die Ports geöffnet sind und ob die CIDR- oder Präfixliste konfiguriert ist. AWS ParallelCluster validiert den Inhalt der CIDR-Block- oder Präfixliste nicht.

- Legen Sie mithilfe von [SecurityGroups](#) und [SlurmQueuesHeadNode/Networking/](#) benutzerdefinierte Sicherheitsgruppen für Clusterknoten fest. [NetworkingSecurityGroups](#) Die benutzerdefinierten Sicherheitsgruppen müssen so konfiguriert werden, dass sie den Datenverkehr zwischen dem Cluster und dem Dateisystem zulassen.

Note

Wenn alle Clusterknoten benutzerdefinierte Sicherheitsgruppen verwenden, wird AWS ParallelCluster nur überprüft, ob die Ports geöffnet sind. AWS ParallelCluster überprüft nicht, ob Quelle und Ziel richtig konfiguriert sind.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

FileCacheSettings

Note

Sie müssen definieren `FileCacheSettings`, ob für angegeben `FileCache` ist [StorageType](#).

(Optional) Die Einstellungen für einen Datei-Cache.

FileCacheSettings:

FileCacheId: *string*

Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.

FileCacheSettings-Eigenschaften

FileCacheId(Erforderlich,String)

Gibt die Datei-Cache-ID eines vorhandenen Datei-Caches an.

Note

- Der Datei-Cache unterstützt keine AWS Batch Scheduler.
- Support für File Cache wurde in AWS ParallelCluster Version 3.7.0 hinzugefügt.
- Das Dateisystem muss einer Sicherheitsgruppe zugeordnet sein, die eingehenden und ausgehenden TCP-Verkehr über Port 988 zulässt.

Stellen Sie sicher, dass Datenverkehr zwischen dem Cluster und dem Dateisystem zulässig ist, indem Sie einen der folgenden Schritte ausführen:

- Konfigurieren Sie die Sicherheitsgruppen des Datei-Cache so, dass der Datenverkehr zum und vom CIDR oder der Präfixliste der Cluster-Subnetze zugelassen wird.

Note

AWS ParallelCluster überprüft, ob die Ports geöffnet sind und ob die CIDR- oder Präfixliste konfiguriert ist. AWS ParallelCluster validiert den Inhalt der CIDR-Block- oder Präfixliste nicht.

- Legen Sie mithilfe von [SecurityGroups](#) und [SlurmQueuesHeadNode/Networking/](#) benutzerdefinierte Sicherheitsgruppen für Clusterknoten fest. [NetworkingSecurityGroups](#) Die benutzerdefinierten Sicherheitsgruppen müssen so konfiguriert werden, dass sie den Datenverkehr zwischen dem Cluster und dem Dateisystem zulassen.

Note

Wenn alle Clusterknoten benutzerdefinierte Sicherheitsgruppen verwenden, wird AWS ParallelCluster nur überprüft, ob die Ports geöffnet sind. AWS ParallelCluster überprüft nicht, ob Quelle und Ziel richtig konfiguriert sind.

Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.

Iam Abschnitt

(Optional) Gibt IAM-Eigenschaften für den Cluster an.

Iam:

Roles:

LambdaFunctionsRole: *string*

PermissionsBoundary: *string*

ResourcePrefix: *string*

Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.

Iam-Eigenschaften

PermissionsBoundary(Fakultativ,String)

Der ARN der IAM-Richtlinie, der als Berechtigungsgrenze für alle Rollen verwendet werden soll, die von AWS ParallelCluster erstellt wurden. Weitere Informationen finden Sie unter [Berechtigungsgrenzen für IAM-Entitäten](#) im IAM-Benutzerhandbuch. Das Format ist `arn:${Partition}:iam:${Account}:policy/${PolicyName}`.

Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.

Roles(Fakultativ)

Gibt Einstellungen für die vom Cluster verwendeten IAM-Rollen an.

Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.

LambdaFunctionsRole(Optional,String)

Der ARN der IAM-Rolle, für AWS Lambda die verwendet werden soll. Dies überschreibt die Standardrolle, die allen Lambda-Funktionen zugewiesen ist, die AWS CloudFormation

benutzerdefinierte Ressourcen unterstützen. Lambda muss als Principal konfiguriert werden, der die Rolle übernehmen darf. Dadurch wird die Rolle der verwendeten Lambda-Funktionen nicht außer Kraft gesetzt. AWS Batch Das Format ist `arn:${Partition}:iam::${Account}:role/${RoleName}`.

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

ResourcePrefix(Fakultativ)

Gibt einen Pfad oder ein Namenspräfix für IAM-Ressourcen an, die von AWS ParallelCluster erstellt wurden.

Das Ressourcenpräfix muss den [von IAM angegebenen Benennungsregeln](#) entsprechen:

- Ein Name kann bis zu 30 Zeichen enthalten.
- Ein Name kann nur eine Zeichenfolge ohne Schrägstrich (/) sein.
- Ein Pfad kann bis zu 512 Zeichen lang sein.
- Ein Pfad muss mit einem Schrägstrich (/) beginnen und enden. Er kann mehrere Schrägstriche (/) zwischen dem Start- und dem Endschrägstrich (/) enthalten. /
- Sie können den Pfad und den Namen kombinieren. /path/name

Geben Sie einen Namen an.

```
Iam:  
  ResourcePrefix: my-prefix
```

Geben Sie einen Pfad an.

```
Iam:  
  ResourcePrefix: /org/dept/team/project/user/
```

Geben Sie einen Pfad und einen Namen an.

```
Iam:  
  ResourcePrefix: /org/dept/team/project/user/my-prefix
```

Wenn Sie angeben/my-prefix, wird ein Fehler zurückgegeben.

```
Iam:
```

```
ResourcePrefix: /my-prefix
```

Es wird ein Konfigurationsfehler zurückgegeben. Ein Pfad muss zwei / s haben. Ein Präfix allein kann nicht / s haben.

Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.

LoginNodes Abschnitt

Note

Support für LoginNodes wurde in AWS ParallelCluster Version 3.7.0 hinzugefügt.

(Optional) Gibt die Konfiguration für den Pool der Anmeldeknoten an.

LoginNodes:

Pools:

- Name: *string*
- Count: *integer*
- InstanceType: *string*
- GracetimePeriod: *integer*
- Image:
 - CustomAmi: *string*
- Ssh:
 - KeyName: *string*
- Networking:
 - SubnetIds:
 - *string*
 - SecurityGroups:
 - *string*
 - AdditionalSecurityGroups:
 - *string*
- Iam:
 - InstanceRole: *string*
 - InstanceProfile: *string*
 - AdditionalIamPolicies:
 - Policy: *string*

Aktualisierungsrichtlinie: Die Rechenflotte muss gestoppt werden, damit diese Einstellung für ein Update geändert werden kann.

LoginNodes-Eigenschaften

Pools-Eigenschaften

Definiert Gruppen von Anmeldeknoten mit derselben Ressourcenkonfiguration. Es kann nur ein einziger Pool angegeben werden.

Pools:

- Name: *string*
- Count: *integer*
- InstanceType: *string*
- GracetimePeriod: *integer*
- Image:
 - CustomAmi: *string*
- Ssh:
 - KeyName: *string*
- Networking:
 - SubnetIds:
 - *string*
 - SecurityGroups:
 - *string*
 - AdditionalSecurityGroups:
 - *string*
- Iam:
 - InstanceRole: *string*
 - InstanceProfile: *string*
 - AdditionalIamPolicies:
 - Policy: *string*

Name(ErforderlichString)

Gibt den Namen des LoginNodes Pools an. Dies wird verwendet, um die LoginNodes Ressourcen zu kennzeichnen.

Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.

Count(ErforderlichInteger)

Gibt die Anzahl der Anmeldeknoten an, die aktiv bleiben sollen.

Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.

InstanceType(ErforderlichString)

Gibt den Amazon EC2 EC2-Instance-Typ an, der für den Login-Knoten verwendet wird. Die Architektur des Instance-Typs muss mit der Architektur übereinstimmen, die für die Slurm InstanceType Einstellung verwendet wurde.

[Richtlinie aktualisieren](#): Diese Einstellung kann geändert werden, wenn der Pool der Anmeldeknoten gestoppt wird.

GracetimePeriod(FakultativInteger)

Gibt die Mindestzeit in Minuten an, die zwischen der Benachrichtigung an den angemeldeten Benutzer, dass ein Anmeldeknoten außer Betrieb genommen werden soll, und dem eigentlichen Stopp-Ereignis vergeht. Gültige Werte für GracetimePeriod liegen zwischen 3 und 120 Minuten. Der -Standardwert beträgt 60 Minuten.

Note

Das auslösende Ereignis beinhaltet Interaktionen zwischen mehreren AWS Diensten. Manchmal können die Netzwerklatenz und die Übertragung der Informationen einige Zeit in Anspruch nehmen, sodass die Übergangszeit aufgrund interner Verzögerungen bei den AWS Diensten länger als erwartet dauern kann.

[Richtlinie aktualisieren](#): Diese Einstellung kann während eines Updates geändert werden.

Image(Fakultativ)

Definiert die Image-Konfiguration für die Anmeldeknoten.

```
Image:  
CustomAmi: String
```

CustomAmi(FakultativString)

Gibt das benutzerdefinierte AML an, das für die Bereitstellung der Anmeldeknoten verwendet wird. Wenn nicht angegeben, wird standardmäßig der in der [HeadNode Abschnitt](#) angegebene Wert verwendet.

[Aktualisierungsrichtlinie](#): Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.

Ssh(Fakultativ)

Definiert die ssh Konfiguration für die Anmeldeknoten.

```
Ssh:  
  KeyName: string
```

KeyName(FakultativString)

Gibt den ssh Schlüssel an, der für die Anmeldung bei den Anmeldeknoten verwendet wird. Wenn nicht angegeben, wird standardmäßig der Wert verwendet, der in der [HeadNode Abschnitt](#) angegeben ist.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

Networking(Erforderlich)

```
Networking:  
  SubnetIds:  
    - string  
  SecurityGroups:  
    - string  
  AdditionalSecurityGroups:  
    - string
```

SubnetIds(Erforderlich[String])

Die ID des vorhandenen Subnetzes, in dem Sie den Pool der Anmeldeknoten bereitstellen. Sie können nur ein Subnetz definieren.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

SecurityGroups(Fakultativ) [String]

Eine Liste von Sicherheitsgruppen, die für den Pool der Anmeldeknoten verwendet werden sollen. Wenn keine Sicherheitsgruppen angegeben sind, AWS ParallelCluster erstellt es Sicherheitsgruppen für Sie.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

AdditionalSecurityGroups(Fakultativ[String])

Eine Liste zusätzlicher Sicherheitsgruppen, die für den Pool der Anmeldeknoten verwendet werden sollen.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

Iam(Fakultativ)

Gibt entweder eine Instanzrolle oder ein Instanzprofil an, das auf den Anmeldeknoten verwendet werden soll, um die Standard-Instanzrolle oder das Instanzprofil für den Cluster zu überschreiben.

```
Iam:  
  InstanceRole: string  
  InstanceProfile: string  
  AdditionalIamPolicies:  
    - Policy: string
```

InstanceProfile(OptionalString)

Gibt ein Instanzprofil an, um das standardmäßige Instanzprofil für den Login-Knoten zu überschreiben. Sie können nicht sowohl InstanceProfile als auch InstanceRole angeben. Das Format ist `arn:Partition:iam::Account:instance-profile/InstanceProfileName`. Wenn dies angegeben ist, können die AdditionalIamPolicies Einstellungen InstanceRole und nicht angegeben werden.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

InstanceRole(FakultativString)

Gibt eine Instanzrolle an, um die standardmäßige Instanzrolle für den Anmeldeknoten zu überschreiben. Sie können nicht sowohl InstanceProfile als auch InstanceRole angeben. Das Format ist `arn:Partition:iam::Account:role/RoleName`. Wenn dies angegeben ist, können die AdditionalIamPolicies Einstellungen S3Access und nicht angegeben werden. Wenn dies angegeben ist, können die AdditionalIamPolicies Einstellungen InstanceProfile und nicht angegeben werden.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

AdditionalIamPolicies(Fakultativ)

```
AdditionalIamPolicies:  
  - Policy: string
```

Eine IAM-Richtlinie Amazon Resource Name (ARN).

Gibt eine Liste von Amazon Resource Names (ARNs) von IAM-Richtlinien für Amazon EC2 an. Diese Liste ist der Root-Rolle angehängt, die für den Anmeldeknoten verwendet wird, zusätzlich zu den Berechtigungen, die für erforderlich sind. AWS ParallelCluster

Ein IAM-Richtliniename und sein ARN sind unterschiedlich. Namen können nicht verwendet werden.

Wenn dies angegeben ist, können die InstanceRole Einstellungen InstanceProfile und nicht angegeben werden. Wir empfehlen Ihnen, diese zu verwenden, AdditionalIamPolicies weil sie zu den erforderlichen Berechtigungen hinzugefügt AdditionalIamPolicies werden und alle erforderlichen Berechtigungen enthalten InstanceRole müssen. AWS ParallelCluster Die erforderlichen Berechtigungen ändern sich häufig von Version zu Version, wenn Funktionen hinzugefügt werden.

Es gibt keinen Standardwert.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

Policy(Erforderlich[String])

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

Monitoring Abschnitt

(Optional) Gibt die Überwachungseinstellungen für den Cluster an.

Monitoring:

Logs:

CloudWatch:

Enabled: *boolean*

RetentionInDays: *integer*

DeletionPolicy: *string*

Rotation:

Enabled: *boolean*

Dashboards:

CloudWatch:

Enabled: *boolean*

DetailedMonitoring: *boolean*

Alarms:

Enabled: *boolean*

Aktualisierungsrichtlinie: Diese Einstellung wird während eines Updates nicht analysiert.

Monitoring-Eigenschaften

Logs(Fakultativ)

Die Protokolleinstellungen für den Cluster.

Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.

CloudWatch(Fakultativ)

Die CloudWatch Logs-Einstellungen für den Cluster.

Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.

Enabled(Erforderlich,Boolean)

Falls `true`, werden Clusterprotokolle in CloudWatch Logs gestreamt. Der Standardwert ist `true`.

Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.

RetentionInDays(Fakultativ,Integer)

Die Anzahl der Tage, für die die Protokollereignisse in den CloudWatch Protokollen aufbewahrt werden sollen. Der Standardwert ist 180. Die unterstützten Werte sind 0, 1, 3, 5, 7, 14, 30, 60, 90, 120, 150, 180, 365, 400, 545, 731, 1827 und 3653. Bei einem Wert von 0 wird die Standardeinstellung für die Aufbewahrung von CloudWatch Protokollen verwendet, d. h. sie läuft nie ab.

Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.

DeletionPolicy(Optional,String)

Gibt an, ob Protokollereignisse in CloudWatch Logs gelöscht werden sollen, wenn der Cluster gelöscht wird. Die möglichen Werte sind `Delete` und `Retain`. Der Standardwert ist `Retain`.

Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.

Rotation(Fakultativ)

Die Einstellungen für die Protokollrotation für den Cluster.

Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.

Enabled(Erforderlich,Boolean)

Fallst`true`, ist die Protokollrotation aktiviert. Der Standardwert ist `true`. Wenn eine AWS ParallelCluster konfigurierte Protokolldatei eine bestimmte Größe erreicht, wird sie rotiert und es wird ein einzelnes Backup gespeichert. Weitere Informationen finden Sie unter [AWS ParallelClusterkonfigurierte Protokollrotation](#).

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

Dashboards(Fakultativ)

Die Dashboard-Einstellungen für den Cluster.

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

CloudWatch(Fakultativ)

Die CloudWatch Dashboard-Einstellungen für den Cluster.

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

Enabled(Erforderlich,Boolean)

Fallst`true`, ist das CloudWatch Dashboard aktiviert. Der Standardwert ist `true`.

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

DetailedMonitoring(Fakultativ,Boolean)

Wenn auf `gesetztttrue`, ist die detaillierte Überwachung für die Compute Fleet EC2-Instances aktiviert. Wenn diese Option aktiviert ist, zeigt die Amazon EC2 EC2-Konsole Diagramme zur Überwachung der Instances in Intervallen von 1 Minute an. Wenn diese Funktion aktiviert ist, fallen zusätzliche Kosten an. Der Standardwert ist `false`.

Weitere Informationen finden Sie unter [Enable or turn off detailed monitoring for your instances](#) im Amazon EC2 User Guide for Linux Instances.

[Aktualisierungsrichtlinie: Die Rechenflotte muss gestoppt werden, damit diese Einstellung für ein Update geändert werden kann.](#)

Note

DetailedMonitoring wird ab AWS ParallelCluster Version 3.6.0 hinzugefügt.

Alarms(Fakultativ)

CloudWatch Alarme für den Cluster.

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

Enabled(Fakultativ)

Falls `true`, werden die CloudWatch Alarme für den Cluster erstellt. Der Standardwert ist `true`.

[Richtlinie aktualisieren: Diese Einstellung kann während eines Updates geändert werden.](#)

Note

Ab AWS ParallelCluster Version 3.8.0 werden die folgenden Alarme für den Head Node erstellt: EC2 Health Check, CPU-/Speicher-/Festplattenauslastung und ein zusammengesetzter Alarm einschließlich aller anderen.

Tags Abschnitt

(Optional), Array Definiert die Tags, die von allen Clusterressourcen verwendet AWS CloudFormation und an diese weitergegeben werden. Weitere Informationen finden Sie unter [dem AWS CloudFormation Ressourcen-Tag](#) im AWS CloudFormation Benutzerhandbuch.

Tags:

- Key: *string*
- Value: *string*

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

Tags-Eigenschaften

Key(Erforderlich,String)

Definiert den Namen des Tags.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

Value(Erforderlich,String)

Definiert den Wert des Tags.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

AdditionalPackages Abschnitt

(Optional) Wird verwendet, um zusätzliche zu installierende Pakete zu identifizieren.

[AdditionalPackages:](#)

[IntelSoftware:](#)

[IntelHpcPlatform:](#) *boolean*

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

IntelSoftware

(Optional) Definiert die Konfiguration für Intel Select-Lösungen.

[IntelSoftware:](#)

[IntelHpcPlatform:](#) *boolean*

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

IntelSoftware-Eigenschaften

IntelHpcPlatform(Optional,Boolean)

Wenn `true`, bedeutet dies, dass die [Endbenutzer-Lizenzvereinbarung](#) für Intel Parallel Studio akzeptiert wurde. Dadurch wird Intel Parallel Studio auf dem Hauptknoten installiert und von den Rechenknoten gemeinsam genutzt. Dadurch verlängert sich die Zeit, die der Hauptknoten für das Bootstrap benötigt, um mehrere Minuten. Die `IntelHpcPlatform` Einstellung wird nur auf Version CentOS 7 unterstützt.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

DirectoryService Abschnitt

Note

Support für `DirectoryService` wurde in AWS ParallelCluster Version 3.1.1 hinzugefügt.

(Optional) Die Verzeichnisdiensteinstellungen für einen Cluster, der den Zugriff mehrerer Benutzer unterstützt.

AWS ParallelCluster verwaltet Berechtigungen, die den Zugriff mehrerer Benutzer auf Cluster mit einem Active Directory (AD) über das Lightweight Directory Access Protocol (LDAP) unterstützen, das vom [System Security Services Daemon \(SSSD\)](#) unterstützt wird. Weitere Informationen finden Sie unter [Was ist AWS Directory Service](#) im AWS Directory Service -Administratorhandbuch.

Wir empfehlen, LDAP über TLS/SSL (kurz LDAPS abgekürzt) zu verwenden, um sicherzustellen, dass potenziell vertrauliche Informationen über verschlüsselte Kanäle übertragen werden.

DirectoryService:

```
DomainName: string
DomainAddr: string
PasswordSecretArn: string
DomainReadOnlyUser: string
LdapTlsCaCert: string
LdapTlsReqCert: string
LdapAccessFilter: string
GenerateSshKeysForUsers: boolean
AdditionalSssdConfigs: dict
```

Aktualisierungsrichtlinie: Die Rechenflotte muss gestoppt werden, damit diese Einstellung für ein Update geändert werden kann.

DirectoryService-Eigenschaften

Note

Wenn Sie die Nutzung AWS ParallelCluster in einem einzelnen Subnetz ohne Internetzugang planen, finden Sie weitere Anforderungen unter [AWS ParallelCluster in einem einzelnen Subnetz ohne Internetzugang ohne Internetzugang ohne Internetzugang ohne Internetzugang](#).

DomainName(Erforderlich,String)

Die Active Directory-Domäne (AD), die Sie für Identitätsinformationen verwenden.

DomainNameakzeptiert sowohl die Formate Fully Qualified Domain Name (FQDN) als auch LDAP Distinguished Name (DN).

- Beispiel für einen FQDN: corp.*example*.com
- Beispiel für einen LDAP-DN: DC=*corp*,DC=*example*,DC=*com*

Diese Eigenschaft entspricht dem aufgerufenen sssd-ldap-Parameter. `ldap_search_base`

[Aktualisierungsrichtlinie: Die Rechenflotte muss gestoppt werden, damit diese Einstellung für ein Update geändert werden kann.](#)

`DomainAddr`(Erforderlich,) `String`

Die URI oder URIs, die auf den AD-Domänencontroller verweisen, der als LDAP-Server verwendet wird. Die URI entspricht dem aufgerufenen SSSD-LDAP-Parameter. `ldap_uri` Der Wert kann eine durch Kommas getrennte Zeichenfolge von URIs sein. Um LDAP zu verwenden, müssen Sie am Anfang `ldap://` jeder URI etwas hinzufügen.

Beispielwerte:

```
ldap://192.0.2.0,ldap://203.0.113.0      # LDAP
ldaps://192.0.2.0,ldaps://203.0.113.0  # LDAPS without support for certificate
verification
ldaps://abcdef01234567890.corp.example.com # LDAPS with support for certificate
verification
192.0.2.0,203.0.113.0                  # AWS ParallelCluster uses LDAPS by
default
```

Wenn Sie LDAPS mit Zertifikatsverifizierung verwenden, müssen die URIs Hostnamen sein.

Wenn Sie LDAPS ohne Zertifikatsverifizierung oder LDAP verwenden, können URIs Hostnamen oder IP-Adressen sein.

Verwenden Sie LDAP über TLS/SSL (LDAPS), um die Übertragung von Passwörtern und anderen vertraulichen Informationen über unverschlüsselte Kanäle zu vermeiden. Wenn AWS ParallelCluster kein Protokoll gefunden wird, wird es am Anfang jeder URI oder `ldaps://` jedes Hostnamens hinzugefügt.

[Aktualisierungsrichtlinie: Die Rechenflotte muss gestoppt werden, damit diese Einstellung für ein Update geändert werden kann.](#)

`PasswordSecretArn`(Erforderlich,`String`)

Der Amazon-Ressourcenname (ARN) des AWS Secrets Manager Geheimnisses, das das `DomainReadOnlyUser` Klartext-Passwort enthält. Der Inhalt des Geheimnisses entspricht dem aufgerufenen SSSD-LDAP-Parameter. `ldap_default_authtok`

Note

Achten Sie beim Erstellen eines Geheimnisses mit der AWS Secrets Manager Konsole darauf, „Andere Art von Geheimnis“ und Klartext auszuwählen und nur den Passworttext in das Geheimnis aufzunehmen.

Weitere Informationen zur Erstellung eines Geheimnisses finden Sie AWS Secrets Manager unter [Create an AWS Secrets Manager Secret](#)

Der LDAP-Client verwendet das Passwort, um sich bei der AD-Domäne zu authentifizieren, `DomainReadOnlyUser` wenn er Identitätsinformationen anfordert.

Ob der Benutzer dazu berechtigt `PasswordSecretArn` ist [DescribeSecret](#), wird überprüft. `PasswordSecretArn` ist gültig, wenn das angegebene Geheimnis existiert. Wenn die Benutzer-IAM-Richtlinie dies nicht beinhaltet `DescribeSecret`, `PasswordSecretArn` nicht validiert wird und eine Warnmeldung angezeigt wird. Weitere Informationen finden Sie unter [AWS ParallelCluster pclusterGrundlegende Benutzerrichtlinie](#).

Wenn sich der Wert des Geheimnisses ändert, wird der Cluster nicht automatisch aktualisiert. Um den Cluster für den neuen geheimen Wert zu aktualisieren, müssen Sie die Rechenflotte mit dem [the section called “pcluster update-compute-fleet”](#) Befehl beenden und dann den folgenden Befehl vom Hauptknoten aus ausführen.

```
$ sudo /opt/parallelcluster/scripts/directory_service/  
update_directory_service_password.sh
```

[Aktualisierungsrichtlinie: Die Rechenflotte muss gestoppt werden, damit diese Einstellung für ein Update geändert werden kann.](#)

`DomainReadOnlyUser(Erforderlich,String)`

Die Identität, die verwendet wird, um die AD-Domäne bei der Authentifizierung von Cluster-Benutzeranmeldungen nach Identitätsinformationen abzufragen. Sie entspricht dem aufgerufenen SSSD-LDAP-Parameter. `ldap_default_bind_dn` Verwenden Sie Ihre AD-Identitätsinformationen für diesen Wert.

Geben Sie die Identität in der Form an, die für den spezifischen LDAP-Client erforderlich ist, der sich auf dem Knoten befindet:

- Microsoft AD:

```
cn=ReadOnlyUser,ou=Users,ou=CORP,dc=corp,dc=example,dc=com
```

- Einfache Anzeige:

```
cn=ReadOnlyUser,cn=Users,dc=corp,dc=example,dc=com
```

[Aktualisierungsrichtlinie: Die Rechenflotte muss gestoppt werden, damit diese Einstellung für ein Update geändert werden kann.](#)

LdapTlsCaCert(Fakultativ,String)

Der absolute Pfad zu einem Zertifikatspaket, das die Zertifikate für jede Zertifizierungsstelle in der Zertifizierungskette enthält, die ein Zertifikat für die Domänencontroller ausgestellt hat. Er entspricht dem aufgerufenen SSSD-LDAP-Parameter. `ldap_tls_cacert`

Ein Zertifikatspaket ist eine Datei, die aus der Verkettung verschiedener Zertifikate im PEM-Format besteht, das in Windows auch als DER Base64-Format bezeichnet wird. Es wird verwendet, um die Identität des AD-Domänencontrollers zu überprüfen, der als LDAP-Server fungiert.

AWS ParallelCluster ist nicht verantwortlich für die erstmalige Platzierung von Zertifikaten auf Knoten. Als Clusteradministrator können Sie das Zertifikat im Hauptknoten manuell konfigurieren, nachdem der Cluster erstellt wurde, oder Sie können ein [Bootstrap-Skript](#) verwenden. Alternativ können Sie ein Amazon Machine Image (AMI) verwenden, das das auf dem Hauptknoten konfigurierte Zertifikat enthält.

[Simple AD](#) bietet keine LDAPS-Unterstützung. Informationen zur Integration eines Simple AD-Verzeichnisses finden Sie unter [So konfigurieren Sie einen LDAPS-Endpunkt für Simple AD](#) im AWS Sicherheitsblog. AWS ParallelCluster

[Aktualisierungsrichtlinie: Die Rechenflotte muss gestoppt werden, damit diese Einstellung für ein Update geändert werden kann.](#)

LdapTlsReqCert(Optional,String)

Gibt an, welche Prüfungen an Serverzertifikaten in einer TLS-Sitzung durchgeführt werden sollen. Es entspricht dem aufgerufenen SSSD-LDAP-Parameter. `ldap_tls_reqcert`

Gültige Werte: `never`, `allow`, `try`, `demand` und `hard`.

never, und try ermöglichen allow, dass Verbindungen auch dann fortgeführt werden können, wenn Probleme mit Zertifikaten gefunden werden.

demand und hard ermöglichen die Fortsetzung der Kommunikation, falls keine Probleme mit Zertifikaten gefunden werden.

Wenn der Clusteradministrator einen Wert verwendet, für den die erfolgreiche Überprüfung des Zertifikats nicht erforderlich ist, wird eine Warnmeldung an den Administrator zurückgegeben. Aus Sicherheitsgründen empfehlen wir, die Zertifikatsüberprüfung nicht zu deaktivieren.

Der Standardwert ist hard.

[Aktualisierungsrichtlinie: Die Rechenflotte muss gestoppt werden, damit diese Einstellung für ein Update geändert werden kann.](#)

LdapAccessFilter(Fakultativ,String)

Gibt einen Filter an, um den Verzeichniszugriff auf eine Untergruppe von Benutzern zu beschränken. Diese Eigenschaft entspricht dem aufgerufenen SSSD-LDAP-Parameter. `ldap_access_filter` Sie können damit Abfragen auf ein AD beschränken, das eine große Anzahl von Benutzern unterstützt.

Dieser Filter kann den Benutzerzugriff auf den Cluster blockieren. Er hat jedoch keinen Einfluss auf die Auffindbarkeit blockierter Benutzer.

Wenn diese Eigenschaft festgelegt ist, wird der SSSD-Parameter `access_provider` auf `ldap` intern von `aws` gesetzt und darf nicht durch die Einstellungen von [DirectoryService/AdditionalSssdConfigs](#) geändert werden.

Wenn diese Eigenschaft weggelassen wird und kein benutzerdefinierter Benutzerzugriff in [DirectoryService](#) angegeben ist [AdditionalSssdConfigs](#), können alle Benutzer im Verzeichnis auf den Cluster zugreifen.

Beispiele:

```
!"(cn=SomeUser*)" # denies access to every user with alias starting with "SomeUser"  
"(cn=SomeUser*)" # allows access to every user with alias starting with "SomeUser"  
"memberOf=cn=TeamOne,ou=Users,ou=CORP,dc=corp,dc=example,dc=com" # allows access  
only to users in group "TeamOne".
```

[Aktualisierungsrichtlinie: Die Rechenflotte muss gestoppt werden, damit diese Einstellung für ein Update geändert werden kann.](#)

GenerateSshKeysForUsers(Optional,Boolean)

Definiert, ob unmittelbar nach ihrer ersten Authentifizierung auf dem Hauptknoten ein SSH-Schlüssel für Clusterbenutzer AWS ParallelCluster generiert wird.

Wenn auf `gesetztrue`, wird für jeden Benutzer nach seiner ersten Authentifizierung auf dem Hauptknoten ein SSH-Schlüssel generiert und gespeichert, falls er nicht existiert.

`USER_HOME_DIRECTORY/.ssh/id_rsa`

Für einen Benutzer, der noch nicht auf dem Hauptknoten authentifiziert wurde, kann die erste Authentifizierung in den folgenden Fällen erfolgen:

- Der Benutzer meldet sich zum ersten Mal mit seinem eigenen Passwort am Hauptknoten an.
- Im Hauptknoten wechselt ein Sudoer zum ersten Mal zum Benutzer: `su USERNAME`
- Im Hauptknoten führt ein Sudoer zum ersten Mal einen Befehl als Benutzer aus: `su -u USERNAME COMMAND`

Benutzer können den SSH-Schlüssel für nachfolgende Anmeldungen am Cluster-Kopfknoten und den Rechenknoten verwenden. Mit AWS ParallelCluster sind Kennwortanmeldungen an Cluster-Rechenknoten standardmäßig deaktiviert. Wenn sich ein Benutzer nicht am Hauptknoten angemeldet hat, werden keine SSH-Schlüssel generiert und der Benutzer kann sich nicht bei Rechenknoten anmelden.

Der Standardwert ist `true`.

[Aktualisierungsrichtlinie: Die Rechenflotte muss gestoppt werden, damit diese Einstellung für ein Update geändert werden kann.](#)

AdditionalSsdConfigs(Optional,Dict)

Ein Diktat von Schlüssel-Wert-Paaren, die SSSD-Parameter und Werte enthalten, die in die SSSD-Konfigurationsdatei auf Cluster-Instances geschrieben werden sollen. Eine vollständige Beschreibung der SSSD-Konfigurationsdatei finden Sie in den Manpages auf der Instanz und den zugehörigen Konfigurationsdateien. SSSD

Die SSSD-Parameter und -Werte müssen mit AWS ParallelCluster der SSSD-Konfiguration kompatibel sein, wie in der folgenden Liste beschrieben.

- `id_provider` ist auf `ldap intern` von gesetzt AWS ParallelCluster und darf nicht geändert werden.

- `access_provider` ist auf `ldap` intern gesetzt, AWS ParallelCluster wenn [DirectoryService](#)/angegeben `LdapAccessFilter` wird, und diese Einstellung darf nicht geändert werden.

Wenn [DirectoryService](#)/weggelassen `LdapAccessFilter` wird, wird auch seine `access_provider` Angabe weggelassen. Wenn Sie beispielsweise `access_provider` auf `simple` in setzen [AdditionalSssdConfigs](#), `LdapAccessFilter` darf [DirectoryService](#)/nicht angegeben werden.

Die folgenden Konfigurationsschnipsel sind Beispiele für gültige Konfigurationen für `AdditionalSssdConfigs`

In diesem Beispiel wird die Debug-Ebene für SSSD-Protokolle aktiviert, die Suchbasis auf eine bestimmte Organisationseinheit beschränkt und das Zwischenspeichern von Anmeldeinformationen deaktiviert.

```
DirectoryService:
  ...
  AdditionalSssdConfigs:
    debug_level: "0xFFF0"
    ldap_search_base: OU=Users,OU=CORP,DC=corp,DC=example,DC=com
    cache_credentials: False
```

Dieses Beispiel spezifiziert die Konfiguration einer SSSD. `simple` `access_provider` Benutzern aus dem `EngineeringTeam` wird Zugriff auf das Verzeichnis gewährt. [DirectoryService/LdapAccessFilter](#) darf in diesem Fall nicht gesetzt werden.

```
DirectoryService:
  ...
  AdditionalSssdConfigs:
    access_provider: simple
    simple_allow_groups: EngineeringTeam
```

[Aktualisierungsrichtlinie: Die Rechenflotte muss gestoppt werden, damit diese Einstellung für ein Update geändert werden kann.](#)

DeploymentSettings Abschnitt

Note

DeploymentSettings wird ab AWS ParallelCluster Version 3.4.0 hinzugefügt.

(Optional) Gibt die Konfiguration der Bereitstellungseinstellungen an.

```
DeploymentSettings:  
  LambdaFunctionsVpcConfig:  
    SecurityGroupIds  
      - string  
    SubnetIds  
      - string  
  DisableSudoAccessForDefaultUser: Boolean  
  DefaultUserHome: string # 'Shared' or 'Local'
```

DeploymentSettings-Eigenschaften

LambdaFunctionsVpcConfig

(Optional) Gibt die VPC-Konfigurationen der AWS Lambda Funktionen an. Weitere Informationen finden Sie unter [AWS Lambda VPC-Konfiguration in AWS ParallelCluster](#).

```
LambdaFunctionsVpcConfig:  
  SecurityGroupIds  
    - string  
  SubnetIds  
    - string
```

LambdaFunctionsVpcConfig properties

SecurityGroupIds(Erforderlich, [String])

Die Liste der Amazon VPC-Sicherheitsgruppen-IDs, die an die Lambda-Funktionen angehängt sind.

Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.

SubnetIds(Erforderlich,) [String]

Die Liste der Subnetz-IDs, die an die Lambda-Funktionen angehängt sind.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

Note

Die Subnetze und Sicherheitsgruppen müssen sich in derselben VPC befinden.

DisableSudoAccessForDefaultUser Eigentum

Note

Diese Konfigurationsoption wird nur mit Slurm-Clustern unterstützt.

(Optional) Wenn `True`, werden die Sudo-Rechte des Standardbenutzers deaktiviert. Dies gilt für alle Knoten im Cluster.

```
# Main DeploymentSettings section in config yaml(applyes to HN, CF and LN)
DeploymentSettings:
  DisableSudoAccessForDefaultUser: True
```

Um den Wert von `DisableSudoAccessForDefaultUser` zu aktualisieren, müssen Sie die Rechenflotte und alle Anmeldeknoten stoppen.

[Aktualisierungsrichtlinie: Die Rechenflotte und die Anmeldeknoten müssen gestoppt werden, damit diese Einstellung für ein Update geändert werden kann.](#)

DefaultUserHome Eigenschaft

Wenn diese Option auf `Shared` gesetzt ist, verwendet der Cluster das Standard-Setup und teilt das Standardbenutzerverzeichnis im gesamten Cluster von `/home/<default user>`.

Wenn diese Option auf `Local` gesetzt ist, wird für den Hauptknoten, die Anmeldeknoten und die Rechenknoten jeweils ein eigenes lokales Standardbenutzerverzeichnis gespeichert `local/home/<default user>`.

Image-Konfigurationsdateien erstellen

AWS ParallelCluster Version 3 verwendet YAML 1.1-Dateien für Build-Image-Konfigurationsparameter. Bitte überprüfen Sie, ob der Einzug korrekt ist, um Konfigurationsfehler zu

vermeiden. Weitere Informationen finden Sie in der YAML 1.1-Spezifikation unter <https://yaml.org/spec/1.1/>

Diese Konfigurationsdateien werden verwendet, um zu definieren, wie Ihre benutzerdefinierten AWS ParallelCluster AMIs mit EC2 Image Builder erstellt werden. Benutzerdefinierte AMI-Erstellungsprozesse werden mit dem `pcluster build-image` Befehl ausgelöst. Einige Beispielkonfigurationsdateien finden Sie unter https://github.com/aws/aws-parallelcluster/tree/release-3.0/cli/tests/pcluster/schemas/test_imagebuilder_schema/test_imagebuilder_schema.

Themen

- [Eigenschaften der Image-Konfigurationsdatei erstellen](#)
- [Build Abschnitt](#)
- [Image Abschnitt](#)
- [DeploymentSettings Abschnitt](#)

Eigenschaften der Image-Konfigurationsdatei erstellen

Region(Fakultativ,String)

Gibt die AWS-Region für den `build-image` Vorgang an. z. B. `us-east-2`.

Build Abschnitt

(Erforderlich) Gibt die Konfiguration an, in der das Image erstellt wird.

```
Build:
  Imds:
    ImdsSupport: string
  InstanceType: string
  SubnetId: string
  ParentImage: string
  Iam:
    InstanceRole: string
    InstanceProfile: string
    CleanupLambdaRole: string
    AdditionalIamPolicies:
      - Policy: string
    PermissionsBoundary: string
```

Components:

- **Type:** *string*
- **Value:** *string*

Tags:

- **Key:** *string*
- **Value:** *string*

SecurityGroupIds:

- *string*

UpdateOsPackages:

- **Enabled:** *boolean*

Build-Eigenschaften

InstanceType(Erforderlich,String)

Gibt den Instanztyp für die Instanz an, die zum Erstellen des Images verwendet wurde.

SubnetId(Optional,String)

Gibt die ID eines vorhandenen Subnetzes an, in dem die Instanz bereitgestellt werden soll, um das Image zu erstellen. Das bereitgestellte Subnetz erfordert Internetzugang.

⚠ Warning

`pcluster build-image` verwendet die Standard-VPC. Wenn die Standard-VPC gelöscht wurde, möglicherweise mithilfe von AWS Control Tower oder AWS Landing Zone, muss die Subnetz-ID angegeben werden.

ParentImage(Erforderlich,) String

Gibt das Basisimage an. Das übergeordnete Image kann entweder ein AWS ParallelCluster Nicht-AMI oder ein offizielles AWS ParallelCluster AMI für dieselbe Version sein. Sie können kein AWS ParallelCluster offizielles oder benutzerdefiniertes AMI aus einer anderen Version von verwenden AWS ParallelCluster. Das Format muss entweder der ARN eines Images `arn:Partition:imagebuilder:Region:Account:image/ImageName/ImageVersion` oder eine AMI-ID sein `ami-12345678`.

SecurityGroupIds(Fakultativ,[String])

Gibt die Liste der Sicherheitsgruppen-IDs für das Bild an.

Imsds

Imsds-Eigenschaften

(Optional) Gibt die Einstellungen für den ImageBuilder EC2-Build- und Testinstanz-Metadatendienst (IMDS) an.

Imsds:

ImsdsSupport: *string*

ImsdsSupport(Optional,) String

Gibt an, welche IMDS-Versionen in den ImageBuilder EC2-Build- und Test-Instances unterstützt werden. Unterstützte Werte sind `v2.0` und `v1.0`. Der Standardwert ist `v2.0`.

Wenn auf gesetzt `ImsdsSupport` ist `v1.0`, werden sowohl IMDSv1 als auch IMDSv2 unterstützt.

Wenn auf gesetzt `ImsdsSupport` ist `v2.0`, wird nur IMDSv2 unterstützt.

Weitere Informationen finden Sie unter [Verwenden von IMDSv2](#) im EC2-Benutzerhandbuch für Linux-Instances.

Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.

Note

Ab AWS ParallelCluster Version 3.7.0 ist der `ImsdsSupport` Standardwert `v2.0`. Wir empfehlen, dass Sie in Ihren Aufrufen für benutzerdefinierte Aktionen `ImsdsSupport` auf `IMDSv1` setzen `v2.0` und dieses durch `IMDSv2` ersetzen.

Support für [Imsds/ImsdsSupport](#) wurde mit AWS ParallelCluster Version 3.3.0 hinzugefügt.

Iam

Iam-Eigenschaften

(Optional) Gibt die IAM-Ressourcen für den Image-Build an.

Iam:

InstanceRole: *string*

InstanceProfile: *string*

```
CleanupLambdaRole: string  
AdditionalIamPolicies:  
- Policy: string  
PermissionsBoundary: string
```

InstanceProfile(Optional,String)

Gibt ein Instanzprofil an, um das Standard-Instanzprofil für die EC2 Image Builder Builder-Instance zu überschreiben. InstanceProfile und InstanceRole und AdditionalIamPolicies können nicht zusammen angegeben werden. Das Format ist `arn:Partition:iam::Account:instance-profile/InstanceProfileName`.

InstanceRole(Fakultativ,String)

Gibt eine Instanzrolle an, um die Standard-Instanzrolle für die EC2 Image Builder Builder-Instance zu überschreiben. InstanceProfile und InstanceRole und AdditionalIamPolicies können nicht zusammen angegeben werden. Das Format ist `arn:Partition:iam::Account:role/RoleName`.

CleanupLambdaRole(Fakultativ,String)

Der ARN der IAM-Rolle, der für die AWS Lambda Funktion verwendet werden soll, die die AWS CloudFormation benutzerdefinierte Ressource unterstützt, die Build-Artefakte nach Abschluss des Builds entfernt. Lambda muss als Principal konfiguriert werden, der die Rolle übernehmen darf. Das Format ist `arn:Partition:iam::Account:role/RoleName`.

AdditionalIamPolicies(Fakultativ)

Gibt zusätzliche IAM-Richtlinien an, die an die EC2 Image Builder Builder-Instance angehängt werden sollen, die zur Erstellung des benutzerdefinierten AMI verwendet wurde.

```
AdditionalIamPolicies:  
- Policy: string
```

Policy(Optional,) [String]

Liste der IAM-Richtlinien. Das Format ist `arn:Partition:iam::Account:policy/PolicyName`.

PermissionsBoundary(Fakultativ,String)

Der ARN der IAM-Richtlinie, der als Berechtigungsgrenze für alle Rollen verwendet werden soll, die von AWS ParallelCluster erstellt wurden. Weitere Informationen zu den Grenzen

von IAM-Berechtigungen finden Sie unter [Berechtigungsgrenzen für IAM-Entitäten](#) im IAM-Benutzerhandbuch. Das Format ist `arn:Partition:iam::Account:policy/PolicyName`.

Components

Components-Eigenschaften

(Optional) Gibt ImageBuilder EC2-Komponenten an, die während des AMI-Build-Prozesses zusätzlich zu den standardmäßig AWS ParallelCluster von bereitgestellten verwendet werden sollen. Solche Komponenten können verwendet werden, um den AMI-Erstellungsprozess anzupassen. Weitere Informationen finden Sie unter [AWS ParallelCluster AMI-Anpassung](#).

Components:

- Type: *string*
- Value: *string*

Type(Fakultativ,String)

Gibt den Typ des Typ-Wert-Paares für die Komponente an. Der Typ kann `arn` oder `script` sein.

Value(Fakultativ,String)

Gibt den Wert des Typ-Wert-Paares für die Komponente an. Wenn `type` gleich `arn` ist, ist dies der ARN einer EC2 Image Builder Builder-Komponente. Wenn `type` gleich `script` ist, ist dies der `https`- oder `s3`-Link, der auf das Skript verweist, das bei der Erstellung der EC2 Image Builder Builder-Komponente verwendet werden soll.

Tags

Tags-Eigenschaften

(Optional) Gibt die Liste der Tags an, die in den Ressourcen festgelegt werden sollen, die zum Erstellen des AMI verwendet werden.

Tags:

- Key: *string*
- Value: *string*

Key(Optional,String)

Definiert den Namen des Tags.

Value(Fakultativ,String)

Definiert den Wert des Tags.

UpdateOsPackages

UpdateOsPackages-Eigenschaften

(Optional) Gibt an, ob das Betriebssystem vor der Installation des AWS ParallelCluster Software-Stacks aktualisiert wird.

UpdateOsPackages:

Enabled: *boolean*

Enabled(Optional,Boolean)

Falls `true`, wird das Betriebssystem aktualisiert und neu gestartet, bevor die AWS ParallelCluster Software installiert wird. Der Standardwert ist `false`.

Note

Wenn diese Option aktiviert `UpdateOsPackages` ist, werden alle verfügbaren Betriebssystempakete aktualisiert, einschließlich des Kernels. Als Kunde sind Sie dafür verantwortlich, zu überprüfen, ob das Update mit den AMI-Abhängigkeiten kompatibel ist, die nicht im Update enthalten sind.

Nehmen wir zum Beispiel an, Sie erstellen ein AMI für AWS ParallelCluster Version X.0, das mit Kernel-Version Y.0 und einigen Komponentenversionen Z.0 ausgeliefert wird. Angenommen, das verfügbare Update enthält die aktualisierte Kernelversion Y.1 ohne Updates für die Komponente Z.0. Es liegt in Ihrer Verantwortung `UpdateOsPackages`, vor der Aktivierung zu überprüfen, ob die Komponente Z.0 den Kernel Y.1 unterstützt.

Image Abschnitt

(Optional) Definiert die Image-Eigenschaften für den Image-Build.

Image:

Name: *string*


```
RootVolume:  
  Size: integer  
  Encrypted: boolean  
  KmsKeyId: string  
Tags:  
  - Key: string  
    Value: string
```

Image-Eigenschaften

Name(Optional,String)

Gibt den Namen des AMI an. Wenn nicht angegeben, wird der Name verwendet, der beim Aufrufen des [pcluster build-image](#) Befehls verwendet wurde.

Tags

Tags-Eigenschaften

(Optional) Gibt Schlüssel-Wert-Paare für das Bild an.

```
Tags:  
  - Key: string  
    Value: string
```

Key(Fakultativ,) String

Definiert den Namen des Tags.

Value(Fakultativ,String)

Definiert den Wert des Tags.

RootVolume

RootVolume-Eigenschaften

(Optional) Gibt die Eigenschaften des Root-Volumes für das Bild an.

```
RootVolume:
```

```
Size: integer  
Encrypted: boolean  
KmsKeyId: string
```

Size(Fakultativ,Integer)

Gibt die Größe des Root-Volumes für das Bild in GiB an. Die Standardgröße entspricht der Größe [ParentImage](#) plus 27 GiB.

Encrypted(Fakultativ,Boolean)

Gibt an, ob das Volume verschlüsselt ist. Der Standardwert ist `false`.

KmsKeyId(Fakultativ,String)

Gibt den ARN des AWS KMS Schlüssels an, der zum Verschlüsseln des Volumes verwendet wurde. Das Format ist "arn:*Partition*:kms:*Region*:*Account*:key/*KeyId*".

DeploymentSettings Abschnitt

(Optional) Gibt die Konfiguration der Bereitstellungseinstellungen an.

```
DeploymentSettings:  
  LambdaFunctionsVpcConfig:  
    SecurityGroupIds  
      - string  
    SubnetIds  
      - string
```

DeploymentSettings-Eigenschaften

LambdaFunctionsVpcConfig

(Optional) Gibt die VPC-Konfigurationen der AWS Lambda Funktionen an. Weitere Informationen finden Sie unter [AWS LambdaVPC-Konfiguration inAWS ParallelCluster](#).

```
LambdaFunctionsVpcConfig:  
  SecurityGroupIds  
    - string  
  SubnetIds  
    - string
```

LambdaFunctionsVpcConfig properties

SecurityGroupIds(Erforderlich,[String])

Die Liste der Amazon VPC-Sicherheitsgruppen-IDs, die an die Lambda-Funktionen angehängt sind.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

SubnetIds(Erforderlich,) [String]

Die Liste der Subnetz-IDs, die an die Lambda-Funktionen angehängt sind.

[Aktualisierungsrichtlinie: Wenn diese Einstellung geändert wird, ist das Update nicht zulässig.](#)

Note

Die Subnetze und Sicherheitsgruppen müssen sich in derselben VPC befinden.

Note

DeploymentSettings wird ab AWS ParallelCluster Version 3.4.0 hinzugefügt.

AWS ParallelCluster-API-Referenz

Dieser Abschnitt enthält Beschreibungen, Syntax und Anwendungsbeispiele für die einzelnen AWS ParallelCluster-API-Aktionen.

Themen

- [Bild erstellen](#)
- [createCluster](#)
- [Cluster löschen](#)
- [deleteClusterInstances](#)
- [Bild löschen](#)
- [Cluster beschreiben](#)
- [describeClusterInstances](#)

- [describeComputeFleet](#)
- [Beschreiben Sie das Bild](#)
- [getClusterLogEreignisse](#)
- [getClusterStackEreignisse](#)
- [getImageLogEreignisse](#)
- [getImageStackEreignisse](#)
- [Cluster auflisten](#)
- [listClusterLogStreams](#)
- [listImageLogStröme](#)
- [Bilder auflisten](#)
- [listOfficialImages](#)
- [Cluster aktualisieren](#)
- [updateComputeFleet](#)

Bild erstellen

Erstellen Sie ein benutzerdefiniertes AWS ParallelCluster Bild in einem AWS-Region.

Themen

- [Erforderliche Syntax](#)
- [Anforderungstext](#)
- [Antwortsyntax](#)
- [Antworttext](#)
- [Beispiel](#)

Erforderliche Syntax

```
POST /v3/images/custom
{
  "imageConfiguration": "string",
  "imageId": "string",
  "dryrun": boolean,
  "region": "string",
```

```
"rollbackOnFailure": boolean,  
"supressValidators": [ "string" ],  
"validationFailureLevel": "string"  
}
```

Anforderungstext

Image-Konfiguration

Die Image-Konfiguration als YAML-Dokument.

Typ: Zeichenfolge

Erforderlich: Ja

imageId

Die ID des zu erstellenden Images.

Typ: Zeichenfolge

Erforderlich: Ja

Trockenlauf

Wenn gesetzt auf `true` führt nur eine Anforderungsvalidierung durch, ohne eine Ressource zu erstellen. Verwenden Sie diesen Parameter, um die Image-Konfiguration zu validieren. Der Standardwert ist `false`.

Typ: boolescher Wert

Required: No

region

Der AWS-Region in dem Sie den Befehl ausführen, in dem das Image erstellt wird.

Typ: Zeichenkette

Required: No

rollbackOnFailure

Wenn gesetzt auf `true` erfolgt ein Rollback des Image-Stacks, wenn das Image nicht erstellt werden kann. Der Standardwert ist `false`.

Typ: boolescher Wert

Required: No

Unterdrücken Sie Validatoren

Identifizieren Sie einen oder mehrere Konfigurationsvalidatoren, die unterdrückt werden sollen.

Typ: Liste von Zeichenketten

Format: (ALL | type:[A-Za-z0-9]+)

Required: No

validationFailureLevel

Die Mindestvalidierungsstufe, die dazu führt, dass die Image-Erstellung fehlschlägt. Der Standardwert ist ERROR.

Typ: Zeichenfolge

Zulässige Werte: INFO | WARNING | ERROR

Required: No

Antwortsyntax

```
{
  "image": {
    "imageId": "string",
    "ec2AmiInfo": {
      "amiId": "string"
    },
    "region": "string",
    "version": "string",
    "cloudformationStackArn": "string",
    "imageBuildStatus": "BUILD_IN_PROGRESS",
    "cloudformationStackStatus": "CREATE_IN_PROGRESS"
  },
  "validationMessages": [
    {
      "id": "string",
      "type": "string",
```

```
    "level": "INFO",  
    "message": "string"  
  }  
]  
}
```

Antworttext

Image

imageId

Die ID des Bildes.

Typ: Zeichenfolge

cloudformationStackArn

Der Amazon-Ressourcenname (ARN) der Haupt-Ressource CloudFormation Stapel.

Typ: Zeichenfolge

cloudformationStackStatus

Der CloudFormation Status des Stapels.

Typ: Zeichenfolge

Zulässige Werte: CREATE_IN_PROGRESS | CREATE_FAILED | CREATE_COMPLETE
| ROLLBACK_IN_PROGRESS | ROLLBACK_FAILED | ROLLBACK_COMPLETE
| DELETE_IN_PROGRESS | DELETE_FAILED | DELETE_COMPLETE |
UPDATE_IN_PROGRESS | UPDATE_COMPLETE_CLEANUP_IN_PROGRESS
| UPDATE_COMPLETE | UPDATE_ROLLBACK_IN_PROGRESS |
UPDATE_ROLLBACK_FAILED | UPDATE_ROLLBACK_COMPLETE_CLEANUP_IN_PROGRESS
| UPDATE_ROLLBACK_COMPLETE

ec2AmiInfo

ami_id

Die EC2-AMI-ID.

Typ: Zeichenfolge

imageBuildStatus

Der Status der Image-Erstellung.

Typ: Zeichenfolge

Zulässige Werte: BUILD_IN_PROGRESS | BUILD_FAILED | BUILD_COMPLETE | DELETE_IN_PROGRESS | DELETE_FAILED | DELETE_COMPLETE

region

DieAWS-Region, in der das Bild eingebaut ist.

Typ: Zeichenfolge

Version

DieAWS ParallelClusterVersion, die zum Erstellen des Images verwendet wurde.

Typ: Zeichenfolge

Bestätigungsnachrichten

Eine Liste von Nachrichten mit einer Validierungsstufe unter `validationFailureLevel`. Die Liste der Nachrichten wird während der Konfigurationsvalidierung gesammelt.

ID

Die Validator-ID.

Typ: Zeichenfolge

level

Die Validierungsebene.

Typ: Zeichenfolge

Zulässige Werte: INFO | WARNING | ERROR

Nachricht

Eine Bestätigungsnachricht.

Typ: Zeichenfolge

type

Der Typ des Validators.

Typ: Zeichenfolge

Beispiel

Python

Anforderung

```
$ build_image(custom-image-id, custom-image-config.yaml)
```

200 Antwort

```
{
  'image': {
    'cloudformation_stack_arn': 'arn:aws:cloudformation:us-
east-1:123456789012:stack/custom-image-id/711b76b0-af81-11ec-a29f-0ee549109f1f',
    'cloudformation_stack_status': 'CREATE_IN_PROGRESS',
    'image_build_status': 'BUILD_IN_PROGRESS',
    'image_id': 'custom-image-id',
    'region': 'us-east-1',
    'version': '3.2.1'
  }
}
```

createCluster

Erstellen Sie einen verwalteten Cluster in einem AWS-Region.

Themen

- [Erforderliche Syntax](#)
- [Anforderungstext](#)
- [Antwortsyntax](#)
- [Antworttext](#)
- [Beispiel](#)

Erforderliche Syntax

```
POST /v3/clusters
{
  "clusterName": "string",
  "clusterConfiguration": "string",
  "dryrun": boolean,
  "region": "string",
  "rollbackOnFailure", boolean,
  "suppressValidators": [ "string" ],
  "validationFailureLevel": "string"
}
```

Anforderungstext

Cluster-Konfiguration

Die Cluster-Konfiguration als YAML-Dokument.

Typ: Zeichenfolge

Erforderlich: Ja

Clustername

Der Name des zu erstellenden Clusters.

Der Name muss mit einem alphabetischen Zeichen beginnen. Der Name kann bis zu 60 Zeichen lang sein. Wenn Slurm die Kontoführung ist aktiviert, der Name kann bis zu 40 Zeichen lang sein.

Typ: Zeichenfolge

Erforderlich: Ja

Trockenlauf

Wenn gesetzt auf `true` führt nur eine Anforderungsvalidierung durch, ohne eine Ressource zu erstellen. Verwenden Sie diesen Parameter, um die Clusterkonfiguration zu validieren. Der Standardwert ist `false`.

Typ: boolescher Wert

Required: No

region

Der AWS-Region in dem sich der Cluster befindet.

Typ: Zeichenkette

Required: No

rollbackOnFailure

Wenn gesetzt auf `true` erfolgt ein Cluster-Stack-Rollback, wenn der Cluster nicht erstellt werden kann. Der Standardwert ist `true`.

Typ: boolescher Wert

Required: No

Unterdrücken Sie Validatoren

Identifizieren Sie einen oder mehrere Konfigurationsvalidatoren, die unterdrückt werden sollen.

Typ: Liste von Zeichenketten

Format: (ALL | type : [A-Za-z0-9]+)

Required: No

validationFailureLevel

Die Mindestvalidierungsstufe, die dazu führt, dass die Clustererstellung fehlschlägt. Der Standardwert ist `ERROR`.

Typ: Zeichenfolge

Zulässige Werte: `INFO` | `WARNING` | `ERROR`

Required: No

Antwortsyntax

```
{
  "cluster": {
    "clusterName": "string",
```

```
"region": "string",
"version": "string",
"cloudformationStackArn": "string",
"cloudformationStackStatus": "CREATE_IN_PROGRESS",
"clusterStatus": "CREATE_IN_PROGRESS",
"scheduler": {
  "type": "string",
  "metadata": {
    "name": "string",
    "version": "string"
  }
},
"validationMessages": [
  {
    "id": "string",
    "type": "string",
    "level": "INFO",
    "message": "string"
  }
]
}
```

Antworttext

Clustername

Der Name des Clusters.

Typ: Zeichenfolge

cloudformationStackArn

Der Amazon-Ressourcenname (ARN) der Haupt-Ressource CloudFormation Stapel.

Typ: Zeichenfolge

cloudformationStackStatus

Typ: Zeichenfolge

Zulässige Werte: CREATE_IN_PROGRESS | CREATE_FAILED | CREATE_COMPLETE
| ROLLBACK_IN_PROGRESS | ROLLBACK_FAILED | ROLLBACK_COMPLETE
| DELETE_IN_PROGRESS | DELETE_FAILED | DELETE_COMPLETE |

UPDATE_IN_PROGRESS | UPDATE_COMPLETE_CLEANUP_IN_PROGRESS |
UPDATE_COMPLETE | UPDATE_ROLLBACK_IN_PROGRESS | UPDATE_ROLLBACK_FAILED |
UPDATE_ROLLBACK_COMPLETE_CLEANUP_IN_PROGRESS | UPDATE_ROLLBACK_COMPLETE

Cluster-Status

Typ: Zeichenfolge

Zulässige Werte: CREATE_IN_PROGRESS | CREATE_FAILED | CREATE_COMPLETE
| DELETE_IN_PROGRESS | DELETE_FAILED | DELETE_COMPLETE |
UPDATE_IN_PROGRESS | UPDATE_COMPLETE | UPDATE_FAILED

region

DieAWS-Regionin dem der Cluster erstellt wurde.

Typ: Zeichenfolge

Scheduler

Metadaten

Die Scheduler-Metadaten

Name

Der Name des Schedulers.

Typ: Zeichenfolge

Version

Die Scheduler-Version.

Typ: Zeichenfolge

type

Der Scheduler-Typ.

Typ: Zeichenfolge

Version

DieAWS ParallelClusterVersion, die zur Erstellung des Clusters verwendet wird.

Typ: Zeichenfolge

validation_messages

Eine Liste von Nachrichten mit einer Validierungsstufe unter `validationFailureLevel`. Die Liste der Nachrichten wird während der Konfigurationsvalidierung gesammelt.

ID

Die ID des Validators.

Typ: Zeichenfolge

level

Typ: Zeichenfolge

Zulässige Werte: INFO | WARNING | ERROR

Nachricht

Eine Bestätigungsnachricht.

Typ: Zeichenfolge

type

Der Typ des Validators.

Typ: Zeichenfolge

Beispiel

Python

Anforderung

```
$ create_cluster(cluster_name_3x, cluster-config.yaml)
```

200 Antwort

```
{  
  'cluster': {  
    'cloudformation_stack_arn': 'arn:aws:cloudformation:us-  
east-1:123456789012:stack/cluster-3x/e0462730-50b5-11ed-99a3-0a5ddc4a34c7',
```

```
'cloudformation_stack_status': 'CREATE_IN_PROGRESS',
'cluster_name': 'cluster-3x',
'cluster_status': 'CREATE_IN_PROGRESS',
'region': 'us-east-1',
'scheduler': {
  'type': 'slurm'
},
'version': '3.2.1'
}
}
```

Cluster löschen

Initiieren Sie das Löschen eines Clusters.

Themen

- [Erforderliche Syntax](#)
- [Anforderungstext](#)
- [Antwortsyntax](#)
- [Antworttext](#)
- [Beispiel](#)

Erforderliche Syntax

```
DELETE /v3/clusters/{clusterName}
{
  "region": "string"
}
```

Anforderungstext

Clustername

Der Name des Clusters

Typ: Zeichenfolge

Erforderlich: Ja

region

Der AWS-Region in dem der Cluster gelöscht wurde.

Typ: Zeichenkette

Erforderlich: Nein

Antwortsyntax

```
{
  "cluster": {
    "clusterName": "string",
    "region": "string",
    "version": "string",
    "cloudformationStackArn": "string",
    "cloudformationStackStatus": "DELETE_IN_PROGRESS",
    "clusterStatus": "DELETE_IN_PROGRESS",
    "scheduler": {
      "type": "string",
      "metadata": {
        "name": "string",
        "version": "string"
      }
    }
  }
}
```

Antworttext

Cluster

Eine Liste von Cluster-Instanzen

Clustername

Der Name eines Clusters.

Typ: Zeichenfolge

cloudformationStackArn

Der Amazon-Ressourcenname (ARN) des Hauptressourcenname CloudFormation Stapel.

Typ: Zeichenfolge

cloudformationStackStatus

Typ: Zeichenfolge

Zulässige Werte: CREATE_IN_PROGRESS | CREATE_FAILED | CREATE_COMPLETE
| ROLLBACK_IN_PROGRESS | ROLLBACK_FAILED | ROLLBACK_COMPLETE
| DELETE_IN_PROGRESS | DELETE_FAILED | DELETE_COMPLETE |
UPDATE_IN_PROGRESS | UPDATE_COMPLETE_CLEANUP_IN_PROGRESS
| UPDATE_COMPLETE | UPDATE_ROLLBACK_IN_PROGRESS |
UPDATE_ROLLBACK_FAILED | UPDATE_ROLLBACK_COMPLETE_CLEANUP_IN_PROGRESS
| UPDATE_ROLLBACK_COMPLETE

Clusterstatus

Typ: Zeichenfolge

Zulässige Werte: CREATE_IN_PROGRESS | CREATE_FAILED | CREATE_COMPLETE
| DELETE_IN_PROGRESS | DELETE_FAILED | DELETE_COMPLETE |
UPDATE_IN_PROGRESS | UPDATE_COMPLETE | UPDATE_FAILED

region

Die AWS-Region in der der Cluster erstellt wurde.

Typ: Zeichenfolge

Scheduler

Metadaten

Die Scheduler-Metadaten.

Name

Der Name des Schedulers.

Typ: Zeichenfolge

Version

Die Scheduler-Version

Typ: Zeichenfolge

type

Der Scheduler-Typ.

Typ: Zeichenfolge

Version

DerAWS ParallelClusterVersion, die zur Erstellung des Clusters verwendet wird.

Typ: Zeichenfolge

Beispiel

Python

Anforderung

```
$ delete_cluster(cluster_name_3x)
```

200 Antwort

```
{
  'cluster': {
    'cloudformation_stack_arn': 'arn:aws:cloudformation:us-
east-1:123456789012:stack/cluster_name_3x/16b49540-ae5-11ec-8e18-0ac1d712b241',
    'cloudformation_stack_status': 'DELETE_IN_PROGRESS',
    'cluster_name': 'cluster_name_3x',
    'cluster_status': 'DELETE_IN_PROGRESS',
    'region': 'us-east-1',
    'version': '3.2.1'
  }
}
```

deleteClusterInstances

Initiieren Sie die erzwungene Kündigung aller Cluster-Rechenknoten. Diese Aktion unterstützt nichtAWS BatchCluster.

Themen

- [Erforderliche Syntax](#)

- [Anforderungstext](#)
- [Antworttext](#)
- [Beispiel](#)

Erforderliche Syntax

```
DELETE /v3/clusters/{clusterName}/instances
{
  "force": boolean,
  "region": "string"
}
```

Anforderungstext

Clustername

Der Name des Clusters

Typ: Zeichenfolge

Erforderlich: Ja

force

Wenn eingestellt auf `true`, erzwingen Sie das Löschen, wenn der Cluster mit dem angegebenen Namen nicht gefunden wird. Der Standardwert ist `false`.

Typ: boolescher Wert

Required: No

region

Das AWS-Region in dem sich der Cluster befindet.

Typ: Zeichenkette

Required: No

Antworttext

Keine

Beispiel

Python

Anforderung

```
$ delete_cluster_instances(cluster_name_3x)
```

200 Antwort

Keine

Bild löschen

Initiieren Sie das Löschen des benutzerdefinierten AWS ParallelCluster Bild.

Themen

- [Erforderliche Syntax](#)
- [Anforderungstext](#)
- [Antwortsyntax](#)
- [Antworttext](#)
- [Beispiel](#)

Erforderliche Syntax

```
DELETE /v3/images/custom/{imageId}
{
  "force": boolean,
  "region": "string"
}
```

Anforderungstext

imageId

Die ID des Bildes.

Typ: Zeichenfolge

Erforderlich: Ja

force

Wenn gesetzt auf `true`, erzwingen Sie das Löschen des AMI. Verwenden Sie diesen Parameter, wenn es Instances gibt, die das AMI verwenden, oder wenn das AMI gemeinsam genutzt wird. Der Standardwert ist `false`.

Typ: boolescher Wert

Required: No

region

Der AWS-Region in dem das Bild erstellt wurde.

Typ: Zeichenkette

Erforderlich: Nein

Antwortsyntax

```
{
  "image": {
    "imageId": "string",
    "ec2AmiInfo": {
      "amiId": "string"
    },
    "region": "string",
    "version": "string",
    "cloudformationStackArn": "string",
    "imageBuildStatus": "DELETE_IN_PROGRESS",
    "cloudformationStackStatus": "DELETE_IN_PROGRESS"
  }
}
```

Antworttext

Image

cloudformationStackArn

Der Amazon-Ressourcenname (ARN) der Hauptressource CloudFormation Stapel.

Typ: Zeichenfolge

cloudformationStackStatus

Der CloudFormation Status des Stapels.

Typ: Zeichenfolge

Zulässige Werte: CREATE_IN_PROGRESS | CREATE_FAILED | CREATE_COMPLETE
| ROLLBACK_IN_PROGRESS | ROLLBACK_FAILED | ROLLBACK_COMPLETE
| DELETE_IN_PROGRESS | DELETE_FAILED | DELETE_COMPLETE |
UPDATE_IN_PROGRESS | UPDATE_COMPLETE_CLEANUP_IN_PROGRESS
| UPDATE_COMPLETE | UPDATE_ROLLBACK_IN_PROGRESS |
UPDATE_ROLLBACK_FAILED | UPDATE_ROLLBACK_COMPLETE_CLEANUP_IN_PROGRESS
| UPDATE_ROLLBACK_COMPLETE

ec2AmiInfo

Ein ID

Die EC2-AMI-ID.

Typ: Zeichenfolge

imageBuildStatus

Der Status der Image-Erstellung.

Typ: Zeichenfolge

Zulässige Werte: BUILD_IN_PROGRESS | BUILD_FAILED | BUILD_COMPLETE |
DELETE_IN_PROGRESS | DELETE_FAILED | DELETE_COMPLETE

imageId

Die ID des Images.

Typ: Zeichenfolge

region

DieAWS-Regionin dem das Bild erstellt wurde.

Typ: Zeichenfolge

Version

DasAWS ParallelClusterVersion, die zum Erstellen des Images verwendet wurde.

Typ: Zeichenfolge

Beispiel

Python

Anforderung

```
$ delete_image(custom-image-id)
```

200 Antwort

```
{
  'image': {
    'image_build_status': 'DELETE_IN_PROGRESS',
    'image_id': 'custom-image-id',
    'region': 'us-east-1',
    'version': '3.2.1'
  }
}
```

Cluster beschreiben

Holen Sie sich detaillierte Informationen zu einem vorhandenen Cluster.

Themen

- [Erforderliche Syntax](#)
- [Anforderungstext](#)
- [Antwortsyntax](#)
- [Antworttext](#)
- [Beispiel](#)

Erforderliche Syntax

```
GET /v3/clusters/{clusterName}
{
```

```
"region": "string"
}
```

Anforderungstext

Clustername

Der Name des Clusters

Typ: Zeichenfolge

Erforderlich: Ja

region

Der AWS-Region in dem sich der Cluster befindet.

Typ: Zeichenkette

Erforderlich: Nein

Antwortsyntax

Note

`failureReason` hat sich geändert zu `failures` beginnend mit AWS ParallelCluster Version 3.5.0.

```
{
  "clusterName": "string",
  "region": "string",
  "version": "string",
  "cloudFormationStackStatus": "CREATE_IN_PROGRESS",
  "clusterStatus": "CREATE_IN_PROGRESS",
  "scheduler": {
    "type": "string",
    "metadata": {
      "name": "string",
      "version": "string"
    }
  }
}
```



```
},
"cloudformationStackArn": "string",
"creationTime": "2019-08-24T14:15:22Z",
"lastUpdatedTime": "2019-08-24T14:15:22Z",
"clusterConfiguration": {
  "url": "string"
},
"computeFleetStatus": "START_REQUESTED",
"tags": [
  {
    "key": "string",
    "value": "string"
  }
],
"headNode": {
  "instanceId": "string",
  "instanceType": "string",
  "launchTime": "2019-08-24T14:15:22Z",
  "privateIpAddress": "string",
  "publicIpAddress": "string",
  "state": "pending"
},
"failures": [
  {
    "failureCode": "string",
    "failureReason": "string"
  }
]
"loginNodes": {
  "status": "string",
  "address": "string",
  "scheme": "string",
  "healthyNodes": integer,
  "unhealthyNodes": integer
}
}
```

Antworttext

Clustername

Der Name des Clusters

Typ: Zeichenfolge

cloudformationStackArn

Der Amazon-Ressourcenname (ARN) der Hauptressource CloudFormation Stapel.

Typ: Zeichenfolge

cloudformationStackStatus

Der CloudFormation Status des Stapels.

Typ: Zeichenfolge

Zulässige Werte: CREATE_IN_PROGRESS | CREATE_FAILED | CREATE_COMPLETE
| ROLLBACK_IN_PROGRESS | ROLLBACK_FAILED | ROLLBACK_COMPLETE
| DELETE_IN_PROGRESS | DELETE_FAILED | DELETE_COMPLETE |
UPDATE_IN_PROGRESS | UPDATE_COMPLETE_CLEANUP_IN_PROGRESS |
UPDATE_COMPLETE | UPDATE_ROLLBACK_IN_PROGRESS | UPDATE_ROLLBACK_FAILED |
UPDATE_ROLLBACK_COMPLETE_CLEANUP_IN_PROGRESS | UPDATE_ROLLBACK_COMPLETE

Cluster-Konfiguration

URL

Die URL der Cluster-Konfigurationsdatei.

Typ: Zeichenfolge

Cluster-Status

Der Clusterstatus.

Typ: Zeichenfolge

Zulässige Werte: CREATE_IN_PROGRESS | CREATE_FAILED | CREATE_COMPLETE
| DELETE_IN_PROGRESS | DELETE_FAILED | DELETE_COMPLETE |
UPDATE_IN_PROGRESS | UPDATE_COMPLETE | UPDATE_FAILED

computeFleetStatus

Der Status der Rechenflotte.

Typ: Zeichenfolge

Zulässige Werte: START_REQUESTED | STARTING | RUNNING | PROTECTED |
STOP_REQUESTED | STOPPING | STOPPED | UNKNOWN | ENABLED | DISABLED

creationTime

Zeitstempel für den Zeitpunkt, an dem der Cluster erstellt wurde.

Typ: DateTime

lastUpdatedTime

Zeitstempel für die letzte Aktualisierung des Clusters.

Typ: DateTime

region

DerAWS-Regionin dem der Cluster erstellt wurde.

Typ: Zeichenfolge

Tags

Die Liste der Tags, die dem Cluster zugeordnet sind.

Schlüssel

Tag-Name.

Typ: Zeichenfolge

Tag (Markierung)

Tag-Wert.

Typ: Zeichenfolge

Version

DerAWS ParallelClusterVersion, die zur Erstellung des Clusters verwendet wird.

Typ: Zeichenfolge

failures

Die Liste der Fehler, wenn der Cluster-Stack aktiviert istCREATE_FAILEDStatus.

Fehlercode

Der Fehlercode, wenn sich der Cluster-Stack befindetCREATE_FAILEDStatus.

Typ: Zeichenfolge

Grund des Fehlers

Der Grund für den Fehler, wenn der Cluster-Stack aktiviert ist `CREATE_FAILED` Status.

Typ: Zeichenfolge

head_node

Der Cluster-Kopfknoten.

instanceId

Die EC2-Instanz-ID.

Typ: Zeichenfolge

instanceType

Der EC2-Instance-Typ.

Typ: Zeichenfolge

launchTime

Der Zeitpunkt, zu dem die EC2-Instance gestartet wurde.

Typ: DateTime

privateIpAddress

Die private Cluster-IP-Adresse.

Typ: Zeichenfolge

publicIpAddress

Die öffentliche Cluster-IP-Adresse.

Typ: Zeichenfolge

state

Der Status der Head-Node-Instance.

Typ: Zeichenfolge

Zulässige Werte: `pending` | `running` | `shutting-down` | `terminated` | `stopping`
| `stopped`

Scheduler

Metadaten

Die Scheduler-Metadaten.

Name

Der Name des Schedulers.

Typ: Zeichenfolge

Version

Die Scheduler-Version.

Typ: Zeichenfolge

LoginNodes

status

Der Status des Login-Knotens.

Typ: Zeichenfolge

Zulässige Werte: `PENDING` | `FAILED` | `ACTIVE`

Adresse

Die Adresse des Anmeldeknotens.

Typ: Zeichenfolge

scheme

Das Anmeldeknotenschema.

Typ: Zeichenfolge

scheme

Anzahl intakter Knoten.

Typ: Ganzzahl

scheme

Die Anzahl der fehlerhaften Knoten.

Typ: Ganzzahl

type

Der Scheduler-Typ.

Typ: Zeichenfolge

Beispiel

Python

Anforderung

```
$ describe_cluster(cluster_name_3x)
```

200 Antwort

```
{
  'cloud_formation_stack_status': 'CREATE_COMPLETE',
  'cloudformation_stack_arn': 'arn:aws:cloudformation:us-east-1:123456789012:stack/
cluster_name_3x/16b49540-ae5-11ec-8e18-0ac1d712b241',
  'cluster_configuration': {
    'url': 'https://parallelcluster-....'
  },
  'cluster_name': 'cluster_name_3x',
  'cluster_status': 'CREATE_COMPLETE',
  'compute_fleet_status': 'RUNNING',
  'creation_time': datetime.datetime(2022, 3, 28, 22, 19, 9, 661000,
tzinfo=tzlocal()),
  'head_node': {
    'instance_id': 'i-abcdef01234567890',
    'instance_type': 't2.micro',
    'launch_time': datetime.datetime(2022, 3, 28, 22, 21, 56, tzinfo=tzlocal()),
    'private_ip_address': '172.31.56.3',
    'public_ip_address': '107.23.100.164',
    'state': 'running'
  },
}
```

```
'last_updated_time': datetime.datetime(2022, 3, 28, 22, 19, 9, 661000,
tzinfo=tzlocal()),
'region': 'us-east-1',
'tags': [
  {
    'key': 'parallelcluster:version', 'value': '3.2.1'
  }
],
'version': '3.2.1'
}
```

describeClusterInstances

Beschreiben Sie die Instanzen, die zu einem Cluster gehören.

Themen

- [Erforderliche Syntax](#)
- [Anforderungstext](#)
- [Antwortsyntax](#)
- [Antworttext](#)
- [Beispiel](#)

Erforderliche Syntax

```
GET /v3/clusters/{clusterName}/instances
{
  "nextToken": "string",
  "nodeType": "string",
  "queueName": "string",
  "region": "string"
}
```

Anforderungstext

Clustername

Der Name des Clusters

Typ: Zeichenfolge

Erforderlich: Ja

nextToken

Ein Token, das für paginierte Anfragen verwendet wird.

Typ: Zeichenkette

Required: No

Knotentyp

Filtern Sie die Instanzen nach Knotentyp.

Typ: Zeichenfolge

Zulässige Werte: HeadNode, ComputeNode, LoginNode

Required: No

Warteschlangenname

Filtert die Instanzen nach dem Namen der Warteschlange.

Typ: Zeichenkette

Required: No

region

DerAWS-Regionin dem sich der Cluster befindet.

Typ: Zeichenkette

Erforderlich: Nein

Antwortsyntax

```
{
  "nextToken": "string",
  "instances": [
    {
      "instanceId": "string",
      "instanceType": "string",
```



```
    "launchTime": "2019-08-24T14:15:22Z",  
    "privateIpAddress": "string",  
    "publicIpAddress": "string",  
    "state": "pending",  
    "nodeType": "HeadNode",  
    "queueName": "string"  
  }  
]  
}
```

Antworttext

-Instances

Die Liste der Cluster-Instanzen.

instanceld

Die EC2-Instanz-ID.

Typ: Zeichenfolge

instanceType

Der EC2-Instance-Typ.

Typ: Zeichenfolge

launchTime

Der Zeitpunkt, zu dem die EC2-Instance gestartet wurde.

Typ: DateTime

Knotentyp

Der Knotentyp.

Typ: Zeichenfolge

Zulässige Werte: HeadNode, ComputeNode, LoginNode

publicIpAddress

Die öffentliche IP-Adresse des Clusters.

Typ: Zeichenfolge

Warteschlangenname

Der Name der Warteschlange, in der die EC2-Instance einen Knoten unterstützt.

Typ: Zeichenfolge

state

Der Status der EC2-Instanz des Knotens.

Typ: Zeichenfolge

Zulässige Werte: pending | running | shutting-down | terminated | stopping
| stopped

nextToken

Ein Token, das für paginierte Anfragen verwendet wird.

Typ: Zeichenfolge

Beispiel

Python

Anforderung

```
$ describe_cluster_instances(cluster_name_3x)
```

200 Antwort

```
{
  'instances': [
    {
      'instance_id': 'i-abcdef01234567890',
      'instance_type': 't2.micro',
      'launch_time': datetime.datetime(2022, 3, 30, 14, 2, 7, tzinfo=tzlocal()),
      'node_type': 'HeadNode',
      'private_ip_address': '192.0.2.5',
      'public_ip_address': '198.51.100.180',
      'state': 'running'
    }
  ]
}
```

```
}  
]  
}
```

describeComputeFleet

Beschreiben Sie den Status der Rechenflotte.

Themen

- [Erforderliche Syntax](#)
- [Anforderungstext](#)
- [Antwortsyntax](#)
- [Antworttext](#)
- [Beispiel](#)

Erforderliche Syntax

```
GET /v3/clusters/{clusterName}/computefleet  
{  
  "region": "string"  
}
```

Anforderungstext

Clustername

Der Name des Clusters

Typ: Zeichenfolge

Erforderlich: Ja

region

DieAWS-Regionin dem sich der Cluster befindet.

Typ: Zeichenkette

Erforderlich: Nein

Antwortsyntax

```
{
  "status": "START_REQUESTED",
  "lastStatusUpdatedTime": "2019-08-24T14:15:22Z"
}
```

Antworttext

status

Typ: Zeichenfolge

Zulässige Werte: START_REQUESTED | STARTING | RUNNING | PROTECTED | STOP_REQUESTED | STOPPING | STOPPED | UNKNOWN | ENABLED | DISABLED

lastStatusUpdatedZeit

Der Zeitstempel, der die Uhrzeit der letzten Statusaktualisierung darstellt.

Typ: DateTime

Beispiel

Python

Anforderung

```
$ describe_compute_fleet(cluster_name_3x)
```

200 Antwort

```
{
  'last_status_updated_time': datetime.datetime(2022, 3, 28, 22, 27, 14,
  tzinfo=tzlocal()),
  'status': 'RUNNING'
}
```

Beschreiben Sie das Bild

Holen Sie sich detaillierte Informationen über ein vorhandenes Image.

Themen

- [Erforderliche Syntax](#)
- [Anforderungstext](#)
- [Antwortsyntax](#)
- [Antworttext](#)
- [Beispiel](#)

Erforderliche Syntax

```
GET /v3/images/custom/{imageId}  
{  
  "region": "string"  
}
```

Anforderungstext

imageId

Die ID des Bildes.

Typ: Zeichenfolge

Erforderlich: Ja

region

DasAWS-Regionin dem das Bild erstellt wurde.

Typ: Zeichenkette

Erforderlich: Nein

Antwortsyntax

```
{  
  "imageId": "string",  
  "region": "string",  
  "version": "string",
```

```
"imageBuildStatus": "BUILD_IN_PROGRESS",
"imageBuildLogsArn": "string",
"cloudformationStackStatus": "CREATE_IN_PROGRESS",
"cloudformationStackStatusReason": "string",
"cloudformationStackArn": "string",
"creationTime": "2019-08-24T14:15:22Z",
"cloudformationStackCreationTime": "2019-08-24T14:15:22Z",
"cloudformationStackTags": [
  {
    "key": "string",
    "value": "string"
  }
],
"imageConfiguration": {
  "url": "string"
},
"imagebuilderImageStatus": "PENDING",
"imagebuilderImageStatusReason": "string",
"ec2AmiInfo": {
  "amiId": "string",
  "tags": [
    {
      "key": "string",
      "value": "string"
    }
  ],
  "amiName": "string",
  "architecture": "string",
  "state": "PENDING",
  "description": "string"
}
}
```

Antworttext

imageId

Die ID des Bildes, für das detaillierte Informationen abgerufen werden sollen.

Typ: Zeichenfolge

imageBuildStatus

Der Status der Image-Erstellung.

Typ: Zeichenfolge

Zulässige Werte: BUILD_IN_PROGRESS | BUILD_FAILED | BUILD_COMPLETE | DELETE_IN_PROGRESS | DELETE_FAILED | DELETE_COMPLETE

Image-Konfiguration

URL

Die URL der Image-Konfigurationsdatei.

Typ: Zeichenfolge

region

DerAWS-Regionin dem das Bild erstellt wurde.

Typ: Zeichenfolge

Version

DasAWS ParallelClusterVersion, die zum Erstellen des Images verwendet wurde.

Typ: Zeichenfolge

cloudformationStackArn

Der Amazon-Ressourcename (ARN) der Hauptversion CloudFormation Stapel.

Typ: Zeichenfolge

cloudformationStackCreationZeit

Der Zeitstempel dafür, wann der CloudFormation Der Stapel wurde erstellt.

Typ: DateTime

cloudformationStackStatus

Der CloudFormation Status des Stapels.

Typ: Zeichenfolge

Zulässige Werte: CREATE_IN_PROGRESS | CREATE_FAILED | CREATE_COMPLETE | ROLLBACK_IN_PROGRESS | ROLLBACK_FAILED | ROLLBACK_COMPLETE | DELETE_IN_PROGRESS | DELETE_FAILED | DELETE_COMPLETE | UPDATE_IN_PROGRESS | UPDATE_COMPLETE_CLEANUP_IN_PROGRESS |

UPDATE_COMPLETE | UPDATE_ROLLBACK_IN_PROGRESS | UPDATE_ROLLBACK_FAILED |
UPDATE_ROLLBACK_COMPLETE_CLEANUP_IN_PROGRESS | UPDATE_ROLLBACK_COMPLETE

cloudformationStackStatusGrund

Der Grund für die CloudFormation Status des Stapels.

Typ: Zeichenfolge

cloudformationStackTags

Die Liste der Tags für CloudFormation stapeln.

Schlüssel

Der Tag-Name.

Typ: Zeichenfolge

Wert

Der Tag-Wert.

Typ: Zeichenfolge

creationTime

Zeitstempel für den Zeitpunkt, an dem das Bild erstellt wurde.

Typ: DateTime

ec2AmiInfo

Ein ID

Die EC2-AMI-ID.

Typ: Zeichenfolge

AMI-Name

Der EC2-AMI-Name.

Typ: Zeichenfolge

Anwendung ansehen

Die EC2-AMI-Architektur.

Typ: Zeichenfolge

state

Der Status des EC2-AMI.

Typ: Zeichenfolge

Zulässige Werte: PENDING | AVAILABLE | INVALID | DEREGISTERED | TRANSIENT
| FAILED | ERROR

Tags

Liste der EC2-AMI-Tags.

Schlüssel

Tag-Name.

Typ: Zeichenfolge

Wert

Tag-Wert.

Typ: Zeichenfolge

imagebuilderImageStatus

Der ImageBuilder Status.

Typ: Zeichenfolge

Zulässige Werte: PENDING | CREATING | BUILDING | TESTING | DISTRIBUTING |
INTEGRATING | AVAILABLE | CANCELLED | FAILED | DEPRECATED | DELETED

imagebuilderImageStatusGrund

Grund für den ImageBuilder Status des Bildes.

Typ: Zeichenfolge

imageBuildLogsArn

Der Amazon-Ressourcenname (ARN) der Protokolle für den Image-Erstellungsprozess.

Typ: Zeichenfolge

Beispiel

Python

Anforderung

```
$ describe_image(custom-image-id)
```

200 Antwort

```
{
  'cloudformation_stack_arn': 'arn:aws:cloudformation:us-east-1:123456789012:stack/
custom-image-id/6accc570-b080-11ec-845e-0e2dc6386985',
  'cloudformation_stack_creation_time': datetime.datetime(2022, 3, 30, 23, 23, 33,
731000, tzinfo=tzlocal()),
  'cloudformation_stack_status': 'CREATE_IN_PROGRESS',
  'cloudformation_stack_tags': [
    {
      'key': 'parallelcluster:version', 'value': '3.2.1'
    },
    {
      'key': 'parallelcluster:image_name',
      'value': 'custom-image-id'
    },
    {
      'key': 'parallelcluster:custom-image-id',
      'value': 'custom-image-id'
    },
    {
      'key': 'parallelcluster:s3_bucket',
      'value': 'parallelcluster-abcdef01234567890-v1-do-not-delete'
    },
    {
      'key': 'parallelcluster:s3_image_dir',
      'value': 'parallelcluster/3.2.1/images/custom-image-id-1234567890abcdef0'
    },
    {
      'key': 'parallelcluster:build_log',
      'value': 'arn:aws:logs:us-east-1:123456789012:log-group:/aws/imagebuilder/
ParallelClusterImage-custom-image-id'
    },
    {
      'key': 'parallelcluster:build_config',
```

```
    'value': 's3://parallelcluster-abcdef01234567890-v1-do-not-delete/
parallelcluster/3.2.1/images/custom-image-id-1234567890abcdef0/configs/image-
config.yaml'
  }
],
'image_build_logs_arn': 'arn:aws:logs:us-east-1:123456789012:log-group:/aws/
imagebuilder/ParallelClusterImage-alinux2-image',
'image_build_status': 'BUILD_IN_PROGRESS',
'image_configuration': {
  'url': 'https://parallelcluster-abcdef01234567890-v1-do-not-
delete.s3.amazonaws.com/parallelcluster/3.2.1/images/custom-image-
id-1234567890abcdef0/configs/image-config.yaml?... '
},
'image_id': 'custom-image-id',
'imagebuilder_image_status': 'PENDING',
'region': 'us-east-1',
'version': '3.2.1'
}
```

getClusterLogEreignisse

Ruft die Ereignisse ab, die einem Protokollstream zugeordnet sind.

Themen

- [Erforderliche Syntax](#)
- [Anforderungstext](#)
- [Antwortsyntax](#)
- [Antworttext](#)
- [Beispiel](#)

Erforderliche Syntax

```
GET /v3/clusters/{clusterName}/logstreams/{logStreamName}
{
  "endTime": datetime,
  "limit": float,
  "nextToken": "string",
  "region": "string",
  "startFromHead": boolean,
```

```
"startTime": datetime
}
```

Anforderungstext

Clustername

Der Name des Clusters

Typ: Zeichenfolge

Erforderlich: Ja

logStreamName

Der Name des Protokollstreams.

Typ: Zeichenfolge

Erforderlich: Ja

endTime

Das Ende des Zeitbereichs, ausgedrückt im ISO 8601-Format. Ereignisse mit einem Zeitstempel, der diesem Zeitpunkt entspricht oder später liegt, sind nicht enthalten.

Typ: DateTime

Format: 2021-01-01T20:00:00Z

Required: No

limit

Die maximale Anzahl der zurückgegebenen Protokollereignisse. Wenn Sie keinen Wert angeben, beträgt das Maximum so viele Protokollereignisse, wie in eine Antwortgröße von 1 MB passen, also bis zu 10.000 Protokollereignisse.

Typ: Gleitkommazahl

Required: No

nextToken

Ein Token, das für paginierte Anfragen verwendet wird.

Typ: Zeichenkette

Required: No

region

Das AWS-Region in dem sich der Cluster befindet.

Typ: Zeichenkette

Required: No

startFromHead

Wenn gesetzt auf `true`, werden die frühesten Protokollereignisse zuerst zurückgegeben.

Wenn der Wert ist `false`, werden die neuesten Protokollereignisse zuerst zurückgegeben. Der Standardwert ist `false`.

Typ: boolescher Wert

Required: No

startTime

Der Beginn des Zeitbereichs, ausgedrückt im ISO 8601-Format. Ereignisse mit einem Zeitstempel, der diesem Zeitpunkt oder einem späteren Zeitpunkt entspricht, sind enthalten.

Typ: DateTime

Format: `2021-01-01T20:00:00Z`

Required: No

Antwortsyntax

```
{
  "nextToken": "string",
  "prevToken": "string",
  "events": [
    {
      "timestamp": "2019-08-24T14:15:22Z",
      "message": "string"
    }
  ]
}
```

Antworttext

Veranstaltungen

Liste der gefilterten Ereignisse.

Nachricht

Die Ereignisnachricht.

Typ: Zeichenfolge

timestamp

Der Ereigniszeitstempel.

Typ: DateTime

nextToken

Ein Token, das für paginierte Anfragen verwendet wird.

Typ: Zeichenfolge

Zurück:Token

Ein Token, das für paginierte Anfragen verwendet wird.

Typ: Zeichenfolge

Beispiel

Python

Anforderung

```
$ get_cluster_log_events(cluster_name_3x, log_stream_name=ip-192-0-2-26.i-  
abcdef01234567890.cfn-init)
```

200 Antwort

```
"events": [  
  {  
    "message": "2022-09-22 16:40:15,127 [DEBUG] CloudFormation client initialized  
with endpoint https://cloudformation.us-east-1.amazonaws.com",
```

```
    "timestamp": "2022-09-22T16:40:15.127Z"
  },
  {
    "message": "2022-09-22 16:40:15,127 [DEBUG] Describing resource
HeadNodeLaunchTemplate in stack cluster_name_3x",
    "timestamp": "2022-09-22T16:40:15.127Z"
  },
  ...
]
```

getClusterStackEreignisse

Rufen Sie die Ereignisse ab, die dem Stack für einen Cluster zugeordnet sind.

Note

Ab Version 3.6.0, AWS ParallelCluster verwendet verschachtelte Stacks, um die Ressourcen zu erstellen, die Warteschlangen und Rechenressourcen zugeordnet sind. Das `GetClusterStackEventsAPI` und die `awscli cluster get-cluster-stack-events` Der Befehl gibt nur die Cluster-Haupt-Stack-Ereignisse zurück. Sie können die Cluster-Stack-Ereignisse, einschließlich derer, die sich auf Warteschlangen und Rechenressourcen beziehen, im CloudFormation Konsole.

Themen

- [Erforderliche Syntax](#)
- [Anforderungstext](#)
- [Antwortsyntax](#)
- [Antworttext](#)
- [Beispiel](#)

Erforderliche Syntax

```
GET /v3/clusters/{clusterName}/stackevents
{
  "nextToken": "string",
  "region": "string"
```

```
}
```

Anforderungstext

Clustername

Der Name des Clusters

Typ: Zeichenfolge

Erforderlich: Ja

nextToken

Ein Token, das für paginierte Anfragen verwendet wird.

Typ: Zeichenkette

Required: No

region

DieAWS-Regionin dem sich der Cluster befindet.

Typ: Zeichenkette

Erforderlich: Nein

Antwortsyntax

```
{
  "nextToken": "string",
  "events": [
    {
      "stackId": "string",
      "eventId": "string",
      "stackName": "string",
      "logicalResourceId": "string",
      "physicalResourceId": "string",
      "resourceType": "string",
      "timestamp": "2019-08-24T14:15:22Z",
      "resourceStatus": "CREATE_IN_PROGRESS",
      "resourceStatusReason": "string",
      "resourceProperties": "string",
    }
  ]
}
```



```
    "clientRequestToken": "string"  
  }  
]  
}
```

Antworttext

Veranstaltungen

Liste der gefilterten Ereignisse.

`clientRequestToken`

Das Token, das an die Aktion übergeben wurde, die dieses Ereignis generiert hat.

Typ: Zeichenfolge

`eventId`

Die eindeutige ID dieses Ereignisses.

Typ: Zeichenfolge

`logicalResourceid`

Der logische Name der in der Vorlage angegebenen Ressource.

Typ: Zeichenfolge

`physicalResourceid`

Der Name oder die eindeutige Kennung, die der physischen Instanz der Ressource zugeordnet ist.

Typ: Zeichenfolge

`Eigenschaften der Ressource`

Ein BLOB der Eigenschaften, die zur Erstellung der Ressource verwendet werden.

Typ: Zeichenfolge

`Ressourcenstatus`

Der Ressourcenstatus.

Typ: Zeichenfolge

Zulässige Werte: CREATE_IN_PROGRESS | CREATE_FAILED | CREATE_COMPLETE | DELETE_IN_PROGRESS | DELETE_FAILED | DELETE_COMPLETE | DELETE_SKIPPED | UPDATE_IN_PROGRESS | UPDATE_FAILED | UPDATE_COMPLETE | IMPORT_FAILED | IMPORT_COMPLETE | IMPORT_IN_PROGRESS | IMPORT_ROLLBACK_IN_PROGRESS | IMPORT_ROLLBACK_FAILED | IMPORT_ROLLBACK_COMPLETE

resourceStatusReason

Eine Erfolgs- oder Fehlschlagsmeldung, die der Ressource zugeordnet ist.

Typ: Zeichenfolge

resourceType

Der Typ der Ressource.

Typ: Zeichenfolge

StackID

Der eindeutige ID-Name der Instanz des Stacks.

Typ: Zeichenfolge

Stack-Name

Der Name, der einem Stack zugeordnet ist.

Typ: Zeichenfolge

timestamp

Der Zeitpunkt, zu dem der Status aktualisiert wurde.

Typ: DateTime

nextToken

Ein Token, das für paginierte Anfragen verwendet wird.

Typ: Zeichenfolge

Beispiel

Python

Anforderung

```
$ get_cluster_stack_events(cluster_name_3x)
```

200 Antwort

```
{
  'events': [
    {
      'event_id': '590b3820-b081-11ec-985e-0a7af5751497',
      'logical_resource_id': 'cluster_name_3x',
      'physical_resource_id': 'arn:aws:cloudformation:us-east-1:123456789012:stack/
cluster_name_3x/11a59710-b080-11ec-b8bd-129def1380e9',
      'resource_status': 'CREATE_COMPLETE',
      'resource_type': 'AWS::CloudFormation::Stack',
      'stack_id': 'arn:aws:cloudformation:us-east-1:123456789012:stack/
cluster_name_3x/11a59710-b080-11ec-b8bd-129def1380e9',
      'stack_name': 'cluster_name_3x',
      'timestamp': datetime.datetime(2022, 3, 30, 23, 30, 13, 268000,
tzinfo=tzlocal())
    },
    ...
  ]
}
```

getImageLogEreignisse

Rufen Sie die Ereignisse ab, die mit einem Image-Build verknüpft sind.

Themen

- [Erforderliche Syntax](#)
- [Anforderungstext](#)
- [Antwortsyntax](#)
- [Antworttext](#)
- [Beispiel](#)

Erforderliche Syntax

```
GET /v3/images/custom/{imageId}/logstreams/{logStreamName}
{
```

```
"endTime": datetime,  
"limit": float,  
"nextToken": "string",  
"region": "string",  
"startFromHead": boolean,  
"startTime": datetime  
}
```

Anforderungstext

imageId

Die ID des Bildes.

Typ: Zeichenfolge

Erforderlich: Ja

logStreamName

Der Name des Logstreams.

Typ: Zeichenfolge

Erforderlich: Ja

endTime

Das Ende des Zeitbereichs, ausgedrückt im ISO 8601-Format. Ereignisse mit einem Zeitstempel, der diesem Zeitpunkt entspricht oder später liegt, sind nicht enthalten.

Typ: DateTime

Format: 2021-01-01T20:00:00Z

Required: No

limit

Die maximale Anzahl der zurückgegebenen Protokollereignisse. Wenn Sie keinen Wert angeben, beträgt das Maximum so viele Protokollereignisse, wie in eine Antwortgröße von 1 MB passen, also bis zu 10.000 Protokollereignisse.

Typ: Gleitkommazahl

Required: No

nextToken

Ein Token, das für paginierte Anfragen verwendet wird.

Typ: Zeichenkette

Required: No

region

Das AWS-Region, das das Bild drin ist.

Typ: Zeichenkette

Required: No

startFromHead

Wenn eingestellt auf `true`, gibt zuerst die frühesten Protokollereignisse zurück. Wenn der Wert auf `false` gesetzt ist, werden zuerst die neuesten Protokollereignisse zurückgegeben. Der Standardwert ist `false`.

Typ: boolescher Wert

Required: No

startTime

Der Beginn des Zeitbereichs, ausgedrückt im ISO 8601-Format. Ereignisse mit einem Zeitstempel, der diesem Zeitpunkt oder einem späteren Zeitpunkt entspricht, sind enthalten.

Typ: DateTime

Format: `2021-01-01T20:00:00Z`

Required: No

Antwortsyntax

```
{
  "nextToken": "string",
  "prevToken": "string",
  "events": [
    {
```

```
    "timestamp": "2019-08-24T14:15:22Z",  
    "message": "string"  
  }  
]  
}
```

Antworttext

Veranstaltungen

Eine Liste gefilterter Ereignisse.

Nachricht

Die Ereignisnachricht.

Typ: Zeichenfolge

timestamp

Der Ereigniszeitstempel.

Typ: DateTime

nextToken

Ein Token, das für paginierte Anfragen verwendet wird.

Typ: Zeichenfolge

Zurück:Token

Ein Token, das für paginierte Anfragen verwendet wird.

Typ: Zeichenfolge

Beispiel

Python

Anforderung

```
$ get_image_log_events(image_id, log_stream_name=3.2.1/1)
```

200 Antwort

```
"events": [  
  {  
    "message": "ExecuteBash: STARTED EXECUTION",  
    "timestamp": "2022-04-05T15:51:20.228Z"  
  },  
  {  
    "message": "ExecuteBash: Created temporary directory: /tmp/1234567890abcdef0",  
    "timestamp": "2022-04-05T15:51:20.228Z"  
  },  
  ...  
]
```

getImageStackEreignisse

Rufen Sie die Ereignisse ab, die dem Stack für einen Image-Build zugeordnet sind.

Themen

- [Erforderliche Syntax](#)
- [Anforderungstext](#)
- [Antwortsyntax](#)
- [Antworttext](#)
- [Beispiel](#)

Erforderliche Syntax

```
GET /v3/images/custom/{imageId}/stackevents  
{  
  "nextToken": "string",  
  "region": "string"  
}
```

Anforderungstext

imageId

Die ID des Bildes.

Typ: Zeichenfolge

Erforderlich: Ja

nextToken

Ein Token, das für paginierte Anfragen verwendet wird.

Typ: Zeichenkette

Required: No

region

DieAWS-Regionin dem das Bild ist.

Typ: Zeichenkette

Erforderlich: Nein

Antwortsyntax

```
{
  "nextToken": "string",
  "events": [
    {
      "stackId": "string",
      "eventId": "string",
      "stackName": "string",
      "logicalResourceId": "string",
      "physicalResourceId": "string",
      "resourceType": "string",
      "timestamp": "2019-08-24T14:15:22Z",
      "resourceStatus": "CREATE_IN_PROGRESS",
      "resourceStatusReason": "string",
      "resourceProperties": "string",
      "clientRequestToken": "string"
    }
  ]
}
```

Antworttext

Veranstaltungen

Eine Liste gefilterter Ereignisse.

clientRequestToken

Das Token, das an die Aktion übergeben wurde, die dieses Ereignis generiert hat.

Typ: Zeichenfolge

eventId

Die eindeutige ID dieses Ereignisses.

Typ: Zeichenfolge

logicalResourceid

Der logische Name der in der Vorlage angegebenen Ressource.

Typ: Zeichenfolge

physicalResourceid

Der Name oder die eindeutige Kennung, die der physischen Instanz der Ressource zugeordnet ist.

Typ: Zeichenfolge

Eigenschaften der Ressource

Ein BLOB der Eigenschaften, die zur Erstellung der Ressource verwendet werden.

Typ: Zeichenfolge

ResourceStatus

Der Ressourcenstatus.

Typ: Zeichenfolge

Zulässige Werte: CREATE_IN_PROGRESS | CREATE_FAILED | CREATE_COMPLETE | DELETE_IN_PROGRESS | DELETE_FAILED | DELETE_COMPLETE | DELETE_SKIPPED | UPDATE_IN_PROGRESS | UPDATE_FAILED | UPDATE_COMPLETE | IMPORT_FAILED | IMPORT_COMPLETE | IMPORT_IN_PROGRESS | IMPORT_ROLLBACK_IN_PROGRESS | IMPORT_ROLLBACK_FAILED | IMPORT_ROLLBACK_COMPLETE

resourceStatusReason

Eine Erfolgs- oder Fehlschlagsmeldung, die der Ressource zugeordnet ist.

Typ: Zeichenfolge

resourceType

Der Typ der Ressource.

Typ: Zeichenfolge

StackID

Der eindeutige ID-Name der Instanz des Stacks.

Typ: Zeichenfolge

Stack-Name

Der Name, der einem Stack zugeordnet ist.

Typ: Zeichenfolge

timestamp

Der Zeitpunkt, zu dem der Status aktualisiert wurde.

Typ: DateTime

nextToken

Ein Token, das für paginierte Anfragen verwendet wird.

Typ: Zeichenfolge

Beispiel

Python

Anforderung

```
$ get_image_stack_events(image_id)
```

200 Antwort

```
{
  'events': [
    {
      'event_id': 'ParallelClusterImage-
CREATE_IN_PROGRESS-2022-03-30T23:26:33.499Z',
      'logical_resource_id': 'ParallelClusterImage',
```

```
    'physical_resource_id': 'arn:aws:imagebuilder:us-east-1:123456789012:image/parallelclusterimage-alinux2-image/3.2.1/1',
    'resource_properties': {
      "InfrastructureConfigurationArn": "arn:aws:imagebuilder:us-east-1:123456789012:infrastructure-configuration/parallelclusterimage-6accc570-b080-11ec-845e-0e2dc6386985",
      "ImageRecipeArn": "arn:aws:imagebuilder:us-east-1:123456789012:image-recipe/parallelclusterimage-alinux2-image/3.2.1",
      "DistributionConfigurationArn": "arn:aws:imagebuilder:us-east-1:123456789012:distribution-configuration/parallelclusterimage-6accc570-b080-11ec-845e-0e2dc6386985",
      "EnhancedImageMetadataEnabled": "false",
      "Tags": {
        "parallelcluster:image_name": "alinux2-image", "parallelcluster:image_id": "alinux2-image"
      }
    },
    'resource_status': 'CREATE_IN_PROGRESS',
    'resource_status_reason': 'Resource creation Initiated',
    'resource_type': 'AWS::ImageBuilder::Image',
    'stack_id': 'arn:aws:cloudformation:us-east-1:123456789012:stack/alinux2-image/6accc570-b080-11ec-845e-0e2dc6386985',
    'stack_name': 'alinux2-image',
    'timestamp': datetime.datetime(2022, 3, 30, 23, 26, 33, 499000, tzinfo=tzlocal())
  },
  ...
]
```

Cluster auflisten

Rufen Sie eine Liste vorhandener Cluster ab.

Themen

- [Erforderliche Syntax](#)
- [Anforderungstext](#)
- [Antwortsyntax](#)
- [Antworttext](#)
- [Beispiel](#)

Erforderliche Syntax

```
GET /v3/clusters
{
  "clusterStatus": "string",
  "nextToken": "string",
  "region": "string"
}
```

Anforderungstext

ClusterStatus

Nach Clusterstatus filtern. Die Standardeinstellung ist „Alle Cluster“.

Typ: Zeichenfolge

Zulässige Werte: CREATE_IN_PROGRESS | CREATE_FAILED | CREATE_COMPLETE
| DELETE_IN_PROGRESS | DELETE_FAILED | UPDATE_IN_PROGRESS |
UPDATE_COMPLETE | UPDATE_FAILED

Erforderlich: nein

nextToken

Ein Token, das für paginierte Anfragen verwendet wird.

Typ: Zeichenkette

Required: No

region

DerAWS-Regionder Cluster.

Typ: Zeichenkette

Erforderlich: Nein

Antwortsyntax

```
{
```

```
"nextToken": "string",
"clusters": [
  {
    "clusterName": "string",
    "region": "string",
    "version": "string",
    "cloudformationStackArn": "string",
    "cloudformationStackStatus": "CREATE_IN_PROGRESS",
    "clusterStatus": "CREATE_IN_PROGRESS",
    "scheduler": {
      "type": "string",
      "metadata": {
        "name": "string",
        "version": "string"
      }
    }
  }
]
}
```

Antworttext

Cluster

cloudformationStackArn

Der Amazon-Ressourcenname (ARN) des Hauptressourcenname CloudFormation Stapel.

Typ: Zeichenfolge

cloudformationStackStatus

Der CloudFormation Status des Stapels.

Typ: Zeichenfolge

Zulässige Werte: CREATE_IN_PROGRESS | CREATE_FAILED | CREATE_COMPLETE
| ROLLBACK_IN_PROGRESS | ROLLBACK_FAILED | ROLLBACK_COMPLETE
| DELETE_IN_PROGRESS | DELETE_FAILED | DELETE_COMPLETE |
UPDATE_IN_PROGRESS | UPDATE_COMPLETE_CLEANUP_IN_PROGRESS
| UPDATE_COMPLETE | UPDATE_ROLLBACK_IN_PROGRESS |
UPDATE_ROLLBACK_FAILED | UPDATE_ROLLBACK_COMPLETE_CLEANUP_IN_PROGRESS
| UPDATE_ROLLBACK_COMPLETE

Clustername

Der Name des Clusters

Typ: Zeichenfolge

Clusterstatus

Der Cluster-Status.

Typ: Zeichenfolge

Zulässige Werte: CREATE_IN_PROGRESS | CREATE_FAILED | CREATE_COMPLETE
| DELETE_IN_PROGRESS | DELETE_FAILED | DELETE_COMPLETE |
UPDATE_IN_PROGRESS | UPDATE_COMPLETE | UPDATE_FAILED

Scheduler

Metadaten

Die Scheduler-Metadaten.

Name

Der Name des Schedulers.

Typ: Zeichenfolge

Version

Die Scheduler-Version.

Typ: Zeichenfolge

type

Der Typ des Schedulers.

Typ: Zeichenfolge

region

Die AWS-Region in dem der Cluster erstellt wurde.

Typ: Zeichenfolge

Version

Die AWS ParallelCluster-Version, die zur Erstellung des Clusters verwendet wird.

Typ: Zeichenfolge

nextToken

Ein Token, das für paginierte Anfragen verwendet wird.

Typ: Zeichenfolge

Beispiel

Python

Anforderung

```
$ list_clusters()
```

200 Antwort

```
{
  'clusters':
  [
    {
      'cloudformation_stack_arn': 'arn:aws:cloudformation:us-
east-1:123456789012:stack/cluster_name_3x/16b49540-ae5-11ec-8e18-0ac1d712b241',
      'cloudformation_stack_status': 'CREATE_COMPLETE',
      'cluster_name': 'cluster_name_3x',
      'cluster_status': 'CREATE_COMPLETE',
      'region': 'us-east-1',
      'version': '3.2.1'
    },
    ...
  ]
}
```

listClusterLogStreams

Rufen Sie die Liste der Log-Streams ab, die einem Cluster zugeordnet sind.

Themen

- [Erforderliche Syntax](#)

- [Anforderungstext](#)
- [Antwortsyntax](#)
- [Antworttext](#)
- [Beispiel](#)

Erforderliche Syntax

```
GET /v3/clusters/{clusterName}/logstreams
{
  "filters": [ "string" ],
  "nextToken": "string",
  "region": "string"
}
```

Anforderungstext

Clustername

Der Name des Clusters

Typ: Zeichenfolge

Erforderlich: Ja

-Filter

Filtert die Log-Streams.

Zulässige Filter sind:

- `private-dns-name`: Die Kurzform des privaten DNS-Namens der Instanz (z. B. `ip-10-0-0-101`).
- `Knotentyp`: Gültiger Wert: `HeadNode`.

Typ: Eindeutiges Zeichenketten-Array

Format: `Name=a,Values=1 Name=b,Values=2,3`

Required: No

nextToken

Ein Token, das für paginierte Anfragen verwendet wird.

Typ: Zeichenkette

Required: No

region

DieAWS-Regionin dem sich der Cluster befindet.

Typ: Zeichenkette

Erforderlich: Nein

Antwortsyntax

```
{
  "nextToken": "string",
  "logStreams": [
    {
      "logStreamName": "string",
      "creationTime": "2019-08-24T14:15:22Z",
      "firstEventTimestamp": "2019-08-24T14:15:22Z",
      "lastEventTimestamp": "2019-08-24T14:15:22Z",
      "lastIngestionTime": "2019-08-24T14:15:22Z",
      "uploadSequenceToken": "string",
      "logStreamArn": "string"
    }
  ]
}
```

Antworttext

LogStreams

Eine Liste von Log-Streams.

creationTime

Die Uhrzeit, zu der der Stream erstellt wurde.

Typ: DateTime

firstEventTimestamp

Die Uhrzeit des ersten Ereignisses des Streams.

Typ: DateTime

lastEventTimestamp

Die Uhrzeit des letzten Ereignisses des Streams. Die lastEventTime Wertaktualisierungen auf einer eventuellen Konsistenzbasis. Die Aktualisierung dauert in der Regel weniger als eine Stunde nach der Aufnahme, kann aber in seltenen Fällen auch länger dauern.

Typ: DateTime

lastIngestionTime

Der Zeitpunkt der letzten Aufnahme.

Typ: DateTime

logStreamArn

Der Amazon-Ressourcenname (ARN) des Log-Streams.

Typ: Zeichenfolge

logStreamName

Name des Log-Streams.

Typ: Zeichenfolge

uploadSequenceToken

Das Sequenz-Token.

Typ: Zeichenfolge

nextToken

Ein Token, das für paginierte Anfragen verwendet wird.

Typ: Zeichenfolge

Beispiel

Python

Anforderung

```
$ list_cluster_log_streams(cluster_name_3x)
```

200 Antwort

```
{
  'log_streams': [
    {
      'creation_time': datetime.datetime(2022, 3, 30, 14, 7, 34, 354000,
tzinfo=tzlocal()),
      'first_event_timestamp': datetime.datetime(2022, 3, 30, 14, 6, 41, 444000,
tzinfo=tzlocal()),
      'last_event_timestamp': datetime.datetime(2022, 3, 30, 14, 25, 55, 462000,
tzinfo=tzlocal()),
      'last_ingestion_time': datetime.datetime(2022, 3, 30, 14, 49, 50, 62000,
tzinfo=tzlocal()),
      'log_stream_arn': 'arn:aws:logs:us-east-1:123456789012:log-group:/aws/
parallelcluster/cluster_name_3x:log-stream:ip-192-0-2-26.i-abcdef01234567890.cfn-
init',
      'log_stream_name': 'ip-192-0-2-26.i-abcdef01234567890.cfn-init',
      ...
      'upload_sequence_token': '####'
    },
    ...
  ]
}
```

listImageLogStröme

Ruft die Liste der Log-Streams ab, die einem Bild zugeordnet sind.

Themen

- [Erforderliche Syntax](#)
- [Anforderungstext](#)
- [Antwortsyntax](#)
- [Antworttext](#)
- [Beispiel](#)

Erforderliche Syntax

```
GET /v3/images/custom/{imageId}/logstreams
{
```

```
"nextToken": "string",  
"region": "string"  
}
```

Anforderungstext

imageId

Die ID des Bildes.

Typ: Zeichenfolge

Erforderlich: Ja

nextToken

Ein Token, das für paginierte Anfragen verwendet wird.

Typ: Zeichenkette

Required: No

region

Das AWS-Region in dem das Bild ist.

Typ: Zeichenkette

Erforderlich: Nein

Antwortsyntax

```
{  
  "nextToken": "string",  
  "logStreams": [  
    {  
      "logStreamName": "string",  
      "creationTime": "2019-08-24T14:15:22Z",  
      "firstEventTimestamp": "2019-08-24T14:15:22Z",  
      "lastEventTimestamp": "2019-08-24T14:15:22Z",  
      "lastIngestionTime": "2019-08-24T14:15:22Z",  
      "uploadSequenceToken": "string",  
      "logStreamArn": "string"  
    }  
  ]  
}
```

```
}  
]  
}
```

Antworttext

Streams protokollieren

Eine Liste von Log-Streams.

`creationTime`

Die Uhrzeit, zu der der Stream erstellt wurde.

Typ: `DateTime`

`firstEventTimestamp`

Die Uhrzeit des ersten Ereignisses im Stream.

Typ: `DateTime`

`lastEventTimestamp`

Die Uhrzeit des letzten Ereignisses des Streams. Die `lastEventTime` Wertaktualisierungen auf einer eventuellen Konsistenzbasis. Die Aktualisierung dauert in der Regel weniger als eine Stunde nach der Aufnahme, kann aber in seltenen Fällen auch länger dauern.

Typ: `DateTime`

`lastIngestionTime`

Der Zeitpunkt der letzten Aufnahme.

Typ: `DateTime`

`logStreamArn`

Der Amazon-Ressourcenname (ARN) des Log-Streams.

Typ: Zeichenfolge

`logStreamName`

Der Name des Protokollstreams.

Typ: Zeichenfolge

uploadSequenceToken

Das Sequenz-Token.

Typ: Zeichenfolge

next_token

Ein Token, das für paginierte Anfragen verwendet wird.

Typ: Zeichenfolge

Beispiel

Python

Anforderung

```
$ list_image_log_streams(custom-image-id)
```

200 Antwort

```
{
  'log_streams': [
    {
      'creation_time': datetime.datetime(2022, 3, 29, 20, 29, 24, 875000,
tzinfo=tzlocal()),
      'first_event_timestamp': datetime.datetime(2022, 3, 29, 20, 29, 24, 775000,
tzinfo=tzlocal()),
      'last_event_timestamp': datetime.datetime(2022, 3, 29, 20, 38, 23, 944000,
tzinfo=tzlocal()),
      'last_ingestion_time': datetime.datetime(2022, 3, 29, 20, 51, 56, 26000,
tzinfo=tzlocal()),
      'log_stream_arn': 'arn:aws:logs:us-east-1:123456789012:log-group:/aws/
imagebuilder/ParallelClusterImage-alinux2-image:log-stream:3.2.1/1',
      'log_stream_name': '3.2.1/1',
      'upload_sequence_token': '####'
    },
    ...
  ]
}
```

Bilder auflisten

Ruft die Liste der vorhandenen benutzerdefinierten Bilder ab.

Themen

- [Erforderliche Syntax](#)
- [Anforderungstext](#)
- [Antwortsyntax](#)
- [Antworttext](#)
- [Beispiel](#)

Erforderliche Syntax

```
GET /images/custom
{
  "imageStatus": "string",
  "nextToken": "string",
  "region": "string"
}
```

Anforderungstext

Bildstatus

Filtert Bilder nach dem angegebenen Status.

Typ: Zeichenfolge

Zulässige Werte: AVAILABLE | PENDING | FAILED

Erforderlich: Ja

nextToken

Ein Token, das für paginierte Anfragen verwendet wird.

Typ: Zeichenkette

Required: No

region

Das AWS-Region die Bilder sind drin.

Typ: Zeichenkette

Erforderlich: Nein

Antwortsyntax

```
{
  "nextToken": "string",
  "images": [
    {
      "imageId": "string",
      "ec2AmiInfo": {
        "amiId": "string"
      },
      "region": "string",
      "version": "string",
      "cloudformationStackArn": "string",
      "imageBuildStatus": "BUILD_IN_PROGRESS",
      "cloudformationStackStatus": "CREATE_IN_PROGRESS"
    }
  ]
}
```

Antworttext

Bilder

Eine Liste von Bildern.

cloudformationStackArn

Der Amazon-Ressourcenname (ARN) der Hauptdatei CloudFormation Stapel.

Typ: Zeichenfolge

cloudformationStackStatus

Der CloudFormation Status des Stapels.

Typ: Zeichenfolge

Zulässige Werte: CREATE_IN_PROGRESS | CREATE_FAILED | CREATE_COMPLETE
| ROLLBACK_IN_PROGRESS | ROLLBACK_FAILED | ROLLBACK_COMPLETE
| DELETE_IN_PROGRESS | DELETE_FAILED | DELETE_COMPLETE |
UPDATE_IN_PROGRESS | UPDATE_COMPLETE_CLEANUP_IN_PROGRESS
| UPDATE_COMPLETE | UPDATE_ROLLBACK_IN_PROGRESS |
UPDATE_ROLLBACK_FAILED | UPDATE_ROLLBACK_COMPLETE_CLEANUP_IN_PROGRESS
| UPDATE_ROLLBACK_COMPLETE

ec2AmiInfo

ami_id

Die EC2-AMI-ID.

Typ: Zeichenfolge

imageBuildStatus

Der Status der Image-Erstellung.

Zulässige Werte: BUILD_IN_PROGRESS | BUILD_FAILED | BUILD_COMPLETE |
DELETE_IN_PROGRESS | DELETE_FAILED | DELETE_COMPLETE

Typ: Zeichenfolge

imageId

Die ID des Bildes.

Typ: Zeichenfolge

region

Die AWS-Region, in der das Bild erstellt wurde.

Typ: Zeichenfolge

Version

Das AWS ParallelCluster-Version, die zum Erstellen des Images verwendet wurde.

Typ: Zeichenfolge

nextToken

Ein Token, das für paginierte Anfragen verwendet wird.

Typ: Zeichenfolge

Beispiel

Python

Anforderung

```
$ list_images("AVAILABLE")
```

200 Antwort

```
{
  'images': [
    {
      'ec2_ami_info': {
        'ami_id': 'ami-abcdef01234567890'
      },
      'image_build_status': 'BUILD_COMPLETE',
      'image_id': 'custom-image',
      'region': 'us-east-1',
      'version': '3.2.1'
    }
  ]
}
```

listOfficialImages

Rufen Sie die Liste von abAWS ParallelClusteroffizielle Bilder.

Themen

- [Erforderliche Syntax](#)
- [Anforderungstext](#)
- [Antwortsyntax](#)
- [Antworttext](#)
- [Beispiel](#)

Erforderliche Syntax

```
GET /v3/images/official
```

```
{
  "architecture": "string",
  "os": "string",
  "region": "string"
}
```

Anforderungstext

Anwendung ansehen

Nach Architektur filtern. Die Standardeinstellung ist keine Filterung.

Typ: Zeichenfolge

Zulässige Werte: x86_64 | arm64

Required: No

os

Nach Betriebssystemdistribution filtern. Die Standardeinstellung ist keine Filterung.

Typ: Zeichenfolge

Zulässige Werte: alinux2 | centos7 | ubuntu2204 | ubuntu2004 | rhe18

Required: No

region

DerAWS-Regionin denen offizielle Bilder aufgeführt sind.

Typ: Zeichenkette

Erforderlich: Nein

Antwortsyntax

```
{
  "images": [
    {
      "architecture": "string",
      "amiId": "string",
      "name": "string",

```

```
    "os": "string",
    "version": "string"
  }
]
```

Antworttext

Bilder

AmiID

Die ID des AMI.

Typ: Zeichenfolge

Anwendung ansehen

Die AMI-Architektur.

Typ: Zeichenfolge

Name

Der Name des AMI.

Typ: Zeichenfolge

os

Das AMI-Betriebssystem.

Typ: Zeichenfolge

Version

Die AWS ParallelCluster-Version.

Typ: Zeichenfolge

Beispiel

Python

Anforderung

```
$ list_official_images()
```

200 Antwort

```
{
  'images': [
    {
      'ami_id': 'ami-015cfefb4e0d6306b2',
      'architecture': 'x86_64',
      'name': 'aws-parallelcluster-3.2.1-ubuntu-2004-lts-hvm-x86_64-202202261505 '
      '2022-02-26T15-08-34.759Z',
      'os': 'ubuntu2004',
      'version': '3.2.1'
    },
    ...
  ]
}
```

Cluster aktualisieren

Aktualisieren Sie den Cluster.

Themen

- [Erforderliche Syntax](#)
- [Anforderungstext](#)
- [Antwortsyntax](#)
- [Antworttext](#)
- [Beispiel](#)

Erforderliche Syntax

```
PUT /v3/clusters/{clusterName}
{
  "clusterConfiguration": "string",
  "dryrun": boolean,
  "forceUpdate": boolean,
  "region": "string",
  "suppressValidators": "string",
```

```
"validationFailureLevel": "string"  
}
```

Anforderungstext

Cluster-Konfiguration

Die Cluster-Konfiguration als YAML-Dokument.

Erforderlich: Ja

Clustername

Der Name des Clusters

Typ: Zeichenfolge

Erforderlich: Ja

Trockenlauf

Wenn eingestellt auf `true` führt nur eine Anforderungsvalidierung durch, ohne eine Ressource zu erstellen. Verwenden Sie diesen Parameter, um die Clusterkonfiguration und die Aktualisierungsanforderungen zu überprüfen. Der Standardwert ist `false`.

Typ: boolescher Wert

Required: No

ForceUpdate

Wenn gesetzt auf `true`, ignorieren Sie die Fehler bei der Überprüfung des Updates und erzwingen Sie das Update. Der Standardwert ist `false`.

Typ: boolescher Wert

Required: No

region

Das AWS-Region in dem sich der Cluster befindet.

Typ: Zeichenkette

Required: No

Unterdrücken Sie Validatoren

Identifiziert einen oder mehrere zu unterdrückende Konfigurationsvalidatoren.

Typ: Zeichenfolge

Format: (ALL | type:[A-Za-z0-9]+)

Required: No

Beispiel für gültige Werte:currentValue,requestedValue,message

validationFailureLevel

Die Mindestvalidierungsstufe, die dazu führen soll, dass das Update fehlschlägt.

Typ: Zeichenfolge

Zulässige Werte: INFO | WARNING | ERROR

Required: No

Antwortsyntax

```
{
  "cluster": {
    "clusterName": "string",
    "region": "string",
    "version": "string",
    "cloudformationStackArn": "string",
    "cloudformationStackStatus": "UPDATE_IN_PROGRESS",
    "clusterStatus": "UPDATE_IN_PROGRESS",
    "scheduler": {
      "type": "string",
      "metadata": {
        "name": "string",
        "version": "string"
      }
    }
  },
  "validationMessages": [
    {
      "id": "string",
```

```
    "type": "string",
    "level": "INFO",
    "message": "string"
  }
],
"changeSet": [
  {
    "parameter": "string",
    "currentValue": "string",
    "requestedValue": "string"
  }
]
}
```

Antworttext

Set ändern

Der Änderungssatz für das Cluster-Update.

Aktueller Wert

Der aktuelle Wert des zu aktualisierenden Parameters.

Typ: Zeichenfolge

Parameter

Der zu aktualisierende Parameter.

Typ: Zeichenfolge

Angeforderter Wert

Der angeforderte Wert für den zu aktualisierenden Parameter.

Typ: Zeichenfolge

Cluster

cloudformationStackArn

Der Amazon-Ressourcenname (ARN) des Hauptressourcennamens CloudFormation Stapel.

Typ: Zeichenfolge

cloudformationStackStatus

Der CloudFormation Status des Stapels.

Typ: Zeichenfolge

Zulässige Werte: CREATE_IN_PROGRESS | CREATE_FAILED | CREATE_COMPLETE
| ROLLBACK_IN_PROGRESS | ROLLBACK_FAILED | ROLLBACK_COMPLETE
| DELETE_IN_PROGRESS | DELETE_FAILED | DELETE_COMPLETE |
UPDATE_IN_PROGRESS | UPDATE_COMPLETE_CLEANUP_IN_PROGRESS
| UPDATE_COMPLETE | UPDATE_ROLLBACK_IN_PROGRESS |
UPDATE_ROLLBACK_FAILED | UPDATE_ROLLBACK_COMPLETE_CLEANUP_IN_PROGRESS
| UPDATE_ROLLBACK_COMPLETE

Clustername

Der Name des Clusters.

Typ: Zeichenfolge

Cluster-Status

Der Clusterstatus.

Typ: Zeichenfolge

Zulässige Werte: CREATE_IN_PROGRESS | CREATE_FAILED | CREATE_COMPLETE
| DELETE_IN_PROGRESS | DELETE_FAILED | DELETE_COMPLETE |
UPDATE_IN_PROGRESS | UPDATE_COMPLETE | UPDATE_FAILED

region

DerAWS-Regionin dem der Cluster erstellt wurde.

Typ: Zeichenfolge

Scheduler

Metadaten

Die Scheduler-Metadaten.

Name

Der Name des Schedulers.

Typ: Zeichenfolge

Version

Die Scheduler-Version.

Typ: Zeichenfolge

type

Der Scheduler-Typ.

Typ: Zeichenfolge

Version

AWS ParallelClusterVersion, die zur Erstellung des Clusters verwendet wird.

Typ: Zeichenfolge

Bestätigungsnachrichten

Eine Liste von Nachrichten mit einer Gültigkeitsstufe `validationFailureLevel`. Die Liste der Nachrichten wird während der Konfigurationsvalidierung gesammelt.

ID

Die ID des Validators.

Typ: Zeichenfolge

level

Die Validierungsebene.

Typ: Zeichenfolge

Zulässige Werte: INFO | WARNING | ERROR

Nachricht

Die Bestätigungsnachricht.

Typ: Zeichenfolge

type

Der Typ des Validators.

Typ: Zeichenfolge

Beispiel

Python

Anforderung

```
$ update_cluster(cluster_name_3x, path/config-file.yaml)
```

200 Antwort

```
{
  'change_set': [
    {
      'current_value': '10',
      'parameter':
'Scheduling.SlurmQueues[queue1].ComputeResources[t2micro].MaxCount',
      'requested_value': '15'
    }
  ],
  'cluster': {
    'cloudformation_stack_arn': 'arn:aws:cloudformation:us-
east-1:123456789012:stack/test-api-cluster/e0462730-50b5-11ed-99a3-0a5ddc4a34c7',
    'cloudformation_stack_status': 'UPDATE_IN_PROGRESS',
    'cluster_name': 'cluster-3x',
    'cluster_status': 'UPDATE_IN_PROGRESS',
    'region': 'us-east-1',
    'scheduler': {
      'type': 'slurm'
    },
    'version': '3.2.1'
  }
}
```

updateComputeFleet

Aktualisieren Sie den Status der Cluster-Rechenflotte.

Themen

- [Erforderliche Syntax](#)
- [Anforderungstext](#)
- [Antwortsyntax](#)
- [Antworttext](#)
- [Beispiel](#)

Erforderliche Syntax

```
PATCH /v3/clusters/{clusterName}/computefleet
{
  "status": "string",
  "region": "string"
}
```

Anforderungstext

Clustername

Der Name des Clusters

Typ: Zeichenfolge

Erforderlich: Ja

status

Der Status der Rechenflotte.

Typ: Zeichenfolge

Zulässige Werte: START_REQUESTED | STOP_REQUESTED | ENABLED | DISABLED

Erforderlich: Ja

region

Der AWS-Region in dem sich der Cluster befindet.

Typ: Zeichenkette

Erforderlich: Nein

Antwortsyntax

```
{
  "status": "START_REQUESTED",
  "lastStatusUpdatedTime": "2019-08-24T14:15:22Z"
}
```

Antworttext

status

Der Status der berechneten Flotte.

Typ: Zeichenfolge

Zulässige Werte: START_REQUESTED | STARTING | RUNNING | PROTECTED | STOP_REQUESTED | STOPPING | STOPPED | UNKNOWN | ENABLED | DISABLED

lastStatusUpdatedZeit

Der Zeitstempel, der die Uhrzeit der letzten Statusaktualisierung darstellt.

Typ: DateTime

Beispiel

Python

Anforderung

```
$ update_compute_fleet(cluster_name_3x, "START_REQUESTED")
```

200 Antwort

```
{
  'last_status_updated_time': datetime.datetime(2022, 3, 28, 22, 27, 14,
  tzinfo=tzlocal()),
  'status': 'START_REQUESTED'
```

}

AWS ParallelClusterPython-Bibliothek-API

Beginnend mitAWS ParallelClusterVersion 3.5.0, auf die Sie zugreifen könnenAWS ParallelClustermit demAWS ParallelClusterPython-Bibliothek. Sie können auf die zugreifenAWS ParallelClusterBibliothek in IhrempcClusterUmgebung oder aus einemAWS LambdaLaufzeit. Erfahren Sie, wie Sie auf die zugreifen könnenAWS ParallelClusterAPI mithilfe derAWS ParallelClusterPython-Bibliothek. DieAWS ParallelClusterDie Python-Bibliothek bietet die gleiche Funktionalität wieAWS ParallelClusterDie API liefert.

DieAWS ParallelClusterDie Operationen und Parameter der Python-Bibliothek entsprechen denen der API-Parameter, wenn sie konvertiert werden insnake_caseohne Großbuchstaben.

Themen

- [AWS ParallelClusterAutorisierung der Python-Bibliothek](#)
- [Installiere dasAWS ParallelClusterPython-Bibliothek](#)
- [Cluster-API-Operationen](#)
- [Berechne die Flotten-API-Operationen](#)
- [Cluster- und Stack-Log-Operationen](#)
- [Image-API-Operationen](#)
- [Image- und Stack-Log-Operationen](#)
- [Beispiel](#)
- [AWS Lambdafür dieAWS ParallelClusterPython-Bibliothek](#)

AWS ParallelClusterAutorisierung der Python-Bibliothek

Geben Sie Anmeldeinformationen an, indem Sie eine der Standardmethoden verwenden, die für boto3 gültig sind. Weitere Informationen finden Sie im[boto3-Dokumentation](#).

Installiere dasAWS ParallelClusterPython-Bibliothek

1. InstallierenpcClusterCLI Version 3.5.0 oder höher, indem Sie den Anweisungen unter folgen[Einrichten AWS ParallelCluster](#).

2. Importiere das `pcluster` Modul und beginnen Sie, die Bibliothek zu verwenden, wie im folgenden Beispiel gezeigt:

```
import pcluster.lib as pc
pc.create_cluster(cluster_name="mycluster", cluster_configuration="config.yaml")
```

Cluster-API-Operationen

Themen

- [list_clusters](#)
- [create_cluster](#)
- [delete_cluster](#)
- [describe_cluster](#)
- [update_cluster](#)

list_clusters

```
list_clusters(region, next_token, cluster_status)
```

Rufen Sie die Liste der vorhandenen Cluster ab.

Parameter:

region

Listet Cluster auf, die auf einem bestimmten Server bereitgestellt wurden AWS-Region.

next_token

Das Token, das für paginierte Anfragen verwendet werden soll.

cluster_status

Filtert nach Clusterstatus. Standardmäßig werden alle Cluster aufgelistet.

Zulässige Werte: CREATE_IN_PROGRESS | CREATE_FAILED | CREATE_COMPLETE |
DELETE_IN_PROGRESS | DELETE_FAILED | UPDATE_IN_PROGRESS | UPDATE_COMPLETE |
UPDATE_FAILED

create_cluster

```
create_cluster(cluster_name, cluster_configuration, region, suppress_validators,  
validation_failure_level, dry_run, rollback_on_failure, wait)
```

Erstellt einen Cluster in einer bestimmten Region.

Parameter:

cluster_name (Erforderlich)

Der Clustername.

cluster_configuration (Erforderlich)

Die Cluster-Konfiguration als Python-Datentyp.

region

Der ClusterAWS-Region.

suppress_validators

Identifiziert einen oder mehrere Validatoren für die Cluster-Konfiguration, die unterdrückt werden sollen.

Format: (ALL | type:[A-Za-z0-9]+)

validation_failure_level

Die Mindestvalidierungsstufe, die dazu führt, dass die Clustererstellung fehlschlägt. Der Standardwert ist ERROR.

Zulässige Werte: INFO | WARNING | ERROR.

dry_run

Führt die Anforderungsvalidierung durch, ohne Ressourcen zu erstellen. Sie können dies verwenden, um die Clusterkonfiguration zu validieren. Der Standardwert ist False.

rollback_on_failure

Wenn gesetzt auf True, AWS ParallelCluster initiiert bei Ausfällen automatisch ein Cluster-Stack-Rollback. Der Standardwert ist True.

wait

Wenn gesetzt auf `True`, AWS ParallelCluster wartet auf den Abschluss des Vorgangs. Der Standardwert ist `False`.

delete_cluster

```
delete_cluster(cluster_name, region, wait)
```

Löscht einen Cluster in einer bestimmten Region.

Parameter:

cluster_name (Erforderlich)

Der Clustername.

region

Der ClusterAWS-Region.

wait

Wenn eingestellt auf `True`, wartet auf den Abschluss des Vorgangs. Der Standardwert ist `False`.

describe_cluster

```
describe_cluster(cluster_name, region)
```

Ruft detaillierte Informationen über einen vorhandenen Cluster ab.

Parameter:

cluster_name (Erforderlich)

Der Clustername.

region

Der ClusterAWS-Region.

update_cluster

```
update_cluster(cluster_name, cluster_configuration, suppress_validators,  
validation_failure_level, region, force_update, dry_run, wait)
```

Aktualisieren Sie einen Cluster in einer bestimmten Region.

Parameter:

cluster_name (Erforderlich)

Der Clustername.

cluster_configuration (Erforderlich)

Die Cluster-Konfiguration als Python-Datentyp.

suppress_validators

Identifiziert einen oder mehrere Validatoren für die Cluster-Konfiguration, die unterdrückt werden sollen.

Format: (ALL | type:[A-Za-z0-9]+)

validation_failure_level

Die Mindestvalidierungsstufe, die dazu führt, dass das Cluster-Update fehlschlägt. Der Standardwert ist ERROR.

Zulässige Werte: INFO | WARNING | ERROR

region

Der ClusterAWS-Region.

dry_run

Führt die Anforderungsvalidierung durch, ohne Ressourcen zu erstellen oder zu aktualisieren. Sie können dies verwenden, um die Clusterkonfiguration zu validieren. Der Standardwert ist `False`.

force_update

Wenn gesetzt auf `True`, erzwingt die Aktualisierung, indem die Fehler bei der Aktualisierungsüberprüfung ignoriert werden. Der Standardwert ist `False`.

wait

Wenn gesetzt auf `True`, wartet auf den Abschluss des Vorgangs. Der Standardwert ist `False`.

Berechne die Flotten-API-Operationen

Themen

- [describe_compute_fleet](#)
- [update_compute_fleet](#)
- [delete_cluster_instances](#)
- [describe_cluster_instances](#)

describe_compute_fleet

```
describe_compute_fleet(cluster_name, region)
```

Beschreiben Sie den Status einer Cluster-Rechenflotte für einen bestimmten Cluster.

Parameter:

cluster_name (Erforderlich)

Der Clustername.

region

Beschreibt den Status der Rechenflotte für einen Cluster, der in einem bestimmten Cluster bereitgestellt wurdeAWS-Region.

update_compute_fleet

```
update_compute_fleet(cluster_name, status, region)
```

Aktualisieren Sie den Status der Cluster-Rechenflotte.

Parameter:

cluster_name (Erforderlich)

Der Clustername.

status (Erforderlich)

Der Status, auf den aktualisiert werden soll.

Zulässige Werte: START_REQUESTED | STOP_REQUESTED | ENABLED | DISABLED

region

Der ClusterAWS-Region.

delete_cluster_instances

```
delete_cluster_instances(cluster_name, region, force)
```

Löscht einen Cluster in einer bestimmten Region.

Parameter:

cluster_name (Erforderlich)

Der Clustername.

region

Der ClusterAWS-Region.

force

Wenn eingestellt auf `True`, erzwingt das Löschen, wenn der Cluster mit dem angegebenen `cluster_name` wurde nicht gefunden. Der Standardwert ist `False`.

describe_cluster_instances

```
describe_cluster_instances(cluster_name, region, next_token, node_type, queue_name)
```

Beschreiben Sie die Instanzen eines Clusters.

Parameter:

cluster_name (Erforderlich)

Der Clustername.

region

Der ClusterAWS-Region.

next_token

Das Token, das für paginierte Anfragen verwendet werden soll.

node_type

Filtert die Instanzen nach `node_type`.

Zulässige Werte: `HeadNode` | `ComputeNode`

queue_name

Filtert die Instanzen nach dem Namen der Warteschlange.

Cluster- und Stack-Log-Operationen

Themen

- [list_cluster_log_streams](#)
- [get_cluster_log_events](#)
- [get_cluster_stack_events](#)

list_cluster_log_streams

```
list_cluster_log_streams(cluster_name, region, filters, next_token)
```

Listet Log-Streams für einen bestimmten Cluster auf.

Parameter:

cluster_name (Erforderlich)

Der Clustername.

region

Der ClusterAWS-Region.

filters

Filtert die Cluster-Log-Streams.

Format: `'Name=a,Values=1 Name=b,Values=2,3'`

Zulässige Filter:

`code-dns-name`

Die Kurzform des privaten DNS-Namens der Instanz; zum Beispiel `ip-10-0-0-101`.

Knotentyp

Der Knotentyp.

Zulässige Werte: `HeadNode`

next_token

Das Token, das für paginierte Anfragen verwendet werden soll.

get_cluster_log_events

```
get_cluster_log_events(cluster_name, log_stream_name, region, next_token,
start_from_head, limit, start_time, end_time)
```

Ruft Protokollereignisse für einen bestimmten Cluster und Protokollstream ab.

Parameter:

cluster_name (Erforderlich)

Der Clustername.

log_stream_name (Erforderlich)

Der Name des Log-Streams.

region

Der ClusterAWS-Region.

next_token

Das Token, das für paginierte Anfragen verwendet werden soll.

start_from_head

Wenn gesetzt auf `True`, AWS ParallelCluster gibt zuerst die frühesten Protokollereignisse zurück. Wenn gesetzt auf `False`, gibt zuerst die neuesten Protokollereignisse zurück. Der Standardwert ist `False`.

limit

Die maximale Anzahl von zurückgegebenen Protokollereignissen. Wenn Sie keinen Wert angeben, entspricht das Maximum der Anzahl von Protokollen, die in eine Antwortgröße von 1 MB passen, also bis zu 10.000 Protokollereignisse.

start_time

Der Beginn des Zeitbereichs für Protokollereignisse, ausgedrückt im Format ISO 8601, z. B. '2021-01-01T20:00:00Z'. Ereignisse mit einem Zeitstempel, der diesem Zeitpunkt entspricht oder später liegt, sind enthalten.

end_time

Das Ende des Zeitbereichs für Protokollereignisse, ausgedrückt im ISO 8601-Format; zum Beispiel '2021-01-01T20:00:00Z'. Ereignisse mit einem Zeitstempel, der diesem Zeitpunkt entspricht oder später liegt, sind nicht enthalten.

get_cluster_stack_events

```
get_cluster_stack_events(cluster_name, region, next_token)
```

Ruft Stack-Ereignisse für einen bestimmten Cluster ab.

Parameter:

cluster_name (Erforderlich)

Der Clustername.

region

Der ClusterAWS-Region.

next_token

Das Token, das für paginierte Anfragen verwendet werden soll.

Image-API-Operationen

Themen

- [list_images](#)
- [build_image](#)
- [delete_image](#)
- [describe_image](#)

list_images

```
list_images(image_status, region, next_token)
```

Rufen Sie die Liste der vorhandenen Bilder ab.

Parameter:

image_status (Erforderlich)

Filtert nach Bildstatus.

Zulässige Werte: AVAILABLE | PENDING | FAILED

region

Listet Bilder auf, die in einem bestimmten Objekt erstellt wurdenAWS-Region.

next_token

Token zur Verwendung für paginierte Anfragen.

build_image

```
build_image(image_configuration, image_id, suppress_validators,  
validation_failure_level, dry_run, rollback_on_failure, region)
```

Erstellen Sie ein benutzerdefiniertesAWS ParallelClusterBild in einer bestimmten Region.

Parameter:

image_configuration (Erforderlich)

Die Bildkonfiguration als Python-Daten.

image_id (Erforderlich)

Die Bild-ID.

suppress_validators

Identifiziert einen oder mehrere Validatoren für die Image-Konfiguration, die unterdrückt werden sollen.

Format: (ALL | type:[A-Za-z0-9]+)

validation_failure_level

Die Mindestvalidierungsstufe, die dazu führt, dass die Image-Erstellung fehlschlägt. Der Standardwert ist ERROR.

Zulässige Werte: INFO | WARNING | ERROR

dry_run

Wenn gesetzt auf `True`, AWS ParallelCluster führt die Anforderungvalidierung durch, ohne Ressourcen zu erstellen. Sie können dies verwenden, um die Image-Konfiguration zu validieren. Der Standardwert ist `False`.

rollback_on_failure

Wenn gesetzt auf `True`, AWS ParallelCluster initiiert bei Fehlern automatisch ein Rollback des Image-Stacks. Der Standardwert ist `False`.

region

Das AWS-Region-Abbild.

delete_image

```
delete_image(image_id, region, force)
```

Löscht ein Bild in einer bestimmten Region.

Parameter:

image_id (Erforderlich)

Die Bild-ID.

region

Das AWS-Region-Abbild.

force

Wenn eingestellt auf `True`, AWS ParallelCluster erzwingt das Löschen, wenn Instances das AMI verwenden oder wenn das AMI gemeinsam genutzt wird. Der Standardwert ist `False`.

describe_image

```
describe_image(image_id, region)
```

Rufen Sie detaillierte Informationen zu einem vorhandenen Image ab.

Parameter:

image_id (Erforderlich)

Die Image-ID.

region

Das AWS-Region-Abbild.

Image- und Stack-Log-Operationen

Themen

- [list_image_log_streams](#)
- [get_image_log_events](#)
- [get_image_stack_events](#)
- [list_official_images](#)

list_image_log_streams

```
list_image_log_streams(image_id, region, next_token)
```

Listet Log-Streams für ein Bild auf.

Parameter:

image_id (Erforderlich)

Die Bild-ID.

region

Das AWS-Region-Abbild.

next_token

Das Token, das für paginierte Anfragen verwendet werden soll.

get_image_log_events

```
get_image_log_events(image_id, log_stream_name, region, next_token, start_from_head,
                    limit, start_time, end_time)
```

Ruft Protokollereignisse für ein bestimmtes Bild und einen bestimmten Protokollstream ab.

Parameter:

image_id (Erforderlich)

Die Bild-ID.

log_stream_name (Erforderlich)

Der Name des Protokollstreams.

region

Das AWS-Region-Abbild.

next_token

Das Token, das für paginierte Anfragen verwendet werden soll.

start_from_head

Wenn gesetzt auf `True`, AWS ParallelCluster gibt zuerst die frühesten Protokollereignisse zurück. Wenn gesetzt auf `False`, gibt zuerst die neuesten Protokollereignisse zurück. Der Standardwert ist `False`.

limit

Die maximale Anzahl von zurückgegebenen Protokollereignissen. Wenn Sie keinen Wert angeben, entspricht das Maximum der Anzahl von Protokollen, die in eine Antwortgröße von 1 MB passen, also bis zu 10.000 Protokollereignisse.

start_time

Der Beginn des Zeitbereichs für Protokollereignisse, ausgedrückt im ISO 8601-Format; zum Beispiel '2021-01-01T20:00:00Z'. Ereignisse mit einem Zeitstempel, der diesem Zeitpunkt entspricht oder später liegt, sind enthalten.

end_time

Das Ende des Zeitbereichs für Protokollereignisse, ausgedrückt im ISO 8601-Format; zum Beispiel '2021-01-01T20:00:00Z'. Ereignisse mit einem Zeitstempel, der diesem Zeitpunkt entspricht oder später liegt, sind nicht enthalten.

get_image_stack_events

```
get_image_stack_events(image_id, region, next_token)
```

Ruft Stack-Ereignisse für ein bestimmtes Bild ab.

Parameter:

image_id (Erforderlich)

Die Bild-ID.

region

Das AWS-Region-Abbild.

next_token

Das Token, das für paginierte Anfragen verwendet werden soll.

list_official_images

```
list_official_images(region, os, architecture)
```

Rufen Sie die Liste der Beamten abAWS ParallelClusterBilder.

Parameter:

region

Das AWS-Region-Abbild.

os

Filtert nach Betriebssystemverteilung. Die Standardeinstellung ist keine Filterung.

architecture

Filtert nach Architektur. Die Standardeinstellung ist keine Filterung.

Beispiel

Themen

- [Erstellen eines -Clusters](#)

Erstellen eines -Clusters

Wenn Sie das folgende Beispielskript ausführen und die angegebenen Eingaben in Ihrer Umgebung gespeichert haben, erstellen Sie einen Cluster. Die Cluster-Konfiguration wird als Python-Datentyp basierend auf dem erstellten [Dokumentation zur Cluster-Konfiguration](#).

```
import os
import pprint
import pcluster.lib as pc
pp = pprint.PrettyPrinter()

HEAD_NODE_SUBNET = os.environ["HEAD_NODE_SUBNET"]
COMPUTE_NODE_SUBNET = os.environ["HEAD_NODE_SUBNET"]
KEY_NAME = os.environ["KEY_NAME"]
CONFIG = {'Image': {'Os': 'alinux2'},
          'HeadNode': {'InstanceType': 't2.large',
                       'Networking': {'SubnetId': HEAD_NODE_SUBNET},
                       'Ssh': {'KeyName': KEY_NAME}},
          'Scheduling': {'Scheduler': 'slurm',
                         'SlurmQueues':
                         [{'Name': 'queue0',
                           'ComputeResources':
```

```
[{'Name': 'queue0-i0', 'InstanceType': 't2.micro',
  'MinCount': 0, 'MaxCount': 10}],
'Networking': {'SubnetIds': [COMPUTE_NODE_SUBNET]}]]}]}
```

```
pp.pprint(pc.create_cluster(cluster_name="mycluster", cluster_configuration=CONFIG))
```

Ausgabe:

```
{'cluster': {'cloudformationStackArn': 'arn:aws:cloudformation:us-
east-2:123456789012:stack/mycluster/00000000-aaaa-1111-999-000000000000',
  'cloudformationStackStatus': 'CREATE_IN_PROGRESS',
  'clusterName': 'mycluster',
  'clusterStatus': 'CREATE_IN_PROGRESS',
  'region': 'us-east-2',
  'scheduler': {'type': 'slurm'},
  'version': '3.7.0'}}
```

AWS Lambda für die AWS ParallelCluster Python-Bibliothek

Sie können eine Lambda-Schicht und eine Laufzeit bereitstellen, um auf die zuzugreifen AWS ParallelCluster Python-Bibliothek. Wir hosten AWS ParallelCluster ZIP-Dateien, die Sie verwenden können, indem Sie den Link zur Zip-Datei eingeben, wie in den folgenden Schritten beschrieben. Lambda verwendet die ZIP-Dateien, um die Laufzeitumgebung so vorzubereiten, dass sie den Zugriff auf die Python-Bibliothek unterstützt. Die AWS ParallelCluster Python-Bibliothek wurde hinzugefügt mit AWS ParallelCluster Version 3.5.0. Sie können die Bibliothek nur für Versionen 3.5.0 und höher verwenden.

Die URL der gehosteten Zip-Datei hat das folgende Format: `s3://aws-region-id-aws-parallelcluster/parallelcluster/3.7.0/layers/aws-parallelcluster/lambda-layer.zip`

Beginnen Sie mit dem Zugriff auf AWS ParallelCluster Python-Bibliothek mit AWS Lambda

Erstellen Sie eine Lambda-Ebene

1. Loggen Sie sich in das AWS Management Console und navigieren Sie zur AWS Lambda-Konsole.
2. Wählen Sie im Navigationsbereich Ebenen, dann Ebene erstellen.

3. Geben Sie einen Namen für Ihre Ebene ein und wählen Sie Laden Sie eine Datei von Amazon S3 hoch.
4. Geben Sie die URL zur Zip-Datei ein: `s3://aws-region-id-aws-parallelcluster/parallelcluster/3.7.0/layers/aws-parallelcluster/lambda-layer.zip`.
5. Für Kompatible Architekturen, wählen Sie die `x86_64` Architektur.
6. Für Kompatible Laufzeiten, wählen Sie die `Python 3.9` Laufzeit.
7. Wählen Sie Erstellen aus.

Verwenden Sie Ihre Lambda-Schicht

1. Wählen Sie im Navigationsbereich der Lambda-Konsole Funktionen, dann Funktion erstellen.
2. Geben Sie einen Namen für die Funktion ein.
3. Für Laufzeit, wählen Sie `Python 3.9` Laufzeit.
4. Für Architektur, wählen Sie `x86_64` Architektur.
5. Wählen Sie Funktion erstellen aus.
6. Nachdem die Funktion erstellt wurde, wählen Sie Ebenen und wählen Fügen Sie eine Ebene hinzu.
7. Wählen Benutzerdefinierte Ebenen und wählen Sie die Ebene aus, die Sie in den vorherigen Schritten erstellt haben.
8. Wählen Sie die Layer-Version aus.
9. Wählen Sie Add (Hinzufügen) aus.
10. Ihr Lambda benötigt Berechtigungen zur Verwaltung von Clustern, die mit erstellt wurden [AWS ParallelCluster](#). Erstellen Sie eine Lambda-Rolle mit den unter aufgeführten Berechtigungen [AWS ParallelCluster pcluster Grundlegende Benutzerrichtlinie](#).

Sie können jetzt darauf zugreifen [AWS ParallelCluster](#) aus der Python-Bibliothek, wie unter beschrieben [AWS ParallelCluster Python-Bibliothek-API](#).

Funktionsweise von AWS ParallelCluster

AWS ParallelCluster wurde nicht nur zur Verwaltung von Clustern entwickelt, sondern auch als Referenz für die Verwendung von AWS Diensten zum Aufbau Ihrer HPC-Umgebung.

Themen

- [AWS ParallelCluster Prozesse](#)
- [AWS Dienste verwendet von AWS ParallelCluster](#)
- [AWS ParallelCluster Interne Verzeichnisse](#)

AWS ParallelCluster Prozesse

Dieser Abschnitt bezieht sich auf Cluster, die mit bereitgestellt werden Slurm. Bei Verwendung mit diesem Scheduler AWS ParallelCluster verwaltet er die Bereitstellung und Entfernung von Rechenknoten durch Interaktion mit dem zugrundeliegenden Job-Scheduler.

Bei HPC-Clustern, die auf basieren AWS Batch, AWS ParallelCluster hängt von den Funktionen ab, die von der AWS Batch für die Compute-Knotenverwaltung bereitgestellt werden.

clustermgtd

Die folgenden Aufgaben werden vom Cluster-Management-Daemon ausgeführt.

- Säuberung inaktiver Partitionen
- Verwaltung von Slurm-Reservierungen und Knoten, die mit Capacity Blocks verknüpft sind (siehe nächster Abschnitt)
- Statisches Kapazitätsmanagement: Stellen Sie sicher, dass die statische Kapazität immer verfügbar und funktionsfähig ist
- Synchronisieren Sie den Scheduler mit Amazon EC2.
- Bereinigung verwaister Instances
- Stellen Sie den Status des Scheduler-Knotens bei einer Amazon EC2 EC2-Terminierung wieder her, die außerhalb des Suspend-Workflows erfolgt
- Verwaltung fehlerhafter Amazon EC2 EC2-Instances (fehlgeschlagene Amazon EC2 EC2-Zustandsprüfungen)

- Verwaltung von geplanten Wartungsereignissen
- Verwaltung fehlerhafter Scheduler-Knoten (fehlgeschlagene Zustandsprüfungen im Scheduler)

Verwaltung von Slurm-Reservierungen und Knoten, die Kapazitätsblöcken zugeordnet sind

ParallelCluster unterstützt On-Demand-Kapazitätsreservierungen (ODCR) und Kapazitätsblöcke für Machine Learning (CB). Im Gegensatz zu ODCR kann CB eine future Startzeit haben und ist zeitgebunden.

Clustermgtd sucht in einer Schleife nach fehlerhaften Knoten und beendet alle ausgefallenen EC2-Instances. Wenn es sich um statische Knoten handelt, werden sie durch neue Instances ersetzt.

ParallelCluster verwaltet statische Knoten, die Kapazitätsblöcken zugeordnet sind, unterschiedlich. AWS ParallelCluster erstellt einen Cluster, auch wenn der CB noch nicht aktiv ist, und Instances werden automatisch gestartet, sobald der CB aktiv ist.

Die Slurm-Knoten, die den Rechenressourcen entsprechen, die CBs zugeordnet sind, die noch nicht aktiv sind, werden solange gewartet, bis die CB-Startzeit erreicht ist. Slurm-Knoten verbleiben in einem Reservierungs-/Wartungsstatus, der dem Slurm-Admin-Benutzer zugeordnet ist, was bedeutet, dass sie Jobs annehmen können, aber Jobs bleiben so lange ausstehend, bis die Slurm-Reservierung entfernt wird.

Clustermgtd erstellt/löscht automatisch Slurm-Reservierungen und setzt die zugehörigen CB-Knoten je nach CB-Status in den Wartungsmodus. Wenn CB aktiv ist, wird die Slurm-Reservierung entfernt, die Knoten werden gestartet und stehen dann für die ausstehenden Jobs oder für die Einreichung neuer Jobs zur Verfügung.

Wenn die CB-Endzeit erreicht ist, werden die Knoten wieder in den Reservierungs-/Wartungszustand versetzt. Es liegt an den Benutzern, die Jobs erneut zu senden/in eine neue Warteschleife/ Rechenressource zu stellen, wenn CB nicht mehr aktiv ist und Instances beendet werden.

clusterstatusmgtd

Der Cluster-Statusmanagement-Daemon verwaltet die Aktualisierung des Compute-Flottenstatus. Jede Minute ruft es den in einer DynamoDB-Tabelle gespeicherten Flottenstatus ab und verwaltet alle STOP/START-Anfragen.

computemgtd

Compute Management Daemon (computemgtd) -Prozesse werden auf jedem der Cluster-Rechenknoten ausgeführt. Alle fünf (5) Minuten bestätigt der Compute Management Daemon, dass der Hauptknoten erreichbar ist und fehlerfrei ist. Wenn fünf (5) Minuten vergehen, in denen der Hauptknoten nicht erreicht werden kann oder nicht fehlerfrei ist, wird der Rechenknoten heruntergefahren.

AWSDienste verwendet von AWS ParallelCluster

Die folgenden Amazon Web Services (AWS) -Dienste werden von verwendetAWS ParallelCluster.

Themen

- [Amazon API Gateway](#)
- [AWS Batch](#)
- [AWS CloudFormation](#)
- [Amazon CloudWatch](#)
- [CloudWatch Amazon-Veranstaltungen](#)
- [CloudWatch Amazon-Protokolle](#)
- [AWS CodeBuild](#)
- [Amazon DynamoDB](#)
- [Amazon Elastic Block Store](#)
- [Amazon Elastic Compute Cloud](#)
- [Amazon Elastic Container Registry](#)
- [Amazon EFS](#)
- [Amazon FSx für Lustre](#)
- [Amazon FSx für ONTAP NetApp](#)
- [Amazon FSx for OpenZFS](#)
- [AWS Identity and Access Management](#)
- [AWS Lambda](#)
- [Amazon RDS](#)
- [Amazon Route 53](#)
- [Amazon Simple Notification Service](#)

- [Amazon Simple Storage Service](#)
- [Amazon VPC](#)
- [Elastic Fabric Adapter](#)
- [EC2 Image Builder](#)
- [NICE DCV](#)

Amazon API Gateway

Amazon API Gateway ist ein AWS Service für die Erstellung, Veröffentlichung, Wartung, Überwachung und Sicherung von REST, HTTP und WebSocket APIs in jeder Größenordnung

AWS ParallelCluster verwendet API Gateway, um die AWS ParallelCluster API zu hosten.

Weitere Informationen AWS Batch dazu finden Sie unter <https://aws.amazon.com/api-gateway/> und <https://docs.aws.amazon.com/apigateway/>.

AWS Batch

AWS Batch ist ein AWS verwalteter Job Scheduler-Dienst. Er stellt dynamisch die optimale Menge und Art von Rechenressourcen (z. B. CPU- oder speicheroptimierte Instances) in Clustern bereit. Diese Ressourcen werden auf der Grundlage der spezifischen Anforderungen Ihrer Batch-Jobs bereitgestellt, einschließlich der Volumenansforderungen. Mit AWS Batch müssen Sie keine zusätzliche Batch-Computing-Software oder Servercluster installieren oder verwalten, um Ihre Jobs effektiv auszuführen.

AWS Batch wird nur mit AWS Batch-Clustern verwendet.

Weitere Informationen zu AWS Batch finden Sie unter <https://aws.amazon.com/batch/> und <https://docs.aws.amazon.com/batch/>.

AWS CloudFormation

AWS CloudFormation ist ein infrastructure-as-code Dienst, der eine gemeinsame Sprache für die Modellierung AWS und Bereitstellung von Anwendungsressourcen von Drittanbietern in Ihrer Cloud-Umgebung bereitstellt. Es ist der Hauptdienst, der verwendet wird AWS ParallelCluster. Jeder Cluster AWS ParallelCluster wird als Stack dargestellt, und alle Ressourcen, die von jedem Cluster benötigt werden, sind in der AWS ParallelCluster AWS CloudFormation Vorlage definiert. In den meisten Fällen entsprechen AWS ParallelCluster CLI-Befehle direkt AWS CloudFormation Stack-

Befehlen wie Create, Update und Delete. Instances, die innerhalb eines Clusters gestartet werden, senden HTTPS-Aufrufe an den AWS CloudFormation Endpunkt, auf AWS-Region dem der Cluster gestartet wurde.

Weitere Informationen zu AWS CloudFormation finden Sie unter <https://aws.amazon.com/cloudformation/> und <https://docs.aws.amazon.com/cloudformation/>.

Amazon CloudWatch

Amazon CloudWatch (CloudWatch) ist ein Überwachungs- und Beobachtbarkeitservice, der Ihnen Daten und umsetzbare Erkenntnisse liefert. Diese Erkenntnisse können verwendet werden, um Ihre Anwendungen zu überwachen, auf Leistungsänderungen und Serviceausnahmen zu reagieren und die Ressourcennutzung zu optimieren. In AWS ParallelCluster CloudWatch wird für ein Dashboard verwendet, um die Schritte zur Erstellung von Docker-Images und die Ausgabe der AWS Batch Jobs zu überwachen und zu protokollieren.

Vor AWS ParallelCluster Version 2.10.0 CloudWatch wurde es nur mit AWS Batch Clustern verwendet.

Weitere Informationen zu finden Sie CloudWatch unter <https://aws.amazon.com/cloudwatch/> und <https://docs.aws.amazon.com/cloudwatch/>.

CloudWatch Amazon-Veranstaltungen

Amazon CloudWatch Events (CloudWatch Events) liefert nahezu in Echtzeit einen Stream von Systemereignissen, die Änderungen an den Ressourcen von Amazon Web Services (AWS) beschreiben. Mit einfachen Regeln, die sich schnell einrichten lassen, können Sie Ereignisse ordnen und sie an eine oder mehrere Zielfunktionen oder Streams weiterleiten. In AWS ParallelCluster wird CloudWatch Events für AWS Batch Jobs verwendet.

Weitere Informationen zu CloudWatch Ereignissen finden Sie unter <https://docs.aws.amazon.com/eventbridge/latest/userguide/eb-cwe-now-eb>.

CloudWatch Amazon-Protokolle

Amazon CloudWatch Logs (CloudWatch Logs) ist eine der Kernfunktionen von Amazon CloudWatch. Sie können damit die Protokolldateien für viele der von verwendeten Komponenten überwachen, speichern, anzeigen und durchsuchenAWS ParallelCluster.

Vor AWS ParallelCluster Version 2.6.0 wurde CloudWatch Logs nur mit AWS Batch Clustern verwendet.

Weitere Informationen finden Sie unter [Integration mit Amazon CloudWatch Logs](#).

AWS CodeBuild

AWS CodeBuild(CodeBuild) ist ein AWS verwalteter Dienst für kontinuierliche Integration, der Quellcode einhält, Tests durchführt und Softwarepakete erstellt, die sofort bereitgestellt werden können. In CodeBuild wird verwendetAWS ParallelCluster, um automatisch und transparent Docker-Images zu erstellen, wenn Cluster erstellt werden.

CodeBuild wird nur mit AWS Batch Clustern verwendet.

Weitere Informationen zu CodeBuild finden Sie unter <https://aws.amazon.com/codebuild/> und <https://docs.aws.amazon.com/codebuild/>.

Amazon DynamoDB

Amazon DynamoDB (DynamoDB) ist ein schneller und flexibler NoSQL-Datenbankservice. Er wird verwendet, um die minimalen Statusinformationen des Clusters zu speichern. Der Hauptknoten verfolgt bereitgestellte Instanzen in einer DynamoDB-Tabelle.

DynamoDB wird nicht mit AWS Batch Clustern verwendet.

[Weitere Informationen zu DynamoDB finden Sie unter https://aws.amazon.com/dynamodb/](https://aws.amazon.com/dynamodb/) und <https://docs.aws.amazon.com/dynamodb/>.

Amazon Elastic Block Store

Amazon Elastic Block Store (Amazon EBS) ist ein leistungsstarker Blockspeicherservice, der persistenten Speicher für gemeinsam genutzte Volumes bereitstellt. Alle Amazon EBS-Einstellungen können durch die Konfiguration übergeben werden. Amazon EBS-Volumes können entweder leer oder aus einem vorhandenen Amazon EBS-Snapshot initialisiert werden.

Weitere Informationen zu Amazon EBS finden Sie unter <https://aws.amazon.com/ebs/> und <https://docs.aws.amazon.com/ebs/>.

Amazon Elastic Compute Cloud

Amazon Elastic Compute Cloud (Amazon EC2) bietet die Rechenkapazität fürAWS ParallelCluster. Die Head- und Compute-Knoten sind Amazon EC2 EC2-Instances. Es kann jeder Instance-Typ ausgewählt werden, der HVM unterstützt. Bei den Kopf- und Rechenknoten kann es sich um unterschiedliche Instance-Typen handeln. Wenn mehrere Warteschlangen verwendet werden,

können außerdem einige oder alle Rechenknoten auch als Spot-Instance gestartet werden. Auf den Instances vorhandene Instance-Speicher-Volumes werden als Striping-LVM-Volume gemountet.

Weitere Informationen zu Amazon EC2 finden Sie unter <https://aws.amazon.com/ec2/> und <https://docs.aws.amazon.com/ec2/>.

Amazon Elastic Container Registry

Amazon Elastic Container Registry (Amazon ECR) ist eine vollständig verwaltete Docker-Container-Registry, die das Speichern, Verwalten und Bereitstellen von Docker-Container-Images vereinfacht. In AWS ParallelCluster speichert Amazon ECR die Docker-Images, die bei der Erstellung von Clustern erstellt werden. Die Docker-Images werden dann von AWS Batch zum Ausführen der Container für die übermittelten Aufgaben verwendet.

Amazon ECR wird nur mit AWS Batch Clustern verwendet.

Weitere Informationen finden Sie unter <https://aws.amazon.com/ecr/> und <https://docs.aws.amazon.com/ecr/>.

Amazon EFS

Amazon Elastic File System (Amazon EFS) bietet ein einfaches, skalierbares und vollständig verwaltetes elastisches NFS-Dateisystem zur Verwendung mit AWS Cloud Services und lokalen Ressourcen. Amazon EFS wird verwendet, wenn die angegebenen [EfsSettings](#) sind. Support für Amazon EFS wurde in AWS ParallelCluster Version 2.1.0 hinzugefügt.

Weitere Informationen zu Amazon EFS finden Sie unter <https://aws.amazon.com/efs/> und <https://docs.aws.amazon.com/efs/>.

Amazon FSx für Lustre

FSx for Lustre bietet ein Hochleistungsdateisystem, das das Open-Source-Dateisystem Lustre verwendet. FSx for Lustre wird verwendet, wenn sie angegeben [FsxLustreSettings-Eigenschaften](#) sind. Support für FSx for Lustre wurde in AWS ParallelCluster Version 2.2.1 hinzugefügt.

[Weitere Informationen zu FSx for Lustre finden Sie unter https://aws.amazon.com/fsx/lustre/ und https://docs.aws.amazon.com/fsx/.](https://aws.amazon.com/fsx/lustre/)

Amazon FSx für ONTAP NetApp

FSx for ONTAP bietet ein vollständig verwaltetes Shared-Storage-System, das auf dem NetApp beliebten ONTAP-Dateisystem basiert. FSx for ONTAP wird verwendet, wenn sie angegeben [FsxOntapSettings-Eigenschaften](#) sind. Support für FSx for ONTAP wurde in AWS ParallelCluster Version 3.2.0 hinzugefügt.

Weitere Informationen zu FSx for ONTAP finden Sie unter <https://aws.amazon.com/fsx/netapp-ontap/> und <https://docs.aws.amazon.com/fsx/>.

Amazon FSx for OpenZFS

FSx for OpenZFS bietet ein vollständig verwaltetes Shared Storage-System, das auf dem beliebten OpenZFS-Dateisystem basiert. fsX for OpenZFS wird verwendet, wenn sie angegeben sind. [FsxOpenZfsSettings-Eigenschaften](#) Support für FSx for OpenZFS wurde in AWS ParallelCluster Version 3.2.0 hinzugefügt.

Weitere Informationen zu FSx for OpenZFS finden Sie unter <https://aws.amazon.com/fsx/openzfs/> und <https://docs.aws.amazon.com/fsx/>.

AWS Identity and Access Management

AWS Identity and Access Management(IAM) wird innerhalb verwendetAWS ParallelCluster, um eine IAM-Rolle mit den geringsten Rechten für Amazon EC2 für die Instance bereitzustellen, die für jeden einzelnen Cluster spezifisch ist. AWS ParallelClusterInstances erhalten nur Zugriff auf die spezifischen API-Aufrufe, die für die Bereitstellung und Verwaltung des Clusters erforderlich sind.

Bei AWS Batch Clustern werden bei der Clustererstellung auch IAM-Rollen für die Komponenten erstellt, die am Prozess der Docker-Image-Erstellung beteiligt sind. Zu diesen Komponenten gehören die Lambda-Funktionen, mit denen Docker-Images zum Amazon ECR-Repository hinzugefügt und daraus gelöscht werden können. Sie enthalten auch die Funktionen, mit denen der Amazon S3 S3-Bucket gelöscht werden kann, der für den Cluster und das CodeBuild Projekt erstellt wurde. Es gibt auch Rollen für AWS Batch-Ressourcen, -Instances und -Aufgaben.

Weitere Informationen zu IAM finden Sie unter <https://aws.amazon.com/iam/> und <https://docs.aws.amazon.com/iam/>.

AWS Lambda

AWS Lambda(Lambda) führt die Funktionen aus, die die Erstellung von Docker-Images orchestrieren. Lambda verwaltet auch die Bereinigung von benutzerdefinierten Cluster-Ressourcen wie Docker-Images, die im Amazon ECR-Repository und in Amazon S3 gespeichert sind.

Weitere Informationen zu Lambda finden Sie unter <https://aws.amazon.com/lambda/> und <https://docs.aws.amazon.com/lambda/>.

Amazon RDS

Amazon Relational Database Service(Amazon RDS) ist ein Webservice, der die Einrichtung, den Betrieb und die Skalierung einer relationalen Datenbank in der AWS Cloud erleichtert.

AWS ParallelClusterverwendet Amazon RDS für AWS Batch undSlurm.

Weitere Informationen zu Amazon RDS finden Sie unter <https://aws.amazon.com/rds/> und <https://docs.aws.amazon.com/rds/>.

Amazon Route 53

Amazon Route 53 (Route 53) wird verwendet, um Hostzonen mit Hostnamen und vollqualifizierten Domainnamen für jeden der Rechenknoten zu erstellen.

Weitere Informationen zu Route 53 finden Sie unter <https://aws.amazon.com/route53/> und <https://docs.aws.amazon.com/route53/>.

Amazon Simple Notification Service

(Amazon SNS) ist ein verwalteter Service, der die Nachrichtenzustellung von Verlagen an Abonnenten (auch bekannt als Produzenten und Verbraucher) ermöglicht.

AWS ParallelClusterverwendet Amazon SNS für API-Hosting.

Weitere Informationen zu Amazon SNS finden Sie unter <https://aws.amazon.com/sns/> und <https://docs.aws.amazon.com/sns/>.

Amazon Simple Storage Service

Amazon Simple Storage Service (Amazon S3) speichert AWS ParallelCluster Vorlagen, die sich jeweils darin befindenAWS-Region. AWS ParallelClusterkann so konfiguriert werden, dass CLI/SDK-Tools Amazon S3 verwenden können.

AWS ParallelCluster erstellt auch einen Amazon S3 S3-Bucket in Ihrem AWS-Konto, um Ressourcen zu speichern, die von Ihren Clustern verwendet werden, wie z. B. die Cluster-Konfigurationsdatei. AWS ParallelCluster verwaltet in jedem, in dem Sie Cluster erstellen AWS-Region, einen Amazon S3 S3-Bucket.

Wenn Sie AWS Batch Cluster verwenden, wird ein Amazon S3 S3-Bucket in Ihrem Konto zum Speichern verwandter Daten verwendet. Der Bucket speichert beispielsweise Artefakte, die entstehen, wenn ein Docker-Image und Skripts aus eingereichten Jobs erstellt werden.

Weitere Informationen finden Sie unter <https://aws.amazon.com/s3/> und <https://docs.aws.amazon.com/s3/>.

Amazon VPC

Eine Amazon VPC definiert ein Netzwerk, das von den Knoten in Ihrem Cluster genutzt wird.

Weitere Informationen zu Amazon VPC finden Sie unter <https://aws.amazon.com/vpc/> und <https://docs.aws.amazon.com/vpc/>.

Elastic Fabric Adapter

Elastic Fabric Adapter (EFA) ist eine Netzwerkschnittstelle für Amazon EC2 EC2-Instances, mit der Kunden Anwendungen ausführen können, die ein hohes Maß an Kommunikation zwischen Knoten erfordern, und das in großem Umfang. AWS

Weitere Informationen zu EC2 Image Builder finden Sie unter <https://aws.amazon.com/hpc/efa/>.

EC2 Image Builder

EC2 Image Builder ist ein vollständig verwalteter AWS Service, mit dem Sie die Erstellung, Verwaltung und Bereitstellung von benutzerdefinierten, sicheren und up-to-date Server-Images automatisieren können.

AWS ParallelCluster verwendet Image Builder, um AWS ParallelCluster Bilder zu erstellen und zu verwalten.

Weitere Informationen zu EC2 Image Builder finden Sie unter <https://aws.amazon.com/image-builder/> und <https://docs.aws.amazon.com/imagebuilder/>.

NICE DCV

NICE DCV ist ein leistungsstarkes Remote-Display-Protokoll, das eine sichere Möglichkeit bietet, Remote-Desktops und Anwendungsstreaming über unterschiedliche Netzwerkbedingungen auf jedem Gerät bereitzustellen. NICE DCV wird verwendet, wenn die [Dcv](#) Einstellungen [HeadNode Abschnitt](#)/angegeben sind. Support für NICE DCV wurde in AWS ParallelCluster Version 2.5.0 hinzugefügt.

Weitere Informationen zu NICE DCV finden Sie unter <https://aws.amazon.com/hpc/dcv/> und <https://docs.aws.amazon.com/dcv/>.

AWS ParallelClusterInterne Verzeichnisse

Es gibt mehrere interne Verzeichnisse, in denen Daten innerhalb des Clusters gemeinsam AWS ParallelCluster genutzt werden. Die folgenden Verzeichnisse werden vom Hauptknoten, den Rechenknoten und den Anmeldeknoten gemeinsam genutzt:

`/opt/slurm`

`/opt/intel`

`/opt/parallelcluster/shared` (only with compute nodes)

`/opt/parallelcluster/shared_login_nodes` (only with login nodes)

`/home` (unless specified in `SharedStorage`)

Note

Standardmäßig werden diese Verzeichnisse auf dem EBS-Volume des Hauptknotens erstellt und als NFS-Exporte an die Rechen- und Anmeldeknoten gemeinsam genutzt. Ab Version AWS ParallelCluster 3.8 können Sie die Erstellung und Verwaltung eines Amazon EFS-Dateisystems zum Hosten und Teilen dieser Verzeichnisse aktivieren AWS ParallelCluster, indem Sie den [SharedStorageType](#) Parameter auf `efs` setzen.

Wenn der Cluster horizontal skaliert wird, können NFS-Exporte über das EBS-Volume zu Leistungsengpässen führen. Mit EFS können Sie NFS-Exporte vermeiden, wenn Ihr Cluster horizontal skaliert wird, und damit verbundene Leistungsengpässe vermeiden.

Tutorials

Die folgenden Tutorials zeigen Ihnen, wie Sie mit AWS ParallelCluster Version 3 loslegen können, und bieten Anleitungen zu bewährten Methoden für einige allgemeine Aufgaben.

Wenn Sie die AWS ParallelCluster Befehlszeilenschnittstelle (CLI) oder API verwenden, zahlen Sie nur für die AWS Ressourcen, die beim Erstellen oder Aktualisieren von AWS ParallelCluster Images und Clustern erstellt werden. Weitere Informationen finden Sie unter [AWS Dienste verwendet von AWS ParallelCluster](#).

Die AWS ParallelCluster Benutzeroberfläche basiert auf einer serverlosen Architektur und Sie können sie in den meisten Fällen innerhalb der Kategorie „AWS Kostenloses Kontingent“ verwenden. Weitere Informationen finden Sie unter [AWS ParallelCluster Die Benutzeroberfläche kostet](#).

Themen

- [Ausführen des ersten Auftrags in AWS ParallelCluster](#)
- [Erstellen eines benutzerdefinierten AWS ParallelCluster-AMIs](#)
- [Integrieren von Active Directory](#)
- [Konfiguration der Verschlüsselung von gemeinsam genutztem Speicher mit einem AWS KMS Schlüssel](#)
- [Ausführung von Jobs in einem Cluster mit mehreren Warteschlangen](#)
- [Verwenden der AWS ParallelCluster-API](#)
- [Einen Cluster mit Slurm Accounting erstellen](#)
- [Zu einer früheren AWS Systems Manager-Dokumentversion zurückkehren](#)
- [Einen Cluster erstellen mit AWS CloudFormation](#)
- [AWS ParallelCluster UI-Integration mit Identity Center](#)

Ausführen des ersten Auftrags in AWS ParallelCluster

Dieses Tutorial führt Sie durch die Ausführung Ihres ersten Hello World-Jobs auf AWS ParallelCluster

Bei der Verwendung der AWS ParallelCluster Befehlszeilenschnittstelle (CLI) oder API, zahlen Sie nur für AWS Ressourcen, die bei der Erstellung oder Aktualisierung erstellt werden

ParallelCluster Bilder und Cluster. Weitere Informationen finden Sie unter [AWS Dienste verwendet von AWS ParallelCluster](#).

Die AWS ParallelCluster Die Benutzeroberfläche basiert auf einer serverlosen Architektur und Sie können sie innerhalb der verwenden AWS Kategorie „Kostenloses Kontingent“ in den meisten Fällen. Weitere Informationen finden Sie unter [AWS ParallelCluster Die Benutzeroberfläche kostet](#).

Voraussetzungen

- AWS ParallelCluster [ist installiert](#).
- Das AWS CLI [ist installiert und konfiguriert](#).
- Sie haben ein [EC2-Schlüsselpaar](#).
- Sie haben eine IAM-Rolle bei [Berechtigungen](#) erforderlich, um das auszuführen `pcluster` CLI.

Überprüfen der Installation

Zuerst überprüfen wir das AWS ParallelCluster ist korrekt installiert und konfiguriert, einschließlich der Abhängigkeit von Node.js.

```
$ node --version
v16.8.0
$ pcluster version
{
  "version": "3.7.0"
}
```

Dabei wird die ausgeführte Version von AWS ParallelCluster zurückgegeben.

Erstellen Sie Ihren ersten Cluster

Jetzt ist es an der Zeit, Ihren ersten Cluster zu erstellen. Da die Workload für dieses Tutorial nicht leistungsintensiv ist, können wir die Standard-Instance-Größe `t2.micro` verwenden. (Für Produktions-Workloads wählen Sie eine Instance-Größe, die Ihren Anforderungen am ehesten entspricht.) Nennen wir Ihren Cluster `hello-world`.

```
$ pcluster create-cluster \
  --cluster-name hello-world \
  --cluster-configuration hello-world.yaml
```

Note

Das AWS-Region zu verwenden muss für die meisten angegeben werden `pcCluster` Befehle. Wenn es nicht in der angegeben ist `AWS_DEFAULT_REGION` Umgebungsvariable oder die `region` Einstellung in der `[default]` Abschnitt der `~/.aws/config` Datei, dann die `--region` Der Parameter muss auf dem angegeben werden `pcCluster` Befehlszeile.

Wenn Sie in der Ausgabe eine Meldung zur Konfiguration erhalten, müssen Sie zur Konfiguration Folgendes ausführen AWS ParallelCluster:

```
$ pcluster configure --config hello-world.yaml
```

Wenn der `pcluster create-cluster` Der Befehl ist erfolgreich, Sie erhalten eine Ausgabe, die der folgenden ähnelt:

```
{
  "cluster": {
    "clusterName": "hello-world",
    "cloudformationStackStatus": "CREATE_IN_PROGRESS",
    "cloudformationStackArn": "arn:aws:cloudformation:xxx:stack/xxx",
    "region": "...",
    "version": "...",
    "clusterStatus": "CREATE_IN_PROGRESS"
  }
}
```

Sie überwachen die Erstellung des Clusters mit:

```
$ pcluster describe-cluster --cluster-name hello-world
```

Der `clusterStatus` Berichte `CREATE_IN_PROGRESS` während der Cluster erstellt wird. Das `clusterStatus` Übergänge zu `CREATE_COMPLETE`, wenn der Cluster erfolgreich erstellt wurde. Die Ausgabe bietet uns auch `publicIpAddress` und `privateIpAddress` unseres Kopfknotens.

Loggen Sie sich in Ihren Hauptknoten ein

Verwenden Sie Ihre OpenSSH-PEM-Datei, um sich bei Ihrem Headnode anzumelden.

```
$ pcluster ssh --cluster-name hello-world -i /path/to/keyfile.pem
```

Sobald Sie angemeldet sind, führen Sie den Befehl `sinfo` aus, um zu bestätigen, dass Ihre Datenverarbeitungsknoten eingerichtet und konfiguriert sind.

```
$ sinfo
PARTITION AVAIL  TIMELIMIT  NODES  STATE NODELIST
queue1*    up        infinite   10    idle~ queue1-dy-queue1t2micro-[1-10]
```

Die Ausgabe zeigt, dass wir in unserem Cluster eine Warteschlange mit bis zu zehn Knoten haben.

Führen Sie Ihren ersten Job mit Slurm aus

Als Nächstes erstellen wir eine Aufgabe, die einen Moment inaktiv ist und dann ihren eigenen Hostnamen ausgibt. Erstellen Sie eine Datei mit dem Namen `hellojob.sh` und den folgenden Inhalten:

```
#!/bin/bash
sleep 30
echo "Hello World from $(hostname)"
```

Als Nächstes übermitteln Sie die Aufgabe mit `sbatch` und überprüfen, dass sie ausgeführt wird.

```
$ sbatch hellojob.sh
Submitted batch job 2
```

Jetzt können Sie Ihre Warteschlange anzeigen und den Status der Aufgabe überprüfen. Die Bereitstellung einer neuen Amazon EC2-Instance wird im Hintergrund gestartet. Sie können den Status der Cluster-Instances mit dem `sinfo`-Befehl.

```
$ squeue
          JOBID PARTITION     NAME     USER ST       TIME  NODES NODELIST(REASON)
           2     queue1 hellojob ec2-user CF       3:30      1 queue1-dy-
queue1t2micro-1
```

Die Ausgabe zeigt, dass der Job weitergeleitet wurde an `queue1`. Warten Sie 30 Sekunden, bis die Aufgabe beendet wird, und führen Sie `squeue` dann erneut aus.

```
$ squeue
```

JOBID	PARTITION	NAME	USER	ST	TIME	NODES	NODELIST(REASON)
-------	-----------	------	------	----	------	-------	------------------

Jetzt, wo sich keine Aufgaben in der Warteschlange befinden, können wir die Ausgabe in unserem aktuellen Verzeichnis überprüfen.

```
$ ls -l
total 8
-rw-rw-r-- 1 ec2-user ec2-user 57 Sep  1 14:25 hellojob.sh
-rw-rw-r-- 1 ec2-user ec2-user 43 Sep  1 14:30 slurm-2.out
```

In der Ausgabe sehen wir ein „out“-Datei. Wir können die Ergebnisse unseres Jobs sehen:

```
$ cat slurm-2.out
Hello World from queue1-dy-queue1t2micro-1
```

Die Ausgabe zeigt auch, dass die Aufgabe auf Instance `queue1-dy-queue1t2micro-1` erfolgreich ausgeführt wurde.

In dem Cluster, den Sie gerade erstellt haben, wird nur das Home-Verzeichnis von allen Knoten des Clusters gemeinsam genutzt.

Weitere Informationen zum Erstellen und Verwenden von Clustern finden Sie unter [Bewährte Methoden](#).

Wenn für Ihre Anwendung gemeinsam genutzte Software, Bibliotheken oder Daten erforderlich sind, sollten Sie die folgenden Optionen in Betracht ziehen:

- Erstellen Sie ein [AWS ParallelCluster-aktiviertes benutzerdefiniertes AMI](#), das Ihre Software enthält, wie unter [beschrieben Erstellen eines benutzerdefinierten AWS ParallelCluster-AMIs](#).
- Benutze die [StorageSettings](#) Option in der [AWS ParallelCluster Konfigurationsdatei](#), um ein gemeinsam genutztes Dateisystem anzugeben und Ihre installierte Software am angegebenen Mount-Speicherort zu speichern.
- Benutzen [Benutzerdefinierte Bootstrap-Aktionen](#) um das Bootstrap-Verfahren für jeden Knoten Ihres Clusters zu automatisieren.

Erstellen eines benutzerdefinierten AWS ParallelCluster-AMIs

Bei der Verwendung der [AWS ParallelCluster Befehlszeilenschnittstelle \(CLI\)](#) oder [API](#), zahlen Sie nur für [AWS Ressourcen](#), die bei der Erstellung oder Aktualisierung erstellt werden [AWS](#)

ParallelCluster Bilder und Cluster. Weitere Informationen finden Sie unter [AWS Dienste verwendet von AWS ParallelCluster](#).

Die AWS ParallelCluster Die Benutzeroberfläche basiert auf einer serverlosen Architektur und Sie können sie innerhalb der verwenden AWS Kategorie „Kostenloses Kontingent“ in den meisten Fällen. Weitere Informationen finden Sie unter [AWS ParallelCluster Die Benutzeroberfläche kostet](#).

Important

Wenn Sie ein benutzerdefiniertes AMI erstellen, müssen Sie die Schritte, die Sie zur Erstellung Ihres benutzerdefinierten AMI verwendet haben, bei jedem neuen wiederholen AWS ParallelCluster veröffentlichen.

Bevor Sie weiterlesen, empfehlen wir Ihnen, zunächst die [Benutzerdefinierte Bootstrap-Aktionen](#) Abschnitt. Stellen Sie fest, ob die Änderungen, die Sie vornehmen möchten, per Skript erstellt und in Zukunft unterstützt werden können AWS ParallelCluster Veröffentlichungen.

Auch wenn das Erstellen eines benutzerdefinierten AMI im Allgemeinen nicht ideal ist, gibt es bestimmte Szenarien, in denen das Erstellen eines benutzerdefinierten AMI für AWS ParallelCluster ist notwendig. In diesem Tutorial wird beschrieben, wie Sie ein benutzerdefiniertes AMI für diese Szenarien erstellen.

Voraussetzungen

- AWS ParallelCluster [ist installiert](#).
- Die AWS CLI [ist installiert und konfiguriert](#).
- Sie haben ein [EC2-Schlüsselpaar](#).
- Sie haben eine IAM-Rolle bei [Berechtigungen](#) erforderlich, um das auszuführen [pcluster](#) CLI und Images erstellen.

So passen Sie das AWS ParallelCluster-AMI an

Es gibt zwei Möglichkeiten, ein benutzerdefiniertes zu erstellen AWS ParallelCluster AMI. Eine dieser beiden Methoden besteht darin, ein neues AMI mit dem zu erstellen AWS ParallelCluster CLI. Bei einer anderen Methode müssen Sie manuelle Änderungen vornehmen, um ein neues AMI zu erstellen, das unter Ihrem AWS-Konto.

Erstellen Sie ein benutzerdefiniertes AWS ParallelCluster AMI

Wenn Sie über ein benutzerdefiniertes AMI und eine maßgeschneiderte Software verfügen, können Sie die erforderlichen Änderungen wie folgt vornehmen. AWS ParallelCluster nutzt den EC2 Image Builder-Dienst, um maßgeschneiderte AMIs zu erstellen. Weitere Informationen finden Sie im [Image Builder-Benutzerhandbuch](#).

Die wichtigsten Punkte:

- Der Vorgang dauert etwa 1 Stunde. Diese Zeit kann variieren, wenn zusätzliche [Build/Components](#) muss zur Build-Zeit installiert werden.
- Das AMI ist mit den Versionen der Hauptkomponenten gekennzeichnet. Dazu gehören der Kernel, der Scheduler und [EFA](#) Fahrer. Eine Teilmenge der Komponentenversionen ist auch in der AMI-Beschreibung aufgeführt.
- Beginnend von AWS ParallelCluster 3.0.0 kann ein neuer Satz von CLI-Befehlen verwendet werden, um den Lebenszyklus von Images zu verwalten. Das beinhaltet [build-image](#), [list-images](#), [describe-image](#), und [delete-image](#).
- Diese Methode ist wiederholbar. Sie können sie erneut ausführen, um AMIs auf dem neuesten Stand zu halten (z. B. Betriebssystemupdates), und sie dann verwenden, wenn Sie einen vorhandenen Cluster aktualisieren.

Note

Wenn Sie diese Methode in der AWS Bei der China-Partition treten möglicherweise Netzwerkfehler auf. Diese Fehler könnten Ihnen beispielsweise in der `pccluster build-image` Befehl, wenn er Pakete herunterlädt von GitHub oder aus einem Betriebssystem-Repository. In diesem Fall empfehlen wir Ihnen, eine der folgenden alternativen Methoden zu verwenden:

1. Folgen Sie dem [Ändern eines AWS ParallelCluster-AMIs](#) Ansatz, der diesen Befehl umgeht.
2. Erstellen Sie das Image in einer anderen Partition und Region, z. B. `us-east-1`, und dann speichern/wiederherstellen, um es in die Region China zu verschieben. Weitere Informationen finden Sie unter [Speichern und Wiederherstellen eines AMI mit S3](#) in der Amazon EC2-Benutzerhandbuch für Linux-Instances.

Schritte:

1. Konfigurieren Sie Ihre AWS-Konto-Anmeldeinformationen, sodass die AWS ParallelCluster-Client-Anrufer bei AWS-API-Operationen in Ihrem Namen. Eine Liste der erforderlichen Berechtigungen finden Sie unter [AWS Identity and Access Management Berechtigungen in AWS ParallelCluster](#).
2. Erstellen Sie eine Basic-AMI-Konfigurationsdatei. Geben Sie dazu die [InstanceType](#) zu verwenden, um das Image zu erstellen und das [ParentImage](#). Diese werden als Ausgangspunkt für die Erstellung des AMI verwendet. Weitere Hinweise zu optionalen Build-Parametern finden Sie unter [Image-Konfiguration](#).

Build:

InstanceType: *<BUILD_INSTANCE_TYPE>*

ParentImage: *<BASE_AMI_ID>*

3. Verwenden Sie den CLI-Befehl `pcluster build-image` um ein AWS ParallelCluster-AMI aus dem AMI, das Sie als Basis angeben.

```
$ pcluster build-image --image-id IMAGE_ID --image-configuration IMAGE_CONFIG.yaml --region REGION
{
  "image": {
    "imageId": "IMAGE_ID",
    "imageBuildStatus": "BUILD_IN_PROGRESS",
    "cloudformationStackStatus": "CREATE_IN_PROGRESS",
    "cloudformationStackArn": "arn:aws:cloudformation:us-east-1:123456789012:stack/IMAGE_ID/abcd1234-ef56-gh78-ij90-1234abcd5678",
    "region": "us-east-1",
    "version": "3.7.0"
  }
}
```

Warning

`pcluster build-image` verwendet die Standard-VPC. Wenn Sie die Standard-VPC löschen mit AWS Control Tower oder AWS Landing Zone, die Subnetz-ID muss in der Image-Konfigurationsdatei angegeben werden. Weitere Informationen finden Sie unter [SubnetId](#).

Eine Liste mit anderen Parametern finden Sie in [pcluster build-image](#) Referenzseite für Befehle. Die Ergebnisse des vorherigen Befehls lauten wie folgt:

- Ein CloudFormation Stack wird auf der Grundlage der Image-Konfiguration erstellt. Der Stack enthält alle EC2 Image Builder-Ressourcen, die für den Build erforderlich sind.
 - Zu den erstellten Ressourcen gehört der offizielle Image Builder AWS ParallelCluster-Komponenten, zu denen benutzerdefinierte Image Builder-Komponenten hinzugefügt werden können. Informationen zum Erstellen benutzerdefinierter Komponenten finden Sie im [Beispiele für benutzerdefinierte AMIs](#) in der Workshop HPC für Kunden aus dem öffentlichen Sektor.
 - EC2 Image Builder startet eine Build-Instanz und wendet das AWS ParallelCluster Kochbuch an, installiert das AWS ParallelCluster Software-Stack und führt die erforderlichen Konfigurationsaufgaben durch. Das AWS ParallelCluster Kochbuch wird zum Erstellen und Bootstrappen verwendet AWS ParallelCluster.
 - Die Instanz wird gestoppt und daraus wird ein neues AMI erstellt.
 - Eine weitere Instanz wird über das neu erstellte AMI gestartet. Während der Testphase führt EC2 Image Builder Tests durch, die in den Image Builder-Komponenten definiert sind.
 - Wenn der Build erfolgreich ist, wird der Stack gelöscht. Wenn der Build fehlschlägt, wird der Stack beibehalten und kann überprüft werden.
4. Sie können den Status des Build-Prozesses überwachen, indem Sie den folgenden Befehl ausführen. Nach Abschluss des Builds können Sie ihn ausführen, um die in der Antwort angegebene AMI-ID abzurufen.

```
$ pcluster describe-image --image-id IMAGE_ID --region REGION

# BEFORE COMPLETE
{
  "imageConfiguration": {
    "url": "https://parallelcluster-1234abcd5678efgh-v1-do-not-
delete.s3.amazonaws.com/parallelcluster/3.7.0/images/IMAGE_ID-abcd1234efgh5678/
configs/image-config.yaml?...",
  },
  "imageId": "IMAGE_ID",
  "imagebuilderImageStatus": "BUILDING",
  "imageBuildStatus": "BUILD_IN_PROGRESS",
  "cloudformationStackStatus": "CREATE_IN_PROGRESS",
  "cloudformationStackArn": "arn:aws:cloudformation:us-east-1:123456789012:stack/
IMAGE_ID/abcd1234-ef56-gh78-ij90-1234abcd5678",
  "region": "us-east-1",
```

```
"version": "3.7.0",
"cloudformationStackTags": [
  {
    "value": "3.7.0",
    "key": "parallelcluster:version"
  },
  {
    "value": "IMAGE_ID",
    "key": "parallelcluster:image_name"
  },
  ...
],
"imageBuildLogsArn": "arn:aws:logs:us-east-1:123456789012:log-group:/aws/
imagebuilder/ParallelClusterImage-IMAGE_ID",
"cloudformationStackCreationTime": "2022-04-05T21:36:26.176Z"
}

# AFTER COMPLETE
{
  "imageConfiguration": {
    "url": "https://parallelcluster-1234abcd5678efgh-v1-do-not-delete.s3.us-
east-1.amazonaws.com/parallelcluster/3.7.0/images/IMAGE_ID-abcd1234efgh5678/configs/
image-config.yaml?Signature=..."
  },
  "imageId": "IMAGE_ID",
  "imageBuildStatus": "BUILD_COMPLETE",
  "region": "us-east-1",
  "ec2AmiInfo": {
    "amiName": "IMAGE_ID 2022-04-05T21-39-24.020Z",
    "amiId": "ami-1234stuv5678wxyz",
    "description": "AWS ParallelCluster AMI for alinux2,
kernel-4.14.238-182.422.amzn2.x86_64, lustre-2.10.8-5.amzn2.x86_64,
efa-1.13.0-1.amzn2.x86_64, dcv-2021.1.10598-1.e17.x86_64, slurm-20-11-8-1",
    "state": "AVAILABLE",
    "tags": [
      {
        "value": "2021.3.11591-1.e17.x86_64",
        "key": "parallelcluster:dcv_version"
      },
      ...
    ],
    "architecture": "x86_64"
  },
  "version": "3.7.0"
}
```

```
}

```

5. Um Ihren Cluster zu erstellen, geben Sie die AMI-ID in das [CustomAmi](#) Feld in Ihrer Cluster-Konfiguration.

Fehlerbehebung und Überwachung des AMI-Erstellungsprozesses

Die Image-Erstellung ist in etwa einer Stunde abgeschlossen. Sie können den Prozess überwachen, indem Sie `pcluster describe-image` Befehle zum Abrufen von Befehlen oder Protokollen.

```
$ pcluster describe-image --image-id IMAGE_ID --region REGION
```

Die `build-image` Befehl erstellt ein CloudFormation Stapel mit allen EC2-Ressourcen, die zum Erstellen des Images erforderlich sind, und startet den EC2 Image Builder-Prozess.

Nach dem Ausführen von `build-image` Befehl, es ist möglich, ihn abzurufen CloudFormation Ereignisse stapeln mit `pcluster get-image-stack-events`. Sie können Ergebnisse filtern mit `--query` Parameter, um die neuesten Ereignisse zu sehen. Weitere Informationen finden Sie unter [Filterung AWS CLI Ausgang](#) in der AWS Command Line Interface Benutzerleitfaden.

```
$ pcluster get-image-stack-events --image-id IMAGE_ID --region REGION --query
"events[0]"
{
  "eventId": "ParallelClusterImage-CREATE_IN_PROGRESS-2022-04-05T21:39:24.725Z",
  "physicalResourceId": "arn:aws:imagebuilder:us-east-1:123456789012:image/
parallelclusterimage-IMAGE_ID/3.7.0/1",
  "resourceStatus": "CREATE_IN_PROGRESS",
  "resourceStatusReason": "Resource creation Initiated",
  "resourceProperties": "{\"InfrastructureConfigurationArn\":
\\\"arn:aws:imagebuilder:us-east-1:123456789012:infrastructure-configuration/
parallelclusterimage-abcd1234-ef56-gh78-ij90-1234abcd5678\\\", \\\"ImageRecipeArn\\\":
\\\"arn:aws:imagebuilder:us-east-1:123456789012:image-recipe/parallelclusterimage-
IMAGE_ID/3.7.0\\\", \\\"DistributionConfigurationArn\\\": \\\"arn:aws:imagebuilder:us-
east-1:123456789012:distribution-configuration/parallelclusterimage-abcd1234-ef56-
gh78-ij90-1234abcd5678\\\", \\\"Tags\\\": {\\\"parallelcluster:image_name\\\": \\\"IMAGE_ID\\\",
\\\"parallelcluster:image_id\\\": \\\"IMAGE_ID\\\"}}\",
  "stackId": "arn:aws:cloudformation:us-east-1:123456789012:stack/IMAGE_ID/abcd1234-
ef56-gh78-ij90-1234abcd5678",
  "stackName": "IMAGE_ID",
  "logicalResourceId": "ParallelClusterImage",
  "resourceType": "AWS::ImageBuilder::Image",
```

```
"timestamp": "2022-04-05T21:39:24.725Z"
}
```

Nach etwa 15 Minuten erscheinen die Stack-Ereignisse im Protokollereigniseintrag, der sich auf die Erstellung von Image Builder bezieht. Sie können jetzt Image-Log-Streams auflisten und die Image Builder-Schritte überwachen, indem Sie [pcluster list-image-log-streams](#) und [pcluster get-image-log-events](#) Befehle.

```
$ pcluster list-image-log-streams --image-id IMAGE_ID --region REGION \
  --query 'logStreams[*].logStreamName'

"3.7.0/1"
]

$ pcluster get-image-log-events --image-id IMAGE_ID --region REGION \
  --log-stream-name 3.7.0/1 --limit 3
{
  "nextToken": "f/36295977202298886557255241372854078762600452615936671762",
  "prevToken": "b/36295977196879805474012299949460899222346900769983430672",
  "events": [
    {
      "message": "ExecuteBash: FINISHED EXECUTION",
      "timestamp": "2022-04-05T22:13:26.633Z"
    },
    {
      "message": "Document arn:aws:imagebuilder:us-east-1:123456789012:component/
parallelclusterimage-test-abcd1234-ef56-gh78-ij90-1234abcd5678/3.7.0/1",
      "timestamp": "2022-04-05T22:13:26.741Z"
    },
    {
      "message": "TOE has completed execution successfully",
      "timestamp": "2022-04-05T22:13:26.819Z"
    }
  ]
}
```

Erkundigen Sie sich weiter bei der [describe-image](#) Befehl, bis Sie das `sehenBUILD_COMPLETE` Status.

```
$ pcluster describe-image --image-id IMAGE_ID --region REGION
{
  "imageConfiguration": {
```

```

    "url": "https://parallelcluster-1234abcd5678efgh-v1-do-not-delete.s3.us-
east-1.amazonaws.com/parallelcluster/3.7.0/images/IMAGE_ID-abcd1234efgh5678/configs/
image-config.yaml?Signature=..."
  },
  "imageId": "IMAGE_ID",
  "imageBuildStatus": "BUILD_COMPLETE",
  "region": "us-east-1",
  "ec2AmiInfo": {
    "amiName": "IMAGE_ID 2022-04-05T21-39-24.020Z",
    "amiId": "ami-1234stuv5678wxyz",
    "description": "AWS ParallelCluster AMI for alinux2,
kernel-4.14.238-182.422.amzn2.x86_64, lustre-2.10.8-5.amzn2.x86_64,
efa-1.13.0-1.amzn2.x86_64, dcv-2021.1.10598-1.el7.x86_64, slurm-20-11-8-1",
    "state": "AVAILABLE",
    "tags": [
      {
        "value": "2021.3.11591-1.el7.x86_64",
        "key": "parallelcluster:dcv_version"
      },
      ...
    ],
    "architecture": "x86_64"
  },
  "version": "3.7.0"
}

```

Wenn Sie ein Problem bei der Erstellung eines benutzerdefinierten AMIs beheben müssen, erstellen Sie ein Archiv der Image-Protokolle, wie in den folgenden Schritten beschrieben.

Es ist möglich, die Protokolle in einem Amazon S3-Bucket oder in einer lokalen Datei zu archivieren, je nachdem `--outputParameter`.

```

$ pcluster export-image-logs --image-id IMAGE_ID --region REGION \
--bucket BUCKET_NAME --bucket-prefix BUCKET_FOLDER
{
  "url": "https://BUCKET_NAME.s3.us-east-1.amazonaws.com/BUCKET-FOLDER/IMAGE_ID-
logs-202209071136.tar.gz?AWSAccessKeyId=..."
}

$ pcluster export-image-logs --image-id IMAGE_ID \
--region REGION --bucket BUCKET_NAME --bucket-prefix BUCKET_FOLDER --output-file /tmp/
archive.tar.gz
{

```

```
"path": "/tmp/archive.tar.gz"  
}
```

Das Archiv enthält CloudWatch Protokollierte Streams im Zusammenhang mit dem Image Builder-Prozess und AWS CloudFormation Ereignisse stapeln. Die Ausführung des Befehls kann mehrere Minuten dauern.

Verwaltung benutzerdefinierter AMIs

Beginnend von AWS ParallelCluster 3.0.0 wurde der CLI ein neuer Befehlssatz hinzugefügt, um den Image-Lebenszyklus zu erstellen, zu überwachen und zu verwalten. Weitere Informationen zu den Befehlen finden Sie unter [pcluster-Befehle](#).

Ändern eines AWS ParallelCluster-AMIs

Diese Methode besteht darin, einen Beamten zu modifizieren AWS ParallelCluster AMI, indem Sie darüber hinaus Anpassungen hinzufügen. Die Basis AWS ParallelCluster AMIs werden mit neuen Versionen aktualisiert. Diese AMIs enthalten alle Komponenten, die erforderlich sind für AWS ParallelCluster um zu funktionieren, wenn es installiert und konfiguriert ist. Sie können mit einer davon als Basis beginnen.

Die wichtigsten Punkte:

- Diese Methode ist schneller als die [build-image](#) Befehl. Es ist jedoch ein manueller Vorgang und nicht automatisch wiederholbar.
- Bei dieser Methode haben Sie keinen Zugriff auf die Befehle zum Abrufen von Protokollen und zur Verwaltung des Image-Lebenszyklus, die über die CLI verfügbar sind.

Schritte:

New EC2 console

1. Suchen Sie das AMI, das dem spezifischen entspricht AWS-Region das du benutzt. Um es zu finden, benutzen Sie das [pcluster list-official-images](#) Befehl mit dem `--region` Parameter zur Auswahl des spezifischen AWS-Region und `--os` und `--architecture` Parameter, um nach dem gewünschten AMI mit dem Betriebssystem und der Architektur zu filtern, die Sie verwenden möchten. Rufen Sie aus der Ausgabe die EC2-Image-ID ab.

2. Melden Sie sich bei der AWS Management Console an und öffnen Sie die Amazon-EC2-Konsole unter <https://console.aws.amazon.com/EC2/>.
3. Wählen Sie im Navigationsbereich Bilder, und dann AMIs. Suchen Sie nach der abgerufenen EC2-Image-ID, wählen Sie das AMI aus und wählen Sie Starten Sie die Instance über das AMI.
4. Scrollen Sie nach unten und wählen Sie Ihre Instanztyp.
5. Wähle deine Schlüsselpaar und Instance starten.
6. Melden Sie sich mit dem Betriebssystembenutzer und Ihrem SSH-Schlüssel.
7. Passen Sie Ihre Instance manuell an Ihre Anforderungen an.
8. Führen Sie den folgenden Befehl aus, um Ihre Instance für die AMI-Erstellung vorzubereiten.

```
sudo /usr/local/sbin/ami_cleanup.sh
```

9. Wählen Sie in der Konsole Status der Instanz und Instance beenden.

Navigiere zu Instanzen, wählen Sie die neue Instance, wählen Sie Status der Instanz, und Instance beenden.

10. Erstellen Sie mithilfe der EC2-Konsole ein neues AMI von der Instance aus oder AWS CLI [Bild erstellen](#).

Von der EC2-Konsole

- a. Wählen Sie im Navigationsbereich Instanzen aus.
 - b. Wählen Sie die Instance aus, die Sie erstellt und geändert haben.
 - c. In Aktionen, wählen Bild und dann Bild erstellen.
 - d. Wählen Sie Create Image aus.
11. Geben Sie die neue AMI-ID in das [Custom AMI](#) Feld in Ihrer Cluster-Konfiguration und erstellen Sie einen Cluster.

Old EC2 console

1. Finden Sie den AWS ParallelCluster AMI, das dem spezifischen entspricht AWS-Region das du benutzt. Um es zu finden, können Sie das benutzen [pcluster list-official-images](#) Befehl mit dem `--region` Parameter zur Auswahl des spezifischen AWS-Region und `--os` und `--architecture` Parameter, um nach dem gewünschten AMI mit dem Betriebssystem und der Architektur zu filtern, die Sie verwenden möchten. Aus der Ausgabe können Sie die EC2-Image-ID abrufen.

2. Melden Sie sich bei der AWS Management Console an und öffnen Sie die Amazon-EC2-Konsole unter <https://console.aws.amazon.com/EC2/>.
3. Wählen Sie im Navigationsbereich Bilder, und dann AMIs. Stellen Sie den Filter ein für Öffentliche Bilder und suchen Sie nach der abgerufenen EC2-Image-ID, wählen Sie das AMI aus und wählen Sie Starten.
4. Wählen Sie Ihren Instanztyp und wählen Sie Weiter: Instanzdetails konfigurieren oder Überprüfen und starten um Ihre Instance zu starten.
5. Wählen Sie Starten, wählen Sie Ihre Schlüsselpaar, und Instances starten.
6. Melden Sie sich bei Ihrer Instance mithilfe des Betriebssystembenutzers und Ihres SSH-Schlüssels an. Weitere Informationen finden Sie unter Instanzen, wählen Sie die neue Instanz aus und Verbinden.
7. Passen Sie Ihre Instance manuell an Ihre Anforderungen an.
8. Führen Sie den folgenden Befehl aus, um Ihre Instance für die AMI-Erstellung vorzubereiten:

```
sudo /usr/local/sbin/ami_cleanup.sh
```

9. Wählen Sie in der EC2-Konsole Instanzen Wählen Sie im Navigationsbereich Ihre neue Instanz aus und wählen Sie Aktionen, Status der Instanz und Stopp.
10. Erstellen Sie mithilfe der EC2-Konsole ein neues AMI von der Instance aus oder AWS CLI [Bild erstellen](#).

Von der EC2-Konsole

- a. Wählen Sie im Navigationsbereich Instances aus.
 - b. Wählen Sie die Instanz aus, die Sie erstellt und geändert haben.
 - c. In Aktionen, wählen Bild und dann Bild erstellen.
 - d. Wählen Sie Create Image aus.
11. Geben Sie die neue AMI-ID in das [CustomAmi](#) Feld in Ihrer Cluster-Konfiguration und erstellen Sie einen Cluster.

Integrieren von Active Directory

In diesem Tutorial erstellen Sie eine Umgebung mit mehreren Benutzern. Diese Umgebung umfasst eine AWS ParallelCluster, die in ein AWS Managed Microsoft AD (Active Directory-) AT integriert ist `corp.example.com`. Sie konfigurieren einen Admin Benutzer für die Verwaltung des

Verzeichnisses, einen ReadOnly Benutzer für das Lesen des Verzeichnisses und einen user000 Benutzer für die Anmeldung am Cluster. Sie können entweder den automatisierten Pfad oder den manuellen Pfad verwenden, um die Netzwerkressourcen, ein Active Directory (AD) und die EC2-Instance zu erstellen, die Sie zur Konfiguration des AD verwenden. Unabhängig vom Pfad ist die Infrastruktur, die Sie erstellen, für die Integration AWS ParallelCluster mit einer der folgenden Methoden vorkonfiguriert:

- LDAPS mit Zertifikatsüberprüfung (als sicherste Option empfohlen)
- LDAPS ohne Zertifikatsverifizierung
- LDAP

LDAP selbst bietet keine Verschlüsselung. Um eine sichere Übertragung potenziell sensibler Informationen zu gewährleisten, empfehlen wir dringend, LDAPS (LDAP über TLS/SSL) für Cluster zu verwenden, die in ADs integriert sind. Weitere Informationen finden Sie unter [Aktivieren serverseitiger LDAPS](#) im Administratorhandbuch. AWS Managed Microsoft AD AWS Directory Service

Nachdem Sie diese Ressourcen erstellt haben, fahren Sie mit der Konfiguration und Erstellung Ihres Clusters fort, der in Ihr Active Directory (AD) integriert ist. Melden Sie sich nach der Erstellung des Clusters als der Benutzer an, den Sie erstellt haben. Weitere Informationen zu der Konfiguration, die Sie in diesem Tutorial erstellen, finden Sie unter [Zugriff mehrerer Benutzer auf Cluster](#) und im Abschnitt [DirectoryService](#)Konfiguration.

In diesem Tutorial wird beschrieben, wie Sie eine Umgebung erstellen, die den Zugriff mehrerer Benutzer auf Cluster unterstützt. In diesem Tutorial wird nicht behandelt, wie Sie ein AWS Directory Service AD erstellen und verwenden. Die Schritte, die Sie AWS Managed Microsoft AD in diesem Tutorial zum Einrichten eines ausführen, dienen nur zu Testzwecken. Sie ersetzen nicht die offizielle Dokumentation und die Best Practices, die Sie unter [AWS Managed Microsoft AD Simple AD](#) im AWS Directory Service Administrationshandbuch finden.

Note

Kennwörter von Verzeichnisbenutzern laufen gemäß den Eigenschaftsdefinitionen der Verzeichniskennwortrichtlinie ab. Weitere Informationen finden Sie unter [Unterstützte Richtlinieneinstellungen](#). Informationen zum Zurücksetzen von Verzeichnispasswörtern mit AWS ParallelCluster finden Sie unter [Wie setze ich ein Benutzerpasswort und abgelaufene Passwörter zurück](#).

Note

Die IP-Adressen der Verzeichnisdomänencontroller können sich aufgrund von Änderungen an den Domänencontrollern und der Verzeichnisverwaltung ändern. Wenn Sie die automatische Schnellerstellungsmethode zum Erstellen der Verzeichnisinfrastruktur gewählt haben, müssen Sie den Load Balancer manuell vor den Verzeichniscontrollern ausrichten, wenn sich die Verzeichnis-IP-Adressen ändern. Wenn Sie die Schnellerstellungsmethode verwenden, werden die Verzeichnis-IP-Adressen nicht automatisch an die Load Balancer angepasst.

Wenn Sie die AWS ParallelCluster Befehlszeilenschnittstelle (CLI) oder API verwenden, zahlen Sie nur für die AWS Ressourcen, die beim Erstellen oder Aktualisieren von AWS ParallelCluster Images und Clustern erstellt werden. Weitere Informationen finden Sie unter [AWS Dienste verwendet von AWS ParallelCluster](#).

Die AWS ParallelCluster Benutzeroberfläche basiert auf einer serverlosen Architektur und Sie können sie in den meisten Fällen innerhalb der Kategorie „AWS Kostenloses Kontingent“ verwenden. Weitere Informationen finden Sie unter [AWS ParallelCluster Die Benutzeroberfläche kostet](#).

Voraussetzungen

- AWS ParallelCluster [ist installiert](#).
- Das AWS CLI [ist installiert und konfiguriert](#).
- Sie haben ein [EC2-Schlüsselpaar](#).
- Sie haben eine IAM-Rolle mit den [für die Ausführung der pcluster CLI erforderlichen Berechtigungen](#).

Ersetzen Sie beim Durcharbeiten des Tutorials *inputs highlighted in red d-abcdef01234567890* beispielsweise *region-id* und durch Ihre eigenen Namen und IDs. Ersetze es *0123456789012* durch deine AWS-Konto Nummer.

Schritt 1: Erstellen Sie die AD-Infrastruktur

Wählen Sie die Registerkarte Automatisiert, um die Active Directory (AD) -Infrastruktur mit einer AWS CloudFormation Schnellerstellungsvorlage zu erstellen.

Wählen Sie die Registerkarte Manuell, um die AD-Infrastruktur manuell zu erstellen.

Automatisiert

1. Melden Sie sich an der AWS Management Console an.
2. Öffnen [Sie CloudFormation Quick Create \(Region US-East-1\)](#), um die folgenden Ressourcen in der CloudFormation Konsole zu erstellen:
 - Eine VPC mit zwei Subnetzen und Routing für den öffentlichen Zugriff, wenn keine VPC angegeben ist.
 - Eine AWS Managed Microsoft AD
 - Eine EC2-Instance, die mit dem AD verknüpft ist und mit der Sie das Verzeichnis verwalten können.
3. Geben Sie im Abschnitt Parameter für die Schnellerstellungs-Stack-Seite Kennwörter für die folgenden Parameter ein:
 - AdminPassword
 - ReadOnlyPassword
 - UserPassword

Notieren Sie sich die Passwörter. Sie verwenden sie später in diesem Tutorial.

4. Geben Sie für DomainName den Wert **corp.example.com** ein.
5. Geben Sie für Keypair den Namen eines EC2-Schlüsselpaars ein.
6. Markieren Sie die Kästchen, um die einzelnen Zugriffsmöglichkeiten unten auf der Seite zu bestätigen.
7. Wählen Sie Stack erstellen aus.
8. Nachdem der CloudFormation Stapel den CREATE_COMPLETE Status erreicht hat, wählen Sie die Registerkarte Ausgaben des Stapels aus. Notieren Sie sich die Namen und IDs der Ausgaberesourcen, da Sie sie in späteren Schritten verwenden müssen. Die Ausgaben enthalten die Informationen, die für die Erstellung des Clusters benötigt werden.

CloudFormation > Stacks > PclusterAD-1234abc

Stacks (3)

Filter by stack name

Active View nested

PclusterAD
2022-05-18 10:27:24 UTC-0700
CREATE_COMPLETE

PclusterAD-abcd123

Delete Update Stack actions Create stack

Stack info Events Resources **Outputs** Parameters Template Change sets

Outputs (10)

Search outputs

Key	Value
DomainAddrLdap	ldap://10.0.111.88,ldap://10.0.222.111
DomainAddrLdaps	ldaps://corp.example.com
DomainCertificateArn	arn:aws:acm:us-east-1:123456789012:certificate/1234abcd-ef56-78gh-ij90-abcd1
DomainCertificateSecretArn	arn:aws:secretsmanager:us-east-1:123456789012:secret:DomainCertificateSecret-F
DomainCertificateSecretReadPolicy	arn:aws:iam::123456789012:policy/DomainCertificateSecretReadPolicy-PclusterAD
DomainName	corp.example.com
DomainReadOnlyUser	cn=ReadOnlyUser,ou=Users,ou=CORP,dc=corp,dc=example,dc=com
PasswordSecretArn	arn:aws:secretsmanager:us-east-1:123456789012:secret>PasswordSecret-PclusterA
PrivateSubnetIds	subnet-1234567890abcdef0,subnet-abcdef01234567890
VpcId	vpc-021345abcdef6789

9. Um die Übungen durchführen zu können [\(Optional\) Schritt 2: AD-Benutzer und -Gruppen verwalten](#), benötigen Sie die Verzeichnis-ID. Wählen Sie Ressourcen und scrollen Sie nach unten, um sich die Verzeichnis-ID zu notieren.
10. Fahren Sie mit [\(Optional\) Schritt 2: AD-Benutzer und -Gruppen verwalten](#) oder fort [Schritt 3: Erstellen Sie den Cluster](#).

Manuell

Erstellen Sie eine VPC für den Verzeichnisdienst mit zwei Subnetzen in verschiedenen Availability Zones und einem AWS Managed Microsoft AD

Erstellen Sie das AD

Note

- Das Verzeichnis und der Domainname sind `corp.example.com`. Der Kurzname ist `CORP`.
- Ändern Sie das Admin Passwort im Skript.

- Die Erstellung des Active Directory (AD) dauert mindestens 15 Minuten.

Verwenden Sie das folgende Python-Skript, um die VPC, Subnetze und AD-Ressourcen in Ihrem lokalen System zu erstellen. AWS-Region Speichern Sie diese Datei unter `ad.py` und führen Sie sie aus.

```
import boto3
import time
from pprint import pprint

vpc_name = "PclusterVPC"
ad_domain = "corp.example.com"
admin_password = "asdfASDF1234"

ec2 = boto3.client("ec2")
ds = boto3.client("ds")
region = boto3.Session().region_name

# Create the VPC, Subnets, IGW, Routes
vpc = ec2.create_vpc(CidrBlock="10.0.0.0/16")["Vpc"]
vpc_id = vpc["VpcId"]
time.sleep(30)
ec2.create_tags(Resources=[vpc_id], Tags=[{"Key": "Name", "Value": vpc_name}])
subnet1 = ec2.create_subnet(VpcId=vpc_id, CidrBlock="10.0.0.0/17",
    AvailabilityZone=f"{region}a")["Subnet"]
subnet1_id = subnet1["SubnetId"]
time.sleep(30)
ec2.create_tags(Resources=[subnet1_id], Tags=[{"Key": "Name", "Value": f"{vpc_name}/
subnet1"}])
ec2.modify_subnet_attribute(SubnetId=subnet1_id, MapPublicIpOnLaunch={"Value": True})
subnet2 = ec2.create_subnet(VpcId=vpc_id, CidrBlock="10.0.128.0/17",
    AvailabilityZone=f"{region}b")["Subnet"]
subnet2_id = subnet2["SubnetId"]
time.sleep(30)
ec2.create_tags(Resources=[subnet2_id], Tags=[{"Key": "Name", "Value": f"{vpc_name}/
subnet2"}])
ec2.modify_subnet_attribute(SubnetId=subnet2_id, MapPublicIpOnLaunch={"Value": True})
igw = ec2.create_internet_gateway()["InternetGateway"]
ec2.attach_internet_gateway(InternetGatewayId=igw["InternetGatewayId"], VpcId=vpc_id)
route_table = ec2.describe_route_tables(Filters=[{"Name": "vpc-id", "Values":
    [vpc_id]}])["RouteTables"][0]
```

```
ec2.create_route(RouteTableId=route_table["RouteTableId"],
  DestinationCidrBlock="0.0.0.0/0", GatewayId=igw["InternetGatewayId"])
ec2.modify_vpc_attribute(VpcId=vpc_id, EnableDnsSupport={"Value": True})
ec2.modify_vpc_attribute(VpcId=vpc_id, EnableDnsHostnames={"Value": True})

# Create the Active Directory
ad = ds.create_microsoft_ad(
    Name=ad_domain,
    Password=admin_password,
    Description="ParallelCluster AD",
    VpcSettings={"VpcId": vpc_id, "SubnetIds": [subnet1_id, subnet2_id]},
    Edition="Standard",
)
directory_id = ad["DirectoryId"]

# Wait for completion
print("Waiting for the directory to be created...")
directories = ds.describe_directories(DirectoryIds=[directory_id])
["DirectoryDescriptions"]
directory = directories[0]
while directory["Stage"] in {"Requested", "Creating"}:
    time.sleep(3)
    directories = ds.describe_directories(DirectoryIds=[directory_id])
["DirectoryDescriptions"]
    directory = directories[0]

dns_ip_addrs = directory["DnsIpAddrs"]

pprint({"directory_id": directory_id,
        "vpc_id": vpc_id,
        "subnet1_id": subnet1_id,
        "subnet2_id": subnet2_id,
        "dns_ip_addrs": dns_ip_addrs})
```

Das Folgende ist eine Beispielausgabe aus dem Python-Skript.

```
{
  "directory_id": "d-abcdef01234567890",
  "dns_ip_addrs": ["192.0.2.254", "203.0.113.237"],
  "subnet1_id": "subnet-021345abcdef6789",
  "subnet2_id": "subnet-1234567890abcdef0",
  "vpc_id": "vpc-021345abcdef6789"
}
```


Notieren Sie sich die Namen und IDs der Ausgaberesourcen. Sie verwenden sie in späteren Schritten.

Fahren Sie nach Abschluss des Skripts mit dem nächsten Schritt fort.

Erstellen Sie eine EC2-Instanz

New EC2 console

1. Melden Sie sich an der AWS Management Console an.
2. [Wenn Sie keine Rolle haben, der die in Schritt 4 aufgeführten Richtlinien zugeordnet sind, öffnen Sie die IAM-Konsole unter <https://console.aws.amazon.com/iam/>](#). Fahren Sie andernfalls mit Schritt 5 fort.
3. Erstellen Sie die ResetUserPassword Richtlinie und ersetzen Sie den rot hervorgehobenen Inhalt durch Ihre AWS-Region ID, Konto-ID und die Verzeichnis-ID aus der Ausgabe des Skripts, das Sie zur Erstellung des AD ausgeführt haben.

ResetUserPassword

```
{
  "Statement": [
    {
      "Action": [
        "ds:ResetUserPassword"
      ],
      "Resource": "arn:aws:ds:region-id:123456789012:directory/d-abcdef01234567890",
      "Effect": "Allow"
    }
  ]
}
```

4. Erstellen Sie eine IAM-Rolle mit den folgenden angehängten Richtlinien.
 - [AWS verwaltete Richtlinie: AmazonSSM. ManagedInstanceCore](#)
 - [AWS verwaltete Richtlinie AmazonSSM DirectoryServiceAccess](#)
 - ResetUserPassword Richtlinie
5. Öffnen Sie die Amazon-EC2-Konsole unter <https://console.aws.amazon.com/ec2/>.
6. Wählen Sie im EC2-Dashboard Launch Instance aus.

7. Wählen Sie unter Anwendungs- und Betriebssystemimages ein aktuelles Amazon Linux 2-AMI aus.
8. Wählen Sie als Instance-Typ t2.micro aus.
9. Wählen Sie für key pair ein Schlüsselpaar aus.
10. Wählen Sie für Netzwerkeinstellungen die Option Bearbeiten aus.
11. Wählen Sie für VPC das Verzeichnis VPC aus.
12. Scrollen Sie nach unten und wählen Sie Erweiterte Details aus.
13. Wählen Sie unter Erweiterte Details unter Verzeichnis für Domänenbeitritt die Option **auscorp.example.com**.
14. Wählen Sie für das IAM-Instanzprofil die Rolle aus, die Sie in Schritt 1 erstellt haben, oder eine Rolle, der die in Schritt 4 aufgeführten Richtlinien beigefügt sind.
15. Wählen Sie unter Zusammenfassung die Option Launch instance aus.
16. Notieren Sie sich die Instanz-ID (z. B. i-1234567890abcdef0) und warten Sie, bis der Start der Instance abgeschlossen ist.
17. Fahren Sie nach dem Start der Instance mit dem nächsten Schritt fort.

Old EC2 console

1. Melden Sie sich an der AWS Management Console an.
2. Wenn Sie keine Rolle haben, der die in Schritt 4 aufgeführten Richtlinien zugeordnet sind, öffnen Sie die IAM-Konsole unter <https://console.aws.amazon.com/iam/>. Fahren Sie andernfalls mit Schritt 5 fort.
3. Erstellen Sie die ResetUserPassword Richtlinie. Ersetzen Sie den rot hervorgehobenen Inhalt durch Ihre AWS-Region AWS-Konto ID, ID und die Verzeichnis-ID aus der Ausgabe des Skripts, das Sie zur Erstellung des Active Directory (AD) ausgeführt haben.

ResetUserPassword

```
{
  "Statement": [
    {
      "Action": [
        "ds:ResetUserPassword"
      ],
      "Resource": "arn:aws:ds:region-id:123456789012:directory/d-
        abcdef01234567890",
```

```
    "Effect": "Allow"
  }
]
}
```

4. Erstellen Sie eine IAM-Rolle mit den folgenden angehängten Richtlinien.
 - [AWSverwaltete Richtlinie: AmazonSSM. ManagedInstanceCore](#)
 - [AWSverwaltete Richtlinie AmazonSSM DirectoryServiceAccess](#)
 - ResetUserPassword-Richtlinie
5. Öffnen Sie die Amazon-EC2-Konsole unter <https://console.aws.amazon.com/ec2/>.
6. Wählen Sie im EC2-Dashboard Launch Instance aus.
7. Wählen Sie unter Anwendungs- und Betriebssystemimages ein aktuelles Amazon Linux 2-AMI aus.
8. Wählen Sie für Instance type die Option t2.micro aus.
9. Wählen Sie für key pair ein Schlüsselpaar aus.
10. Wählen Sie für Netzwerkeinstellungen die Option Bearbeiten aus.
11. Wählen Sie unter Netzwerkeinstellungen, VPC, das Verzeichnis VPC aus.
12. Scrollen Sie nach unten und wählen Sie Erweiterte Details aus.
13. Wählen Sie unter Erweiterte Details unter Verzeichnis für Domänenbeitritt die Option **auscorp.example.com**.
14. Wählen Sie unter Erweiterte Details, Instanzprofil, die Rolle aus, die Sie in Schritt 1 erstellt haben, oder eine Rolle, der die in Schritt 4 aufgeführten Richtlinien beigefügt sind.
15. Wählen Sie unter Zusammenfassung die Option Instanz starten aus.
16. Notieren Sie sich die Instanz-ID (z. B. i-1234567890abcdef0) und warten Sie, bis der Start der Instance abgeschlossen ist.
17. Fahren Sie nach dem Start der Instance mit dem nächsten Schritt fort.

Verbinden Sie Ihre Instance mit dem AD

1. Connect zu Ihrer Instance her und treten Sie dem AD-Realm als **beiadmin**.

Führen Sie die folgenden Befehle aus, um eine Verbindung mit der Instanz herzustellen.

```
$ INSTANCE_ID="i-1234567890abcdef0"
```

```
$ PUBLIC_IP=$(aws ec2 describe-instances \  
--instance-ids $INSTANCE_ID \  
--query "Reservations[0].Instances[0].PublicIpAddress" \  
--output text)
```

```
$ ssh -i ~/.ssh/keys/keypair.pem ec2-user@$PUBLIC_IP
```

2. Installieren Sie die erforderliche Software und treten Sie dem Realm bei.

```
$ sudo yum -y install sssd realmd oddjob oddjob-mkhomedir adcli samba-common samba-  
common-tools krb5-workstation openldap-clients polycycoreutils-python
```

3. Ersetzen Sie das Admin-Passwort durch Ihr **admin** Passwort.

```
$ ADMIN_PW="asdfASDF1234"
```

```
$ echo $ADMIN_PW | sudo realm join -U Admin corp.example.com  
Password for Admin:
```

Wenn der vorherige Vorgang erfolgreich war, sind Sie dem Realm beigetreten und können mit dem nächsten Schritt fortfahren.

Fügen Sie Benutzer zum AD hinzu

1. Erstellen Sie den ReadOnlyUser und einen weiteren Benutzer.

In diesem Schritt verwenden Sie die Tools [adcli](#) und [openldap-Clients](#), die Sie in einem vorherigen Schritt installiert haben.

```
$ echo $ADMIN_PW | adcli create-user -x -U Admin --domain=corp.example.com --  
display-name=ReadOnlyUser ReadOnlyUser
```

```
$ echo $ADMIN_PW | adcli create-user -x -U Admin --domain=corp.example.com --  
display-name=user000 user000
```

2. Stellen Sie sicher, dass die Benutzer erstellt wurden:

Die Verzeichnis-DNS-IP-Adressen sind Ausgaben des Python-Skripts.

```
$ DIRECTORY_IP="192.0.2.254"
```

```
$ ldapsearch -x -h $DIRECTORY_IP -D Admin -w $ADMIN_PW -b
"cn=ReadOnlyUser,ou=Users,ou=CORP,dc=corp,dc=example,dc=com"
```

```
$ ldapsearch -x -h $DIRECTORY_IP -D Admin -w $ADMIN_PW -b
"cn=user000,ou=Users,ou=CORP,dc=corp,dc=example,dc=com"
```

Wenn Sie einen Benutzer mit dem erstellenad-cli, ist der Benutzer standardmäßig deaktiviert.

3. Setzen Sie die Benutzerkennwörter von Ihrem lokalen Computer aus zurück und aktivieren Sie sie:

Melden Sie sich von Ihrer EC2-Instance ab.

Note

- ro-p@ssw0rd ist das Passwort von ReadOnlyUser, abgerufen von AWS Secrets Manager.
- user-p@ssw0rd ist das Passwort eines Cluster-Benutzers, das angegeben wird, wenn Sie eine Verbindung (ssh) mit dem Cluster herstellen.

Das directory-id ist eine Ausgabe des Python-Skripts.

```
$ DIRECTORY_ID="d-abcdef01234567890"
```

```
$ aws ds reset-user-password \
--directory-id $DIRECTORY_ID \
--user-name "ReadOnlyUser" \
--new-password "ro-p@ssw0rd" \
--region "region-id"
```

```
$ aws ds reset-user-password \
--directory-id $DIRECTORY_ID \
```

```
--user-name "user000" \  
--new-password "user-p@ssw0rd" \  
--region "region-id"
```

4. Fügen Sie das Passwort zu einem Secrets Manager Manager-Geheimnis hinzu.

Nachdem Sie nun ein Passwort erstellt `ReadOnlyUser` und festgelegt haben, speichern Sie es in einem geheimen Ordner, der für die Validierung von Anmeldungen AWS ParallelCluster verwendet wird.

Verwenden Sie Secrets Manager, um ein neues Geheimnis zu erstellen, das das Passwort für `ReadOnlyUser` als Wert enthält. Das geheime Wertformat darf nur Klartext sein (kein JSON-Format). Notieren Sie sich den geheimen ARN für future Schritte.

```
$ aws secretsmanager create-secret --name "ADSecretPassword" \  
--region region_id \  
--secret-string "ro-p@ssw0rd" \  
--query ARN \  
--output text  
arn:aws:secretsmanager:region-id:123456789012:secret:ADSecretPassword-1234
```

Einrichtung von LDAPS mit Zertifikatsverifizierung (empfohlen)

Notieren Sie sich die Ressourcen-IDs. Sie verwenden sie später schrittweise.

1. Generieren Sie lokal ein Domänenzertifikat.

```
$ PRIVATE_KEY="corp-example-com.key"  
CERTIFICATE="corp-example-com.crt"  
printf ".\n.\n.\n.\n.\n.\ncorp.example.com\n.\n" | openssl req -x509 -sha256 -nodes -  
newkey rsa:2048 -keyout $PRIVATE_KEY -days 365 -out $CERTIFICATE
```

2. Speichern Sie das Zertifikat in Secrets Manager, damit es später innerhalb des Clusters abgerufen werden kann.

```
$ aws secretsmanager create-secret --name example-cert \  
--secret-string file://$CERTIFICATE \  
--region region-id  
{  
  "ARN": "arn:aws:secretsmanager:region-id:123456789012:secret:example-  
cert-123abc",
```

```

    "Name": "example-cert",
    "VersionId": "14866070-092a-4d5a-bcdd-9219d0566b9c"
  }

```

3. Fügen Sie der IAM-Rolle, die Sie erstellt haben, um die EC2-Instance mit der AD-Domäne zu verbinden, die folgende Richtlinie hinzu.

PutDomainCertificateSecrets

```

{
  "Statement": [
    {
      "Action": [
        "secretsmanager:PutSecretValue"
      ],
      "Resource": [
        "arn:aws:secretsmanager:region-id:123456789012:secret:example-cert-123abc",
      ],
      "Effect": "Allow"
    }
  ]
}

```

4. Importieren Sie das Zertifikat in AWS Certificate Manager (ACM).

```

$ aws acm import-certificate --certificate fileb://$CERTIFICATE \
  --private-key fileb://$PRIVATE_KEY \
  --region region-id
{
  "CertificateArn": "arn:aws:acm:region-
id:123456789012:certificate/343db133-490f-4077-b8d4-3da5bfd89e72"
}

```

5. Erstellen Sie und den Load Balancer, der vor den Active Directory-Endpunkten platziert wird.

```

$ aws elbv2 create-load-balancer --name CorpExampleCom-NLB \
  --type network \
  --scheme internal \
  --subnets subnet-1234567890abcdef0 subnet-021345abcdef6789 \
  --region region-id
{
  "LoadBalancers": [

```

```

{
  "LoadBalancerArn": "arn:aws:elasticloadbalancing:region-
id:123456789012:loadbalancer/net/CorpExampleCom-NLB/3afe296bf4ba80d4",
  "DNSName": "CorpExampleCom-NLB-3afe296bf4ba80d4.elb.region-id.amazonaws.com",
  "CanonicalHostedZoneId": "Z2IF0LAFXWL04F",
  "CreatedTime": "2022-05-05T12:56:55.988000+00:00",
  "LoadBalancerName": "CorpExampleCom-NLB",
  "Scheme": "internal",
  "VpcId": "vpc-021345abcdef6789",
  "State": {
    "Code": "provisioning"
  },
  "Type": "network",
  "AvailabilityZones": [
    {
      "ZoneName": "region-idb",
      "SubnetId": "subnet-021345abcdef6789",
      "LoadBalancerAddresses": []
    },
    {
      "ZoneName": "region-ida",
      "SubnetId": "subnet-1234567890abcdef0",
      "LoadBalancerAddresses": []
    }
  ],
  "IpAddressType": "ipv4"
}
]
}

```

6. Erstellen Sie die Zielgruppe, die auf die Active Directory-Endpunkte abzielt.

```

$ aws elbv2 create-target-group --name CorpExampleCom-Targets --protocol TCP \
--port 389 \
--target-type ip \
--vpc-id vpc-021345abcdef6789 \
--region region-id
{
  "TargetGroups": [
    {
      "TargetGroupArn": "arn:aws:elasticloadbalancing:region-
id:123456789012:targetgroup/CorpExampleCom-Targets/44577c583b695e81",
      "TargetGroupName": "CorpExampleCom-Targets",
      "Protocol": "TCP",

```



```

    "Port": 389,
    "VpcId": "vpc-021345abcdef6789",
    "HealthCheckProtocol": "TCP",
    "HealthCheckPort": "traffic-port",
    "HealthCheckEnabled": true,
    "HealthCheckIntervalSeconds": 30,
    "HealthCheckTimeoutSeconds": 10,
    "HealthyThresholdCount": 3,
    "UnhealthyThresholdCount": 3,
    "TargetType": "ip",
    "IpAddressType": "ipv4"
  }
]
}

```

7. Registrieren Sie die Active Directory (AD) -Endpunkte in der Zielgruppe.

```

$ aws elbv2 register-targets --target-group-arn
arn:aws:elasticloadbalancing:region-id:123456789012:targetgroup/CorpExampleCom-
Targets/44577c583b695e81 \
--targets Id=192.0.2.254,Port=389 Id=203.0.113.237,Port=389 \
--region region-id

```

8. Erstellen Sie den LB-Listener mit dem Zertifikat.

```

$ aws elbv2 create-listener --load-balancer-arn
arn:aws:elasticloadbalancing:region-id:123456789012:loadbalancer/net/
CorpExampleCom-NLB/3afe296bf4ba80d4 \
--protocol TLS \
--port 636 \
--default-actions
Type=forward,TargetGroupArn=arn:aws:elasticloadbalancing:region-
id:123456789012:targetgroup/CorpExampleCom-Targets/44577c583b695e81 \
--ssl-policy ELBSecurityPolicy-TLS-1-2-2017-01 \
--certificates CertificateArn=arn:aws:acm:region-
id:123456789012:certificate/343db133-490f-4077-b8d4-3da5bfd89e72 \
--region region-id
"Listeners": [
{
  "ListenerArn": "arn:aws:elasticloadbalancing:region-id:123456789012:listener/
net/CorpExampleCom-NLB/3afe296bf4ba80d4/a8f9d97318743d4b",
  "LoadBalancerArn": "arn:aws:elasticloadbalancing:region-
id:123456789012:loadbalancer/net/CorpExampleCom-NLB/3afe296bf4ba80d4",

```

```

    "Port": 636,
    "Protocol": "TLS",
    "Certificates": [
      {
        "CertificateArn": "arn:aws:acm:region-
id:123456789012:certificate/343db133-490f-4077-b8d4-3da5bfd89e72"
      }
    ],
    "SslPolicy": "ELBSecurityPolicy-TLS-1-2-2017-01",
    "DefaultActions": [
      {
        "Type": "forward",
        "TargetGroupArn": "arn:aws:elasticloadbalancing:region-
id:123456789012:targetgroup/CorpExampleCom-Targets/44577c583b695e81",
        "ForwardConfig": {
          "TargetGroups": [
            {
              "TargetGroupArn": "arn:aws:elasticloadbalancing:region-
id:123456789012:targetgroup/CorpExampleCom-Targets/44577c583b695e81"
            }
          ]
        }
      }
    ]
  }
}

```

- Erstellen Sie die Hosting-Zone, damit die Domain innerhalb der Cluster-VPC auffindbar ist.

```

$ aws route53 create-hosted-zone --name corp.example.com \
  --vpc VPCRegion=region-id,VPCId=vpc-021345abcdef6789 \
  --caller-reference "ParallelCluster AD Tutorial"
{
  "Location": "https://route53.amazonaws.com/2013-04-01/hostedzone/
Z09020002B5MZQNXMSJUB",
  "HostedZone": {
    "Id": "/hostedzone/Z09020002B5MZQNXMSJUB",
    "Name": "corp.example.com.",
    "CallerReference": "ParallelCluster AD Tutorial",
    "Config": {
      "PrivateZone": true
    }
  },
  "ResourceRecordSetCount": 2
}

```

```

},
"ChangeInfo": {
  "Id": "/change/C05533343BF3IKSORW1TQ",
  "Status": "PENDING",
  "SubmittedAt": "2022-05-05T13:21:53.863000+00:00"
},
"VPC": {
  "VPCRegion": "region-id",
  "VPCId": "vpc-021345abcdef6789"
}
}

```

10. Erstellen Sie eine Datei, die **recordset-change.json** den folgenden Inhalt hat. **HostedZoneId** ist die kanonische Hosting-Zonen-ID des Load Balancers.

```

{
  "Changes": [
    {
      "Action": "CREATE",
      "ResourceRecordSet": {
        "Name": "corp.example.com",
        "Type": "A",
        "Region": "region-id",
        "SetIdentifier": "example-active-directory",
        "AliasTarget": {
          "HostedZoneId": "Z2IF0LAFXWL04F",
          "DNSName": "CorpExampleCom-NLB-3afe296bf4ba80d4.elb.region-
id.amazonaws.com",
          "EvaluateTargetHealth": true
        }
      }
    }
  ]
}

```

11. Sendet die Änderung der Datensatzgruppe an die Hosting-Zone, diesmal unter Verwendung der Hosting-Zonen-ID.

```

$ aws route53 change-resource-record-sets --hosted-zone-id Z09020002B5MZQNMSJUB \
--change-batch file://recordset-change.json
{
  "ChangeInfo": {
    "Id": "/change/C0137926I56R3GC7XW2Y",

```

```
"Status": "PENDING",
"SubmittedAt": "2022-05-05T13:40:36.553000+00:00"
}
}
```

12. Erstellen Sie ein Richtliniendokument **policy.json** mit dem folgenden Inhalt.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Action": [
        "secretsmanager:GetSecretValue"
      ],
      "Resource": [
        "arn:aws:secretsmanager:region-id:123456789012:secret:example-cert-abc123"
      ],
      "Effect": "Allow"
    }
  ]
}
```

13. Erstellen Sie ein Richtliniendokument, das **policy.json** den folgenden Inhalt hat.

```
$ aws iam create-policy --policy-name ReadCertExample \
  --policy-document file://policy.json
{
  "Policy": {
    "PolicyName": "ReadCertExample",
    "PolicyId": "ANPAUUXUVBC42VZSI4LDY",
    "Arn": "arn:aws:iam::123456789012:policy/ReadCertExample-efg456",
    "Path": "/",
    "DefaultVersionId": "v1",
    "AttachmentCount": 0,
    "PermissionsBoundaryUsageCount": 0,
    "IsAttachable": true,
    "CreateDate": "2022-05-05T13:42:18+00:00",
    "UpdateDate": "2022-05-05T13:42:18+00:00"
  }
}
```

14. Folgen Sie weiterhin den Schritten unter [\(Optional\) Schritt 2: AD-Benutzer und -Gruppen verwalten](#) oder [Schritt 3: Erstellen Sie den Cluster](#).

(Optional) Schritt 2: AD-Benutzer und -Gruppen verwalten

In diesem Schritt verwalten Sie Benutzer und Gruppen von einer EC2 Amazon Linux 2-Instance aus, die mit der Active Directory (AD) -Domäne verbunden ist.

Wenn Sie dem automatisierten Pfad gefolgt sind, starten Sie die AD-verbundene Instance neu und melden Sie sich bei ihr an, die im Rahmen der Automatisierung erstellt wurde.

Wenn Sie dem manuellen Pfad gefolgt sind, starten Sie die Instanz neu und melden Sie sich bei der Instanz an, die Sie in den vorherigen Schritten erstellt und dem AD hinzugefügt haben.

In diesen Schritten verwenden Sie die Tools [adcli](#) und [openldap-clients](#), die im Rahmen eines vorherigen Schritts in der Instanz installiert wurden.

Melden Sie sich bei einer EC2-Instance an, die mit der AD-Domäne verbunden ist

1. Wählen Sie in der EC2-Konsole die unbenannte EC2-Instance aus, die in den vorherigen Schritten erstellt wurde. Der Instanzstatus könnte „Gestoppt“ lauten.
2. Wenn der Instanzstatus Gestoppt lautet, wählen Sie Instanzstatus und dann Instanz starten aus.
3. Nachdem die Statusprüfungen bestanden wurden, wählen Sie die Instance aus und klicken Sie auf Connect and SSH in the instance.

Benutzer und Gruppen verwalten, wenn Sie bei einer EC2 Amazon Linux 2-Instance angemeldet sind, die dem AD beigetreten ist

Wenn Sie die `adcli` Befehle mit der `-U "Admin"` Option ausführen, werden Sie aufgefordert, das Admin AD-Passwort einzugeben. Sie geben das Admin AD-Passwort als Teil der `ldapsearch` Befehle an.

1. Erstellen eines Benutzers.

```
$ adcli create-user "clusteruser" --domain "corp.example.com" -U "Admin"
```

- Legen Sie ein Benutzerkennwort fest.

```
$ aws --region "region-id" ds reset-user-password --directory-id "d-
abcdef01234567890" --user-name "clusteruser" --new-password "new-p@ssw0rd"
```

- Erstellen Sie eine Gruppe.

```
$ adcli create-group "clusterteam" --domain "corp.example.com" -U "Admin"
```

- Fügt einen Benutzer zu einer Gruppe hinzu.

```
$ adcli add-member "clusterteam" "clusteruser" --domain "corp.example.com" -U
"Admin"
```

- Beschreiben Sie Benutzer und Gruppen.

Beschreiben Sie alle Benutzer.

```
$ ldapsearch "(&(objectClass=user))" -x -h "192.0.2.254" -b
"DC=corp,DC=example,DC=com" -D
"CN=Admin,OU=Users,OU=CORP,DC=corp,DC=example,DC=com" -w "p@ssw0rd"
```

Beschreiben Sie einen bestimmten Benutzer.

```
$ ldapsearch "(&(objectClass=user)(cn=clusteruser))"
-x -h "192.0.2.254" -b "DC=corp,DC=example,DC=com" -D
"CN=Admin,OU=Users,OU=CORP,DC=corp,DC=example,DC=com" -w "p@ssw0rd"
```

Beschreiben Sie alle Benutzer mit einem Namensmuster.

```
$ ldapsearch "(&(objectClass=user)(cn=user*))" -x -h "192.0.2.254" -b
"DC=corp,DC=example,DC=com" -D
"CN=Admin,OU=Users,OU=CORP,DC=corp,DC=example,DC=com" -w "p@ssw0rd"
```

Beschreiben Sie alle Benutzer, die Teil einer bestimmten Gruppe sind.

```
$ ldapsearch "(&(objectClass=user)
(memberOf=CN=clusterteam,OU=Users,OU=CORP,DC=corp,DC=example,DC=com))"
-x -h "192.0.2.254" -b "DC=corp,DC=example,DC=com" -D
"CN=Admin,OU=Users,OU=CORP,DC=corp,DC=example,DC=com" -w "p@ssw0rd"
```

Beschreiben Sie alle Gruppen

```
$ ldapsearch "objectClass=group" -x -h "192.0.2.254" -b "DC=corp,DC=example,DC=com"
-D "CN=Admin,OU=Users,OU=CORP,DC=corp,DC=example,DC=com" -w "p@ssw0rd"
```

Beschreiben Sie eine bestimmte Gruppe

```
$ ldapsearch "(&(objectClass=group)(cn=clusterteam))"
-x -h "192.0.2.254" -b "DC=corp,DC=example,DC=com" -D
"CN=Admin,OU=Users,OU=CORP,DC=corp,DC=example,DC=com" -w "p@ssw0rd"
```

6. Einen Benutzer aus einer Gruppe entfernen.

```
$ adcli remove-member "clusterteam" "clusteruser" --domain "corp.example.com" -U
"Admin"
```

7. Einen Benutzer löschen.

```
$ adcli delete-user "clusteruser" --domain "corp.example.com" -U "Admin"
```

8. Löscht eine Gruppe.

```
$ adcli delete-group "clusterteam" --domain "corp.example.com" -U "Admin"
```

Schritt 3: Erstellen Sie den Cluster

Wenn Sie die EC2-Instance nicht verlassen haben, tun Sie dies jetzt.

Die Umgebung ist so eingerichtet, dass ein Cluster erstellt wird, der Benutzer anhand des Active Directory (AD) authentifizieren kann.

Erstellen Sie eine einfache Clusterkonfiguration und geben Sie die Einstellungen an, die für die Verbindung mit dem AD relevant sind. Weitere Informationen finden Sie im Abschnitt [DirectoryService](#).

Wählen Sie eine der folgenden Clusterkonfigurationen aus und kopieren Sie sie in eine Datei mit dem Namen `ldaps_config.yaml`, `ldaps_nocert_config.yaml`, oder `ldap_config.yaml`.

Wir empfehlen Ihnen, die LDAPS-Konfiguration mit Zertifikatüberprüfung zu wählen. Wenn Sie diese Konfiguration wählen, müssen Sie auch das Bootstrap-Skript in eine Datei mit dem Namen `active-directory.head.post.sh` kopieren und Sie müssen es in einem Amazon S3 S3-Bucket speichern, wie in der Konfigurationsdatei angegeben.

LDAPS mit Konfiguration zur Zertifikatsverifizierung (empfohlen)

Note

Die folgenden Komponenten müssen geändert werden.

- `KeyName`: Eines Ihrer EC2-Tastenpaare.
- `SubnetId` / `SubnetIds`: Eine der Subnetz-IDs, die in der Ausgabe des CloudFormation Quick Create Stacks (automatisiertes Tutorial) oder des Python-Skripts (manuelles Tutorial) angegeben wurden.
- `Region`: Die Region, in der Sie die AD-Infrastruktur erstellt haben.
- `DomainAddr`: Diese IP-Adresse ist eine der DNS-Adressen Ihres AD-Dienstes.
- `PasswordSecretArn`: Der Amazon-Ressourcenname (ARN) des Geheimnisses, das das Passwort für den `enthältDomainReadOnlyUser` enthält.
- `BucketName`: Der Name des Buckets, der das Bootstrap-Skript enthält.
- `AdditionalPolicies/Policy`: Der Amazon-Ressourcenname (ARN) der gelesenen Domain-Zertifizierungsrichtlinie `ReadCertExample`.
- `CustomActions/OnNodeConfigured/Args`: Der geheime Amazon-Ressourcenname (ARN), der die Domain-Zertifizierungsrichtlinie enthält.

Aus Sicherheitsgründen empfehlen wir, die `AllowedIps` Konfiguration `HeadNode/Ssh` zu verwenden, um den SSH-Zugriff auf den Hauptknoten zu beschränken.

```
Region: region-id
Image:
  Os: alinux2
HeadNode:
  InstanceType: t2.micro
Networking:
  SubnetId: subnet-abcdef01234567890
```



```
Ssh:
  KeyName: keypair
Iam:
  AdditionalIamPolicies:
    - Policy: arn:aws:iam::123456789012:policy/ReadCertExample
  S3Access:
    - BucketName: my-bucket
      EnableWriteAccess: false
      KeyName: bootstrap/active-directory/active-directory.head.post.sh
  CustomActions:
    OnNodeConfigured:
      Script: s3://my-bucket/bootstrap/active-directory/active-directory.head.post.sh
      Args:
        - arn:aws:secretsmanager:region-id:123456789012:secret:example-cert-123abc
        - /opt/parallelcluster/shared/directory_service/domain-certificate.crt
  Scheduling:
    Scheduler: slurm
    SlurmQueues:
      - Name: queue0
        ComputeResources:
          - Name: queue0-t2-micro
            InstanceType: t2.micro
            MinCount: 1
            MaxCount: 10
        Networking:
          SubnetIds:
            - subnet-abcdef01234567890
  DirectoryService:
    DomainName: corp.example.com
    DomainAddr: ldaps://corp.example.com
    PasswordSecretArn: arn:aws:secretsmanager:region-id:123456789012:secret:ADSecretPassword-1234
    DomainReadOnlyUser: cn=ReadOnlyUser,ou=Users,ou=CORP,dc=corp,dc=example,dc=com
    LdapTlsCaCert: /opt/parallelcluster/shared/directory_service/domain-certificate.crt
    LdapTlsReqCert: hard
```

Bootstrap-Skript

Nachdem Sie die Bootstrap-Datei erstellt haben und bevor Sie sie in Ihren S3-Bucket hochladen, führen Sie den Befehl aus, `chmod +x active-directory.head.post.sh` um die Ausführungsberechtigung zu erteilenAWS ParallelCluster.

```
#!/bin/bash
```

```
set -e

CERTIFICATE_SECRET_ARN="$1"
CERTIFICATE_PATH="$2"

[[ -z $CERTIFICATE_SECRET_ARN ]] && echo "[ERROR] Missing CERTIFICATE_SECRET_ARN" &&
exit 1
[[ -z $CERTIFICATE_PATH ]] && echo "[ERROR] Missing CERTIFICATE_PATH" && exit 1

source /etc/parallelcluster/cfnconfig
REGION="${cfn_region:?}"

mkdir -p $(dirname $CERTIFICATE_PATH)
aws secretsmanager get-secret-value --region $REGION --secret-id
$CERTIFICATE_SECRET_ARN --query SecretString --output text > $CERTIFICATE_PATH
```

LDAPS ohne Konfiguration zur Zertifikatsverifizierung

Note

Die folgenden Komponenten müssen geändert werden.

- **KeyName:** Eines Ihrer EC2-Tastenpaare.
- **SubnetId / SubnetIds:** Eine der Subnetz-IDs, die in der Ausgabe des CloudFormation Quick Create Stacks (automatisiertes Tutorial) oder des Python-Skripts (manuelles Tutorial) enthalten sind.
- **Region:** Die Region, in der Sie die AD-Infrastruktur erstellt haben.
- **DomainAddr:** Diese IP-Adresse ist eine der DNS-Adressen Ihres AD-Dienstes.
- **PasswordSecretArn:** Der Amazon-Ressourcenname (ARN) des Geheimnisses, das das Passwort für den `enthältDomainReadOnlyUser` enthält.

Aus Sicherheitsgründen empfehlen wir, die `AllowedIps` Konfiguration `HeadNode /Ssh/` zu verwenden, um den SSH-Zugriff auf den Hauptknoten zu beschränken.

```
Region: region-id
Image:
Os: alinux2
```

```
HeadNode:
  InstanceType: t2.micro
  Networking:
    SubnetId: subnet-abcdef01234567890
  Ssh:
    KeyName: keypair
Scheduling:
  Scheduler: slurm
  SlurmQueues:
    - Name: queue0
      ComputeResources:
        - Name: queue0-t2-micro
          InstanceType: t2.micro
          MinCount: 1
          MaxCount: 10
      Networking:
        SubnetIds:
          - subnet-abcdef01234567890
DirectoryService:
  DomainName: corp.example.com
  DomainAddr: ldaps://corp.example.com
  PasswordSecretArn: arn:aws:secretsmanager:region-
id:123456789012:secret:ADSecretPassword-1234
  DomainReadOnlyUser: cn=ReadOnlyUser,ou=Users,ou=CORP,dc=corp,dc=example,dc=com
  LdapTlsReqCert: never
```

LDAP-Konfiguration

Note

Die folgenden Komponenten müssen geändert werden.

- `KeyName`: Eines Ihrer EC2-Tastenpaare.
- `SubnetId` / `SubnetIds`: Eine der Subnetz-IDs, die in der Ausgabe des CloudFormation Quick Create Stacks (automatisiertes Tutorial) oder des Python-Skripts (manuelles Tutorial) angegeben wurden.
- `Region`: Die Region, in der Sie die AD-Infrastruktur erstellt haben.
- `DomainAddr`: Diese IP-Adresse ist eine der DNS-Adressen Ihres AD-Dienstes.
- `PasswordSecretArn`: Der Amazon-Ressourcenname (ARN) des Geheimnisses, das das Passwort für den `enthältDomainReadOnlyUser`.

Aus Sicherheitsgründen empfehlen wir, die AllowedIps Konfiguration HeadNode /Ssh/ zu verwenden, um den SSH-Zugriff auf den Hauptknoten zu beschränken.

```

Region: region-id
Image:
  Os: alinux2
HeadNode:
  InstanceType: t2.micro
  Networking:
    SubnetId: subnet-abcdef01234567890
  Ssh:
    KeyName: keypair
Scheduling:
  Scheduler: slurm
  SlurmQueues:
    - Name: queue0
      ComputeResources:
        - Name: queue0-t2-micro
          InstanceType: t2.micro
          MinCount: 1
          MaxCount: 10
      Networking:
        SubnetIds:
          - subnet-abcdef01234567890
DirectoryService:
  DomainName: dc=corp,dc=example,dc=com
  DomainAddr: ldap://192.0.2.254,ldap://203.0.113.237
  PasswordSecretArn: arn:aws:secretsmanager:region-id:123456789012:secret:ADSecretPassword-1234
  DomainReadOnlyUser: cn=ReadOnlyUser,ou=Users,ou=CORP,dc=corp,dc=example,dc=com
  AdditionalSssdConfigs:
    ldap_auth_disable_tls_never_use_in_production: True

```

Erstellen Sie den Cluster mit dem folgenden Befehl.

```

$ pcluster create-cluster --cluster-name "ad-cluster" --cluster-configuration "./ldaps_config.yaml"
{
  "cluster": {
    "clusterName": "pcluster",

```

```
"cloudformationStackStatus": "CREATE_IN_PROGRESS",
"cloudformationStackArn": "arn:aws:cloudformation:region-id:123456789012:stack/ad-cluster/1234567-abcd-0123-def0-abcdef0123456",
"region": "region-id",
"version": 3.7.0,
"clusterStatus": "CREATE_IN_PROGRESS"
}
}
```

Schritt 4: Stellen Sie als Benutzer eine Verbindung zum Cluster her

Sie können den Status des Clusters mit den folgenden Befehlen ermitteln.

```
$ pcluster describe-cluster -n ad-cluster --region "region-id" --query "clusterStatus"
```

Die Ausgabe sieht wie folgt aus.

```
"CREATE_IN_PROGRESS" / "CREATE_COMPLETE"
```

Wenn der Status erreicht ist "CREATE_COMPLETE", melden Sie sich mit dem erstellten Benutzernamen und Passwort an.

```
$ HEAD_NODE_IP=$(pcluster describe-cluster -n "ad-cluster" --region "region-id" --query headNode.publicIpAddress | xargs echo)
```

```
$ ssh user000@$HEAD_NODE_IP
```

Sie können sich auch ohne das Passwort anmelden, indem Sie den SSH Schlüssel angeben, der für den neuen Benutzer erstellt wurde /home/user000@HEAD_NODE_IP/.ssh/id_rsa.

Wenn der ssh Befehl erfolgreich war, haben Sie als Benutzer, der für die Verwendung von Active Directory (AD) authentifiziert wurde, erfolgreich eine Verbindung zum Cluster hergestellt.

Schritt 5: Bereinigen

1. Löschen Sie den Cluster von Ihrem lokalen Computer.

```
$ pcluster delete-cluster --cluster-name "ad-cluster" --region "region-id"
```

```
{
  "cluster": {
    "clusterName": "ad-cluster",
    "cloudformationStackStatus": "DELETE_IN_PROGRESS",
    "cloudformationStackArn": "arn:aws:cloudformation:region-id:123456789012:stack/ad-cluster/1234567-abcd-0123-def0-abcdef0123456",
    "region": "region-id",
    "version": "3.7.0",
    "clusterStatus": "DELETE_IN_PROGRESS"
  }
}
```

2. Überprüfen Sie den Fortschritt beim Löschen des Clusters.

```
$ pcluster describe-cluster --cluster-name "ad-cluster" --region "region-id" --
query "clusterStatus"
"DELETE_IN_PROGRESS"
```

Nachdem der Cluster erfolgreich gelöscht wurde, fahren Sie mit dem nächsten Schritt fort.

Automatisiert

Löschen Sie die Active Directory-Ressourcen

1. Von <https://console.aws.amazon.com/cloudformation/>.
2. Klicken Sie im Navigationsbereich auf Stacks.
3. Wählen Sie aus der Liste der Stacks den AD-Stack aus (z. B. `pcluster-ad`).
4. Wählen Sie Löschen aus.

Manuell

1. Löschen Sie die EC2-Instance.
 - a. Wählen Sie [unter https://console.aws.amazon.com/ec2/](https://console.aws.amazon.com/ec2/) im Navigationsbereich Instances aus.
 - b. Wählen Sie aus der Liste der Instanzen die Instanz aus, die Sie erstellt haben, um dem Verzeichnis Benutzer hinzuzufügen.
 - c. Wählen Sie Instanzstatus und anschließend Instanz beenden aus.

2. Löschen Sie die gehostete Zone.

- a. Erstellen Sie eine `recordset-delete.json` mit dem folgenden Inhalt. In diesem Beispiel `HostedZoneId` ist dies die kanonische Hosting-Zonen-ID des Load Balancers.

```
{
  "Changes": [
    {
      "Action": "DELETE",
      "ResourceRecordSet": {
        "Name": "corp.example.com",
        "Type": "A",
        "Region": "region-id",
        "SetIdentifier": "pcluster-active-directory",
        "AliasTarget": {
          "HostedZoneId": "Z2IFOLAFXWL04F",
          "DNSName": "CorpExampleCom-NLB-3afe296bf4ba80d4.elb.region-id.amazonaws.com",
          "EvaluateTargetHealth": true
        }
      }
    }
  ]
}
```

- b. Übermitteln Sie die Änderung der Datensatzgruppe mithilfe der Hosting-Zonen-ID an die gehostete Zone.

```
$ aws route53 change-resource-record-sets --hosted-zone-id Z09020002B5MZQNXSJUB \
  --change-batch file://recordset-delete.json
{
  "ChangeInfo": {
    "Id": "/change/C04853642A0TH2TJ5NLNI",
    "Status": "PENDING",
    "SubmittedAt": "2022-05-05T14:25:51.046000+00:00"
  }
}
```

- c. Löschen Sie die gehostete Zone.

```
$ aws route53 delete-hosted-zone --id Z09020002B5MZQNXSJUB
{
```

```
"ChangeInfo": {
  "Id": "/change/C0468051QFABTVHMDEG9",
  "Status": "PENDING",
  "SubmittedAt": "2022-05-05T14:26:13.814000+00:00"
}
```

3. Löschen Sie den LB-Listener.

```
$ aws elbv2 delete-listener \
  --listener-arn arn:aws:elasticloadbalancing:region-id:123456789012:listener/net/
  CorpExampleCom-NLB/3afe296bf4ba80d4/a8f9d97318743d4b --region region-id
```

4. Löschen Sie die Zielgruppe.

```
$ aws elbv2 delete-target-group \
  --target-group-arn arn:aws:elasticloadbalancing:region-
id:123456789012:targetgroup/CorpExampleCom-Targets/44577c583b695e81 --
  region region-id
```

5. Löschen Sie den Load Balancer.

```
$ aws elbv2 delete-load-balancer \
  --load-balancer-arn arn:aws:elasticloadbalancing:region-
id:123456789012:loadbalancer/net/CorpExampleCom-NLB/3afe296bf4ba80d4 --
  region region-id
```

6. Löschen Sie die Richtlinie, die der Cluster verwendet, um das Zertifikat aus Secrets Manager zu lesen.

```
$ aws iam delete-policy --policy-arn arn:aws:iam::123456789012:policy/
  ReadCertExample
```

7. Löschen Sie das Geheimnis, das das Domänenzertifikat enthält.

```
$ aws secretsmanager delete-secret \
  --secret-id arn:aws:secretsmanager:region-id:123456789012:secret:example-
  cert-123abc \
  --region region-id
{
  "ARN": "arn:aws:secretsmanager:region-id:123456789012:secret:example-cert-123abc",
  "Name": "example-cert",
  "DeletionDate": "2022-06-04T16:27:36.183000+02:00"
```



```
}

```

8. Löschen Sie das Zertifikat aus ACM.

```
$ aws acm delete-certificate \
  --certificate-arn arn:aws:acm:region-id:123456789012:certificate/343db133-490f-4077-b8d4-3da5bfd89e72 --region region-id
```

9. Löschen Sie die Active Directory-Ressourcen (AD).

a. Rufen Sie die folgenden Ressourcen-IDs aus der Ausgabe des Python-Skripts `abad.py`:

- ANZEIGEN-ID
- AD-Subnetz-IDs
- UND VPC-ID

b. Löschen Sie das Verzeichnis, indem Sie den folgenden Befehl ausführen.

```
$ aws ds delete-directory --directory-id d-abcdef0123456789 --region region-id
{
  "DirectoryId": "d-abcdef0123456789"
}
```

c. Listet die Sicherheitsgruppen in der VPC auf.

```
$ aws ec2 describe-security-groups --filters '[{"Name":"vpc-id","Values":["vpc-07614ade95ebad1bc"]}]' --region region-id
```

d. Löschen Sie die benutzerdefinierte Sicherheitsgruppe.

```
$ aws ec2 delete-security-group --group-id sg-021345abcdef6789 --region region-id
```

e. Löschen Sie die Subnetze.

```
$ aws ec2 delete-subnet --subnet-id subnet-1234567890abcdef --region region-id
```

```
$ aws ec2 delete-subnet --subnet-id subnet-021345abcdef6789 --region region-id
```

f. Beschreiben Sie das Internet-Gateway.

```
$ aws ec2 describe-internet-gateways \
```

```

--filters Name=attachment.vpc-id,Values=vpc-021345abcdef6789 \
--region region-id
{
  "InternetGateways": [
    {
      "Attachments": [
        {
          "State": "available",
          "VpcId": "vpc-021345abcdef6789"
        }
      ],
      "InternetGatewayId": "igw-1234567890abcdef",
      "OwnerId": "123456789012",
      "Tags": []
    }
  ]
}

```

- g. Trennen Sie das Internet-Gateway.

```

$ aws ec2 detach-internet-gateway \
--internet-gateway-id igw-1234567890abcdef \
--vpc-id vpc-021345abcdef6789 \
--region region-id

```

- h. Löschen Sie das Internet-Gateway.

```

$ aws ec2 delete-internet-gateway \
--internet-gateway-id igw-1234567890abcdef \
--region region-id

```

- i. Löschen Sie die VPC.

```

$ aws ec2 delete-vpc \
--vpc-id vpc-021345abcdef6789 \
--region region-id

```

- j. Löschen Sie das Geheimnis, das das ReadOnlyUser Passwort enthält.

```

$ aws secretsmanager delete-secret \
--secret-id arn:aws:secretsmanager:region-
id:123456789012:secret:ADSecretPassword-1234" \

```

```
--region region-id
```

Konfiguration der Verschlüsselung von gemeinsam genutztem Speicher mit einem AWS KMS Schlüssel

Erfahren Sie, wie Sie einen vom Kunden verwalteten AWS KMS Schlüssel zur Verschlüsselung und zum Schutz Ihrer Daten in den Cluster-Dateispeichersystemen einrichten, für AWS ParallelCluster die sie konfiguriert sind.

Wenn Sie die AWS ParallelCluster Befehlszeilenschnittstelle (CLI) oder die API verwenden, zahlen Sie nur für die AWS Ressourcen, die beim Erstellen oder Aktualisieren von AWS ParallelCluster Images und Clustern erstellt oder aktualisiert werden. Weitere Informationen finden Sie unter [AWSDienste verwendet von AWS ParallelCluster](#).

Die AWS ParallelCluster Benutzeroberfläche basiert auf einer serverlosen Architektur und Sie können sie in den meisten Fällen innerhalb der Kategorie AWS Free Tier verwenden. Weitere Informationen finden Sie unter [AWS ParallelClusterDie Benutzeroberfläche kostet](#).

AWS ParallelClusterunterstützt die folgenden Konfigurationsoptionen für gemeinsam genutzten Speicher:

- [SharedStorage](#) / [EbsSettings](#) / [KmsKeyId](#)
- [SharedStorage](#) / [EfsSettings](#) / [KmsKeyId](#)
- [SharedStorage](#) / [FsxLustreSettings](#) / [KmsKeyId](#)

Sie können diese Optionen verwenden, um einen vom Kunden verwalteten AWS KMS Schlüssel für die Verschlüsselung gemeinsam genutzter Speichersysteme von Amazon EBS, Amazon EFS und FSx for Lustre bereitzustellen. Um sie zu verwenden, müssen Sie eine IAM-Richtlinie für Folgendes erstellen und konfigurieren:

- [HeadNode](#) / [Iam](#) / [AdditionalIamPolicies](#) / [Policy](#)
- [Scheduler](#) / [SlurmQueues](#) / [Iam](#) / [AdditionalIamPolicies](#) / [Policy](#)

Voraussetzungen

- AWS ParallelCluster [ist installiert](#).

- Der AWS CLI [ist installiert und konfiguriert](#).
- Sie haben ein [EC2-Schlüsselpaar](#).
- Sie haben eine IAM-Rolle mit den [Berechtigungen](#), die für die Ausführung der [pcluster](#) CLI erforderlich sind.

Themen

- [Erstellen der -Richtlinie](#)
- [Konfiguration und Erstellung des Clusters](#)

Erstellen der -Richtlinie

Erstellen Sie eine Richtlinie.

1. Gehen Sie zur IAM-Konsole: <https://console.aws.amazon.com/iam/home>.
2. Wählen Sie Policies (Richtlinien).
3. Wählen Sie Create Policy (Richtlinie erstellen) aus.
4. Wählen Sie die Registerkarte JSON und fügen Sie die folgende Richtlinie ein. Stellen Sie sicher, dass Sie alle Vorkommen von **123456789012** durch Ihre AWS-Konto ID und den Amazon Resource Name (ARN) sowie AWS-Region durch Ihren eigenen ersetzen.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "kms:DescribeKey",
        "kms:ReEncrypt*",
        "kms:CreateGrant",
        "kms:Decrypt"
      ],
      "Resource": [
        "arn:aws:kms:region-id:123456789012:key/abcd1234-ef56-gh78-ij90-
abcd1234efgh5678"
      ]
    }
  ]
}
```

```
}
```

5. Geben Sie für dieses Tutorial der Richtlinie `ParallelClusterKmsPolicy` einen Namen und wählen Sie dann Richtlinie erstellen aus.
6. Notieren Sie sich die Richtlinie ARN. Sie benötigen es, um Ihren Cluster zu konfigurieren.

Konfiguration und Erstellung des Clusters

Im Folgenden finden Sie ein Beispiel für eine Clusterkonfiguration, die ein gemeinsam genutztes Amazon Elastic Block Store-Dateisystem mit Verschlüsselung beinhaltet.

```
Region: eu-west-1
Image:
  Os: alinux2
HeadNode:
  InstanceType: t2.micro
  Networking:
    SubnetId: subnet-abcdef01234567890
  Ssh:
    KeyName: my-ssh-key
  Iam:
    AdditionalIamPolicies:
      - Policy: arn:aws:iam::123456789012:policy/ParallelClusterKmsPolicy
Scheduling:
  Scheduler: slurm
  SlurmQueues:
    - Name: q1
    ComputeResources:
      - Name: t2micro
        InstanceType: t2.micro
        MinCount: 0
        MaxCount: 10
    Networking:
      SubnetIds:
        - subnet-abcdef01234567890
    Iam:
      AdditionalIamPolicies:
        - Policy: arn:aws:iam::123456789012:policy/ParallelClusterKmsPolicy
SharedStorage:
  - MountDir: /shared/ebs1
```

```
Name: shared-efs1
StorageType: Ebs
EbsSettings:
  Encrypted: True
  KmsKeyId: abcd1234-ef56-gh78-ij90-abcd1234efgh5678
```

Ersetzen Sie die Elemente in rotem Text durch Ihre eigenen Werte. Erstellen Sie dann einen Cluster, der Ihren AWS KMS Schlüssel verwendet, um Ihre Daten in Amazon EBS zu verschlüsseln.

Die Konfiguration ist für Amazon EFS- und FSx for Lustre-Dateisysteme ähnlich.

Die Amazon SharedStorage EFS-Konfiguration sieht wie folgt aus.

```
...
SharedStorage:
  - MountDir: /shared/efs1
    Name: shared-efs1
    StorageType: Efs
    EfsSettings:
      Encrypted: True
      KmsKeyId: abcd1234-ef56-gh78-ij90-abcd1234efgh5678
```

Die SharedStorage Konfiguration von FSx for Lustre sieht wie folgt aus.

```
...
SharedStorage:
  - MountDir: /shared/fsx1
    Name: shared-fsx1
    StorageType: FsxLustre
    FsxLustreSettings:
      StorageCapacity: 1200
      DeploymentType: PERSISTENT_1
      PerUnitStorageThroughput: 200
      KmsKeyId: abcd1234-ef56-gh78-ij90-abcd1234efgh5678
```

Ausführung von Jobs in einem Cluster mit mehreren Warteschlangen

In diesem Tutorial erfahren Sie, wie Sie Ihre erste "Hello World"-Job am AWS ParallelCluster mit [Modus mit mehreren Warteschlangen](#).

Bei der Verwendung des AWS ParallelCluster Befehlszeilenschnittstelle (CLI) oder API, Sie zahlen nur für AWS Ressourcen, die bei der Erstellung oder Aktualisierung erstellt werden AWS ParallelCluster Bilder und Cluster. Weitere Informationen finden Sie unter [AWS Dienste verwendet von AWS ParallelCluster](#).

Die AWS ParallelCluster Die Benutzeroberfläche basiert auf einer serverlosen Architektur und Sie können sie innerhalb der verwenden AWS Kategorie „Kostenloses Kontingent“ in den meisten Fällen. Weitere Informationen finden Sie unter [AWS ParallelCluster Die Benutzeroberfläche kostet](#).

Voraussetzungen

- AWS ParallelCluster [ist installiert](#).
- Das AWS CLI [ist installiert und konfiguriert](#).
- Sie haben ein [EC2-Schlüsselpaar](#).
- Sie haben eine IAM-Rolle bei der [Berechtigungen](#) die erforderlich sind, um das auszuführen `pcluster` CLI.

Konfigurieren Sie Ihren Cluster

Stellen Sie zunächst sicher, dass AWS ParallelCluster ist korrekt installiert, indem Sie den folgenden Befehl ausführen.

```
$ pcluster version
```

Mehr über `pcluster version` erfahren Sie unter [pcluster version](#).

Dieser Befehl gibt die laufende Version von zurück AWS ParallelCluster.


Führen Sie als Nächstes aus `pcluster configure` um eine grundlegende Konfigurationsdatei zu generieren. Folgen Sie allen Anweisungen, die diesem Befehl folgen.

```
$ pcluster configure --config multi-queue-mode.yaml
```

Weitere Informationen zum Befehl `pcluster configure` finden Sie unter [pcluster configure](#).

Nachdem Sie diesen Schritt abgeschlossen haben, erscheint eine grundlegende Konfigurationsdatei mit dem Namen `multi-queue-mode.yaml` erscheint. Diese Datei enthält eine grundlegende Clusterkonfiguration.

Im nächsten Schritt ändern Sie Ihre neue Konfigurationsdatei und starten einen Cluster mit mehreren Warteschlangen.

 Note

Für einige Instanzen, die in diesem Tutorial verwendet werden, gilt das kostenlose Kontingent nicht.

Ändern Sie für dieses Tutorial Ihre Konfigurationsdatei so, dass sie der folgenden Konfiguration entspricht. Die rot hervorgehobenen Elemente stellen die Werte Ihrer Konfigurationsdatei dar. Behalten Sie Ihre eigenen Werte bei.

```
Region: region-id
Image:
  Os: alinux2
HeadNode:
  InstanceType: c5.xlarge
Networking:
  SubnetId: subnet-abcdef01234567890
Ssh:
  KeyName: yourkeypair
Scheduling:
  Scheduler: slurm
SlurmQueues:
- Name: spot
  ComputeResources:
- Name: c5xlarge
  InstanceType: c5.xlarge
  MinCount: 1
  MaxCount: 10
- Name: t2micro
  InstanceType: t2.micro
  MinCount: 1
  MaxCount: 10
Networking:
  SubnetIds:
- subnet-abcdef01234567890
- Name: ondemand
  ComputeResources:
- Name: c52xlarge
  InstanceType: c5.2xlarge
```



```
MinCount: 0
MaxCount: 10
Networking:
SubnetIds:
- subnet-021345abcdef6789
```

Erstellen Ihres -Clusters

Erstellen Sie einen Cluster mit einem Namen `multi-queue-cluster` basierend auf Ihrer Konfigurationsdatei.

```
$ pcluster create-cluster --cluster-name multi-queue-cluster --cluster-configuration
multi-queue-mode.yaml
{
  "cluster": {
    "clusterName": "multi-queue-cluster",
    "cloudformationStackStatus": "CREATE_IN_PROGRESS",
    "cloudformationStackArn": "arn:aws:cloudformation:eu-west-1:123456789012:stack/
multi-queue-cluster/1234567-abcd-0123-def0-abcdef0123456",
    "region": "eu-west-1",
    "version": "3.7.0",
    "clusterStatus": "CREATE_IN_PROGRESS"
  }
}
```

Weitere Informationen zum Befehl `pcluster create-cluster` finden Sie unter [pcluster create-cluster](#).

Führen Sie den folgenden Befehl aus, um den Status des Clusters zu überprüfen.

```
$ pcluster list-clusters
{
  "cluster": {
    "clusterName": "multi-queue-cluster",
    "cloudformationStackStatus": "CREATE_IN_PROGRESS",
    "cloudformationStackArn": "arn:aws:cloudformation:eu-west-1:123456789012:stack/
multi-queue-cluster/1234567-abcd-0123-def0-abcdef0123456",
    "region": "eu-west-1",
    "version": "3.7.0",
    "clusterStatus": "CREATE_IN_PROGRESS"
  }
}
```

Wenn der Cluster erstellt wird, `clusterStatus` Feld zeigt `CREATE_COMPLETE`.

Loggen Sie sich in den Hauptknoten ein

Verwenden Sie Ihre private SSH-Schlüsseldatei, um sich beim Hauptknoten anzumelden.

```
$ pcluster ssh --cluster-name multi-queue-cluster -i ~/path/to/yourkeyfile.pem
```

Mehr über `pcluster ssh` erfahren Sie unter [pcluster ssh](#).

Führen Sie nach dem Einloggen den `sinfo` Befehl, um zu überprüfen, ob Ihre Scheduler-Warteschlangen eingerichtet und konfiguriert sind.

Für weitere Informationen über `sinfo`, siehe [sinfo](#) in der Slurm Dokumentation.

```
$ sinfo
PARTITION AVAIL  TIMELIMIT  NODES  STATE NODELIST
spot*      up    infinite   18  idle~ spot-dy-c5xlarge-[1-9],spot-dy-t2micro-[1-9]
spot*      up    infinite    2  idle  spot-st-c5xlarge-1,spot-st-t2micro-1
ondemand   up    infinite   10  idle~ ondemand-dy-c52xlarge-[1-10]
```

Die Ausgabe zeigt, dass Sie einen haben `2.micro` und ein `c5.xlarge` Rechenknoten im `idle` Status, die in Ihrem Cluster verfügbar sind.

Alle anderen Knoten befinden sich im Energiesparmodus, der durch den `~` Suffix im Knotenstatus, ohne dass sie von EC2-Instances unterstützt werden. Die Standardwarteschlange wird durch ein `*` gekennzeichnet `*` Suffix nach dem Namen der Warteschlange. `spot` ist Ihre Standard-Job-Warteschlange.

Job im Modus mit mehreren Warteschlangen ausführen

Versuchen Sie als Nächstes, einen Job für eine Weile im Ruhezustand auszuführen. Der Job gibt später seinen eigenen Hostnamen aus. Stellen Sie sicher, dass dieses Skript vom aktuellen Benutzer ausgeführt werden kann.

```
$ tee <<EOF hellojob.sh
#!/bin/bash
sleep 30
echo "Hello World from \$(hostname)"
EOF
```

```
$ chmod +x hellojob.sh
$ ls -l hellojob.sh
-rwxrwxr-x 1 ec2-user ec2-user 57 Sep 23 21:57 hellojob.sh
```

Senden Sie den Job mit dem `batch`-Befehl. Fordere zwei Knoten für diesen Job an mit dem `-N 2` Option, und stellen Sie sicher, dass der Job erfolgreich gesendet wurde. Für weitere Informationen über `batch`, siehe [batch](#) in der Slurm-Dokumentation.

```
$ sbatch -N 2 --wrap "srun hellojob.sh"
Submitted batch job 1
```

Sie können Ihre Warteschlange einsehen und den Status des Jobs überprüfen mit dem `queue`-Befehl. Da Sie keine bestimmte Warteschlange angegeben haben, ist die Standardwarteschlange (`spot`) verwendet. Für weitere Informationen über `queue`, siehe [queue](#) in der Slurm-Dokumentation.

```
$ queue
JOBID PARTITION   NAME     USER  ST        TIME  NODES NODELIST(REASON)
   1      spot     wrap ec2-user  R         0:10     2 spot-st-c5xlarge-1,spot-st-t2micro-1
```

Die Ausgabe zeigt, dass die Aufgabe zurzeit ausgeführt wird. Warte, bis der Job beendet ist. Das dauert etwa 30 Sekunden. Dann laufe `queue` noch einmal.

```
$ queue
JOBID PARTITION   NAME     USER  ST        TIME  NODES NODELIST(REASON)
```

Nachdem alle Jobs in der Warteschlange abgeschlossen sind, suchen Sie nach der Ausgabedatei mit dem Namen `slurm-1.out` in Ihrem aktuellen Verzeichnis.

```
$ cat slurm-1.out
Hello World from spot-st-t2micro-1
Hello World from spot-st-c5xlarge-1
```

Die Ausgabe zeigt, dass der Job erfolgreich ausgeführt wurde auf `spot-st-t2micro-1` und `spot-st-c5xlarge-1` Knoten.

Reichen Sie nun denselben Job ein, indem Sie mit den folgenden Befehlen Einschränkungen für bestimmte Instanzen angeben.

```
$ sbatch -N 3 -p spot -C "[c5.xlarge*1&t2.micro*2]" --wrap "srun hellojob.sh"
```

```
Submitted batch job 2
```

Sie haben diese Parameter verwendet fürsbatch:

- `-N 3`— fordert drei Knoten an.
- `-p spot`— sendet den Job an denspotWarteschlange. Sie können einen Job auch bei der einreichenondemandWarteschlange, indem Sie Folgendes angeben`-p ondemand`.
- `-C "[c5.xlarge*1&t2.micro*2]"`— spezifiziert die spezifischen Knotenbeschränkungen für diesen Job. Dies erfordert einenc5.xlargeKnoten und zweit2.microKnoten, die für diesen Job verwendet werden sollen.

Führen Sie das `ssinfo`Befehl zum Anzeigen der Knoten und Warteschlangen. Warteschlangen inAWS ParallelClusterwerden Partitionen genannt inSlurm.

```
$ sinfo
PARTITION AVAIL  TIMELIMIT  NODES  STATE NODELIST
spot*      up    infinite   1    alloc# spot-dy-t2micro-1
spot*      up    infinite  17    idle~  spot-dy-c5xlarge-[2-10],spot-dy-t2micro-[2-9]
spot*      up    infinite   1    mix   spot-st-c5xlarge-1
spot*      up    infinite   1    alloc spot-st-t2micro-1
ondemand   up    infinite  10    idle~  ondemand-dy-c52xlarge-[1-10]
```

Die Knoten werden hochgefahren. Dies wird angezeigt durch den#Suffix für den Knotenstatus. Führen Sie den `ssqueue`Befehl, um Informationen über die Jobs im Cluster anzuzeigen.

```
$ squeue
JOBID PARTITION   NAME     USER ST       TIME  NODES NODELIST(REASON)
   2    spot     wrap    ec2-user CF        0:04     3  spot-dy-c5xlarge-1,spot-dy-
t2micro-1,spot-st-t2micro-1
```

Ihr Job ist in derCF(CONFIGURING) Status, der darauf wartet, dass die Instances hochskaliert werden und dem Cluster beitreten.

Nach etwa drei Minuten sind die Knoten verfügbar und der Job geht inR(RUNNING) Bundesstaat.

```
$ squeue
JOBID PARTITION   NAME     USER ST       TIME  NODES NODELIST(REASON)
   2    spot     wrap    ec2-user R        0:07     3  spot-dy-t2micro-1,spot-st-
c5xlarge-1,spot-st-t2micro-1
```

Der Job ist beendet und alle drei Knoten befinden sich in `idle` Bundesstaat.

```
$ squeue
JOBID PARTITION      NAME      USER ST      TIME  NODES NODELIST(REASON)
$ sinfo
PARTITION AVAIL  TIMELIMIT  NODES  STATE NODELIST
spot*      up    infinite   17   idle~ spot-dy-c5xlarge-[1-9],spot-dy-t2micro-[2-9]
spot*      up    infinite    3   idle  spot-dy-t2micro-1,spot-st-c5xlarge-1,spot-st-t2micro-1
ondemand   up    infinite   10   idle~ ondemand-dy-c52xlarge-[1-10]
```

Wenn sich keine Jobs mehr in der Warteschlange befinden, suchen Sie `nachslurm-2.out` in Ihrem lokalen Verzeichnis.

```
$ cat slurm-2.out
Hello World from spot-st-t2micro-1
Hello World from spot-dy-t2micro-1
Hello World from spot-st-c5xlarge-1
```

Dies ist der endgültige Status des Clusters.

```
$ sinfo
PARTITION AVAIL  TIMELIMIT  NODES  STATE NODELIST
spot*      up    infinite   17   idle~ spot-dy-c5xlarge-[1-9],spot-dy-t2micro-[2-9]
spot*      up    infinite    3   idle  spot-dy-t2micro-1,spot-st-c5xlarge-1,spot-st-t2micro-1
ondemand   up    infinite   10   idle~ ondemand-dy-c52xlarge-[1-10]
```

Nachdem Sie sich vom Cluster abgemeldet haben, können Sie den Vorgang aufräumen, indem Sie folgenden Befehl ausführen `pcluster delete-cluster`. Weitere Informationen finden Sie unter [pcluster list-clusters](#) und [pcluster delete-cluster](#).

```
$ pcluster list-clusters
{
  "clusters": [
    {
      "clusterName": "multi-queue-cluster",
      "cloudformationStackStatus": "CREATE_COMPLETE",
      "cloudformationStackArn": "arn:aws:cloudformation:eu-west-1:123456789012:stack/multi-queue-cluster/1234567-abcd-0123-def0-abcdef0123456",
      "region": "eu-west-1",
      "version": "3.1.4",
    }
  ]
}
```

```
    "clusterStatus": "CREATE_COMPLETE"
  }
]
}
$ pcluster delete-cluster -n multi-queue-cluster
{
  "cluster": {
    "clusterName": "multi-queue-cluster",
    "cloudformationStackStatus": "DELETE_IN_PROGRESS",
    "cloudformationStackArn": "arn:aws:cloudformation:eu-west-1:123456789012:stack/multi-queue-cluster/1234567-abcd-0123-def0-abcdef0123456",
    "region": "eu-west-1",
    "version": "3.1.4",
    "clusterStatus": "DELETE_IN_PROGRESS"
  }
}
```

Verwenden der AWS ParallelCluster-API

In diesem Tutorial erstellen und testen Sie die API mit [Amazon API Gateway](#) und einer AWS ParallelCluster CloudFormation Vorlage. Anschließend verwenden Sie den Beispielclient, der auf [GitHub](#) verfügbar ist, um die API zu verwenden. Weitere Informationen zur Verwendung der API finden Sie unter [AWS ParallelCluster-API](#).

Dieses Tutorial wurde aus dem Workshop [HPC For Public Sector](#) Customers entnommen.

Wenn Sie die AWS ParallelCluster Befehlszeilenschnittstelle (CLI) oder die API verwenden, zahlen Sie nur für die AWS Ressourcen, die beim Erstellen oder Aktualisieren von AWS ParallelCluster Images und Clustern erstellt oder aktualisiert werden. Weitere Informationen finden Sie unter [AWS Dienste verwendet von AWS ParallelCluster](#).

Die AWS ParallelCluster Benutzeroberfläche basiert auf einer serverlosen Architektur und Sie können sie in den meisten Fällen innerhalb der Kategorie AWS Free Tier verwenden. Weitere Informationen finden Sie unter [AWS ParallelCluster Die Benutzeroberfläche kostet](#).

Voraussetzungen

- Der AWS CLI ist in Ihrer Computerumgebung [installiert](#) und konfiguriert.
- AWS ParallelCluster ist in einer virtuellen Umgebung installiert. Weitere Informationen finden [Sie unter Installation AWS ParallelCluster in einer virtuellen Umgebung](#).
- Sie haben ein [EC2-Schlüsselpaar](#).

- Sie haben eine IAM-Rolle mit den [Berechtigungen](#), die für die Ausführung der [pcluster](#) CLI erforderlich sind.

Schritt 1: Erstellen Sie die API mit Amazon API Gateway

Bleiben Sie in Ihrem privaten Benutzerverzeichnis und aktivieren Sie Ihre virtuelle Umgebung:

1. Installieren Sie einen hilfreichen JSON-Befehlszeilenprozessor.

```
$ sudo yum groupinstall -y "Development Tools"
sudo yum install -y jq python3-devel
```

2. Führen Sie den folgenden Befehl aus, um Ihre AWS ParallelCluster Version abzurufen und sie einer Umgebungsvariablen zuzuweisen.

```
$ PCLUSTER_VERSION=$(pcluster version | jq -r '.version')
echo "export PCLUSTER_VERSION=${PCLUSTER_VERSION}" |tee -a ~/.bashrc
```

3. Erstellen Sie eine Umgebungsvariable und weisen Sie ihr Ihre Regions-ID zu.

```
$ export AWS_DEFAULT_REGION="us-east-1"
echo "export AWS_DEFAULT_REGION=${AWS_DEFAULT_REGION}" |tee -a ~/.bashrc
```

4. Führen Sie die folgenden Befehle aus, um die API bereitzustellen.

```
API_STACK_NAME="pc-api-stack"
echo "export API_STACK_NAME=${API_STACK_NAME}" |tee -a ~/.bashrc
```

```
aws cloudformation create-stack \
  --region ${AWS_DEFAULT_REGION} \
  --stack-name ${API_STACK_NAME} \
  --template-url https://${AWS_DEFAULT_REGION}-aws-parallelcluster.s3.
  ${AWS_DEFAULT_REGION}.amazonaws.com/parallelcluster/${PCLUSTER_VERSION}/api/
  parallelcluster-api.yaml \
  --capabilities CAPABILITY_NAMED_IAM CAPABILITY_AUTO_EXPAND \
  --parameters ParameterKey=EnableIamAdminAccess,ParameterValue=true

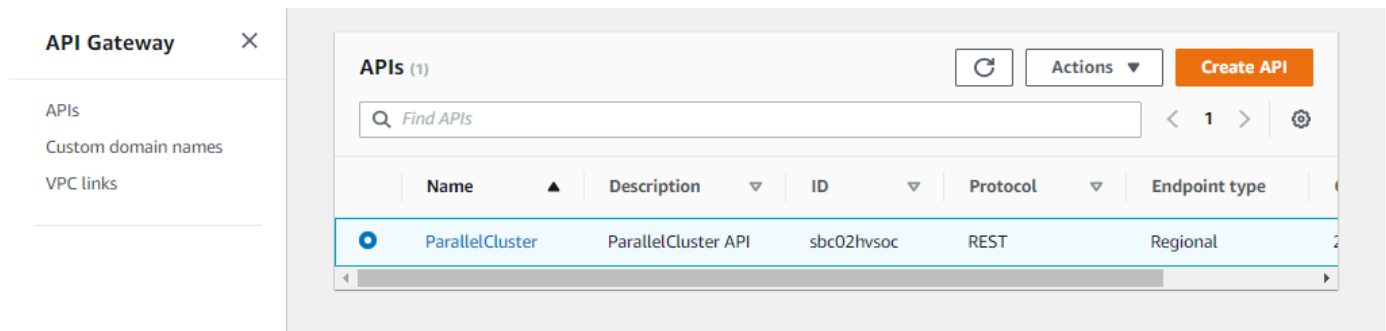
{
  "StackId": "arn:aws:cloudformation:us-east-1:123456789012:stack/my-api-
  stack/abcd1234-ef56-gh78-ei90-1234abcd5678"
```

```
}
```

Nachdem der Vorgang abgeschlossen ist, fahren Sie mit dem nächsten Schritt fort.

Schritt 2: Testen Sie die API in der Amazon API Gateway-Konsole

1. Melden Sie sich an der AWS Management Console an.
2. Navigieren Sie zur [Amazon API Gateway-Konsole](#).
3. Wählen Sie Ihr API-Deployment.



4. Wählen Sie Stufen und wählen Sie eine Phase aus.

Amazon API Gateway APIs > ParallelCluster (sbc02hvsoc) > Stages > prod Show all hints ?

APIs Stages **Create** prod Stage Editor Delete Stage Configure Tags

Invoke URL: <https://sbc02hvsoc.execute-api.us-east-1.amazonaws.com/prod>

Settings Logs/Tracing Stage Variables SDK Generation Export

Deployment History Documentation History Canary

Cache Settings

Enable API cache

Default Method Throttling

Choose the default throttling level for the methods in this stage. Each method in this stage will respect these rate and burst settings. Your current account level throttling rate is **10000** requests per second with a burst of 5000 requests. [Read more about API Gateway throttling](#)

Enable throttling ⓘ

Rate requests per second

Burst requests

Web Application Firewall (WAF) [Learn more.](#)

Select the Web ACL to be applied to this stage.

Web ACL [Create Web ACL](#)

Client Certificate

Select the client certificate that API Gateway will use to call your integration endpoints in this stage.

Certificate

5. Notieren Sie sich die URL, die API Gateway für den Zugriff auf Ihre API oder den Aufruf Ihrer API bereitstellt. Es ist blau hervorgehoben.
6. Wählen Sie Ressourcen und wählen Sie **GET** unter **/clusters**.
7. Wählen Sie das TEST-Symbol und scrollen Sie dann nach unten und wählen Sie das TEST-Symbol.

APIs > ParallelCluster (sbc02hvsoc) > Resources > /v3/clusters (ulfkw2) > GET

Resources Actions **/v3/clusters - GET - Method Execution**

The screenshot displays the AWS ParallelCluster console interface for testing an API endpoint. On the left, a tree view shows the resource hierarchy: / > /v3 > /clusters > GET. The 'GET' method is selected. In the center, a 'Client' box contains a 'TEST' button with a lightning bolt icon. To the right, the 'Method Request' box shows the following details: Auth: AWS IAM, ARN: arn:aws:execute-api:us-east-1:123456789012:sbc02hvsoc*/GET/v3/clusters, and Query Strings: region, nextToken, clusterStatus. Below it, the 'Method Response' box contains the text 'Select an integration response.' Arrows indicate the flow of data from the client to the request and from the response back to the client.

Die Antwort auf Ihre `/clusters` GET wird angezeigt.

APIs > ParallelCluster (sbc02hvsoc) > Resources > /v3/clusters (ulfkw2) > GET

Show all hints ?

Resources Actions

Method Execution /v3/clusters - GET - Method Test

Make a test call to your method. When you make a test call, API Gateway skips authorization and directly invokes your method

Path
No path parameters exist for this resource. You can define path parameters by using the syntax `{myPathParam}` in a resource path.

Request: /v3/clusters
Status: 200
Latency: 3203 ms
Response Body

Query Strings
{clusters}

Headers
{clusters}
Use a colon (:) to separate header name and value, and new lines to declare multiple headers. eg. `Accept:application/json`.

Stage Variables
No stage variables exist for this method.

Client Certificate
No client certificates have been generated.

```
{
  "clusters": [
    {
      "cloudformationStackArn": "arn:aws:cloudformation:us-east-1:123456789012:stack/test-cluster/4450d850-b684-11ec-84a7-0a047567c9f3",
      "cloudformationStackStatus": "CREATE_COMPLETE",
      "clusterName": "test-cluster",
      "clusterStatus": "CREATE_COMPLETE",
      "region": "us-east-1",
      "version": "3.1.2"
    }
  ]
}
```

Response Headers

```
{"Content-Length": "360", "X-Amzn-Trace-Id": "Root=1-62686455-c1cf243417b2721e33822ac5;Sampled=1", "Content-Type": "application/json"}
```

Logs

Schritt 3: Bereiten Sie einen Beispielclient vor und testen Sie ihn, um die API aufzurufen

Klonen Sie den AWS ParallelCluster Quellcode `cd` in das `api` Verzeichnis und installieren Sie die Python-Clientbibliotheken.

- ```
$ git clone -b v${PCLUSTER_VERSION} https://github.com/aws/aws-parallelcluster aws-parallelcluster-v${PCLUSTER_VERSION}
cd aws-parallelcluster-v${PCLUSTER_VERSION}/api
```

```
$ pip3 install client/src
```

2. Navigieren Sie zurück zu Ihrem Home-Benutzerverzeichnis.
3. Exportieren Sie die API Gateway-Basis-URL, die der Client bei der Ausführung verwendet.

```
$ export PCLUSTER_API_URL=$(aws cloudformation describe-stacks
--stack-name ${API_STACK_NAME} --query 'Stacks[0].Outputs[?
OutputKey==`ParallelClusterApiInvokeUrl`].OutputValue' --output text)
echo "export PCLUSTER_API_URL=${PCLUSTER_API_URL}" |tee -a ~/.bashrc
```

4. Exportieren Sie einen Clusternamen, den der Client verwendet, um einen Cluster zu erstellen.

```
$ export CLUSTER_NAME="test-api-cluster"
echo "export CLUSTER_NAME=${CLUSTER_NAME}" |tee -a ~/.bashrc
```

5. Führen Sie die folgenden Befehle aus, um die Anmeldeinformationen zu speichern, die der Beispielclient für den Zugriff auf die API verwendet.

```
$ export PCLUSTER_API_USER_ROLE=$(aws cloudformation describe-
stacks --stack-name ${API_STACK_NAME} --query 'Stacks[0].Outputs[?
OutputKey==`ParallelClusterApiUserRole`].OutputValue' --output text)
echo "export PCLUSTER_API_USER_ROLE=${PCLUSTER_API_USER_ROLE}" |tee -a ~/.bashrc
```

## Schritt 4: Client-Codeskript kopieren und Clustertests ausführen

1. Kopieren Sie den folgenden Beispiel-Client-Code `test_pcluster_client.py` in Ihr Home-Benutzerverzeichnis. Der Client-Code fordert Folgendes an:
  - Erstellen Sie den -Cluster.
  - Beschreiben Sie den Cluster.
  - Listet die Cluster auf.
  - Beschreiben Sie die Rechenflotte.
  - Beschreiben Sie die Cluster-Instanzen.

```
Copyright 2021 Amazon.com, Inc. or its affiliates. All Rights Reserved.
SPDX-License-Identifier: MIT-0
#
```

```
Permission is hereby granted, free of charge, to any person obtaining a copy of
this
software and associated documentation files (the "Software"), to deal in the
Software
without restriction, including without limitation the rights to use, copy,
modify,
merge, publish, distribute, sublicense, and/or sell copies of the Software, and
to
permit persons to whom the Software is furnished to do so.
#
THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR
IMPLIED,
INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A
PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR
COPYRIGHT
HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION
OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE
SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.
#
Author: Evan F. Bollig (Github: bollig)

import time, datetime
import os
import pcluster_client
from pprint import pprint
from pcluster_client.api import (
 cluster_compute_fleet_api,
 cluster_instances_api,
 cluster_operations_api
)
from pcluster_client.model.create_cluster_request_content import
 CreateClusterRequestContent
from pcluster_client.model.cluster_status import ClusterStatus
region=os.environ.get("AWS_DEFAULT_REGION")

Defining the host is optional and defaults to http://localhost
See configuration.py for a list of all supported configuration parameters.
configuration = pcluster_client.Configuration(
 host = os.environ.get("PCLUSTER_API_URL")
)
cluster_name=os.environ.get("CLUSTER_NAME")

Enter a context with an instance of the API client
with pcluster_client.ApiClient(configuration) as api_client:
```

```
cluster_ops = cluster_operations_api.ClusterOperationsApi(api_client)
fleet_ops = cluster_compute_fleet_api.ClusterComputeFleetApi(api_client)
instance_ops = cluster_instances_api.ClusterInstancesApi(api_client)

Create cluster
build_done = False
try:
 with open('cluster-config.yaml', encoding="utf-8") as f:
 body = CreateClusterRequestContent(cluster_name=cluster_name,
cluster_configuration=f.read())
 api_response = cluster_ops.create_cluster(body, region=region)
except pcluster_client.ApiException as e:
 print("Exception when calling create_cluster: %s\n" % e)
 build_done = True
time.sleep(60)

Confirm cluster status with describe_cluster
while not build_done:
 try:
 api_response = cluster_ops.describe_cluster(cluster_name,
region=region)
 pprint(api_response)
 if api_response.cluster_status == ClusterStatus('CREATE_IN_PROGRESS'):
 print('. . . working . . .', end='', flush=True)
 time.sleep(60)
 elif api_response.cluster_status == ClusterStatus('CREATE_COMPLETE'):
 print('READY!')
 build_done = True
 else:
 print('ERROR!!!!')
 build_done = True
 except pcluster_client.ApiException as e:
 print("Exception when calling describe_cluster: %s\n" % e)

List clusters
try:
 api_response = cluster_ops.list_clusters(region=region)
 pprint(api_response)
except pcluster_client.ApiException as e:
 print("Exception when calling list_clusters: %s\n" % e)

DescribeComputeFleet
try:
```

```

 api_response = fleet_ops.describe_compute_fleet(cluster_name,
region=region)
 pprint(api_response)
except pcluster_client.ApiException as e:
 print("Exception when calling compute fleet: %s\n" % e)

DescribeClusterInstances
try:
 api_response = instance_ops.describe_cluster_instances(cluster_name,
region=region)
 pprint(api_response)
except pcluster_client.ApiException as e:
 print("Exception when calling describe_cluster_instances: %s\n" % e)

```

## 2. Erstellen Sie eine Clusterkonfiguration.

```
$ pcluster configure --config cluster-config.yaml
```

3. Die API-Clientbibliothek erkennt automatisch Konfigurationsdetails aus Ihren Umgebungsvariablen (z. B. `AWS_ACCESS_KEY_ID`, `AWS_SECRET_ACCESS_KEY`, oder `AWS_SESSION_TOKEN`) oder `$HOME/.aws`. Mit dem folgenden Befehl wird Ihre aktuelle IAM-Rolle auf die angegebene `ParallelClusterApiUserRole` geändert.

```
$ eval $(aws sts assume-role --role-arn ${PCLUSTER_API_USER_ROLE} --role-session-name ApiTestSession | jq -r '.Credentials | "export AWS_ACCESS_KEY_ID=\(.AccessKeyId)\nexport AWS_SECRET_ACCESS_KEY=\(.SecretAccessKey)\nexport AWS_SESSION_TOKEN=\(.SessionToken)\n"')
```

Fehler, auf den Sie achten sollten:

Wenn Sie einen ähnlichen Fehler wie den folgenden sehen, haben Sie bereits davon ausgegangen, dass der `ParallelClusterApiUserRole` und Ihr `AWS_SESSION_TOKEN` abgelaufen ist.

```

An error occurred (AccessDenied) when calling the AssumeRole operation:
User: arn:aws:sts::XXXXXXXXXXXX:assumed-role/ParallelClusterApiUserRole-XXXXXXXX-XXXX-XXXX-XXXX-XXXXXXXXXXXX/ApiTestSession
is not authorized to perform: sts:AssumeRole on resource:
arn:aws:iam::XXXXXXXXXXXX:role/ParallelClusterApiUserRole-XXXXXXXX-XXXX-XXXX-XXXX-XXXXXXXXXXXX

```

Löschen Sie die Rolle und führen Sie dann den `aws sts assume-role` Befehl erneut aus, um die `ParallelClusterApiUserRole` zu verwenden.

```
$ unset AWS_SESSION_TOKEN
unset AWS_SECRET_ACCESS_KEY
unset AWS_ACCESS_KEY_ID
```

Um Ihre aktuellen Benutzerberechtigungen für den API-Zugriff bereitzustellen, müssen Sie [die Ressourcenrichtlinie erweitern](#).

4. Führen Sie den folgenden Befehl aus, um den Beispielclient zu testen.

```
$ python3 test_pcluster_client.py
{'cluster_configuration': 'Region: us-east-1\n'
 'Image:\n'
 ' Os: alinux2\n'
 'HeadNode:\n'
 ' InstanceType: t2.micro\n'
 ' Networking . . . :\n'
 ' SubnetId: subnet-1234567890abcdef0\n'
 ' Ssh:\n'
 ' KeyName: adpc\n'
 'Scheduling:\n'
 ' Scheduler: slurm\n'
 ' SlurmQueues:\n'
 ' - Name: queue1\n'
 ' ComputeResources:\n'
 ' - Name: t2micro\n'
 ' InstanceType: t2.micro\n'
 ' MinCount: 0\n'
 ' MaxCount: 10\n'
 ' Networking . . . :\n'
 ' SubnetIds:\n'
 ' - subnet-1234567890abcdef0\n',
 'cluster_name': 'test-api-cluster'}
{'cloud_formation_stack_status': 'CREATE_IN_PROGRESS',
 'cloudformation_stack_arn': 'arn:aws:cloudformation:us-east-1:123456789012:stack/test-api-cluster/abcd1234-ef56-gh78-ij90-1234abcd5678',
 'cluster_configuration': {'url': 'https://parallelcluster-021345abcdef6789-v1-do-not-delete...'},
 'cluster_name': 'test-api-cluster',
 'cluster_status': 'CREATE_IN_PROGRESS',
```



```

'compute_fleet_status': 'UNKNOWN',
'creation_time': datetime.datetime(2022, 4, 28, 16, 18, 47, 972000,
tzinfo=tzlocal()),
'last_updated_time': datetime.datetime(2022, 4, 28, 16, 18, 47, 972000,
tzinfo=tzlocal()),
'region': 'us-east-1',
'tags': [{'key': 'parallelcluster:version', 'value': '3.1.3'}],
'version': '3.1.3'
.
.
.
. . . working . . . {'cloud_formation_stack_status': 'CREATE_COMPLETE',
'cloudformation_stack_arn': 'arn:aws:cloudformation:us-east-1:123456789012:stack/
test-api-cluster/abcd1234-ef56-gh78-ij90-1234abcd5678',
'cluster_configuration': {'url': 'https://parallelcluster-021345abcdef6789-v1-do-
not-delete...'},
'cluster_name': 'test-api-cluster',
'cluster_status': 'CREATE_COMPLETE',
'compute_fleet_status': 'RUNNING',
'creation_time': datetime.datetime(2022, 4, 28, 16, 18, 47, 972000,
tzinfo=tzlocal()),
'head_node': {'instance_id': 'i-abcdef01234567890',
'instance_type': 't2.micro',
'launch_time': datetime.datetime(2022, 4, 28, 16, 21, 46,
tzinfo=tzlocal()),
'private_ip_address': '172.31.27.153',
'public_ip_address': '52.90.156.51',
'state': 'running'},
'last_updated_time': datetime.datetime(2022, 4, 28, 16, 18, 47, 972000,
tzinfo=tzlocal()),
'region': 'us-east-1',
'tags': [{'key': 'parallelcluster:version', 'value': '3.1.3'}],
'version': '3.1.3'
READY!

```

## Schritt 5: Client-Codeskript kopieren und Cluster löschen

1. Kopieren Sie den folgenden Beispiel-Clientcode in `delete_cluster_client.py`. Der Client-Code fordert das Löschen des Clusters an.

```

Copyright 2021 Amazon.com, Inc. or its affiliates. All Rights Reserved.
SPDX-License-Identifier: MIT-0

```

```
#
Permission is hereby granted, free of charge, to any person obtaining a copy of
this
software and associated documentation files (the "Software"), to deal in the
Software
without restriction, including without limitation the rights to use, copy,
modify,
merge, publish, distribute, sublicense, and/or sell copies of the Software, and
to
permit persons to whom the Software is furnished to do so.
#
THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR
IMPLIED,
INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A
PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR
COPYRIGHT
HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION
OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE
SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.
#
Author: Evan F. Bollig (Github: bollig)

import time, datetime
import os
import pcluster_client
from pprint import pprint
from pcluster_client.api import (
 cluster_compute_fleet_api,
 cluster_instances_api,
 cluster_operations_api
)
from pcluster_client.model.create_cluster_request_content import
 CreateClusterRequestContent
from pcluster_client.model.cluster_status import ClusterStatus
region=os.environ.get("AWS_DEFAULT_REGION")

Defining the host is optional and defaults to http://localhost
See configuration.py for a list of all supported configuration parameters.
configuration = pcluster_client.Configuration(
 host = os.environ.get("PCLUSTER_API_URL")
)
cluster_name=os.environ.get("CLUSTER_NAME")

Enter a context with an instance of the API client
```

```

with pcluster_client.ApiClient(configuration) as api_client:
 cluster_ops = cluster_operations_api.ClusterOperationsApi(api_client)

 # Delete the cluster
 gone = False
 try:
 api_response = cluster_ops.delete_cluster(cluster_name, region=region)
 except pcluster_client.ApiException as e:
 print("Exception when calling delete_cluster: %s\n" % e)
 time.sleep(60)

 # Confirm cluster status with describe_cluster
 while not gone:
 try:
 api_response = cluster_ops.describe_cluster(cluster_name,
region=region)
 pprint(api_response)
 if api_response.cluster_status == ClusterStatus('DELETE_IN_PROGRESS'):
 print('. . . working . . .', end='', flush=True)
 time.sleep(60)
 except pcluster_client.ApiException as e:
 gone = True
 print("DELETE COMPLETE or Exception when calling describe_cluster: %s
\n" % e)

```

2. Führen Sie den folgenden Befehl aus, um den Cluster zu löschen.

```

$ python3 delete_cluster_client.py
{'cloud_formation_stack_status': 'DELETE_IN_PROGRESS',
'cloudformation_stack_arn': 'arn:aws:cloudformation:us-east-1:123456789012:stack/
test-api-cluster/abcd1234-ef56-gh78-ij90-1234abcd5678',
'cluster_configuration': {'url': 'https://parallelcluster-021345abcdef6789-v1-do-
not-delete...'},
'cluster_name': 'test-api-cluster',
'cluster_status': 'DELETE_IN_PROGRESS',
'compute_fleet_status': 'UNKNOWN',
'creation_time': datetime.datetime(2022, 4, 28, 16, 50, 47, 943000,
tzinfo=tzlocal()),
'head_node': {'instance_id': 'i-abcdef01234567890',
'instance_type': 't2.micro',
'launch_time': datetime.datetime(2022, 4, 28, 16, 53, 48,
tzinfo=tzlocal()),
'private_ip_address': '172.31.17.132',
'public_ip_address': '34.201.100.37',

```

```

 'state': 'running'},
 'last_updated_time': datetime.datetime(2022, 4, 28, 16, 50, 47, 943000,
 tzinfo=tzlocal()),
 'region': 'us-east-1',
 'tags': [{'key': 'parallelcluster:version', 'value': '3.1.3'}],
 'version': '3.1.3'}
 .
 .
 . . . working . . . {'cloud_formation_stack_status': 'DELETE_IN_PROGRESS',
 'cloudformation_stack_arn': 'arn:aws:cloudformation:us-east-1:123456789012:stack/
 test-api-cluster/abcd1234-ef56-gh78-ij90-1234abcd5678',
 'cluster_configuration': {'url': 'https://parallelcluster-021345abcdef6789-v1-do-
 not-delete...'},
 'cluster_name': 'test-api-cluster',
 'cluster_status': 'DELETE_IN_PROGRESS',
 'compute_fleet_status': 'UNKNOWN',
 'creation_time': datetime.datetime(2022, 4, 28, 16, 50, 47, 943000,
 tzinfo=tzlocal()),
 'last_updated_time': datetime.datetime(2022, 4, 28, 16, 50, 47, 943000,
 tzinfo=tzlocal()),
 'region': 'us-east-1',
 'tags': [{'key': 'parallelcluster:version', 'value': '3.1.3'}],
 'version': '3.1.3'}
 . . . working . . . DELETE COMPLETE or Exception when calling describe_cluster:
 (404)
 Reason: Not Found
 .
 .
 .
 HTTP response body: {"message": "Cluster 'test-api-cluster' does not exist or
 belongs to an incompatible ParallelCluster major version."}

```

3. Wenn Sie mit dem Testen fertig sind, deaktivieren Sie die Umgebungsvariablen.

```

$ unset AWS_SESSION_TOKEN
unset AWS_SECRET_ACCESS_KEY
unset AWS_ACCESS_KEY_ID

```

## Schritt 6: Bereinigen

Sie können das AWS Management Console oder verwenden AWS CLI, um Ihre API zu löschen.

1. Wählen Sie in der AWS CloudFormation Konsole den API-Stack aus und wählen Sie dann Löschen.
2. Führen Sie den folgenden Befehl aus, wenn Sie den verwenden AWS CLI.

Verwenden von AWS CloudFormation

```
$ aws cloudformation delete-stack --stack-name ${API_STACK_NAME}
```

## Einen Cluster mit Slurm Accounting erstellen

Erfahren Sie, wie Sie einen Cluster mit Slurm Accounting konfigurieren und erstellen. Weitere Informationen finden Sie unter [Slurm Abrechnung mit AWS ParallelCluster](#).

Wenn Sie die AWS ParallelCluster Befehlszeilenschnittstelle (CLI) oder API verwenden, zahlen Sie nur für die AWS Ressourcen, die beim Erstellen oder Aktualisieren von AWS ParallelCluster Images und Clustern erstellt werden. Weitere Informationen finden Sie unter [AWS Dienste verwendet von AWS ParallelCluster](#).

Die AWS ParallelCluster Benutzeroberfläche basiert auf einer serverlosen Architektur und Sie können sie in den meisten Fällen innerhalb der Kategorie „AWS Kostenloses Kontingent“ verwenden. Weitere Informationen finden Sie unter [AWS ParallelCluster Die Benutzeroberfläche kostet](#).

In diesem Tutorial verwenden Sie eine [CloudFormation Schnellerstellungsvorlage \(us-east-1\)](#), um eine serverlose Datenbank [Amazon Aurora](#) für MySQL zu erstellen. Die Vorlage weist CloudFormation an, alle erforderlichen Komponenten für die Bereitstellung einer Amazon Aurora serverlosen Datenbank auf derselben VPC wie der Cluster zu erstellen. Die Vorlage erstellt auch eine grundlegende Netzwerk- und Sicherheitskonfiguration für die Verbindung zwischen dem Cluster und der Datenbank.

### Note

Ab Version 3.3.0 wird AWS ParallelCluster die Slurm Abrechnung mit dem Cluster-Konfigurationsparameter [SlurmSettings/Database](#) unterstützt.

**Note**

Die Schnellerstellungsvorlage dient als Beispiel. Diese Vorlage deckt nicht alle möglichen Anwendungsfälle für einen Slurm Buchhaltungsdatenbankserver ab. Es liegt in Ihrer Verantwortung, einen Datenbankserver mit der Konfiguration und Kapazität zu erstellen, die für Ihre Produktionsworkloads geeignet sind.

Voraussetzungen:

- AWS ParallelCluster [ist installiert](#).
- Das AWS CLI [ist installiert und konfiguriert](#).
- Sie haben ein [EC2-Schlüsselpaar](#).
- Sie haben eine IAM-Rolle mit den [Berechtigungen, die für](#) die Ausführung der [pcluster](#) CLI erforderlich sind.
- Die Region, in der Sie die Schnellerstellungsvorlage bereitstellen, unterstützt Amazon Aurora MySQL Serverless v2. Weitere Informationen finden Sie unter [Aurora Serverless v2 mit Aurora MySQL](#).

## Schritt 1: Erstellen Sie die VPC und die Subnetze für AWS ParallelCluster

Um die bereitgestellte CloudFormation Vorlage für die Slurm Accounting-Datenbank verwenden zu können, müssen Sie die VPC für den Cluster bereit haben. Sie können dies manuell oder als Teil des [Konfigurieren und erstellen Sie einen Cluster mit der AWS ParallelCluster Befehlszeilenschnittstelle](#) Verfahrens tun. Wenn Sie es bereits verwendet haben AWS ParallelCluster, haben Sie möglicherweise eine VPC für die Bereitstellung des Clusters und des Datenbankservers bereit.

## Schritt 2: Erstellen Sie den Datenbank-Stack

Verwenden Sie die [CloudFormation Schnellerstellungsvorlage \(us-east-1\), um einen Datenbank-Stack](#) für die Buchhaltung zu erstellen. Slurm Die Vorlage erfordert folgende Eingaben:

- Anmeldeinformationen für den Datenbankserver, insbesondere den Admin-Benutzernamen und das Passwort.
- Größe des Amazon Aurora serverlosen Clusters. Dies hängt von der erwarteten Clusterbelastung ab.

- Netzwerkparameter, insbesondere die Ziel-VPC und die Subnetze oder CIDR-Blöcke für die Erstellung der Subnetze.

Wählen Sie die entsprechenden Anmeldeinformationen und die Größe für Ihren Datenbankserver aus. Für die Netzwerkoptionen müssen Sie dieselbe VPC verwenden, auf der der AWS ParallelCluster Cluster bereitgestellt wird. Sie können die Subnetze für die Datenbank erstellen und sie als Eingabe an die Vorlage übergeben. Oder stellen Sie zwei unzusammenhängende CIDR-Blöcke für die beiden Subnetze bereit und lassen Sie die CloudFormation Vorlage die beiden Subnetze für CIDR-Blöcke erstellen. Stellen Sie sicher, dass sich die CIDR-Blöcke nicht mit vorhandenen Subnetzen überschneiden. Wenn sich die CIDR-Blöcke mit vorhandenen Subnetzen überschneiden, kann der Stapel nicht erstellt werden.

Die Erstellung des Datenbankservers dauert mehrere Minuten.

### Schritt 3: Erstellen Sie einen Cluster mit aktivierter Slurm Kontoführung

Die bereitgestellte CloudFormation Vorlage generiert einen CloudFormation Stapel mit einigen definierten Ausgaben. Von der AWS Management Console aus können Sie die Ausgaben auf der Registerkarte Ausgaben in der CloudFormation Stapelansicht anzeigen. Um die Slurm Abrechnung zu aktivieren, müssen einige dieser Ausgaben in der AWS ParallelCluster Cluster-Konfigurationsdatei verwendet werden:

- `DatabaseHost`: Wird für den `Uri` Cluster-Konfigurationsparameter `SlurmSettingsDatabase//` verwendet.
- `DatabaseAdminUser`: Wird für den `UserName` Cluster-Konfigurationsparameterwert `SlurmSettingsDatabase//` verwendet.
- `DatabaseSecretArn`: Wird für den `PasswordSecretArn` Cluster-Konfigurationsparameter `SlurmSettingsDatabase//` verwendet.
- `DatabaseClientSecurityGroup`: Dies ist die Sicherheitsgruppe, die an den Hauptknoten des Clusters angehängt ist, der im `SecurityGroups` Konfigurationsparameter `HeadNode/Networking` definiert ist.

Aktualisieren Sie die Database Parameter Ihrer Cluster-Konfigurationsdatei mit den Ausgabewerten. Verwenden Sie die `pcluster` CLI, um den Cluster zu erstellen.

```
$ pcluster create-cluster -n cluster-3.x -c path/to/cluster-config.yaml
```

Nachdem der Cluster erstellt wurde, können Sie damit beginnen, Slurm Accounting-Befehle wie `sacctmgr` oder `zu verwendensacct`.

## Zu einer früheren AWS Systems Manager-Dokumentversion zurückkehren

Erfahren Sie, wie Sie zu einer früheren AWS Systems Manager-Dokumentversion zurückkehren. Weitere Informationen zu SSM-Dokumenten finden Sie unter [AWSSystems Manager-Dokumente](#) im AWSSystems Manager-Benutzerhandbuch.

Wenn Sie die AWS ParallelCluster Befehlszeilenschnittstelle (CLI) oder die API verwenden, zahlen Sie nur für die AWS Ressourcen, die beim Erstellen oder Aktualisieren von AWS ParallelCluster Images und Clustern erstellt oder aktualisiert werden. Weitere Informationen finden Sie unter [AWS Dienste verwendet von AWS ParallelCluster](#).

Die AWS ParallelCluster Benutzeroberfläche basiert auf einer serverlosen Architektur und Sie können sie in den meisten Fällen innerhalb der Kategorie AWS Free Tier verwenden. Weitere Informationen finden Sie unter [AWS ParallelCluster Die Benutzeroberfläche kostet](#).

Voraussetzungen:

- Und AWS-Konto mit Berechtigungen zur Verwaltung von SSM-Dokumenten.
- Der AWS CLI [ist installiert und konfiguriert](#).

## Zu einer früheren SSM-Dokumentversion zurückkehren

1. Führen Sie in Ihrem Terminal den folgenden Befehl aus, um die Liste der vorhandenen SSM-Dokumente abzurufen, die Sie besitzen.

```
$ aws ssm list-documents --document-filter "key=Owner,value=Self"
```

2. Setzt ein SSM-Dokument auf eine frühere Version zurück. In diesem Beispiel kehren wir zu einer früheren Version des `SessionManagerRunShell` Dokuments zurück. Sie können das `SessionManagerRunShell` SSM-Dokument verwenden, um jede SSM-Shell-Sitzung, die Sie initiieren, anzupassen.
  - a. Finden Sie den `DocumentVersion` Parameter für `SessionManagerRunShell` indem Sie den folgenden Befehl ausführen:



```
$ aws ssm describe-document --name "SSM-SessionManagerRunShell"
{
 "Document": {
 "Hash": "...",
 "HashType": "Sha256",
 "Name": "SSM-SessionManagerRunShell",
 "Owner": "123456789012",
 "CreateDate": "2023-02-20T19:04:32.390000+00:00",
 "Status": "Active",
 "DocumentVersion": "1",
 "Parameters": [
 {
 "Name": "linuxcmd",
 "Type": "String",
 "Description": "The command to run on connection...",
 "DefaultValue": "if [-d '/opt/parallelcluster']; then
source /opt/parallelcluster/cfnconfig; sudo su - $cfn_cluster_user; fi; /bin/
bash"
 }
],
 "PlatformTypes": [
 "Windows",
 "Linux",
 "MacOS"
],
 "DocumentType": "Session",
 "SchemaVersion": "1.0",
 "LatestVersion": "2",
 "DefaultVersion": "1",
 "DocumentFormat": "JSON",
 "Tags": []
 }
}
```

Die neueste Version ist 2.

- b. Kehren Sie zur vorherigen Version zurück, indem Sie den folgenden Befehl ausführen:

```
$ aws ssm delete-document --name "SSM-SessionManagerRunShell" --document-
version 2
```

3. Stellen Sie sicher, dass die Dokumentversion wiederhergestellt wurde, indem Sie den `describe-document` Befehl erneut ausführen:

```
$ aws ssm describe-document --name "SSM-SessionManagerRunShell"
{
 "Document": {
 "Hash": "...",
 "HashType": "Sha256",
 "Name": "SSM-SessionManagerRunShell",
 "Owner": "123456789012",
 "CreateDate": "2023-02-20T19:04:32.390000+00:00",
 "Status": "Active",
 "DocumentVersion": "1",
 "Parameters": [
 {
 "Name": "linuxcmd",
 "Type": "String",
 "Description": "The command to run on connection...",
 "DefaultValue": "if [-d '/opt/parallelcluster']; then source /opt/parallelcluster/cfnconfig; sudo su - $cfn_cluster_user; fi; /bin/bash"
 }
],
 "PlatformTypes": [
 "Windows",
 "Linux",
 "MacOS"
],
 "DocumentType": "Session",
 "SchemaVersion": "1.0",
 "LatestVersion": "1",
 "DefaultVersion": "1",
 "DocumentFormat": "JSON",
 "Tags": []
 }
}
```

Die neueste Version ist 1.

# Einen Cluster erstellen mit AWS CloudFormation

Erfahren Sie, wie Sie einen Cluster mit einer AWS ParallelCluster CloudFormation benutzerdefinierten Ressource erstellen. Weitere Informationen finden Sie unter [AWS CloudFormation benutzerdefinierte Ressource](#).

Bei der Nutzung AWS ParallelCluster zahlen Sie nur für die AWS Ressourcen, die beim Erstellen oder Aktualisieren von AWS ParallelCluster Images und Clustern erstellt werden. Weitere Informationen finden Sie unter [AWS-Dienste verwendet von AWS ParallelCluster](#).

Voraussetzungen:

- Das AWS CLI [ist installiert und konfiguriert](#).
- Ein [EC2-Schlüsselpaar](#).
- Eine IAM-Rolle mit den [Berechtigungen, die für](#) die Ausführung der [pcluster](#) CLI erforderlich sind.

## Clustererstellung mit einem CloudFormation Schnellerstellungsstapel

In diesem Tutorial verwenden Sie einen Schnellerstellungsstapel, um eine CloudFormation Vorlage bereitzustellen, die einen Cluster und die folgenden Ressourcen erstellt: AWS

- Ein CloudFormation Root-Stack, der mithilfe eines CloudFormation Schnellerstellungsstapels erstellt wurde.
- Verschachtelte CloudFormation Stacks, die Standardrichtlinien, eine Standard-VPC-Einrichtung und einen benutzerdefinierten Ressourcenanbieter enthalten.
- Ein Beispiel für einen AWS ParallelCluster Cluster-Stack und einen Cluster, bei dem Sie sich anmelden und Jobs ausführen können.

Erstellen Sie einen Cluster mit AWS CloudFormation

1. Melden Sie sich bei der an AWS Management Console.
2. Öffnen Sie den [Link zur CloudFormation Schnellerstellung](#), um die folgenden Ressourcen in der CloudFormation Konsole zu erstellen:
  - Ein verschachtelter CloudFormation Stack mit einer VPC mit einem öffentlichen und einem privaten Subnetz für den Betrieb des Cluster-Kopfknosens bzw. der Rechenknoten.

- Ein verschachtelter CloudFormation Stack mit einer AWS ParallelCluster benutzerdefinierten Ressource für die Verwaltung des Clusters.
- Ein verschachtelter CloudFormation Stack mit den Standardrichtlinien für die Verwaltung des Clusters.
- Ein CloudFormation Root-Stack für die verschachtelten Stacks.
- Ein AWS ParallelCluster Cluster mit dem Slurm Scheduler und einer definierten Anzahl von Rechenknoten.

CloudFormation > Stacks > Create stack

## Quick create stack

### Template

Template URL  
https://pcluster-cfn-us-east-2.s3.amazonaws.com/parallelcluster/3.5.0/templates/custom\_resource/cluster-1-click.yaml

Stack description  
AWS ParallelCluster CloudFormation Cluster

### Stack name

Stack name  
cluster-0

Stack name can include letters (A-Z and a-z), numbers (0-9), and dashes (-).

### Parameters

Parameters are defined in your template and allow you to input custom values when you create or update a stack.

AvailabilityZone  
Availability zone where instances will be launched  
us-east-2a

KeyName  
KeyPair to login to the head node  
Select AWS::EC2::KeyPair::KeyName

### Capabilities

**The following resource(s) require capabilities: [AWS::CloudFormation::Stack]**

This template contains Identity and Access Management (IAM) resources. Check that you want to create each of these resources and that they have the minimum required permissions. In addition, they have custom names. Check that the custom names are unique within your AWS account. [Learn more](#)

For this template, AWS CloudFormation might require an unrecognized capability: {0}. Check the capabilities of these resources. [Learn more](#)

I acknowledge that AWS CloudFormation might create IAM resources with custom names.

I acknowledge that AWS CloudFormation might require the following capability: CAPABILITY\_AUTO\_EXPAND

Cancel Create change set Create stack

3. Geben Sie im Abschnitt Stack-Parameter für die Schnellerstellung Werte für die folgenden Parameter ein:
  - a. Geben Sie für KeyNameden Namen Ihres EC2-Schlüsselpaars key pair.

- b. Wählen Sie für AvailabilityZone eine AZ für Ihre Clusterknoten, us-east-1a z. B.
4. Markieren Sie die Kästchen, um die einzelnen Zugriffsmöglichkeiten unten auf der Seite zu bestätigen.
5. Wählen Sie Stack erstellen aus.
6. Warten Sie, bis der CloudFormation Stack den CREATE\_COMPLETE Status erreicht hat.

## Clustererstellung mit der AWS CloudFormation Befehlszeilenschnittstelle (CLI)

In diesem Tutorial verwenden Sie die AWS Befehlszeilenschnittstelle (CLI) CloudFormation, um eine CloudFormation Vorlage bereitzustellen, die einen Cluster erstellt.

Erstellen Sie die folgenden AWS Ressourcen:

- Ein CloudFormation Root-Stack, der mithilfe eines CloudFormation Schnellerstellungsstapels erstellt wurde.
- Verschachtelte CloudFormation Stacks, die Standardrichtlinien, Standard-VPC-Setup und einen benutzerdefinierten Ressourcenanbieter enthalten.
- Ein Beispiel für einen AWS ParallelCluster Cluster-Stack und einen Cluster, bei dem Sie sich anmelden und Jobs ausführen können.

Ersetzen Sie *rot hervorgehobene Eingaben*, wie z. B. *Tastenpaar*, durch Ihre eigenen Werte.

Erstellen Sie einen Cluster mit AWS CloudFormation

1. Erstellen Sie eine CloudFormation Vorlage `cluster_template.yaml` mit dem folgenden Namen:

```
AWSTemplateFormatVersion: '2010-09-09'
Description: > AWS ParallelCluster CloudFormation Template

Parameters:
 KeyName:
 Description: KeyPair to login to the head node
 Type: AWS::EC2::KeyPair::KeyName
```

```
AvailabilityZone:
 Description: Availability zone where instances will be launched
 Type: AWS::EC2::AvailabilityZone::Name
 Default: us-east-2a

Mappings:
 ParallelCluster:
 Constants:
 Version: 3.7.0

Resources:
 PclusterClusterProvider:
 Type: AWS::CloudFormation::Stack
 Properties:
 TemplateURL: !Sub
 - https://${AWS::Region}-aws-parallelcluster.s3.${AWS::Region}.
 ${AWS::URLSuffix}/parallelcluster/${Version}/templates/custom_resource/cluster.yaml
 - { Version: !FindInMap [ParallelCluster, Constants, Version] }

 PclusterVpc:
 Type: AWS::CloudFormation::Stack
 Properties:
 Parameters:
 PublicCIDR: 10.0.0.0/24
 PrivateCIDR: 10.0.16.0/20
 AvailabilityZone: !Ref AvailabilityZone
 TemplateURL: !Sub
 - https://${AWS::Region}-aws-parallelcluster.s3.${AWS::Region}.
 ${AWS::URLSuffix}/parallelcluster/${Version}/templates/networking/public-private-
 ${Version}.cf.json
 - { Version: !FindInMap [ParallelCluster, Constants, Version] }

 PclusterCluster:
 Type: Custom::PclusterCluster
 Properties:
 ServiceToken: !GetAtt [PclusterClusterProvider , Outputs.ServiceToken]
 ClusterName: !Sub 'c-${AWS::StackName}'
 ClusterConfiguration:
 Image:
 Os: alinux2
 HeadNode:
 InstanceType: t2.medium
 Networking:
 SubnetId: !GetAtt [PclusterVpc , Outputs.PublicSubnetId]
```

```
Ssh:
 KeyName: !Ref KeyName
Scheduling:
 Scheduler: slurm
 SlurmQueues:
 - Name: queue0
 ComputeResources:
 - Name: queue0-cr0
 InstanceType: t2.micro
 Networking:
 SubnetIds:
 - !GetAtt [PclusterVpc , Outputs.PrivateSubnetId]
Outputs:
 HeadNodeIp:
 Description: The Public IP address of the HeadNode
 Value: !GetAtt [PclusterCluster, headNode.publicIpAddress]
```

2. Führen Sie den folgenden AWS CLI-Befehl aus, um den CloudFormation Stack für die Clustererstellung und -verwaltung bereitzustellen.

```
$ aws cloudformation deploy --template-file ./cluster_template.yaml \
 --stack-name mycluster \
 --parameter-overrides KeyName=keypair \
 AvailabilityZone=us-east-2b \
 --capabilities CAPABILITY_NAMED_IAM CAPABILITY_AUTO_EXPAND
```

## CloudFormation Cluster-Ausgabe anzeigen

Sehen Sie sich die CloudFormation Cluster-Ausgabe an, um nützliche Cluster-Details zu erhalten. Die hinzugefügte `ValidationMessages` Eigenschaft ermöglicht den Zugriff auf Bestätigungsmeldungen aus Clustererstellungs- und -aktualisierungsvorgängen.

1. Navigieren Sie zur [CloudFormation Konsole](#) und wählen Sie den Stack aus, der Ihre AWS ParallelCluster benutzerdefinierte Ressource enthält.
2. Wählen Sie Stack-Details und dann die Registerkarte Ausgaben aus.



| Key                | Value                                                                                                                                                                                                                   | Description                                            |
|--------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------|
| HeadNodeIp         | 1.2.3.4                                                                                                                                                                                                                 | The Public IP address of the HeadNode                  |
| ValidationMessages | [[{"level": "WARNING", "type": "KeyPairValidator", "message": "If you do not specify a key pair, you can't connect to the instance unless you choose an AMI that is configured to allow users another way to log in"}]] | Any warnings from cluster create or update operations. |

Bestätigungsnachrichten werden möglicherweise gekürzt. Weitere Informationen zum Abrufen von Protokollen finden Sie unter [Fehlerbehebung für AWS ParallelCluster](#).

## Greifen Sie auf Ihren Cluster zu

Greifen Sie auf den Cluster zu.

**ssh** in den Cluster-Kopfknoten

1. Rufen Sie nach Abschluss der CloudFormation Stack-Bereitstellung die IP-Adresse des Hauptknotens mit dem folgenden Befehl ab:

```
$ HEAD_NODE_IP=$(aws cloudformation describe-stacks --stack-name=mycluster --query "Stacks|[0].Outputs[?OutputKey=='HeadNodeIp']|[0].OutputValue" --output=text)
```

Sie können die IP-Adresse des Hauptknotens auch über den HeadNodeIpParameter auf der Registerkarte Cluster-Stack-Ausgaben in der CloudFormation Konsole abrufen.

Die IP-Adresse des Hauptknotens finden Sie hier, da sie im Outputs Abschnitt der CloudFormation Clustervorlage speziell für diesen Beispielcluster hinzugefügt wurde.

2. Connect zum Cluster-Kopfknoten her, indem Sie den folgenden Befehl ausführen:

```
$ ssh -i keyname.pem ec2-user@$HEAD_NODE_IP
```

## Bereinigen

Löschen Sie den Cluster.

1. Führen Sie den folgenden AWS CLI-Befehl aus, um den CloudFormation Stack und den Cluster zu löschen.

```
$ aws cloudformation delete-stack --stack-name=mycluster
```

2. Überprüfen Sie den Status des Stack-Löschvorgangs, indem Sie den folgenden Befehl ausführen.

```
$ aws cloudformation describe-stacks --stack-name=mycluster
```

## AWS ParallelClusterUI-Integration mit Identity Center

Das Ziel dieses Tutorials besteht darin, zu demonstrieren, wie die AWS ParallelCluster Benutzeroberfläche in IAM Identity Center integriert werden kann, um eine Single-Sign-On-Lösung zu schaffen, die Benutzer in Active Directory vereint und mit Clustern gemeinsam genutzt werden kann.

### AWS ParallelCluster

Bei der Nutzung AWS ParallelCluster zahlen Sie nur für die AWS Ressourcen, die bei der Erstellung oder Aktualisierung von AWS ParallelCluster Images und Clustern entstehen. Weitere Informationen finden Sie unter [AWSDienste verwendet von AWS ParallelCluster](#).

Voraussetzungen:

- Eine vorhandene AWS ParallelCluster Benutzeroberfläche, die gemäß den Anweisungen [hier installiert werden kann](#).
- Ein vorhandenes verwaltetes Active Directory, vorzugsweise eines, das Sie auch für die [Integration](#) verwenden werdenAWS ParallelCluster.

## IAM Identity Center aktivieren

Wenn Sie bereits über ein Identity Center verfügen, das mit Ihrem AWS Managed Microsoft AD (Active Directory) verbunden ist, können Sie dieses verwenden und mit dem Abschnitt Hinzufügen Ihrer Anwendung zu IAM Identity Center fortfahren.

Wenn Sie noch kein Identity Center mit einem verbunden habenAWS Managed Microsoft AD, gehen Sie wie folgt vor, um es einzurichten.

### Identity Center aktivieren

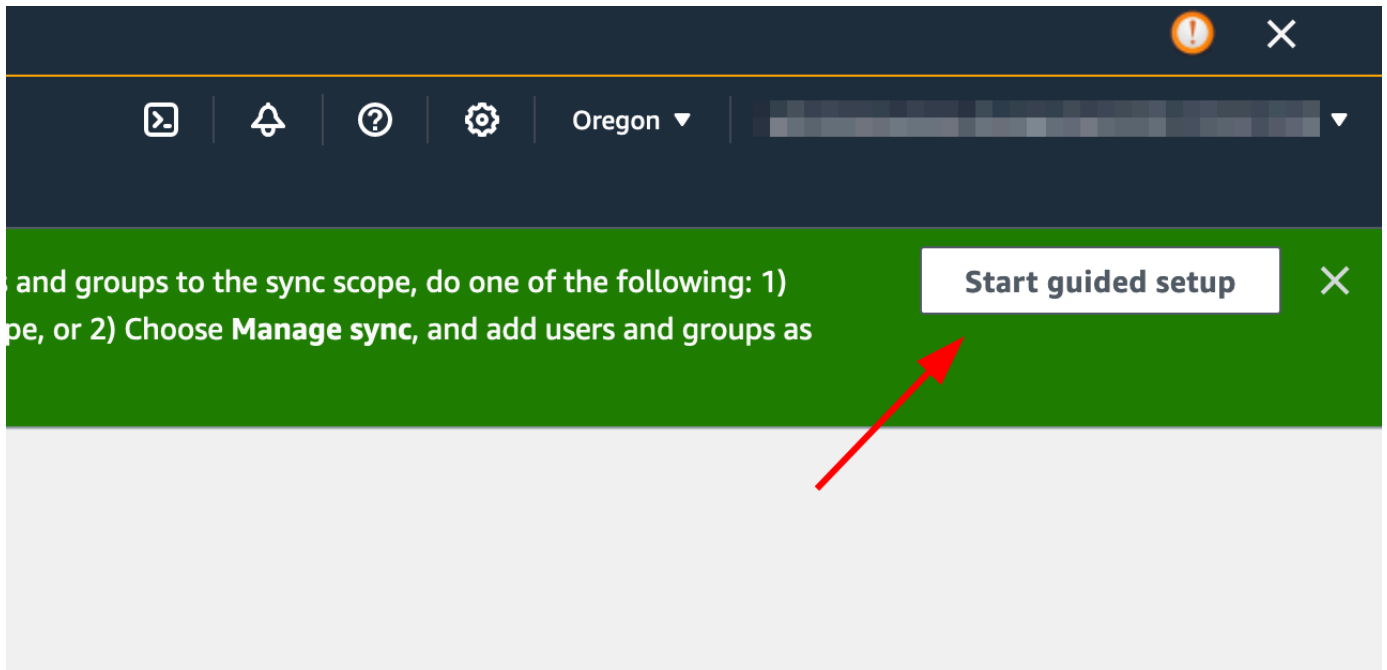
1. Navigieren Sie in der Konsole zu IAM Identity Center. (Stellen Sie sicher, dass Sie sich in der Region befinden, in der Sie Ihre haben AWS Managed Microsoft AD.)
2. Klicken Sie auf die Schaltfläche Aktivieren. Möglicherweise werden Sie gefragt, ob Sie Organisationen aktivieren möchten. Dies ist eine Voraussetzung, damit Sie sie aktivieren können. Hinweis: Dadurch erhalten Sie eine E-Mail an den Administrator Ihres Kontos mit einer Bestätigungs-E-Mail. Folgen Sie zur Bestätigung dem Link.

### Identity Center mit Managed AD verbinden

1. Auf der nächsten Seite sollten Sie nach der Aktivierung von Identity Center die empfohlenen Einrichtungsschritte sehen. Wählen Sie unter Schritt 1 die Option Wählen Sie Ihre Identitätsquelle aus.
2. Klicken Sie im Abschnitt Identitätsquelle auf das Dropdownmenü Aktionen (oben rechts) und wählen Sie dann Identitätsquelle ändern aus.
3. Wählen Sie Active Directory aus.
4. Wählen Sie unter Existierende Verzeichnisse Ihr Verzeichnis aus.
5. Klicken Sie auf Weiter.
6. Überprüfen Sie Ihre Änderungen, scrollen Sie nach unten, geben Sie zur Bestätigung ACCEPT in das Textfeld ein und klicken Sie dann auf Identitätsquelle ändern.
7. Warte, bis die Änderungen abgeschlossen sind. Dann solltest du oben ein grünes Banner sehen.

### Synchronisieren von Benutzern und Gruppen mit Identity Center

1. Klicken Sie im grünen Banner auf Geführte Installation starten (Schaltfläche oben rechts)



2. Klicken Sie unter Attributzuordnungen konfigurieren auf Weiter
3. Geben Sie im Abschnitt Synchronisierungsbereich konfigurieren den Namen der Benutzer ein, die mit Identity Center synchronisiert werden sollen, und klicken Sie dann auf Hinzufügen
4. Wenn Sie mit dem Hinzufügen von Benutzern und Gruppen fertig sind, klicken Sie auf Weiter

**Users**

Groups

**User**

corp.pcluster.com
▼

🔍

Add

**Added users and groups (4)** Remove

|                          | Username / Group name | Type | Domain            |
|--------------------------|-----------------------|------|-------------------|
| <input type="checkbox"/> | user1                 | User | corp.pcluster.com |
| <input type="checkbox"/> | user2                 | User | corp.pcluster.com |
| <input type="checkbox"/> | admin1                | User | corp.pcluster.com |
| <input type="checkbox"/> | admin2                | User | corp.pcluster.com |

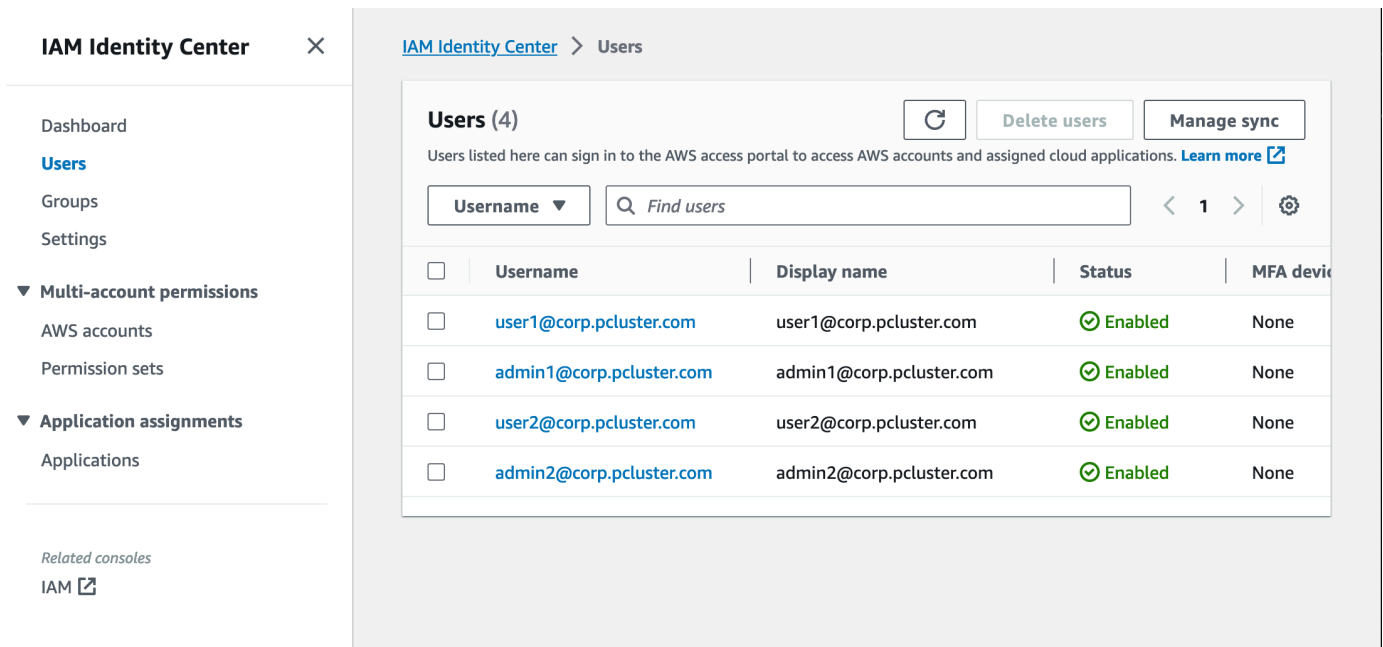
Cancel

Previous

Next

5. Überprüfen Sie Ihre Änderungen und klicken Sie dann auf Konfiguration speichern
6. Wenn auf dem nächsten Bildschirm eine Warnung angezeigt wird, dass Benutzer nicht synchronisiert werden, wählen Sie oben rechts die Schaltfläche Synchronisierung fortsetzen aus.
7. Um Benutzer zu aktivieren, wählen Sie als Nächstes auf der linken Seite auf der Registerkarte Benutzer einen Benutzer aus und klicken Sie dann auf Benutzerzugriff aktivieren > Benutzerzugriff aktivieren

Hinweis: Möglicherweise müssen Sie Synchronisierung fortsetzen auswählen, wenn oben ein Warnbanner angezeigt wird, und dann warten, bis die Benutzer synchronisiert sind (versuchen Sie mit der Schaltfläche „Aktualisieren“, ob sie bereits synchronisiert sind).



The screenshot shows the IAM Identity Center console interface. On the left is a navigation menu with options like Dashboard, Users, Groups, Settings, Multi-account permissions, and Application assignments. The main content area is titled 'IAM Identity Center > Users' and displays 'Users (4)'. It includes a search bar, a table of users, and action buttons like 'Delete users' and 'Manage sync'. The table lists four users with their usernames, display names, status (Enabled), and MFA device (None).

| <input type="checkbox"/> | Username                                 | Display name             | Status  | MFA device |
|--------------------------|------------------------------------------|--------------------------|---------|------------|
| <input type="checkbox"/> | <a href="#">user1@corp.pcluster.com</a>  | user1@corp.pcluster.com  | Enabled | None       |
| <input type="checkbox"/> | <a href="#">admin1@corp.pcluster.com</a> | admin1@corp.pcluster.com | Enabled | None       |
| <input type="checkbox"/> | <a href="#">user2@corp.pcluster.com</a>  | user2@corp.pcluster.com  | Enabled | None       |
| <input type="checkbox"/> | <a href="#">admin2@corp.pcluster.com</a> | admin2@corp.pcluster.com | Enabled | None       |

## Ihre Anwendung zum IAM Identity Center hinzufügen

Sobald Sie Ihre Benutzer mit IAM Identity Center synchronisiert haben, müssen Sie eine neue Anwendung hinzufügen. Dadurch wird konfiguriert, welche SSO-fähigen Anwendungen in Ihrem IAM Identity Center-Portal verfügbar sein werden. In diesem Fall fügen wir AWS ParallelCluster UI als Anwendung hinzu, während IAM Identity Center der Identitätsanbieter ist.

Im nächsten Schritt wird die AWS ParallelCluster Benutzeroberfläche als Anwendung in IAM Identity Center hinzugefügt. AWS ParallelCluster UI ist ein Webportal, das dem Benutzer hilft, seine Cluster zu verwalten. Weitere Informationen finden Sie unter [AWS ParallelClusterUI](#).

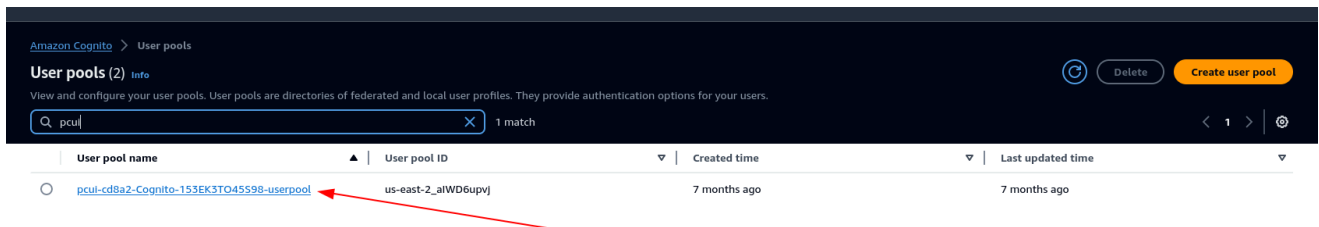
### Einrichtung der Anwendung in Identity Center

1. Unter IAM Identity Center > Anwendungen (klicken Sie in der linken Menüleiste auf Anwendungen)
2. Klicken Sie auf Anwendung hinzufügen
3. Wählen Sie Benutzerdefinierte SAML 2.0-Anwendung hinzufügen
4. Klicken Sie auf Weiter
5. Wählen Sie den Anzeigenamen und die Beschreibung aus, die Sie verwenden möchten (z. B. PCUI und AWS ParallelCluster UI)

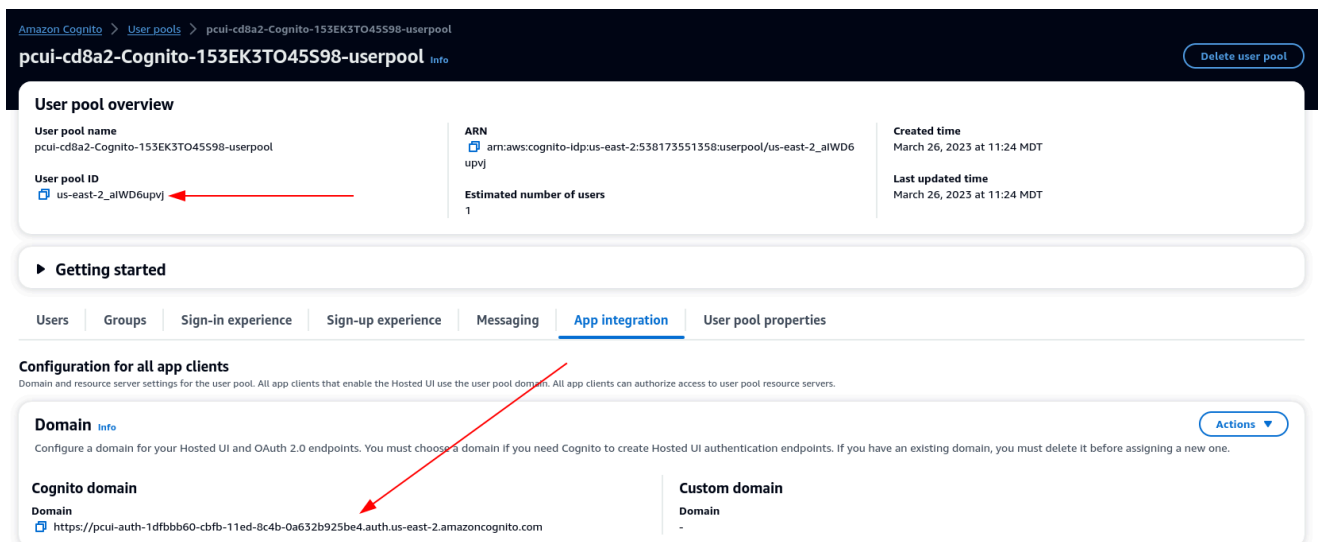
6. Kopieren Sie unter IAM Identity Center-Metadaten den Link für die IAM Identity Center SAML-Metadatendatei und speichern Sie ihn für später. Diese wird bei der Konfiguration von SSO in der Web-App verwendet
7. Geben Sie unter Anwendungseigenschaften in der Start-URL der Anwendung Ihre PCUI-Adresse ein. Diese finden Sie, indem Sie zur CloudFormation Konsole gehen, den Stack auswählen, der der PCUI entspricht (z. B. parallelcluster-ui), und auf der Registerkarte Outputs nach uiUrl suchen ParallelCluster

z. B. <https://m2iwazsi1j.execute-api.us-east-1.amazonaws.com>

8. Wählen Sie unter Anwendungsmetadaten die Option Manuelles Eingeben Ihrer Metadatenwerte aus. Geben Sie dann die folgenden Werte ein.
  - a. Wichtig: Achten Sie darauf, die Werte für das Domänenpräfix, die Region und die Benutzerpool-ID durch Informationen zu ersetzen, die für Ihre Umgebung spezifisch sind.
  - b. Das Domain-Präfix, die Region und die Benutzerpool-ID erhalten Sie, indem Sie die Konsole Amazon Cognito > Benutzerpools öffnen.



- c. Wählen Sie den Benutzerpool aus, der PCUI entspricht (der einen Benutzerpoolnamen wie PCUI-CD8A2-Cognito-153EK3TO45S98-UserPool haben wird)
- d. Navigieren Sie zu App Integration



9. <domain-prefix>URL des Application Assertion Consumer Service (ACS): <https://.auth.<region>.amazoncognito.com/saml2/idpresponse>

SAML-Anwendungszielgruppe: urn:amazon:cognito:sp: <userpool-id>

10. Wählen Sie Absenden aus. Rufen Sie dann die Detailseite für die Anwendung auf, die Sie hinzugefügt haben.
11. Wählen Sie die Dropdownliste Aktionen aus und wählen Sie Attributzuordnungen bearbeiten aus. Geben Sie dann die folgenden Attribute an.
- Benutzerattribut in der Anwendung: Betreff (Hinweis: Betreff ist vorausgefüllt.) → Ordnet diesem Zeichenkettenwert oder Benutzerattribut in IAM Identity Center zu: \$ {user:email}, Format: emailAddress
  - Benutzerattribut in der Anwendung: E-Mail → Ordnet diesem Zeichenfolgenwert oder Benutzerattribut in IAM Identity Center zu: \$ {user:email}, Format: nicht spezifiziert

Attribute mappings for PCUI

Attributes you map here become part of the SAML assertion that is sent to the application. You can choose which user attributes in your application map to corresponding user attributes in your connected directory. [Learn more](#)

| User attribute in the application | Maps to this string value or user attribute in IAM Identity Center | Format       |        |
|-----------------------------------|--------------------------------------------------------------------|--------------|--------|
| Subject                           | `\${user:email}`                                                   | emailAddress |        |
| email                             | `\${user:email}`                                                   | unspecified  | Remove |

[Add new attribute mapping](#)


[Cancel](#) [Save changes](#)

12. Speichern Sie Ihre Änderungen.
13. Wählen Sie die Schaltfläche „Benutzer zuweisen“ und weisen Sie dann Ihren Benutzer der Anwendung zu. Dies sind die Benutzer in Ihrem Active Directory, die Zugriff auf die PCUI-Schnittstelle haben werden.

[IAM Identity Center](#) > [Applications](#) > PCUI

## PCUI

**Details** Actions ▾

|                                                                                     |                      |                                       |
|-------------------------------------------------------------------------------------|----------------------|---------------------------------------|
|  | Display name<br>PCUI | Description<br>AWS ParallelCluster UI |
|-------------------------------------------------------------------------------------|----------------------|---------------------------------------|

**Assigned users (1)** [Remove access](#) [Assign Users](#)

The following users and groups from your connected directory can access this application. [Learn more](#)

| <input type="checkbox"/> | User/Group name                          | Type |
|--------------------------|------------------------------------------|------|
| <input type="checkbox"/> | <a href="#">admin1@corp.pcluster.com</a> | User |



## Konfigurieren Sie IAM Identity Center als SAML-IdP in Ihrem Benutzerpool

1. Wählen Sie in Ihren Benutzerpool-Einstellungen die Option **Anmeldeerfahrung** > **Identitätsanbieter hinzufügen** aus

The screenshot shows the AWS IAM Identity Center console for a user pool named 'parallelcluster-ui-Cognito-ZQRNCGGD3MHU-userpool'. The 'Sign-in experience' tab is selected, and a red arrow points to the 'Add identity provider' button in the 'Federated identity provider sign-in' section. The console displays the user pool overview, getting started options, and a table for federated identity providers (currently empty).

**User pool overview**

|                                                                           |                                                                                      |                                                           |
|---------------------------------------------------------------------------|--------------------------------------------------------------------------------------|-----------------------------------------------------------|
| <b>User pool name</b><br>parallelcluster-ui-Cognito-ZQRNCGGD3MHU-userpool | <b>ARN</b><br>arn:aws:cognito-idp:us-east-1:538173551358:userpool/us-east-1_Bgyu7Lz6 | <b>Created time</b><br>November 5, 2023 at 12:32 MST      |
| <b>User pool ID</b><br>us-east-1_Bgyu7LLz6                                | <b>Estimated number of users</b><br>1                                                | <b>Last updated time</b><br>November 5, 2023 at 12:32 MST |

**Getting started**

Users | Groups | **Sign-in experience** | Sign-up experience | Messaging | App integration | User pool properties

**Cognito user pool sign-in** Info  
Users can sign in using their email address, phone number, or user name. User attributes, group memberships, and security settings will be stored and configured in your user pool.

**Cognito user pool sign-in options**  
Email

**Federated identity provider sign-in (0)** Info  
Your app users can sign-in through external social identity providers like Facebook, Google, Amazon, or Apple, and through your on-prem directories via SAML or Open ID Connect.

Search identity providers by name

| Identity provider     | Identity provider type | Created time | Last updated time |
|-----------------------|------------------------|--------------|-------------------|
| No identity providers |                        |              |                   |

Add identity provider

2. Wählen Sie einen SAML-IdP
3. Geben Sie als Anbieternamen an IdentityCenter
4. Wählen Sie unter Metadaten-Dokumentquelle die Option URL für den Endpunkt des Metadatendokuments eingeben aus und geben Sie die URL ein, die Sie bei der Einrichtung der Anwendung von Identity Center kopiert haben
5. Wählen Sie unter „Attribute“ für E-Mail die Option E-Mail

### SAML

Configure a SAML 2.0 identity provider for your user pool.

#### Register your app with your SAML provider

To connect a SAML provider to Cognito, add your user pool as a relying party or application with your SAML 2.0 identity provider, and upload a metadata document to Cognito.

---

#### Set up SAML federation with this user pool

**Provider name** [Info](#)  
Enter a friendly name for your SAML 2.0 identity provider.

**Identifiers - optional** [Info](#)  
Enter identifiers for this provider. Identifiers can be used to redirect users to the correct IdP in multitenant apps.

Separate each identifier by a comma

**Sign-out flow** [Info](#)  
 Add sign-out flow  
Enable simultaneous sign-out from the SAML provider and Cognito.

**Metadata document source** [Info](#)  
Provide a SAML metadata document. This document is issued by your SAML provider. It includes the issuer's name, expiration information, and keys that can be used to validate the response from the identity provider.

Upload metadata document  
 Enter metadata document endpoint URL

**Enter metadata document endpoint URL** [Info](#)

---

#### Map attributes between your SAML provider and your user pool

Your required attributes are mapped to the equivalent SAML attributes. Each attribute you add must be mapped to a SAML attribute.

| User pool attribute | SAML attribute |
|---------------------|----------------|
| email               | email          |

[Add another attribute](#)

[Cancel](#) [Add identity provider](#)

## 6. Wählen Sie Identitätsanbieter hinzufügen aus.

### Integrieren Sie den IdP in den Benutzerpool-App-Client

1. Wählen Sie als Nächstes im Bereich App-Integration Ihres Benutzerpools den Client aus, der unter App-Client-Liste aufgeführt ist

Amazon Cognito > User pools > parallelcluster-ui-Cognito-ZQRNCGGD3MHU-userpool

## parallelcluster-ui-Cognito-ZQRNCGGD3MHU-userpool Info

[Delete user pool](#)

### User pool overview

|                                                                           |                                                                                      |                                                           |
|---------------------------------------------------------------------------|--------------------------------------------------------------------------------------|-----------------------------------------------------------|
| <b>User pool name</b><br>parallelcluster-ui-Cognito-ZQRNCGGD3MHU-userpool | <b>ARN</b><br>arn:aws:cognito-idp:us-east-1:123456789012:userpool/us-east-1_Bgyu7Lz6 | <b>Created time</b><br>November 5, 2023 at 12:32 MST      |
| <b>User pool ID</b><br>us-east-1_Bgyu7Lz6                                 | <b>Estimated number of users</b><br>1                                                | <b>Last updated time</b><br>November 5, 2023 at 12:32 MST |

### Getting started

Users | Groups | Sign-in experience | Sign-up experience | Messaging | **App integration** | User pool properties

### Configuration for all app clients

Domain and resource server settings for the user pool. All app clients that enable the Hosted UI use the user pool domain. All app clients can authorize access to user pool resource servers.

#### Domain Info

Configure a domain for your Hosted UI and OAuth 2.0 endpoints. You must choose a domain if you need Cognito to create Hosted UI authentication endpoints. If you have an existing domain, you must delete it before assigning a new one.

**Cognito domain**

Domain  
https://pcul-auth-06f29200-7c12-11ee-b755-0e11297ecb0d.auth.us-east-1.amazoncognito.com

**Custom domain**

Domain  
-

[Actions](#)

#### Resource servers (0) Info

Configure resource servers. A resource server is a remote server that authorizes access based on OAuth 2.0 scopes in an access token.

| Resource server name | Resource server identifier | Custom scopes |
|----------------------|----------------------------|---------------|
| No resource servers  |                            |               |

[Create resource server](#)

#### App client defaults

Hosted UI customization and advanced security settings for the user pool. You can customize the Hosted UI and advanced security in app clients to override the defaults.

#### Hosted UI customization Info

Customize the hosted sign-up and sign-in pages to match your app's style and branding by uploading your own logo and customized CSS.

**Logo**  
-

**Custom CSS**  
-

[Edit](#)

#### Advanced security Info

Configure advanced security features, including Cognito's automatic responses to suspicious user activity. Advanced security adds cost to your bill. [See pricing](#)

**Status**  
 Disabled

[Enable](#)

#### App client list

The app clients that integrate your apps with your user pool. Configure client overrides to user pool default configurations, and configure Amazon Pinpoint analytics.

#### App clients and analytics (1) Info

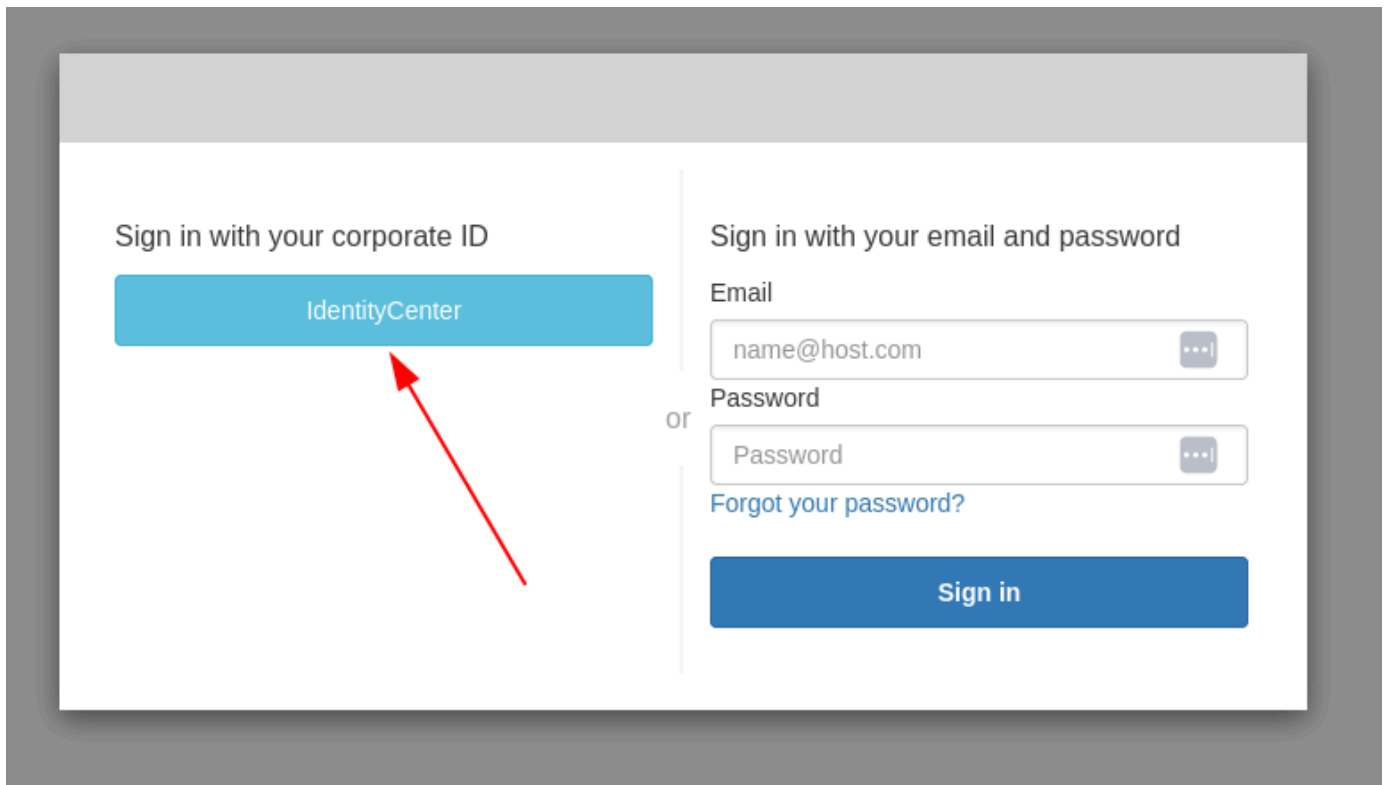
Configure an app client. App clients are the user pool authentication resources attached to your app. Select an app client to configure the permitted authentication actions for an app.

| App client name                                                     | Client ID |
|---------------------------------------------------------------------|-----------|
| <input type="radio"/> <a href="#">CognitoAppClient-ETqXbqL5wRVs</a> | ...       |

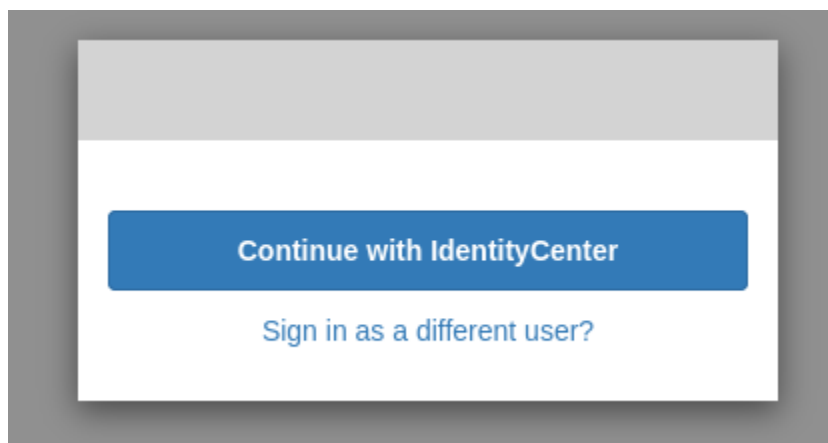
2. Wählen Sie unter Gehostete Benutzeroberfläche die Option Bearbeiten
3. Wählen Sie unter Identitätsanbieter IdentityCenterebenfalls aus.
4. Wählen Sie Save Changes (Änderungen speichern)

Bestätigen Sie Ihr Setup

1. Als Nächstes validieren wir das Setup, das wir gerade erstellt haben, indem wir uns bei PCUI anmelden. Melden Sie sich bei Ihrem PCUI-Portal an und Sie sollten nun eine Option sehen, mit der Sie sich mit Ihrer Unternehmens-ID anmelden können:



2. Wenn Sie auf die IdentityCenterSchaltfläche klicken, gelangen Sie zur IAM Identity Center IdP-Anmeldung, gefolgt von einer Seite mit Ihren Anwendungen, die PCUI enthält. Öffnen Sie diese Anwendung.
3. Sobald Sie zum folgenden Bildschirm gelangen, wurde Ihr Benutzer dem Cognito-Benutzerpool hinzugefügt.



## Machen Sie Ihren Benutzer zum Administrator

1. Navigieren Sie nun zur Konsole Amazon Cognito > Benutzerpools und wählen Sie den neu erstellten Benutzer aus, der das Präfix identitycenter haben soll.

The screenshot shows the AWS IAM console interface for a Cognito user pool. The top navigation bar includes 'Amazon Cognito', 'User pools', and the specific user pool name 'parallecluster-ui-Cognito-ZQRNCGGD3MHU-userpool'. Below this is a 'User pool overview' card with details like 'User pool name', 'ARN', 'User pool ID', 'Created time', and 'Estimated number of users'. The main section is titled 'Users (2)' and contains a table of users. A red arrow points to the user 'identitycenter\_admin1@corp.pcluster.c...' in the table. Below the table is an 'Import users (0)' section.

| User name                                | Email address            | Email verified | Confirmation status | Status  |
|------------------------------------------|--------------------------|----------------|---------------------|---------|
| c8f7eccc-e647-4227-afa2-080274ebfefe     | user@amazon.com          | Yes            | Confirmed           | Enabled |
| identitycenter_admin1@corp.pcluster.c... | admin1@corp.pcluster.com | No             | External provider   | Enabled |

2. Wählen Sie unter Gruppenmitgliedschaften die Option Benutzer zur Gruppe hinzufügen aus, wählen Sie Administrator aus und klicken Sie auf Hinzufügen.
3. Wenn Sie nun auf Weiter mit klicken, werden IdentityCenter Sie zur AWS ParallelCluster UI-Seite weitergeleitet.

# Fehlerbehebung für AWS ParallelCluster

Die AWS ParallelCluster Community unterhält eine Wiki-Seite, die viele Tipps zur Fehlerbehebung im [AWS ParallelCluster GitHub Wiki](#) enthält. Eine Liste der bekannten Probleme finden Sie unter [Bekannte Probleme](#).

## Themen

- [Es wird versucht, einen Cluster zu erstellen](#)
- [Ich versuche, einen Job auszuführen](#)
- [Es wird versucht, einen Cluster zu aktualisieren](#)
- [Ich versuche, auf Speicher zuzugreifen](#)
- [Es wird versucht, einen Cluster zu löschen](#)
- [Ich versuche, den AWS ParallelCluster API-Stack zu aktualisieren](#)
- [Fehler bei der Initialisierung von Compute-Knoten werden angezeigt](#)
- [Anleitung ster-ster-ster-ster-ster-ster-ster-ster-](#)
- [Behebung von Problemen bei der Clusterbereitstellung](#)
- [Behebung von Skalierungsproblemen](#)
- [Probleme beim Platzieren von Gruppen und beim Starten von Instances](#)
- [Verzeichnisse, die nicht ersetzt werden können](#)
- [Behebung von Problemen in NICE DCV](#)
- [Behebung von Problemen in Clustern mit AWS Batch Integration](#)
- [Problembehandlung bei der Mehrbenutzerintegration mit Active Directory](#)
- [Behebung von Problemen mit benutzerdefinierten AMIs](#)
- [Fehlerbehebung bei einem Timeout für ein Cluster-Update, wenn es nicht läuft cfn-hup](#)
- [Fehlerbehebung im Netzwerk](#)
- [Das Cluster-Update ist bei der benutzerdefinierten Aktion onNodeUpdated fehlgeschlagen](#)
- [Fehler bei der benutzerdefinierten Slurm Konfiguration werden angezeigt](#)
- [Cluster-Alarme](#)
- [Zusätzliche Unterstützung](#)

## Es wird versucht, einen Cluster zu erstellen

Wenn Sie AWS ParallelCluster Version 3.5.0 und höher zum Erstellen eines Clusters verwenden und die Clustererstellung mit der `--rollback-on-failure` Einstellung auf `fehlgeschlagen` ist `false`, verwenden Sie den [`pcluster describe-cluster`](#) CLI-Befehl, um Status- und Fehlerinformationen abzurufen. In diesem Fall ist `clusterStatus` `CREATE_FAILED` die erwartete `pcluster describe-cluster` Ausgabe. Suchen Sie im `failures` Abschnitt in der Ausgabe nach dem `failureCode` und `failureReason`. Suchen Sie dann im folgenden Abschnitt nach dem passenden `failureCode`, um zusätzliche Hilfe zur Fehlerbehebung zu erhalten. Weitere Informationen finden Sie unter [`pcluster describe-cluster`](#).

In den folgenden Abschnitten empfehlen wir Ihnen, die Protokolle auf dem Hauptknoten zu überprüfen, z. B. die `/var/log/chef-client.log` Dateien `/var/log/cfn-init.log` und. Weitere Informationen zu AWS ParallelCluster Protokollen und deren Anzeige finden Sie unter [Wichtige Protokolle für das Debuggen](#) und [Protokolle abrufen und aufbewahren](#).

Wenn Sie noch keinen `failureCode` haben, navigieren Sie zur AWS CloudFormation Konsole, um den Cluster-Stack anzuzeigen. Suchen Sie `Status Reason` nach Fehlern auf anderen Ressourcen, um weitere Fehlerdetails zu finden. `HeadNodeWaitCondition` Weitere Informationen finden Sie unter [AWS CloudFormationEreignisse anzeigen auf CREATE\\_FAILED](#). Überprüfen Sie die `/var/log/chef-client.log` Dateien `/var/log/cfn-init.log` und auf dem Hauptknoten.

### **failureCode ist OnNodeConfiguredExecutionFailure**

- Warum ist es gescheitert?

Sie haben in `OnNodeConfigured` der Konfiguration im Abschnitt „Hauptknoten“ ein benutzerdefiniertes Skript zur Erstellung eines Clusters bereitgestellt. Das benutzerdefinierte Skript konnte jedoch nicht ausgeführt werden.

- Wie löst man das Problem?

In der `/var/log/cfn-init.log` Datei erfahren Sie mehr über den Fehler und darüber, wie Sie das Problem in Ihrem benutzerdefinierten Skript beheben können. Gegen Ende dieses Protokolls werden nach der `Running command runpostinstall` Meldung möglicherweise Informationen zur Ausführung des `OnNodeConfigured` Skripts angezeigt.

## **failureCode ist OnNodeConfiguredDownloadFailure**

- Warum ist es gescheitert?

Sie haben in `OnNodeConfigured` der Konfiguration im Abschnitt „Hauptknoten“ ein benutzerdefiniertes Skript zur Erstellung eines Clusters bereitgestellt. Das benutzerdefinierte Skript konnte jedoch nicht heruntergeladen werden.

- Wie löst man das Problem?

Stellen Sie sicher, dass die URL gültig und der Zugriff korrekt konfiguriert ist. Weitere Informationen zur Konfiguration von benutzerdefinierten Bootstrap-Skripten finden Sie unter [Benutzerdefinierte Bootstrap-Aktionen](#).

Überprüfen Sie die `/var/log/cfn-init.log` Datei. Gegen Ende dieses Protokolls werden nach der Running command `runpostinstall` Meldung möglicherweise Ausführungsinformationen zur `OnNodeConfigured` Skriptverarbeitung, einschließlich des Herunterladens, angezeigt.

## **failureCode ist OnNodeConfiguredFailure**

- Warum ist es fehlgeschlagen?

Sie haben in `OnNodeConfigured` der Konfiguration im Abschnitt „Hauptknoten“ ein benutzerdefiniertes Skript zur Erstellung eines Clusters bereitgestellt. Die Verwendung des benutzerdefinierten Skripts schlug jedoch in der Clusterbereitstellung fehl. Eine unmittelbare Ursache kann nicht ermittelt werden und es sind weitere Untersuchungen erforderlich.

- Wie löst man das Problem?

Überprüfe die `/var/log/cfn-init.log` Datei. Gegen Ende dieses Protokolls werden nach der Running command `runpostinstall` Meldung möglicherweise Ausführungsinformationen zur `OnNodeConfigured` Skriptverarbeitung angezeigt.

## **failureCode ist OnNodeStartExecutionFailure**

- Warum ist es fehlgeschlagen?



Sie haben in `OnNodeStart` der Konfiguration im Abschnitt „Hauptknoten“ ein benutzerdefiniertes Skript zur Erstellung eines Clusters bereitgestellt. Das benutzerdefinierte Skript konnte jedoch nicht ausgeführt werden.

- Wie löst man das Problem?

In der `/var/log/cfn-init.log` Datei erfahren Sie mehr über den Fehler und darüber, wie Sie das Problem in Ihrem benutzerdefinierten Skript beheben können. Gegen Ende dieses Protokolls werden nach der `Running command runpreinstall` Meldung möglicherweise Informationen zur Ausführung des `OnNodeStart` Skripts angezeigt.

## **failureCode ist OnNodeStartDownloadFailure**

- Warum ist es gescheitert?

Sie haben in `OnNodeStart` der Konfiguration im Abschnitt „Hauptknoten“ ein benutzerdefiniertes Skript zur Erstellung eines Clusters bereitgestellt. Das benutzerdefinierte Skript konnte jedoch nicht heruntergeladen werden.

- Wie löst man das Problem?

Stellen Sie sicher, dass die URL gültig und der Zugriff korrekt konfiguriert ist. Weitere Informationen zur Konfiguration von benutzerdefinierten Bootstrap-Skripten finden Sie unter [Benutzerdefinierte Bootstrap-Aktionen](#).

Überprüfen Sie die `/var/log/cfn-init.log` Datei. Gegen Ende dieses Protokolls werden nach der `Running command runpreinstall` Meldung möglicherweise Ausführungsinformationen zur `OnNodeStart` Skriptverarbeitung, einschließlich des Herunterladens, angezeigt.

## **failureCode ist OnNodeStartFailure**

- Warum ist es fehlgeschlagen?

Sie haben in der Konfiguration im Abschnitt `OnNodeStart` des Hauptknotens ein benutzerdefiniertes Skript zur Erstellung eines Clusters bereitgestellt. Die Verwendung des benutzerdefinierten Skripts schlug jedoch in der Clusterbereitstellung fehl. Eine unmittelbare Ursache kann nicht ermittelt werden und es sind weitere Untersuchungen erforderlich.

- Wie löst man das Problem?

Überprüfe die `/var/log/cfn-init.log` Datei. Gegen Ende dieses Protokolls werden nach der Running command `runpreinstall` Meldung möglicherweise Ausführungsinformationen zur `OnNodeStart` Skriptverarbeitung angezeigt.

## **failureCode ist EbsMountFailure**

- Warum ist es fehlgeschlagen?

Das in der Clusterkonfiguration definierte EBS-Volume konnte nicht bereitgestellt werden.

- Wie löst man das Problem?

Überprüfen Sie die `/var/log/chef-client.log` Datei auf Fehlerdetails.

## **failureCode ist EfsMountFailure**

- Warum ist es gescheitert?

Das in der Cluster-Konfiguration definierte Amazon EFS-Volume konnte nicht bereitgestellt werden.

- Wie löst man das Problem?

Wenn Sie ein vorhandenes Amazon EFS-Dateisystem definiert haben, stellen Sie sicher, dass Datenverkehr zwischen dem Cluster und dem Dateisystem zulässig ist. Weitere Informationen finden Sie unter [SharedStorage/EfsSettings/FileSystemId](#).

Überprüfen Sie die `/var/log/chef-client.log` Datei auf Fehlerdetails.

## **failureCode ist FsxMountFailure**

- Warum ist es gescheitert?

Das in der Cluster-Konfiguration definierte Amazon FSx-Dateisystem konnte nicht bereitgestellt werden.

- Wie löst man das Problem?

Wenn Sie ein vorhandenes Amazon FSx-Dateisystem definiert haben, stellen Sie sicher, dass Datenverkehr zwischen dem Cluster und dem Dateisystem zulässig ist. Weitere Informationen finden Sie unter [SharedStorage/FsxLustreSettings/FileSystemId](#).

Überprüfen Sie die `/var/log/chef-client.log` Datei auf Fehlerdetails.

## **failureCode ist RaidMountFailure**

- Warum ist es gescheitert?

Die in der Cluster-Konfiguration definierten RAID-Volumes konnten nicht bereitgestellt werden.

- Wie löst man das Problem?

Überprüfen Sie die `/var/log/chef-client.log` Datei auf Fehlerdetails.

## **failureCode ist AmiVersionMismatch**

- Warum ist es gescheitert?

Die AWS ParallelCluster Version, die zum Erstellen des benutzerdefinierten AMI verwendet wurde, unterscheidet sich von der AWS ParallelCluster Version, die zur Konfiguration des Clusters verwendet wurde. Sehen Sie sich in der CloudFormation Konsole die CloudFormation Cluster-Stack-Details an und klicken Sie auf die `HeadNodeWaitCondition`, um zusätzliche Informationen zu den AWS ParallelCluster Versionen und dem AMI zu erhalten. Weitere Informationen finden Sie unter [AWS CloudFormationEreignisse anzeigen auf CREATE\\_FAILED](#).

- Wie löst man das Problem?

Stellen Sie sicher, dass es sich bei der AWS ParallelCluster Version, die zur Erstellung des benutzerdefinierten AMI verwendet wurde, um dieselbe AWS ParallelCluster Version handelt, die zur Konfiguration des Clusters verwendet wurde. Sie können entweder die benutzerdefinierte AMI-Version oder die `pccluster` CLI-Version ändern, um sie identisch zu machen.

## **failureCode ist InvalidAmi**

- Warum ist es gescheitert?

Das benutzerdefinierte AMI ist ungültig, da es nicht mit erstellt wurde AWS ParallelCluster.

- Wie löst man das Problem?

Verwenden Sie den `pcluster build-image` Befehl, um ein AMI zu erstellen, indem Sie Ihr AMI zum übergeordneten Image machen. Weitere Informationen finden Sie unter [pcluster build-image](#).

## **failureCode** lautet „**failureReason** Fehler **HeadNodeBootstrapFailure** beim Einrichten des Hauptknotens“.

- Warum ist es gescheitert?

Eine unmittelbare Ursache kann nicht ermittelt werden und zusätzliche Untersuchungen sind erforderlich. Es könnte beispielsweise sein, dass sich der Cluster im geschützten Status befindet, und dies könnte darauf zurückzuführen sein, dass die statische Rechenflotte nicht bereitgestellt werden konnte.

- Wie löst man das Problem?

Überprüfen Sie die `/var/log/chef-client.log`. Datei auf Fehlerdetails.

### Note

Wenn Sie eine `RuntimeError` Ausnahme sehen `Cluster state has been set to PROTECTED mode due to failures detected in static node provisioning`, befindet sich der Cluster im geschützten Status. Weitere Informationen finden Sie unter [So debuggen Sie den geschützten Modus](#).

## **failureCode** hat das Timeout **HeadNodeBootstrapFailure** bei der **failureReason** Clustererstellung überschritten.

- Warum ist es gescheitert?

Standardmäßig gibt es ein Zeitlimit von 30 Minuten, bis die Clustererstellung abgeschlossen ist. Wenn die Clustererstellung nicht innerhalb dieses Zeitrahmens abgeschlossen wurde, schlägt die Clustererstellung mit einem Timeoutfehler fehl. Bei der Clustererstellung

kann es aus verschiedenen Gründen zu einem Timeout kommen. Timeoutfehler können beispielsweise durch einen Fehler bei der Erstellung eines Hauptknotens, ein Netzwerkproblem, benutzerdefinierte Skripts, deren Ausführung im Hauptknoten zu lange dauert, einen Fehler in einem benutzerdefinierten Skript, das in Rechenknoten ausgeführt wird, oder lange Wartezeiten bei der Bereitstellung von Rechenknoten verursacht werden. Eine unmittelbare Ursache kann nicht ermittelt werden und es sind zusätzliche Untersuchungen erforderlich.

- Wie löst man das Problem?

Einzelheiten zum Fehler finden Sie in den `/var/log/chef-client.log` Dateien `/var/log/cfn-init.log` und. Weitere Informationen zu AWS ParallelCluster Protokollen und deren Abruf finden Sie unter [Wichtige Protokolle für das Debuggen](#) und [Protokolle abrufen und aufbewahren](#).

Möglicherweise finden Sie in diesen Protokollen Folgendes.

- Ich sehe **Waiting for static fleet capacity provisioning** fast das Ende des **chef-client.log**

Dies deutet darauf hin, dass bei der Clustererstellung eine Zeitüberschreitung aufgetreten ist, als auf das Hochfahren statischer Knoten gewartet wurde. Weitere Informationen finden Sie unter [Fehler bei der Initialisierung von Compute-Knoten werden angezeigt](#).

- Das Skript für Seeing **OnNodeConfigured** oder **OnNodeStart** Head Node ist am Ende des **cfn-init.log**

Dies weist darauf hin, dass die Ausführung des Skripts `OnNodeConfigured` oder des `OnNodeStart` benutzerdefinierten Skripts lange gedauert hat und einen Timeoutfehler verursacht hat. Überprüfen Sie Ihr benutzerdefiniertes Skript auf Probleme, die dazu führen könnten, dass es über einen längeren Zeitraum ausgeführt wird. Wenn die Ausführung Ihres benutzerdefinierten Skripts viel Zeit in Anspruch nimmt, sollten Sie erwägen, das Timeout-Limit zu ändern, indem Sie Ihrer Cluster-Konfigurationsdatei einen `DevSettings` Abschnitt hinzufügen, wie im folgenden Beispiel gezeigt:

```
DevSettings:
 Timeouts:
 HeadNodeBootstrapTimeout: 1800 # default setting: 1800 seconds
```

- Die Protokolle können nicht gefunden werden, oder der Hauptknoten wurde nicht erfolgreich erstellt

Es ist möglich, dass der Hauptknoten nicht erfolgreich erstellt wurde und die Protokolle nicht gefunden werden können. Sehen Sie sich in der CloudFormation Konsole die Cluster-Stack-Details an, um nach weiteren Fehlerdetails zu suchen.

## **failureCode** lautet „**failureReasonFehler HeadNodeBootstrapFailure** beim Bootstrapping des Hauptknotens“.

- Warum ist es gescheitert?

Eine unmittelbare Ursache kann nicht ermittelt werden und zusätzliche Untersuchungen sind erforderlich.

- Wie löst man das Problem?

Überprüfen Sie die `/var/log/chef-client.log` Dateien `/var/log/cfn-init.log` und.

## **failureCode** ist **ResourceCreationFailure**

- Warum ist es gescheitert?

Die Erstellung einiger Ressourcen ist während der Clustererstellung fehlgeschlagen. Der Fehler kann aus verschiedenen Gründen auftreten. Fehler bei der Ressourcenerstellung können beispielsweise durch Kapazitätsprobleme oder eine falsch konfigurierte IAM-Richtlinie verursacht werden.

- Wie kann das Problem behoben werden?

Sehen Sie sich in der CloudFormation Konsole den Cluster-Stack an, um nach weiteren Informationen zu Fehlern bei der Ressourcenerstellung zu suchen.

## **failureCode** ist **ClusterCreationFailure**

- Warum ist es gescheitert?

Eine unmittelbare Ursache kann nicht ermittelt werden und zusätzliche Untersuchungen sind erforderlich.

- Wie löst man das Problem?

Sehen Sie sich in der CloudFormation Konsole den Cluster-Stack an und suchen Sie `HeadNodeWaitCondition` nach weiteren Fehlerdetails. Status Reason

Überprüfen Sie die `/var/log/chef-client.log` Dateien `/var/log/cfn-init.log` und.

## **WaitCondition timed out...**Im CloudFormation Stapel sehen

Weitere Informationen finden Sie unter [failureCode hat das Timeout HeadNodeBootstrapFailure bei der failureReason Clustererstellung überschritten.](#)

## **Resource creation cancelled**Im CloudFormation Stapel sehen

Weitere Informationen finden Sie unter [failureCode ist ResourceCreationFailure.](#)

## Sehen **Failed to run cfn-init...** oder andere Fehler im AWS CloudFormation Stapel

Weitere Fehlerdetails finden Sie unter `/var/log/cfn-init.log` und `/var/log/chef-client.log`.

## Sehen **chef-client.log** endet mit **INFO: Waiting for static fleet capacity provisioning**

Dies hängt mit dem Timeout bei der Clustererstellung zusammen, wenn auf das Hochfahren statischer Knoten gewartet wird. Weitere Informationen finden Sie unter [Fehler bei der Initialisierung von Compute-Knoten werden angezeigt.](#)

## Sehen **Failed to run preinstall or postinstall in cfn-init.log**

Sie haben ein `OnNodeConfigured OnNodeStart` OR-Skript im `HeadNode` Abschnitt Cluster-Konfiguration. Das Skript funktioniert nicht richtig. Suchen Sie in der `/var/log/cfn-init.log` Datei nach benutzerdefinierten Skriptfehlerdetails.

**This AMI was created with xxx, but is trying to be used with xxx...** Im CloudFormation Stapel sehen

Weitere Informationen finden Sie unter [failureCode ist AmiVersionMismatch](#).

**This AMI was not baked by AWS ParallelCluster...** Im CloudFormation Stapel sehen

Weitere Informationen finden Sie unter [failureCode ist InvalidAmi](#).

Der **pcluster create-cluster** Befehl Seeing kann nicht lokal ausgeführt werden

Suchen Sie `~/parallelcluster/pcluster-cli.log` in Ihrem lokalen Dateisystem nach Fehlerdetails.

## Zusätzliche Unterstützung

Folgen Sie den Anleitungen zur Fehlerbehebung unter [Behebung von Problemen bei der Clusterbereitstellung](#).

Prüfen Sie, ob Ihr Szenario unter [GitHub Bekannte Probleme unter](#) AWS ParallelCluster on behandelt wird GitHub.

Zusätzliche Unterstützung finden Sie unter [Zusätzliche Unterstützung](#).

Ich versuche, einen Job auszuführen

**srun** Der interaktive Job schlägt mit einem Fehler fehl **srun: error: fwd\_tree\_thread: can't find address for <host>, check slurm.conf**

- Warum ist es gescheitert?

Sie haben den `srun` Befehl zum Senden eines Jobs ausgeführt und dann die Größe einer Warteschlange erhöht, indem Sie den `pcluster update-cluster` Befehl verwendet haben, ohne die Slurm Daemons nach Abschluss des Updates neu zu starten.



Slurm organisiert Slurm Daemons in einer Baumhierarchie, um die Kommunikation zu optimieren. Diese Hierarchie wird nur aktualisiert, wenn die Daemons starten.

Angenommensrun, Sie starten einen Job und führen dann den `pcluster update-cluster` Befehl aus, um die Warteschlange zu vergrößern. Neue Rechenknoten werden im Rahmen des Updates gestartet. Stellt Ihren Job dann in eine Slurm Warteschlange auf einen der neuen Rechenknoten. In diesem Fall erkennen sowohl die Slurm Daemons als auch die neuen Rechenknoten `srun` nicht. `srun` gibt einen Fehler zurück, weil die neuen Knoten nicht erkannt werden.

- Wie löst man das Problem?

Starten Sie die Slurm Daemons auf allen Rechenknoten neu und verwenden Sie sie dann, `srun` um Ihren Job einzureichen. Sie können den Neustart der Slurm Daemons planen, indem Sie den `scontrol reboot` Befehl ausführen, mit dem die Rechenknoten neu gestartet werden. Weitere Informationen finden Sie in der Dokumentation unter [scontrol reboot](#). Slurm Sie können die Slurm Daemons auf den Rechenknoten auch manuell neu starten, indem Sie einen Neustart der entsprechenden Dienste anfordern. `systemd`

## Der Job steckt im **CF** Status mit dem **squeue** Befehl fest

Dies könnte ein Problem beim Einschalten dynamischer Knoten sein. Weitere Informationen finden Sie unter [Fehler bei der Initialisierung von Compute-Knoten werden angezeigt](#).

## Großaufträge ausführen und sehen **nfsd: too many open connections, consider increasing the number of threads in /var/log/messages**

Bei einem Netzwerkdateisystem nimmt die I/O-Wartezeit ebenfalls zu, wenn die Netzwerkgrenzen erreicht werden. Dies kann zu Soft-Lockups führen, da das Netzwerk zum Schreiben von Daten sowohl für Netzwerk- als auch für I/O-Metriken verwendet wird.

Bei Instances der 5. Generation verwenden wir den ENA-Treiber, um Paketzähler verfügbar zu machen. Diese Zähler zählen die Pakete, die dadurch AWS geformt werden, dass das Netzwerk die Bandbreitengrenzen der Instanz erreicht. Sie können diese Zähler überprüfen, um festzustellen, ob sie größer als 0 sind. Wenn dies der Fall ist, haben Sie Ihre Bandbreitenlimits überschritten.

Sie können diese Zähler anzeigen, indem Sie Folgendes ausführen `ethtool -S eth0 | grep exceeded`.

Eine Überschreitung der Netzwerkgrenzen ist häufig darauf zurückzuführen, dass zu viele NFS-Verbindungen unterstützt werden. Dies ist eines der ersten Dinge, die Sie überprüfen sollten, wenn Sie Netzwerkgrenzen erreichen oder überschreiten.

Die folgende Ausgabe zeigt beispielsweise gelöschte Pakete:

```
$ ethtool -S eth0 | grep exceeded
bw_in_allowance_exceeded: 38750610
bw_out_allowance_exceeded: 1165693
pps_allowance_exceeded: 103
conntrack_allowance_exceeded: 0
linklocal_allowance_exceeded: 0
```

Um diese Meldung zu vermeiden, sollten Sie erwägen, den Instanztyp des Hauptknotens in einen leistungsfähigeren Instanztyp zu ändern. Erwägen Sie, Ihren Datenspeicher in Dateisysteme mit gemeinsam genutztem Speicher zu verschieben, die nicht als NFS-Freigabe exportiert werden, z. B. Amazon EFS oder Amazon FSx. Weitere Informationen finden Sie unter [Gemeinsamer Speicher](#) und die [Best Practices](#) im AWS ParallelCluster Wiki unter. GitHub

## Einen MPI-Job ausführen

### Debug-Modus aktivieren

Informationen zum Aktivieren des OpenMPI-Debug-Modus finden Sie unter [Welche Steuerelemente hat Open MPI, die beim Debuggen helfen](#).

[Informationen zum Aktivieren des IntelMPI-Debug-Modus finden Sie unter Andere Umgebungsvariablen.](#)

### Anzeige **MPI\_ERRORS\_ARE\_FATAL** und **OPAL ERROR** in der Jobausgabe

Diese Fehlercodes stammen aus der MPI-Schicht in Ihrer Anwendung. Informationen zum Abrufen von MPI-Debug-Logs aus Ihrer Anwendung finden Sie unter. [Debug-Modus aktivieren](#)

Eine mögliche Ursache für diesen Fehler ist, dass Ihre Anwendung für eine bestimmte MPI-Implementierung wie OpenMPI kompiliert wurde und Sie versuchen, sie mit einer anderen MPI-

Implementierung wie IntelMPI auszuführen. Stellen Sie sicher, dass Sie Ihre Anwendung mit derselben MPI-Implementierung kompilieren und ausführen.

## Verwendung mit `mpirun` deaktiviertem verwaltetem DNS

Bei Clustern, die mit [SlurmSettings](#) der Einstellung [DisableManagedDns/Dns/](#) und [UseEc2Hostnames](#) auf `true` erstellt wurden, wird der Slurm Knotenname nicht vom [DNS](#) aufgelöst. Slurm kann MPI-Prozesse booten, wenn sie `nodenames` nicht aktiviert sind und wenn der MPI-Job in einem Kontext ausgeführt wird. Slurm Wir empfehlen, die Anweisungen im [MPI-Benutzerhandbuch zur Ausführung von Slurm MPI-Jobs](#) mit zu befolgen. Slurm

## Es wird versucht, einen Cluster zu aktualisieren

**`pcluster update-cluster`** Der Befehl kann nicht lokal ausgeführt werden

Suchen Sie `~/parallelcluster/pcluster-cli.log` in Ihrem lokalen Dateisystem nach Fehlerdetails.

**`clusterStatus`** Das Sehen erfolgt **`UPDATE_FAILED`** mit einem **`pcluster describe-cluster`** Befehl

Wenn das Cluster-Stack-Update rückgängig gemacht wurde, überprüfen Sie die `/var/log/chef-client.log` Datei auf Fehlerdetails.

Prüfen Sie, ob Ihr Problem unter [GitHub Bekannte Probleme unter](#) AWS ParallelCluster on erwähnt wird GitHub.

## Das Cluster-Update hat das Zeitlimit überschritten

Dies könnte ein Problem sein, das damit zusammenhängt, dass es `cf-hup` nicht ausgeführt wird. Wenn der `cf-hup` Dämon durch einen externen Grund beendet wird, wird er nicht automatisch neu gestartet. Wenn er `cf-hup` nicht ausgeführt wird, startet der CloudFormation Stack den Aktualisierungsvorgang während eines Cluster-Updates wie erwartet, aber der Aktualisierungsvorgang ist auf dem Hauptknoten nicht aktiviert, und bei der Stack-Bereitstellung kommt es irgendwann zu einem Timeout. Weitere Informationen finden Sie unter [Fehlerbehebung bei einem Timeout für ein Cluster-Update, wenn es nicht läuft cf-hup](#) So beheben Sie das Problem und beheben es.

## Ich versuche, auf Speicher zuzugreifen

### Verwenden eines externen Amazon FSx for Lustre-Dateisystems

Stellen Sie sicher, dass Datenverkehr zwischen dem Cluster und dem Dateisystem zulässig ist. Das Dateisystem muss einer Sicherheitsgruppe zugeordnet sein, die eingehenden und ausgehenden TCP-Verkehr über die Ports 988, 1021, 1022 und 1023 zulässt. Weitere Informationen zum Einrichten von Sicherheitsgruppen finden Sie unter [FileSystemId](#)

### Verwenden eines externen Amazon Elastic File System-Dateisystems

Stellen Sie sicher, dass Datenverkehr zwischen dem Cluster und dem Dateisystem zulässig ist. Das Dateisystem muss einer Sicherheitsgruppe zugeordnet sein, die eingehenden und ausgehenden TCP-Verkehr über die Ports 988, 1021, 1022 und 1023 zulässt. Weitere Informationen zum Einrichten von Sicherheitsgruppen finden Sie unter [FileSystemId](#)

## Es wird versucht, einen Cluster zu löschen

Der **pcluster delete-cluster** Befehl kann nicht lokal ausgeführt werden

Überprüfen Sie die `~/.parallelcluster/pcluster-cli.log` Datei in Ihrem lokalen Dateisystem.

## Der Cluster-Stack kann nicht gelöscht werden

Wenn der Cluster-Stack nicht gelöscht werden kann, überprüfen Sie die CloudFormation Stack-Ereignismeldung.

Prüfen Sie, ob Ihr Problem [GitHub unter Bekannte Probleme unter](#) AWS ParallelCluster on erwähnt wird GitHub.

## Ich versuche, den AWS ParallelCluster API-Stack zu aktualisieren

Prüfen Sie, ob Ihr Problem [GitHub unter Bekannte Probleme unter](#) AWS ParallelCluster on erwähnt wird GitHub.

# Fehler bei der Initialisierung von Compute-Knoten werden angezeigt

## Ich sehe rein **Node bootstrap errorclustermgtd.log**

Das Problem hängt damit zusammen, dass Rechenknoten nicht booten können. Informationen zum Debuggen eines Problems im geschützten Clustermodus finden Sie unter. [So debuggen Sie den geschützten Modus](#)

## Ich habe On-Demand-Kapazitätsreservierungen (ODCRs) oder zonale Reserved Instances konfiguriert

ODCRs, die Instances mit mehreren Netzwerkschnittstellen wie P4d, P4de und Trainium (Trn) enthalten AWS

Überprüfen Sie in der Cluster-Konfigurationsdatei, ob sich der in einem öffentlichen Subnetz HeadNode befindet und ob sich die Rechenknoten in einem privaten Subnetz befinden.

ODCRs sind gezielte ODCRs

Ich sehe, **Unable to read file '/opt/slurm/etc/pcluster/run\_instances\_overrides.json'**. obwohl ich es bereits eingerichtet habe, indem ich die Anweisungen **/opt/slurm/etc/pcluster/run\_instances\_overrides.json** in befolge [Erstellen Sie Instances mit ODCR Starten Sie Instances](#)

Wenn Sie die AWS ParallelCluster Versionen 3.1.1 bis 3.2.1 mit gezielten ODCRs verwenden und auch die JSON-Datei [run instances override verwenden, ist die JSON-Datei](#) möglicherweise nicht richtig formatiert. Sie könnten einen Fehler wie den folgenden sehen `clustermgtd.log`:

```
Unable to read file '/opt/slurm/etc/pcluster/run_instances_overrides.json'.
Using default: {} in /var/log/parallelcluster/clustermgtd.
```

Stellen Sie sicher, dass das JSON-Dateiformat korrekt ist, indem Sie Folgendes ausführen:

```
$ echo /opt/slurm/etc/pcluster/run_instances_overrides.json | jq
```

Zeigt **Found RunInstances parameters override. anclustermgtd.log**, wenn die Clustererstellung fehlgeschlagen ist oder **slurm\_resume.log** wenn die Ausführung des Jobs fehlgeschlagen ist

Wenn Sie [Run-Instances verwenden, die die JSON-Datei überschreiben](#), überprüfen Sie, ob Sie den Warteschlangennamen und den Namen der Rechenressourcen in der `/opt/slurm/etc/pcluster/run_instances_overrides.json` Datei korrekt angegeben haben.

Ich sehe **An error occurred (InsufficientInstanceCapacity)slurm\_resume.log**, wenn ich einen Job nicht ausführen kann oder **clustermgtd.log** wenn ich keinen Cluster erstellen kann

Verwenden von PG-ODCR (Placement Group ODCR)

Wenn Sie ein ODCR mit einer zugehörigen Platzierungsgruppe erstellen, muss derselbe Platzierungsgruppenname in der Konfigurationsdatei verwendet werden. Geben Sie den [Namen der entsprechenden Platzierungsgruppe](#) in der Cluster-Konfiguration ein.

Verwendung zonaler Reserved Instances

Wenn Sie zonale Reserved Instances mit `PlacementGroup/Enabled` bis `true` in der Cluster-Konfiguration verwenden, wird möglicherweise ein Fehler wie der folgende angezeigt:

```
We currently do not have sufficient trn1.32xlarge capacity in the Availability Zone
you requested (us-east-1d). Our system will be working on provisioning additional
capacity.
You can currently get trn1.32xlarge capacity by not specifying an Availability Zone in
your request or choosing us-east-1a, us-east-1b, us-east-1c, us-east-1e, us-east-1f.
```

Dieser Fehler tritt möglicherweise auf, weil sich die zonalen Reserved Instances nicht in derselben UC (oder Spine) befinden, was bei der Verwendung von Platzierungsgruppen zu Fehlern bei unzureichender Kapazität (ICEs) führen kann. Sie können diesen Fall überprüfen, indem Sie die `PlacementGroup` Gruppeneinstellung in der Clusterkonfiguration deaktivieren, um festzustellen, ob der Cluster die Instances zuweisen kann.

Ich sehe **An error occurred (VcpuLimitExceeded)slurm\_resume.log**, wenn ich einen Job nicht ausführen kann, oder wenn ich keinen Cluster erstellen kann **clustermgtd.log**

Überprüfen Sie die vCPU-Limits in Ihrem Konto für den spezifischen EC2-Instance-Typ, den Sie verwenden. Wenn Sie keine oder weniger vCPUs sehen, als Sie anfordern, fordern Sie eine Erhöhung Ihrer Limits an. Informationen darüber, wie Sie aktuelle Limits anzeigen und neue Limits anfordern können, finden Sie unter [Amazon EC2-Servicekontingente](#) im Amazon EC2 EC2-Benutzerhandbuch für Linux-Instances.

Ich sehe **An error occurred (InsufficientInstanceCapacity)slurm\_resume.log**, wenn ich einen Job nicht ausführen kann, oder wenn ich keinen Cluster erstellen kann **clustermgtd.log**

Sie haben ein Problem mit unzureichender Kapazität. Folgen Sie <https://aws.amazon.com/premiumsupport/knowledge-center/ec2-insufficient-capacity-errors/>, um das Problem zu beheben.

Ich sehe, dass sich die Knoten im **DOWN** Zustand von befinden **Reason (Code:InsufficientInstanceCapacity)...**

Sie haben ein Problem mit unzureichender Kapazität. Folgen Sie <https://aws.amazon.com/premiumsupport/knowledge-center/ec2-insufficient-capacity-errors/>, um das Problem zu beheben. Weitere Informationen AWS ParallelCluster zum schnellen Failover-Modus für unzureichende Kapazität finden Sie unter. [Slurmschneller Cluster-Failover mit unzureichender Kapazität](#)

Seht rein **cannot change locale (en\_US.utf-8) because it has an invalid nameslurm\_resume.log**

Dies kann passieren, wenn der yum Installationsvorgang nicht erfolgreich war und die Gebietsschemaeinstellungen inkonsistent geblieben sind. Dies kann beispielsweise der Fall sein, wenn ein Benutzer den Installationsvorgang beendet.

Gehen Sie wie folgt vor, um die Ursache zu überprüfen:

- Führen Sie `su - pcluster-admin`.

Die Shell zeigt einen Fehler an, `cannot change locale...no such file or directory` z. B.

- Führen Sie `localedef --list`.

Gibt eine leere Liste zurück oder enthält nicht das Standardgebietsschema.

- Überprüfen Sie den letzten yum Befehl mit `yum history und yum history info #ID`. Hat die letzte ID Return-Code: Success?

Wenn die letzte ID nicht vorhanden ist Return-Code: Success, wurden die Skripts nach der Installation möglicherweise nicht erfolgreich ausgeführt.

Um das Problem zu beheben, versuchen Sie, das Gebietsschema mit neu zu erstellen. `yum reinstall glibc-all-langpacks` Nach der Neuerstellung `su - pcluster-admin` wird kein Fehler oder keine Warnung angezeigt, wenn das Problem behoben ist.

## Keines der vorherigen Szenarien trifft auf meine Situation zu

Informationen zur Behebung von Problemen bei der Initialisierung von Compute-Knoten finden Sie unter [Behebung von Problemen bei der Knoteninitialisierung](#).

Prüfen Sie, ob Ihr Szenario unter [GitHub Bekannte Probleme](#) unter AWS ParallelCluster on GitHub behandelt wird.

Zusätzliche Unterstützung finden Sie unter [Zusätzliche Unterstützung](#).

## Anleitung ster-ster-ster-ster-ster-ster-ster-ster-

Cluster-Integritätsmetriken werden ab AWS ParallelCluster Version 3.6.0 zum AWS ParallelCluster CloudWatch Amazon-Dashboard hinzugefügt. In den folgenden Abschnitten finden Sie mehr über die Integrität, die Sie zur Fehlerbehebung und Fehlerbehebung.

Themen

- [Diagramm mit Fehlern bei der Instanzbereitstellung anzeigen](#)
- [Das Diagramm „Ungesunde Instance-Fehler“ anzeigen](#)



- [Das Diagramm zur Leerlaufzeit der Rechenflotte anzeigen](#)

## Diagramm mit Fehlern bei der Instanzbereitstellung anzeigen

Wenn Sie in der `Instance Provisioning Errors` Grafik einen Wert ungleich Null sehen, bedeutet dies, dass die EC2-Instanz zur Unterstützung von Slurm-Knoten auf der `CreateFleet` OR-API nicht gestartet werden konnte. `RunInstance`

### Sehen **IAMPolicyErrors**

- Was ist passiert?

Eine Reihe von Instanzen konnte nicht gestartet werden, was auf unzureichende Berechtigungen mit Fehlercode zurückzuführen ist `UnauthorizedOperation`.

- Wie löst man das?

Wenn Sie ein benutzerdefiniertes [InstanceRole](#) oder konfiguriert haben [InstanceProfile](#), überprüfen Sie Ihre IAM-Richtlinien und stellen Sie sicher, dass Sie die richtigen Anmeldeinformationen verwenden.

Überprüfen Sie die `clustermgtd` Datei auf Fehlerdetails für statische Knoten. Überprüfen Sie die `slurm_resume.log` Datei auf Fehlerdetails für dynamische Knoten. Verwenden Sie die Details, um mehr über die fehlenden Berechtigungen zu erfahren, die hinzugefügt werden müssen.

### Sehen **VcpuLimitErrors**

- Was ist passiert?

AWS ParallelClusterInstances konnten nicht gestartet werden, da das vCPU-Limit auf Ihrer AWS-Konto für einen bestimmten EC2-Instance-Typ, den Sie für Cluster-Rechenknoten konfiguriert haben, erreicht wurde.

- Wie löst man das?

Suchen Sie in der `clustermgtd` Datei für statische Knoten nach dem `VcpuLimitExceeded` Fehler und suchen Sie in der `slurm_resume.log` Datei nach dynamischen Knoten, um weitere Informationen zu erhalten. Um dieses Problem zu beheben, können Sie vCPU. Weitere Informationen zum Anzeigen von aktuellen Limits und zum Anstellen einer Verbindung mit Linux-

Instances finden Sie unter [Amazon EC2-Servizen](#) im Amazon EC2 EC2-Benutzerhandbuch für Linux-Instances.

## Sehen **VolumeLimitErrors**

- Was ist passiert?

Sie haben Ihr Amazon EBS-Volumenlimit für Ihr AWS-Konto Gerät erreicht und AWS ParallelCluster können keine Instances mit dem Fehlercode `InsufficientVolumeCapacity` oder `VolumeLimitExceeded` starten.

- Wie löst man das?

Überprüfen Sie die `clustermgtd` Datei auf statische Knoten und überprüfen Sie die `slurm_resume.log` Datei auf dynamische Knoten, um weitere Details zur Volumenbegrenzung zu erhalten. Um dieses Problem zu lösen, können Sie ein anderes verwendenAWS-Region, vorhandene Volumes bereinigen oder sich an das AWS Support Center wenden, um eine Anfrage zur Erhöhung Ihres Amazon EBS-Volumenlimits einzureichen.

## Sehen **InsufficientCapacityErrors**

- Was ist passiert?

AWS ParallelClusterverfügt nicht über genügend Kapazität, um EC2-Instances auf Back-Knoten zu starten.

- Wie löst man das?

Überprüfen Sie die `clustermgtd` Datei auf statische Knoten und überprüfen Sie die `slurm_resume.log` Datei auf dynamische Knoten, um unzureichende Kapazitätsfehlerdetails zu erhalten. Folgen Sie den Anweisungen [unter `https://aws.amazon.com/premiumsupport/knowledge-center/ec2-insufficient-capacity-errors/`](https://aws.amazon.com/premiumsupport/knowledge-center/ec2-insufficient-capacity-errors/), um das Problem zu beheben.

## **OtherInstanceLaunchFailures**

- Was ist passiert?

Die EC2-Instance zur Sicherung von Rechenknoten konnte nicht mit der `CreateFleet` oder `RunInstance` API gestartet werden.

- Wie löst man das?

Überprüfen Sie die `clustermgtd` Datei auf statische Knoten und überprüfen Sie die `slurm_resume.log` Datei auf dynamische Knoten, um Fehlerdetails zu erhalten.

## Das Diagramm „Ungesunde Instance-Fehler“ anzeigen

- Was ist passiert?

Eine Reihe von Recheninstanzen wurden gestartet, aber später als fehlerhaft beendet.

- Wie löst man das?

Weitere Informationen zur Fehlerbehebung bei [Behebung unerwarteter Knotenersetzungen und -beendigungen](#)

## Sehen **InstanceBootstrapTimeoutError**

- Was ist passiert?

Eine Instanz kann dem Cluster nicht innerhalb der `resume_timeout` (für dynamische Knoten) oder `node_replacement_timeout` (für statische Knoten) beitreten. Dies kann vorkommen, wenn das Netzwerk nicht richtig für die Rechenknoten konfiguriert ist, oder es kann vorkommen, dass die Ausführung benutzerdefinierter Skripts, die auf dem Rechenknoten ausgeführt werden, zu lange dauert.

- Wie löst man das?

Suchen Sie bei dynamischen Knoten im `clustermgtd` log (`/var/log/parallelcluster/clustermgtd`) nach der IP-Adresse des Rechenknotens und nach Fehlern wie den folgenden:

```
Node bootstrap error: Resume timeout expires for node
```

Suchen Sie bei statischen Knoten im `clustermgtd` log (`/var/log/parallelcluster/clustermgtd`) nach der IP-Adresse des Rechenknotens und nach Fehlern wie den folgenden:

```
Node bootstrap error: Replacement timeout expires for node ... in replacement.
```

Weitere Informationen finden Sie in der `/var/log/cloud-init-output.log` Datei auf Fehler. Sie können problematische IP-Adressen von Rechenknoten aus den `clustermgtd` und `slurm_resume` Protokolldateien abrufen.

## Sehen **EC2HealthCheckErrors**

- Was ist passiert?

Eine Instance hat eine EC2-Integritätsprüfung nicht bestanden.

- Wie löst man das?

Weitere Informationen [zur Fehlerbehebung](#).

## Sehen **ScheduledEventHealthCheckErrors**

- Was ist passiert?

Eine Instance hat eine Zustandsprüfung für geplante EC2-Ereignisse nicht bestanden, und sie ist fehlerhaft.

- Wie löst man das?

Informationen zur Behebung dieses Problems finden Sie unter [Geplante Ereignisse für Ihre Instances](#).

## Sehen **NoCorrespondingInstanceErrors**

- Was ist passiert?

AWS ParallelCluster kann keine Instanzen finden, die Knoten unterstützen. Die Knoten haben sich wahrscheinlich während Bootstrap-Operationen selbst beendet. [SlurmQueues/CustomActions/OnNodeStart|OnNodeConfigured](#) Skript- oder Netzwerkfehler erzeugen können `NoCorrespondingInstanceErrors`.

- Wie löst man das?

Weitere Informationen finden Sie unter `/var/log/cloud-init-output.log` für den Rechenknoten.

## Das Diagramm zur Leerlaufzeit der Rechenflotte anzeigen

Es wird ein Wert **MaxDynamicNodeIdleTime** angezeigt, der deutlich länger als der Scaledown-Schwellenwert für die Leerlaufzeit ist

- Was ist passiert?

Ihre Instance wird nicht ordnungsgemäß beendet. `MaxDynamicNodeIdleTime` zeigt die maximale Zeit in Sekunden an, während der ein dynamischer Knoten, der von einer EC2-Instance unterstützt wird, inaktiv ist. Der Schwellenwert für die Idle Time Scaledown wird vom [ScaledownIdleTime](#) Cluster-Konfigurationsparameter abgeleitet. Wenn ein Rechenknoten länger als Sekunden im Leerlauf war, wird der Knoten heruntergefahren und Slurm die Backing-Instance AWS ParallelCluster beendet. In diesem Fall verhindert etwas die Beendigung der Instanz.

- Wie löst man das?

Weitere Informationen zu diesem Problem finden Sie [Ersetzen, Beenden oder Herunterfahren problematischer Instanzen und Knoten](#) unter [Behebung von Skalierungsproblemen](#).

## Behebung von Problemen bei der Clusterbereitstellung

Wenn Ihr Cluster nicht erstellt werden kann und die Stack-Erstellung rückgängig gemacht wird, können Sie die Protokolldateien durchsuchen, um das Problem zu diagnostizieren. Die Fehlermeldung sieht wahrscheinlich wie die folgende Ausgabe aus:

```
$ pcluster create-cluster --cluster-name mycluster --region eu-west-1 \
--cluster-configuration cluster-config.yaml
{
 "cluster": {
 "clusterName": "mycluster",
 "cloudformationStackStatus": "CREATE_IN_PROGRESS",
 "cloudformationStackArn": "arn:aws:cloudformation:eu-west-1:xxx:stack/
mycluster/1bf6e7c0-0f01-11ec-a3b9-024fcc6f3387",
 "region": "eu-west-1",
 "version": "3.7.0",
 "clusterStatus": "CREATE_IN_PROGRESS"
 }
}

$ pcluster describe-cluster --cluster-name mycluster --region eu-west-1
```

```
{
 "creationTime": "2021-09-06T11:03:47.696Z",
 ...
 "cloudFormationStackStatus": "ROLLBACK_IN_PROGRESS",
 "clusterName": "mycluster",
 "computeFleetStatus": "UNKNOWN",
 "cloudFormationStackArn": "arn:aws:cloudformation:eu-west-1:xxx:stack/
mycluster/1bf6e7c0-0f01-11ec-a3b9-024fcc6f3387",
 "lastUpdatedTime": "2021-09-06T11:03:47.696Z",
 "region": "eu-west-1",
 "clusterStatus": "CREATE_FAILED"
}
```

## Themen

- [AWS CloudFormationEreignisse anzeigen auf CREATE\\_FAILED](#)
- [Verwenden Sie die CLI, um Protokollstreams anzuzeigen](#)
- [Erstellen Sie den ausgefallenen Cluster erneut mit rollback-on-failure](#)

## AWS CloudFormationEreignisse anzeigen auf **CREATE\_FAILED**

Sie können die Konsole oder die AWS ParallelCluster CLI verwenden, um CloudFormation Ereignisse bei CREATE\_FAILED Fehlern anzuzeigen und so die Ursache zu finden.

### Themen

- [Ereignisse in der CloudFormation Konsole anzeigen](#)
- [Verwenden Sie die CLI, um CloudFormation Ereignisse anzuzeigen und zu filtern](#)  
[CREATE\\_FAILED](#)

## Ereignisse in der CloudFormation Konsole anzeigen

Weitere Informationen zur Ursache des "CREATE\_FAILED" Status finden Sie in der CloudFormation Konsole.

CloudFormation Fehlermeldungen von der Konsole aus anzeigen.

1. Melden Sie sich bei der an AWS Management Console und navigieren Sie zu <https://console.aws.amazon.com/cloudformation>.
2. Wählen Sie den Stack mit dem Namen *cluster\_name* aus.

3. Wählen Sie die Registerkarte Ereignisse.
4. Überprüfen Sie den Status der Ressource, die nicht erstellt werden konnte, indem Sie die Liste der Ressourcenereignisse nach der logischen ID durchsuchen. Wenn eine Unteraufgabe nicht erstellt werden konnte, gehen Sie rückwärts vor, um das fehlgeschlagene Ressourcenereignis zu finden.
5. Wenn Sie beispielsweise die folgende Statusmeldung sehen, müssen Sie Instance-Typen verwenden, die Ihr aktuelles vCPU-Limit nicht überschreiten, oder mehr vCPU-Kapazität anfordern.

```
2022-02-04 16:09:44 UTC-0800 HeadNode CREATE_FAILED You have requested more vCPU
capacity than your current vCPU limit of 0 allows
 for the instance bucket that the specified instance type belongs to. Please
visit http://aws.amazon.com/contact-us/ec2-request to request an adjustment to
this limit.
 (Service: AmazonEC2; Status Code: 400; Error Code: VcpuLimitExceeded; Request
ID: a9876543-b321-c765-d432-dcba98766789; Proxy: null).
```

## Verwenden Sie die CLI, um CloudFormation Ereignisse anzuzeigen und zu filtern **CREATE\_FAILED**

Um das Problem bei der Clustererstellung zu diagnostizieren, können Sie den [pcluster get-cluster-stack-events](#) Befehl verwenden, indem Sie nach dem CREATE\_FAILED Status filtern. Weitere Informationen finden Sie im [AWS Command Line Interface Benutzerhandbuch](#) unter [Filtern der AWS CLI Ausgabe](#).

```
$ pcluster get-cluster-stack-events --cluster-name mycluster --region eu-west-1 \
 --query 'events[?resourceStatus==`CREATE_FAILED`]'
[
 {
 "eventId": "3ccdedd0-0f03-11ec-8c06-02c352fe2ef9",
 "physicalResourceId": "arn:aws:cloudformation:eu-west-1:xxx:stack/
mycluster/1bf6e7c0-0f02-11ec-a3b9-024fcc6f3387",
 "resourceStatus": "CREATE_FAILED",
 "resourceStatusReason": "The following resource(s) failed to create: [HeadNode].",
 "stackId": "arn:aws:cloudformation:eu-west-1:xxx:stack/
mycluster/1bf6e7c0-0f02-11ec-a3b9-024fcc6f3387",
 "stackName": "mycluster",
 "logicalResourceId": "mycluster",
```

```

 "resourceType": "AWS::CloudFormation::Stack",
 "timestamp": "2021-09-06T11:11:51.780Z"
 },
 {
 "eventId": "HeadNode-CREATE_FAILED-2021-09-06T11:11:50.127Z",
 "physicalResourceId": "i-04e91cc1f4ea796fe",
 "resourceStatus": "CREATE_FAILED",
 "resourceStatusReason": "Received FAILURE signal with UniqueId
i-04e91cc1f4ea796fe",
 "resourceProperties": "{\"LaunchTemplate\":{\"Version\":\"1\"},\"LaunchTemplateId
\": \"lt-057d2b1e687f05a62\"}}",
 "stackId": "arn:aws:cloudformation:eu-west-1:xxx:stack/
mycluster/1bf6e7c0-0f02-11ec-a3b9-024fcc6f3387",
 "stackName": "mycluster",
 "logicalResourceId": "HeadNode",
 "resourceType": "AWS::EC2::Instance",
 "timestamp": "2021-09-06T11:11:50.127Z"
 }
]

```

Im vorherigen Beispiel lag der Fehler an der Einrichtung des Hauptknotens.

## Verwenden Sie die CLI, um Protokollstreams anzuzeigen

Um diese Art von Problem zu debuggen, können Sie die vom Hauptknoten aus verfügbaren Log-Streams auflisten, [pcluster list-cluster-log-streams](#) indem Sie nach dem Inhalt der Log-Streams filtern `node-type` und diese anschließend analysieren.

```

$ pcluster list-cluster-log-streams --cluster-name mycluster --region eu-west-1 \
--filters 'Name=node-type,Values=HeadNode'
{
 "logStreams": [
 {
 "logStreamArn": "arn:aws:logs:eu-west-1:xxx:log-group:/aws/parallelcluster/
mycluster-202109061103:log-stream:ip-10-0-0-13.i-04e91cc1f4ea796fe.cfn-init",
 "logStreamName": "ip-10-0-0-13.i-04e91cc1f4ea796fe.cfn-init",
 ...
 },
 {
 "logStreamArn": "arn:aws:logs:eu-west-1:xxx:log-group:/aws/parallelcluster/
mycluster-202109061103:log-stream:ip-10-0-0-13.i-04e91cc1f4ea796fe.chef-client",
 "logStreamName": "ip-10-0-0-13.i-04e91cc1f4ea796fe.chef-client",

```



```

 ...
 },
 {
 "logStreamArn": "arn:aws:logs:eu-west-1:xxx:log-group:/aws/parallelcluster/
mycluster-202109061103:log-stream:ip-10-0-0-13.i-04e91cc1f4ea796fe.cloud-init",
 "logStreamName": "ip-10-0-0-13.i-04e91cc1f4ea796fe.cloud-init",
 ...
 },
 ...
]
}

```

Die beiden wichtigsten Protokollströme, die Sie verwenden können, um Initialisierungsfehler zu finden, sind die folgenden:

- `cfn-init` ist das Protokoll für das `cfn-init` Skript. Überprüfen Sie zuerst diesen Log-Stream. Sie werden den Command `chef failed` Fehler wahrscheinlich in diesem Protokoll sehen. In den Zeilen unmittelbar vor dieser Zeile finden Sie weitere Einzelheiten zu der Fehlermeldung. Weitere Informationen finden Sie unter [cfn-init](#).
- `cloud-init` ist [das Protokoll für Cloud-Init](#). Wenn Sie nichts darin sehen, versuchen Sie als `cfn-init` Nächstes, in diesem Protokoll nachzuschauen.

Sie können den Inhalt des Protokollstreams abrufen, indem Sie die folgende Option verwenden [pcluster get-cluster-log-events](#) (beachten Sie die `--limit 5` Option, die Anzahl der abgerufenen Ereignisse zu begrenzen):

```

$ pcluster get-cluster-log-events --cluster-name mycluster \
 --region eu-west-1 --log-stream-name ip-10-0-0-13.i-04e91cc1f4ea796fe.cfn-init \
 --limit 5
{
 "nextToken": "f/36370880979637159565202782352491087067973952362220945409/s",
 "prevToken": "b/36370880752972385367337528725601470541902663176996585497/s",
 "events": [
 {
 "message": "2021-09-06 11:11:39,049 [ERROR] Unhandled exception during build:
Command runpostinstall failed",
 "timestamp": "2021-09-06T11:11:39.049Z"
 },
 {
 "message": "Traceback (most recent call last):\n File \"/opt/aws/bin/
cfn-init\", line 176, in <module>\n worklog.build(metadata, configSets)\n

```

```

File \"/usr/lib/python3.7/site-packages/cfnbootstrap/construction.py\", line
135, in build\n Contractor(metadata).build(configSets, self)\n File \"/
usr/lib/python3.7/site-packages/cfnbootstrap/construction.py\", line 561, in
build\n self.run_config(config, worklog)\n File \"/usr/lib/python3.7/
site-packages/cfnbootstrap/construction.py\", line 573, in run_config\n
CloudFormationCarpenter(config, self._auth_config).build(worklog)\n File \"/usr/
lib/python3.7/site-packages/cfnbootstrap/construction.py\", line 273, in build\n
self._config.commands)\n File \"/usr/lib/python3.7/site-packages/cfnbootstrap/
command_tool.py\", line 127, in apply\n raise ToolError(u\"Command %s failed\" %
name)\",
 \"timestamp\": \"2021-09-06T11:11:39.049Z\"
 },
 {
 \"message\": \"cfnbootstrap.construction_errors.ToolError: Command runpostinstall
failed\",
 \"timestamp\": \"2021-09-06T11:11:39.049Z\"
 },
 {
 \"message\": \"2021-09-06 11:11:49,212 [DEBUG] CloudFormation client initialized
with endpoint https://cloudformation.eu-west-1.amazonaws.com\",
 \"timestamp\": \"2021-09-06T11:11:49.212Z\"
 },
 {
 \"message\": \"2021-09-06 11:11:49,213 [DEBUG] Signaling resource HeadNode in stack
mycluster with unique ID i-04e91cc1f4ea796fe and status FAILURE\",
 \"timestamp\": \"2021-09-06T11:11:49.213Z\"
 }
]
}

```

Im vorherigen Beispiel wurde der Fehler durch einen `runpostinstall` Fehler verursacht. Er steht also in engem Zusammenhang mit dem Inhalt des benutzerdefinierten Bootstrap-Skripts, das im `OnNodeConfigured` Konfigurationsparameter verwendet wurde. [CustomActions](#)

## Erstellen Sie den ausgefallenen Cluster erneut mit **rollback-on-failure**

AWS ParallelCluster erstellt CloudWatch Cluster-Protokollstreams in Protokollgruppen. Sie können diese Protokolle in der CloudWatch Konsole „Benutzerdefinierte Dashboards“ oder „Protokollgruppen“ anzeigen. Weitere Informationen finden Sie unter [Integration mit Amazon CloudWatch Logs](#) und [CloudWatchAmazon-Dashboard](#). Wenn keine Protokollstreams verfügbar sind, kann der Fehler durch das [CustomActions](#) benutzerdefinierte Bootstrap-Skript oder ein AMI-Problem verursacht werden.

Um das Erstellungsproblem in diesem Fall zu diagnostizieren, erstellen Sie den Cluster erneut, einschließlich des `--rollback-on-failure` Parameters `pcluster create-cluster`, der auf `false` gesetzt ist. Verwenden Sie dann SSH, um den Cluster anzuzeigen, wie im Folgenden gezeigt:

```
$ pcluster create-cluster --cluster-name mycluster --region eu-west-1 \
 --cluster-configuration cluster-config.yaml --rollback-on-failure false
{
 "cluster": {
 "clusterName": "mycluster",
 "cloudformationStackStatus": "CREATE_IN_PROGRESS",
 "cloudformationStackArn": "arn:aws:cloudformation:eu-west-1:xxx:stack/
mycluster/1bf6e7c0-0f01-11ec-a3b9-024fcc6f3387",
 "region": "eu-west-1",
 "version": "3.7.0",
 "clusterStatus": "CREATE_IN_PROGRESS"
 }
}
$ pcluster ssh --cluster-name mycluster
```

Nachdem Sie beim Hauptknoten angemeldet sind, sollten Sie drei primäre Protokolldateien finden, anhand derer Sie den Fehler finden können.

- `/var/log/cfn-init.log` ist das Protokoll für das `cfn-init` Skript. Überprüfen Sie zuerst dieses Protokoll. Es ist wahrscheinlich, dass Ihnen ein Fehler wie `Command chef failed` in diesem Protokoll angezeigt wird. Sehen Sie sich die Zeilen unmittelbar vor dieser Zeile an, um genauere Informationen zu der Fehlermeldung zu erhalten. Weitere Informationen finden Sie unter [cfn-init](#).
- `/var/log/cloud-init.log` ist das Protokoll für `Cloud-Init`. Wenn Sie nichts darin sehen, versuchen Sie als `cfn-init.log` Nächstes, in diesem Protokoll nachzuschauen.
- `/var/log/cloud-init-output.log` ist die Ausgabe von Befehlen, die von `cloud-init` ausgeführt wurden. Dies beinhaltet die Ausgabe von `cfn-init`. In den meisten Fällen müssen Sie sich dieses Protokoll nicht ansehen, um diese Art von Problem zu beheben.

## Behebung von Skalierungsproblemen

Dieser Abschnitt ist relevant für Cluster, die mit AWS ParallelCluster Version 3.0.0 und höher mit dem Slurm-Job-Scheduler installiert wurden. Weitere Informationen zur Konfiguration mehrerer Warteschlangen finden Sie unter [Konfiguration mehrerer Warteschlangen](#)

Wenn bei einem Ihrer laufenden Cluster Probleme auftreten, versetzen Sie den Cluster in einen STOPPED Zustand, indem Sie den folgenden Befehl ausführen, bevor Sie mit der Problembehandlung beginnen. Dadurch werden unerwartete Kosten vermieden.

```
$ pcluster update-compute-fleet --cluster-name mycluster \
 --status STOP_REQUESTED
```

Sie können die auf den Clusterknoten verfügbaren Protokolldatenströme auflisten, indem Sie den [pcluster list-cluster-log-streams](#) Befehl verwenden und mit einem private-dns-name der ausfallenden Knoten oder dem Hauptknoten filtern:

```
$ pcluster list-cluster-log-streams --cluster-name mycluster --region eu-west-1 \
 --filters 'Name=private-dns-name,Values=ip-10-0-0-101'
```

Anschließend können Sie den Inhalt des Log-Streams abrufen, um ihn zu analysieren, indem Sie den [pcluster get-cluster-log-events](#) Befehl verwenden und die --log-stream-name entsprechenden Logs an eines der im folgenden Abschnitt genannten Schlüsselprotokolle übergeben:

```
$ pcluster get-cluster-log-events --cluster-name mycluster \
 --region eu-west-1 --log-stream-name ip-10-0-0-13.i-04e91cc1f4ea796fe.cfn-init
```

AWS ParallelCluster erstellt CloudWatch Cluster-Protokollstreams in Protokollgruppen. Sie können diese Protokolle in der CloudWatch Konsole „Benutzerdefinierte Dashboards“ oder „Protokollgruppen“ anzeigen. Weitere Informationen finden Sie unter [Integration mit Amazon CloudWatch Logs](#) und [CloudWatchAmazon-Dashboard](#).

## Themen

- [Wichtige Protokolle für das Debuggen](#)
- [Es InsufficientInstanceCapacity wird ein Fehler angezeigt, slurm\\_resume.log wenn ich einen Job nicht ausführen kann oder clustermgtd.log wenn ich keinen Cluster erstellen kann](#)
- [Behebung von Problemen bei der Knoteninitialisierung](#)
- [Behebung unerwarteter Knotenersetzungen und -beendigungen](#)
- [Ersetzen, Beenden oder Herunterfahren problematischer Instanzen und Knoten](#)
- [Status der Warteschlange \(Partition\) Inactive](#)
- [Behebung anderer bekannter Knoten- und Jobprobleme](#)

## Wichtige Protokolle für das Debuggen

Die folgende Tabelle bietet einen Überblick über die Schlüsselprotokolle für den Hauptknoten:

- `/var/log/cfn-init.log`- Dies ist das AWS CloudFormation Init-Protokoll. Es enthält alle Befehle, die bei der Einrichtung einer Instanz ausgeführt wurden. Verwenden Sie es, um Initialisierungsprobleme zu beheben.
- `/var/log/chef-client.log`- Dies ist das Chef-Client-Protokoll. Es enthält alle Befehle, die über Chef/Cinc ausgeführt wurden. Verwenden Sie es, um Initialisierungsprobleme zu beheben.
- `/var/log/parallelcluster/slurm_resume.log`- Das ist ein ResumeProgram Protokoll. Es startet Instanzen für dynamische Knoten. Verwenden Sie es, um Probleme beim Start dynamischer Knoten zu beheben.
- `/var/log/parallelcluster/slurm_suspend.log`- Das ist das SuspendProgram Protokoll. Es wird aufgerufen, wenn Instanzen für dynamische Knoten beendet werden. Verwenden Sie es, um Probleme mit der Kündigung dynamischer Knoten zu beheben. Wenn Sie dieses Protokoll überprüfen, sollten Sie auch das `clustermgtd` Protokoll überprüfen.
- `/var/log/parallelcluster/clustermgtd`- Das ist das `clustermgtd` Protokoll. Es läuft als zentraler Daemon, der die meisten Clusteroperationen verwaltet. Verwenden Sie ihn, um Probleme beim Starten, Beenden oder Clusterbetrieb zu beheben.
- `/var/log/slurmctld.log`- Dies ist das Protokoll des Slurm-Control-Daemons. AWS ParallelCluster trifft keine Skalierungsentscheidungen. Vielmehr versucht es nur, Ressourcen bereitzustellen, um die Slurm-Anforderungen zu erfüllen. Es ist nützlich bei Problemen mit der Skalierung und Zuweisung, bei Problemen mit dem Job und bei Problemen mit dem Terminplaner.
- `/var/log/parallelcluster/compute_console_output`- In diesem Protokoll wird die Konsolenausgabe einer Stichprobe von statischen Rechenknoten aufgezeichnet, die unerwartet beendet wurden. Verwenden Sie dieses Protokoll, wenn statische Rechenknoten beendet werden und die Rechenknotenprotokolle in CloudWatch nicht verfügbar sind. Der `compute_console_output` log Inhalt, den Sie erhalten, ist derselbe, wenn Sie die EC2-Konsole verwenden oder AWS CLI die Ausgabe der Instance-Konsole abrufen.

Dies sind die wichtigsten Protokolle für die Rechenknoten:

- `/var/log/cloud-init-output.log`- Dies ist das [Cloud-Init-Protokoll](#). Es enthält alle Befehle, die bei der Einrichtung einer Instanz ausgeführt wurden. Verwenden Sie es, um Initialisierungsprobleme zu beheben.

- `/var/log/parallelcluster/computemgtd`- Das ist das `computemgtd` Protokoll. Es läuft auf jedem Rechenknoten und überwacht den Knoten für den seltenen Fall, dass der `clustermgtd` Daemon auf dem Hauptknoten offline ist. Verwenden Sie ihn, um unerwartete Terminierungsprobleme zu beheben.
- `/var/log/slurmd.log`- Dies ist das Slurm-Compute-Daemon-Protokoll. Verwenden Sie es, um Probleme mit der Initialisierung und Rechenfehlern zu beheben.

Es **InsufficientInstanceCapacity** wird ein Fehler angezeigt, **slurm\_resume.log** wenn ich einen Job nicht ausführen kann oder **clustermgtd.log** wenn ich keinen Cluster erstellen kann

Wenn der Cluster einen Slurm Scheduler verwendet, liegt ein Problem mit unzureichender Kapazität vor. Wenn bei einer Anfrage zum Starten einer Instanz nicht genügend Instances verfügbar sind, wird ein `InsufficientInstanceCapacity` Fehler zurückgegeben.

Bei der statischen Instance-Kapazität finden Sie den Fehler im `clustermgtd` Protokoll unter `/var/log/parallelcluster/clustermgtd`.

Für die dynamische Instanzkapazität finden Sie den Fehler im `ResumeProgram` Protokoll unter `/var/log/parallelcluster/slurm_resume.log`.

Die Meldung sieht dem folgenden Beispiel ähnlich:

```
An error occurred (InsufficientInstanceCapacity) when calling the RunInstances/
CreateFleet operation...
```

Je nach Anwendungsfall sollten Sie eine der folgenden Methoden in Betracht ziehen, um diese Art von Fehlermeldungen zu vermeiden:

- Deaktivieren Sie die Platzierungsgruppe, falls sie aktiviert ist. Weitere Informationen finden Sie unter [Probleme beim Platzieren von Gruppen und beim Starten von Instances](#).
- Reservieren Sie Kapazität für die Instances und starten Sie sie mit ODCR (On-Demand-Kapazitätsreservierungen). Weitere Informationen finden Sie unter [Erstellen Sie Instances mit ODCR Starten Sie Instances](#).
- Konfigurieren Sie mehrere Rechenressourcen mit unterschiedlichen Instanztypen. Wenn für Ihren Workload kein bestimmter Instance-Typ erforderlich ist, können Sie ein schnelles Failover mit

unzureichender Kapazität mit mehreren Rechenressourcen nutzen. Weitere Informationen finden Sie unter [Slurmschneller Cluster-Failover mit unzureichender Kapazität](#).

- Konfigurieren Sie mehrere Instanztypen in derselben Rechenressource und nutzen Sie die Zuweisung mehrerer Instanztypen. Weitere Informationen zur Konfiguration mehrerer Instanzen finden Sie unter [Zuweisung mehrerer Instanztypen mit Slurm](#) und [Scheduling/SlurmQueues/ComputeResources/Instances](#).
- Verschieben Sie die Warteschlange in eine andere Availability Zone, indem Sie die Subnetz-ID in der Cluster-Konfiguration ändern [Scheduling/SlurmQueues/Networking/SubnetIds](#).
- Wenn Ihre Arbeitslast nicht eng miteinander verknüpft ist, verteilen Sie die Warteschlange auf verschiedene Availability Zones. Weitere Informationen zur Konfiguration mehrerer Subnetze finden Sie unter [Scheduling/SlurmQueuesNetworking/SubnetIds](#).

## Behebung von Problemen bei der Knoteninitialisierung

In diesem Abschnitt wird beschrieben, wie Sie Probleme mit der Knoteninitialisierung beheben können. Dazu gehören Probleme, bei denen der Knoten nicht gestartet, eingeschaltet oder einem Cluster nicht beitreten kann.

### Themen

- [Hauptknoten](#)
- [Datenverarbeitungsknoten](#)

### Hauptknoten

Anwendbare Protokolle:

- `/var/log/cfn-init.log`
- `/var/log/chef-client.log`
- `/var/log/parallelcluster/clustermgtd`
- `/var/log/parallelcluster/slurm_resume.log`
- `/var/log/slurmctld.log`

Überprüfen Sie die `/var/log/cfn-init.log` `/var/log/chef-client.log` AND-Protokolle oder die entsprechenden Protokolldatenströme. Diese Protokolle enthalten alle Aktionen, die bei

der Einrichtung des Hauptknotens ausgeführt wurden. Bei den meisten Fehlern, die während der Installation auftreten, sollten sich Fehlermeldungen im `/var/log/chef-client.log` Protokoll befinden. Wenn in der Konfiguration des Clusters `OnNodeStart` oder `OnNodeConfigured` - Skripts angegeben sind, überprüfen Sie anhand der Protokollmeldungen, ob das Skript erfolgreich ausgeführt wird.

Wenn ein Cluster erstellt wird, muss der Hauptknoten warten, bis die Rechenknoten dem Cluster beitreten, bevor er dem Cluster beitreten kann. Aus diesem Grund schlägt auch der Hauptknoten fehl, wenn die Rechenknoten dem Cluster nicht beitreten können. Je nachdem, welche Art von Compute Notes Sie verwenden, können Sie eines der folgenden Verfahren anwenden, um diese Art von Problem zu beheben:

## Datenverarbeitungsknoten

- Anwendbare Protokolle:
  - `/var/log/cloud-init-output.log`
  - `/var/log/slurmd.log`
- Wenn ein Rechenknoten gestartet wird, überprüfen Sie zunächst `/var/log/cloud-init-output.log`, ob dieser die Setup-Protokolle enthalten sollte, die dem `/var/log/chef-client.log` Protokoll auf dem Hauptknoten ähneln. Bei den meisten Fehlern, die während des Setups auftreten, sollten sich Fehlermeldungen im `/var/log/cloud-init-output.log` Protokoll befinden. Wenn in der Clusterkonfiguration Skripts vor oder nach der Installation angegeben sind, überprüfen Sie, ob sie erfolgreich ausgeführt wurden.
- Wenn Sie ein benutzerdefiniertes AMI mit Änderung der Slurm Konfiguration verwenden, liegt möglicherweise ein Slurm verwandter Fehler vor, der verhindert, dass der Compute-Knoten dem Cluster beitrifft. Suchen Sie im Protokoll nach Fehlern im Zusammenhang mit dem `/var/log/slurmd.log` Scheduler.

## Dynamische Rechenknoten:

- Suchen Sie in `ResumeProgram log` (`/var/log/parallelcluster/slurm_resume.log`) nach dem Namen Ihres Rechenknotens, um zu sehen, ob der Knoten jemals aufgerufen `ResumeProgram` wurde. (Falls der Node nie aufgerufen `ResumeProgram` wurde, können Sie anhand des `slurmctld Logs` (`/var/log/slurmctld.log`) nachsehen, ob Slurm jemals versucht hat, `ResumeProgram` mit dem Node aufzurufen).
- Beachten Sie, dass falsche Berechtigungen für dazu führen `ResumeProgram` können, dass der Fehler `ResumeProgram` automatisch fehlschlägt. Wenn Sie ein benutzerdefiniertes AMI mit



Änderungen am ResumeProgram Setup verwenden, überprüfen Sie, ob das dem slurm Benutzer ResumeProgram gehört und über die 744 (rwxr--r--) -Berechtigung verfügt.

- Wenn aufgerufen ResumeProgram wird, überprüfen Sie, ob eine Instance für den Knoten gestartet wurde. Wenn keine Instance gestartet wurde, wird eine Fehlermeldung angezeigt, die den Startfehler beschreibt.
- Wenn die Instance gestartet wird, ist möglicherweise ein Problem während des Einrichtungsvorgangs aufgetreten. Sie sollten die entsprechende private IP-Adresse und Instanz-ID aus dem ResumeProgram Protokoll sehen. Darüber hinaus können Sie sich die entsprechenden Setup-Protokolle für die jeweilige Instanz ansehen. Weitere Informationen zur Behebung eines Setup-Fehlers mit einem Compute-Knoten finden Sie im nächsten Abschnitt.

Statische Rechenknoten:

- Prüfen Sie im Protokoll clustermgtd (/var/log/parallelcluster/clustermgtd), ob Instanzen für den Knoten gestartet wurden. Wenn sie nicht gestartet wurden, sollte eine klare Fehlermeldung angezeigt werden, in der der Startfehler detailliert beschrieben wird.
- Wenn die Instanz gestartet wird, liegt während des Einrichtungsvorgangs ein Problem vor. Sie sollten die entsprechende private IP-Adresse und Instanz-ID aus dem ResumeProgram Protokoll sehen. Darüber hinaus können Sie sich die entsprechenden Setup-Protokolle für die jeweilige Instanz ansehen.

Rechenknoten, die von Spot-Instances unterstützt werden:

- Wenn Sie Spot-Instances zum ersten Mal verwenden und der Job im Status PD (ausstehend) verbleibt, überprüfen Sie die /var/log/parallelcluster/slurm\_resume.log Datei noch einmal. Sie werden wahrscheinlich einen Fehler wie den folgenden finden:

```
2022-05-20 13:06:24,796 - [slurm_plugin.common:add_instances_for_nodes] - ERROR -
Encountered exception when launching instances for nodes (x1) ['spot-dy-t2micro-2']:
An error occurred (AuthFailure.ServiceLinkedRoleCreationNotPermitted) when calling
the RunInstances operation: The provided credentials do not have permission to
create the service-linked role for EC2 Spot Instances.
```

Wenn Sie Spot-Instances verwenden, muss in Ihrem Konto eine AWSServiceRoleForEC2Spot serviceverknüpfte Rolle vorhanden sein. Führen Sie den folgenden Befehl aus AWS CLI, um diese Rolle in Ihrem Konto mithilfe von zu erstellen:

```
$ aws iam create-service-linked-role --aws-service-name spot.amazonaws.com
```

Weitere Informationen finden Sie [Arbeiten mit Spot-Instances](#) im AWS ParallelCluster Benutzerhandbuch und unter [Service-verknüpfte Rolle für Spot-Instance-Anfragen](#) im Amazon EC2 EC2-Benutzerhandbuch für Linux-Instances.

## Behebung unerwarteter Knotenersetzungen und -beendigungen

In diesem Abschnitt wird weiter untersucht, wie Sie Probleme im Zusammenhang mit Knoten beheben können, insbesondere wenn ein Knoten ersetzt oder unerwartet beendet wird.

- Anwendbare Protokolle:
  - `/var/log/parallelcluster/clustermgtd`(Kopfknoten)
  - `/var/log/slurmctld.log`(Kopfknoten)
  - `/var/log/parallelcluster/computemgtd`(Rechenknoten)

### Knoten wurden unerwartet ersetzt oder beendet

- Prüfen Sie im `clustermgtd` Protokoll (`/var/log/parallelcluster/clustermgtd`), ob ein Knoten `clustermgtd` ersetzt oder beendet wurde. Beachten Sie, dass alle normalen Wartungsaktionen für Knoten `clustermgtd` behandelt werden.
- Wenn der Knoten `clustermgtd` ersetzt oder beendet wurde, sollte eine Meldung erscheinen, in der detailliert beschrieben wird, warum diese Aktion auf dem Knoten ausgeführt wurde. Wenn der Grund mit dem Scheduler zusammenhängt (z. B. weil der Knoten aktiv istDOWN), schauen Sie im `slurmctld` Protokoll nach, um weitere Informationen zu erhalten. Wenn der Grund mit Amazon EC2 zusammenhängt, sollte es eine informative Nachricht geben, in der das Problem im Zusammenhang mit Amazon EC2 beschrieben wird, das den Austausch erforderlich machte.
- Wenn der Knoten `clustermgtd` nicht beendet wurde, überprüfen Sie zunächst, ob es sich um eine erwartete Kündigung durch Amazon EC2 handelt, genauer gesagt um eine Spot-Terminierung. `computemgtd`, der auf einem Rechenknoten ausgeführt wird, kann einen Knoten auch beenden, wenn er als `clustermgtd` fehlerhaft eingestuft wird. Prüfen Sie `computemgtd log` (`/var/log/parallelcluster/computemgtd`), um zu sehen, ob der Knoten `computemgtd` beendet wurde.

### Knoten sind ausgefallen

- Schauen Sie in `slurmctld log (/var/log/slurmctld.log)` nach, warum ein Job oder ein Knoten fehlgeschlagen ist. Beachten Sie, dass Jobs automatisch in die Warteschlange gestellt werden, wenn ein Knoten ausfällt.
- Wenn `slurm_resume` gemeldet wird, dass der Knoten gestartet wurde, und nach einigen Minuten `clustermgtd` meldet, dass es keine entsprechende Instance in Amazon EC2 für diesen Knoten gibt, schlägt der Knoten möglicherweise während der Einrichtung fehl. Gehen Sie wie folgt vor, um das Protokoll von einem Compute (`/var/log/cloud-init-output.log`) abzurufen:
  - Reichen Sie einen Job ein, um einen neuen Knoten hochfahren zu lassenSlurm.
  - Nachdem der Knoten gestartet wurde, aktivieren Sie den Kündigungsschutz mit diesem Befehl.

```
$ aws ec2 modify-instance-attribute --instance-id i-1234567890abcdef0 --disable-api-termination
```

- Rufen Sie mit diesem Befehl die Konsolenausgabe vom Knoten ab.

```
$ aws ec2 get-console-output --instance-id i-1234567890abcdef0 --output text
```

## Ersetzen, Beenden oder Herunterfahren problematischer Instanzen und Knoten

- Anwendbare Protokolle:
  - `/var/log/parallelcluster/clustermgtd(Kopfknoden)`
  - `/var/log/parallelcluster/slurm_suspend.log(Kopfknoden)`
- Bearbeitet in den meisten Fällen `clustermgtd` alle erwarteten Aktionen zur Instanzbeendigung. Sehen Sie im `clustermgtd` Protokoll nach, warum ein Knoten nicht ersetzt oder beendet werden konnte.
- Wenn dynamische Knoten ausfallen[SlurmSettingsEigenschaften](#), schauen Sie im `SuspendProgram` Protokoll nach, ob der spezifische Knoten als Argument aufgerufen `SuspendProgram` wurde. `slurmctld` Beachten Sie, dass tatsächlich `SuspendProgram` keine Aktion ausgeführt wird. Vielmehr protokolliert es nur, wenn es aufgerufen wird. Das Beenden und `NodeAddr` Zurücksetzen aller Instanzen erfolgt von `clustermgtd`. Slurm versetzt Knoten danach `SuspendTimeout` automatisch wieder in einen `POWER_SAVING` Zustand.
- Wenn Rechenknoten aufgrund von Bootstrap-Fehlern ständig ausfallen, überprüfen Sie, ob sie mit [SlurmCluster-geschützter Modus](#) aktivierter Option gestartet werden. Wenn der geschützte Modus

nicht aktiviert ist, ändern Sie die Einstellungen für den geschützten Modus, um den geschützten Modus zu aktivieren. Beheben Sie Fehler und korrigieren Sie das Bootstrap-Skript.

## Status der Warteschlange (Partition) **Inactive**

Wenn Sie das Programm ausführen `sinfo` und in der Ausgabe Warteschlangen mit dem `AVAIL` Status von `inact` angezeigt werden, wurde Ihr Cluster möglicherweise [SlurmCluster-geschützter Modus](#) aktiviert und die Warteschlange wurde für einen vordefinierten Zeitraum auf den `INACTIVE` Status gesetzt.

## Behebung anderer bekannter Knoten- und Jobprobleme

Ein anderes bekanntes Problem besteht darin, dass AWS ParallelCluster möglicherweise keine Jobs zugewiesen oder Skalierungsentscheidungen getroffen werden können. Bei dieser Art von Problem werden Ressourcen AWS ParallelCluster nur gemäß den Slurm-Anweisungen gestartet, beendet oder verwaltet. Überprüfen Sie bei diesen Problemen das `slurmctl` Protokoll, um sie zu beheben.

## Probleme beim Platzieren von Gruppen und beim Starten von Instances

Verwenden Sie eine Platzierungsgruppe, um die niedrigste Latenz zwischen den Knoten zu erzielen. Eine Platzierungsgruppe stellt sicher, dass sich Ihre Instances auf demselben Netzwerk-Backbone befinden. Wenn bei einer Anfrage nicht genügend Instances verfügbar sind, wird ein `InsufficientInstanceCapacity` Fehler zurückgegeben. Um die Wahrscheinlichkeit zu verringern, dass dieser Fehler bei der Verwendung von Cluster-Placement-Gruppen auftritt, setzen Sie den `Enabled` Parameter `SlurmQueuesNetworkingPlacementGroup` auf `false`.

Für zusätzliche Kontrolle über den Kapazitätzugriff sollten Sie erwägen, [Instances mit ODCR \(On-Demand-Kapazitätsreservierungen\) zu starten](#).

Weitere Informationen finden Sie unter [Problembehandlung bei Instance-Startproblemen und Platzierungsgruppen, Rollen und Einschränkungen](#) im Amazon EC2 EC2-Benutzerhandbuch für Linux-Instances.

## Verzeichnisse, die nicht ersetzt werden können

Die folgenden Verzeichnisse werden von den Knoten gemeinsam genutzt und können nicht ersetzt werden.

- `/home-` Dazu gehört der Standard-Home-Ordner des Benutzers (`/home/ec2_user` auf Amazon Linux und RedHat CentOS und `/home/ubuntu` auf Ubuntu). `/home/centos`
- `/opt/intel-` Dazu gehören Intel MPI, Intel Parallel Studio und zugehörige Dateien.
- `/opt/slurm-` Dazu gehören Slurm Workload Manager und zugehörige Dateien. (Bedingt, nur wenn `Scheduler: slurm.`)

## Behebung von Problemen in NICE DCV

### Themen

- [Logs für NICE DCV](#)
- [Probleme mit Ubuntu NICE DCV](#)

### Logs für NICE DCV

Die Protokolle für NICE DCV werden in Dateien im `/var/log/dcv/` Verzeichnis geschrieben. Die Überprüfung dieser Protokolle kann bei der Behebung von Problemen hilfreich sein.

Der Instance-Typ sollte mindestens 1,7 Gibibyte (GiB) RAM haben, um NICE DCV ausführen zu können. Nano- und Mikroinstanztypen verfügen nicht über genügend Arbeitsspeicher, um NICE DCV auszuführen.

AWS ParallelCluster erstellt NICE-DCV-Protokollstreams in Protokollgruppen. Sie können diese Protokolle in der CloudWatch Konsole „Benutzerdefinierte Dashboards“ oder „Protokollgruppen“ anzeigen. Weitere Informationen finden Sie unter [Integration mit Amazon CloudWatch Logs](#) und [CloudWatchAmazon-Dashboard](#).

### Probleme mit Ubuntu NICE DCV

Wenn Sie Gnome Terminal über eine NICE-DCV-Sitzung auf Ubuntu ausführen, haben Sie möglicherweise nicht automatisch Zugriff auf die Benutzerumgebung, die über die Login-Shell verfügbar AWS ParallelCluster ist. Die Benutzerumgebung bietet Umgebungsmodul wie `openmpi` oder `intelmpi` und andere Benutzereinstellungen.

Die Standardeinstellungen von Gnome Terminal verhindern, dass die Shell als Login-Shell gestartet wird. Das bedeutet, dass Shell-Profile nicht automatisch bezogen werden und die AWS ParallelCluster Benutzerumgebung nicht geladen wird.

Gehen Sie wie folgt vor, um das Shell-Profil korrekt zu beziehen und auf die AWS ParallelCluster Benutzerumgebung zuzugreifen:

- Ändern Sie die standardmäßigen Terminaleinstellungen:
  1. Wählen Sie im Gnome-Terminal das Menü Bearbeiten.
  2. Wählen Sie Einstellungen und dann Profile aus.
  3. Wählen Sie „Befehl“ und anschließend „Befehl als Login-Shell ausführen“.
  4. Öffnen Sie ein neues Terminal.
- Verwenden Sie die Befehlszeile, um die verfügbaren Profile abzurufen:

```
$ source /etc/profile && source $HOME/.bashrc
```

## Behebung von Problemen in Clustern mit AWS Batch Integration

Dieser Abschnitt ist relevant für Cluster mit AWS Batch Scheduler-Integration.

Themen

- [Probleme mit dem Hauptknoten](#)
- [Probleme mit der Rechenleistung](#)
- [Fehlschläge Job](#)
- [Verbindungstimeout bei Endpunkt-URL-Fehler](#)

### Probleme mit dem Hauptknoten

Sie können Probleme mit der Einrichtung des Kopfknotens auf die gleiche Weise wie bei einem Slurm Cluster beheben (mit Ausnahme Slurm bestimmter Protokolle). Weitere Informationen zu diesen Problemen finden Sie unter [Hauptknoten](#).

### Probleme mit der Rechenleistung

AWS Batchverwaltet die Skalierungs- und Rechenaspekte Ihrer Dienste. Wenn Sie auf Probleme im Zusammenhang mit der Datenverarbeitung stoßen, finden Sie in der Dokumentation AWS Batch [zur Fehlerbehebung](#) Hilfe.

## Fehlschläge Job

Wenn ein Job fehlschlägt, können Sie den [awsbout](#) Befehl ausführen, um die Jobausgabe abzurufen. Sie können den [awsbstat](#) Befehl auch ausführen, um einen Link zu den von Amazon gespeicherten Jobprotokollen zu erhalten CloudWatch.

## Verbindungstimeout bei Endpunkt-URL-Fehler

Wenn parallel Jobs mit mehreren Knoten mit folgendem Fehler fehlschlagen: `Connect timeout on endpoint URL`

- Überprüfen Sie im `awsbout` Ausgabelog, ob der Job parallel zur Ausgabe mehrere Knoten hat: `Detected 3/3 compute nodes. Waiting for all compute nodes to start.`
- Überprüfen Sie, ob das Subnetz der Rechenknoten öffentlich ist.

parallel Jobs mit mehreren Knoten unterstützen nicht die Verwendung von öffentlichen Subnetzen bei der Verwendung AWS Batch von. AWS ParallelCluster Verwenden Sie ein privates Subnetz für Ihre Rechenknoten und Jobs. Weitere Informationen finden Sie im AWS BatchBenutzerhandbuch unter [Überlegungen zur Rechenumgebung](#). Informationen zur Konfiguration eines privaten Subnetzes für Ihre Rechenknoten finden Sie unter [AWS ParallelClustermitAWS Batch Scheduler](#).

## Problembehandlung bei der Mehrbenutzerintegration mit Active Directory

Dieser Abschnitt ist relevant für Cluster, die in ein Active Directory integriert sind.

Wenn die Active Directory-Integrationsfunktion nicht wie erwartet funktioniert, können die SSSD-Protokolle nützliche Diagnoseinformationen liefern. Diese Protokolle befinden sich `/var/log/sss` auf Clusterknoten. Standardmäßig werden sie auch in der CloudWatch Amazon-Protokollgruppe eines Clusters gespeichert.

### Themen

- [Active Directory-spezifische Fehlerbehebung](#)
- [Aktivieren Sie den Debug-Modus](#)
- [Wie wechselt man von LDAPS zu LDAP](#)
- [Wie deaktiviere ich die Überprüfung von LDAPS-Serverzertifikaten](#)

- [Wie melde ich mich mit einem SSH-Schlüssel statt mit einem Passwort an](#)
- [Wie setze ich ein Benutzerpasswort und abgelaufene Passwörter zurück](#)
- [Wie verifiziert man die beigetretene Domäne](#)
- [Wie behebt man Probleme mit Zertifikaten](#)
- [Wie kann überprüft werden, ob die Integration mit Active Directory funktioniert](#)
- [Wie behebt man Probleme bei der Anmeldung bei Rechenknoten](#)
- [Bekannte Probleme mit SimCenter StarCCM+-Jobs in einer Mehrbenutzerumgebung](#)
- [Bekannte Probleme bei der Benutzernamenauflösung](#)
- [Wie löst man Probleme beim Erstellen von Home-Verzeichnissen](#)

## Active Directory-spezifische Fehlerbehebung

Dieser Abschnitt ist relevant für die Problembehandlung, die für einen Active Directory-Typ spezifisch ist.

### Simple AD

- Der `DomainReadOnlyUser` Wert muss mit der Basissuche im Simple AD AD-Verzeichnis für Benutzer übereinstimmen:

```
cn=ReadOnlyUser, cn=Users, dc=corp, dc=example, dc=com
```

Hinweis `cn` für `Users`.

- Der Standard-Admin-Benutzer ist `Administrator`.
- `Ldapsearch` erfordert den NetBIOS-Namen vor dem Benutzernamen.

`Ldapsearch` Die Syntax muss wie folgt lauten:

```
$ ldapsearch -x -D "corp\\Administrator" -w "Password" -H ldap://192.0.2.103 \
-b "cn=Users, dc=corp, dc=example, dc=com"
```

### AWS Managed Microsoft AD

- Der `DomainReadOnlyUser` Wert muss mit der Basissuche im AWS Managed Microsoft AD Verzeichnis für Benutzer übereinstimmen:



```
cn=ReadOnlyUser,ou=Users,ou=CORP,dc=corp,dc=example,dc=com
```

- Der Standard-Admin-Benutzer ist Admin.
- Ldapsearch Die Syntax muss wie folgt lauten:

```
$ ldapsearch -x -D "Admin" -w "Password" -H ldap://192.0.2.103 \
 -b "ou=Users,ou=CORP,dc=corp,dc=example,dc=com"
```

## Aktivieren Sie den Debug-Modus

Debug-Protokolle von SSSD können nützlich sein, um Probleme zu beheben. Um den Debug-Modus zu aktivieren, müssen Sie den Cluster mit den folgenden Änderungen an der Clusterkonfiguration aktualisieren:

```
DirectoryService:
 AdditionalSssdConfigs:
 debug_level: "0x1ff"
```

## Wie wechselt man von LDAPS zu LDAP

Von der Umstellung von LDAPS (LDAP mit TLS/SSL) auf LDAP wird abgeraten, da LDAP allein keine Verschlüsselung bietet. Dennoch kann es für Testzwecke und zur Fehlerbehebung nützlich sein.

Sie können die vorherige Konfiguration des Clusters wiederherstellen, indem Sie den Cluster mit der vorherigen Konfigurationsdefinition aktualisieren.

Um von LDAPS zu LDAP zu wechseln, müssen Sie den Cluster mit den folgenden Änderungen in der Clusterkonfiguration aktualisieren:

```
DirectoryService:
 LdapTlsReqCert: never
 AdditionalSssdConfigs:
 ldap_auth_disable_tls_never_use_in_production: True
```

## Wie deaktiviere ich die Überprüfung von LDAPS-Serverzertifikaten

Zu Test- oder Fehlerbehebungszwecken kann es nützlich sein, die Überprüfung des LDAPS-Serverzertifikats auf dem Hauptknoten vorübergehend zu deaktivieren.

Sie können die vorherige Konfiguration des Clusters wiederherstellen, indem Sie den Cluster mit der vorherigen Konfigurationsdefinition aktualisieren.

Um die Überprüfung des LDAPS-Serverzertifikats zu deaktivieren, müssen Sie den Cluster mit den folgenden Änderungen in der Clusterkonfiguration aktualisieren:

```
DirectoryService:
 LdapTlsReqCert: never
```

## Wie melde ich mich mit einem SSH-Schlüssel statt mit einem Passwort an

Der SSH-Schlüssel wird erstellt, `/home/$user/.ssh/id_rsa` nachdem Sie sich zum ersten Mal mit einem Passwort angemeldet haben. Um sich mit dem SSH-Schlüssel anzumelden, müssen Sie sich mit Ihrem Passwort anmelden, den SSH-Schlüssel lokal kopieren und ihn dann wie gewohnt passwortlos für SSH verwenden:

```
$ ssh -i $LOCAL_PATH_TO_SSH_KEY $username@$head_node_ip
```

## Wie setze ich ein Benutzerpasswort und abgelaufene Passwörter zurück

Wenn ein Benutzer den Zugriff auf einen Cluster verliert, ist sein [AWS Managed Microsoft AD](#) Passwort möglicherweise abgelaufen.

Um das Passwort zurückzusetzen, führen Sie den folgenden Befehl mit einem Benutzer und einer Rolle aus, die über Schreibberechtigungen für das Verzeichnis verfügen:

```
$ aws ds reset-user-password \
 --directory-id "d-abcdef01234567890" \
 --user-name "USER_NAME" \
 --new-password "NEW_PASSWORD" \
 --region "region-id"
```

Wenn Sie das Passwort für [DirectoryService](#)/zurücksetzen [DomainReadOnlyUser](#):

1. Achten Sie darauf, das [DirectoryService/PasswordSecretArn](#) Secret mit dem neuen Passwort zu aktualisieren.
2. Aktualisieren Sie den Cluster für den neuen geheimen Wert:
  - a. Stoppen Sie die Rechenflotte mit dem `pcluster update-compute-fleet` Befehl.
  - b. Führen Sie den folgenden Befehl innerhalb des Cluster-Hauptknotens aus.

```
$ sudo /opt/parallelcluster/scripts/directory_service/
update_directory_service_password.sh
```

Nach dem Zurücksetzen des Kennworts und dem Cluster-Update sollte der Clusterzugriff des Benutzers wiederhergestellt werden.

Weitere Informationen finden Sie unter [Benutzerpasswort zurücksetzen](#) im AWS Directory ServiceAdministratorhandbuch.

## Wie verifiziert man die beigetretene Domäne

Der folgende Befehl muss von einer Instanz aus ausgeführt werden, die mit der Domäne verknüpft ist, nicht vom Hauptknoten aus.

```
$ realm list corp.example.com \
type: kerberos \
realm-name: CORP.EXAMPLE.COM \
domain-name: corp.example.com \
configured: kerberos-member \
server-software: active-directory \
client-software: sssd \
required-package: oddjob \
required-package: oddjob-mkhomedir \
required-package: sssd \
required-package: adcli \
required-package: samba-common-tools \
login-formats: %U \
login-policy: allow-realm-logins
```

## Wie behebt man Probleme mit Zertifikaten

Wenn die LDAPS-Kommunikation nicht funktioniert, kann dies an Fehlern in der TLS-Kommunikation liegen, die wiederum auf Probleme mit Zertifikaten zurückzuführen sein können.

Hinweise zu Zertifikaten:

- Das in der Clusterkonfiguration angegebene Zertifikat `LdapTlsCaCert` muss ein Paket von PEM-Zertifikaten sein, das die Zertifikate für die gesamte Zertifizierungskette (CA) enthält, die Zertifikate für die Domänencontroller ausgestellt hat.

- Ein Paket von PEM-Zertifikaten ist eine Datei, die aus der Verkettung von PEM-Zertifikaten besteht.
- Ein Zertifikat im PEM-Format (normalerweise in Linux verwendet) entspricht einem Zertifikat im Base64-DER-Format (normalerweise von Windows exportiert).
- Wenn das Zertifikat für Domänencontroller von einer untergeordneten Zertifizierungsstelle ausgestellt wird, muss das Zertifikatspaket das Zertifikat sowohl der untergeordneten Zertifizierungsstelle als auch der Stammzertifizierungsstelle enthalten.

Schritte zur Fehlerbehebung bei der Überprüfung:

Bei den folgenden Überprüfungsschritten wird davon ausgegangen, dass die Befehle vom Cluster-Hauptknoten aus ausgeführt werden und dass der Domänencontroller unter erreichbar ist **SERVER:PORT**.

Gehen Sie folgendermaßen vor, um ein Problem im Zusammenhang mit Zertifikaten zu beheben:

Schritte zur Überprüfung:

1. Überprüfen Sie die Verbindung zu den Active Directory-Domänencontrollern:

Stellen Sie sicher, dass Sie eine Verbindung zu einem Domänencontroller herstellen können. Wenn dieser Schritt erfolgreich ist, ist die SSL-Verbindung zum Domänencontroller erfolgreich und das Zertifikat wird überprüft. Ihr Problem hat nichts mit Zertifikaten zu tun.

Wenn dieser Schritt fehlschlägt, fahren Sie mit der nächsten Überprüfung fort.

```
$ openssl s_client -connect SERVER:PORT -CAfile PATH_TO_CA_BUNDLE_CERTIFICATE
```

2. Überprüfen Sie die Überprüfung des Zertifikats:

Stellen Sie sicher, dass das lokale Zertifizierungsstellenzertifikatspaket das vom Domänencontroller bereitgestellte Zertifikat validieren kann. Wenn dieser Schritt erfolgreich ist, hängt Ihr Problem nicht mit Zertifikaten zusammen, sondern mit anderen Netzwerkproblemen.

Wenn dieser Schritt fehlschlägt, fahren Sie mit der nächsten Überprüfung fort.

```
$ openssl verify -verbose -
CAfile PATH_TO_CA_BUNDLE_CERTIFICATE PATH_TO_A_SERVER_CERTIFICATE
```

3. Überprüfen Sie das von den Active Directory-Domänencontrollern bereitgestellte Zertifikat:

Stellen Sie sicher, dass der Inhalt des von den Domänencontrollern bereitgestellten Zertifikats den Erwartungen entspricht. Wenn dieser Schritt erfolgreich ist, haben Sie wahrscheinlich Probleme mit dem CA-Zertifikat, das zur Überprüfung der Controller verwendet wird. Fahren Sie mit dem nächsten Schritt zur Fehlerbehebung fort.

Schlägt dieser Schritt fehl, müssen Sie das für die Domänencontroller ausgestellte Zertifikat korrigieren und die Schritte zur Fehlerbehebung erneut ausführen.

```
$ openssl s_client -connect SERVER:PORT -showcerts
```

#### 4. Überprüfen Sie den Inhalt eines Zertifikats:

Stellen Sie sicher, dass der Inhalt des von den Domänencontrollern bereitgestellten Zertifikats den Erwartungen entspricht. Wenn dieser Schritt erfolgreich ist, haben Sie wahrscheinlich Probleme mit dem CA-Zertifikat, das zur Überprüfung der Controller verwendet wird. Fahren Sie mit dem nächsten Schritt zur Fehlerbehebung fort.

Schlägt dieser Schritt fehl, müssen Sie das für die Domänencontroller ausgestellte Zertifikat korrigieren und die Schritte zur Fehlerbehebung erneut ausführen.

```
$ openssl s_client -connect SERVER:PORT -showcerts
```

#### 5. Überprüfen Sie den Inhalt des lokalen CA-Zertifikatspakets:

Stellen Sie sicher, dass der Inhalt des lokalen Zertifizierungsstellenzertifikatspakets, das zur Überprüfung des Zertifikats der Domänencontroller verwendet wird, den Erwartungen entspricht. Wenn dieser Schritt erfolgreich ist, haben Sie wahrscheinlich Probleme mit dem Zertifikat, das von den Domänencontrollern bereitgestellt wird.

Wenn dieser Schritt fehlschlägt, müssen Sie das für die Domänencontroller ausgestellte CA-Zertifikatspaket korrigieren und die Schritte zur Fehlerbehebung erneut ausführen.

```
$ openssl x509 -in PATH_TO_A_CERTIFICATE -text
```

## Wie kann überprüft werden, ob die Integration mit Active Directory funktioniert

Wenn die folgenden beiden Prüfungen erfolgreich sind, funktioniert die Integration mit dem Active Directory.

Prüfungen:

1. Sie können Benutzer finden, die im Verzeichnis definiert sind:

Aus dem Hauptknoten des Clusters heraus, `alsec2-user`:

```
$ getent passwd $ANY_AD_USER
```

2. Sie können per SSH auf den Hauptknoten zugreifen und das Benutzerpasswort angeben:

```
$ ssh $ANY_AD_USER@$HEAD_NODE_IP
```

Wenn Prüfung eins fehlschlägt, gehen wir davon aus, dass auch Prüfung zwei fehlschlägt.

Zusätzliche Prüfungen zur Fehlerbehebung:

- Stellen Sie sicher, dass der Benutzer im Verzeichnis vorhanden ist.
- Aktivieren Sie die [Debug-Protokollierung](#).
- Erwägen Sie, die Verschlüsselung vorübergehend zu deaktivieren, indem Sie [von LDAPS zu LDAP wechseln, um LDAPS-Probleme](#) auszuschließen.

## Wie behebt man Probleme bei der Anmeldung bei Rechenknoten

Dieser Abschnitt ist relevant für die Anmeldung bei Rechenknoten in Clustern, die in Active Directory integriert sind.

Mit AWS ParallelCluster sind Kennwortanmeldungen an Cluster-Rechenknoten standardmäßig deaktiviert.

Alle Benutzer müssen ihren eigenen SSH-Schlüssel verwenden, um sich bei Rechenknoten anzumelden.

Benutzer können ihren SSH-Schlüssel nach der ersten Authentifizierung (z. B. Anmeldung) im Hauptknoten abrufen, sofern dies in der Clusterkonfiguration aktiviert [GenerateSshKeysForUsers](#) ist.

Wenn sich Benutzer zum ersten Mal auf dem Hauptknoten authentifizieren, können sie SSH-Schlüssel abrufen, die automatisch für sie als Verzeichnisbenutzer generiert werden. Home-Verzeichnisse für den Benutzer werden ebenfalls erstellt. Dies kann auch passieren, wenn ein Sudo-Benutzer zum ersten Mal zu einem Benutzer im Hauptknoten wechselt.

Wenn sich ein Benutzer nicht am Hauptknoten angemeldet hat, werden keine SSH-Schlüssel generiert und der Benutzer kann sich nicht bei Rechenknoten anmelden.

## Bekannte Probleme mit SimCenter StarCCM+-Jobs in einer Mehrbenutzerumgebung

Dieser Abschnitt bezieht sich auf Jobs, die mit der Computational Fluid Dynamics Software Simcenter StarCCM+ von Siemens in einer Mehrbenutzerumgebung gestartet wurden.

Wenn Sie StarCCM+ v16-Jobs ausführen, die für die Verwendung des eingebetteten IntelMPI konfiguriert sind, werden die MPI-Prozesse standardmäßig mithilfe von SSH gebootet.

Aufgrund eines bekannten [Slurm-Fehlers, der dazu führt, dass die Benutzernamen-Auflösung falsch ist, können Jobs mit einem Fehler](#) wie `error setting up the bootstrap proxies` Dieser Fehler betrifft nur die AWS ParallelCluster Versionen 3.1.1 und 3.1.2.

Um dies zu verhindern, zwingen Sie IntelMPI, Slurm als MPI-Bootstrap-Methode zu verwenden. [Exportieren Sie die Umgebungsvariable I\\_MPI\\_HYDRA\\_BOOTSTRAP=slurm in das Job-Skript, das StarCCM+ startet, wie in der offiziellen IntelMPI-Dokumentation beschrieben.](#)

## Bekannte Probleme bei der Benutzernamenauflösung

Dieser Abschnitt ist relevant für das Abrufen von Benutzernamen innerhalb von Jobs.

Aufgrund eines bekannten [Fehlers in Slurm](#) könnte der innerhalb eines Jobprozesses abgerufene Nutzernamen lauten, `nobody` wenn Sie einen Job ohne ausführen. `slurm` Dieser Fehler betrifft nur die AWS ParallelCluster Versionen 3.1.1 und 3.1.2.

Wenn Sie den Befehl beispielsweise `sbatch --wrap 'slurm id'` als Verzeichnisbenutzer ausführen, wird der richtige Benutzername zurückgegeben. Wenn Sie den jedoch `sbatch --`

`wrap 'id'` als Verzeichnisbenutzer ausführen, wird `nobody` möglicherweise der Benutzername zurückgegeben.

Sie können die folgenden Probleumlösungen verwenden.

1. Starten Sie Ihren Job `'sbatch'`, wenn möglich, mit `'srun'` statt mit.
2. Aktivieren Sie die SSSD-Enumeration, indem Sie die [AdditionalSssdConfigs](#) Cluster-Konfiguration wie folgt festlegen.

```
AdditionalSssdConfigs:
 enumerate: true
```

## Wie löst man Probleme beim Erstellen von Home-Verzeichnissen

Dieser Abschnitt ist relevant für Probleme bei der Erstellung von Home-Verzeichnissen.

Wenn Sie Fehler wie den im folgenden Beispiel gezeigten sehen, wurde kein Home-Verzeichnis für Sie erstellt, als Sie sich zum ersten Mal am Hauptknoten angemeldet haben. Oder es wurde kein Home-Verzeichnis für Sie erstellt, als Sie zum ersten Mal im Hauptknoten von einem Sudoer zu einem Active Directory-Benutzer wechselten.

```
$ ssh AD_USER@$HEAD_NODE_IP
/opt/parallelcluster/scripts/generate_ssh_key.sh failed: exit code 1

 __| __|_)
 _| (/ Amazon Linux 2 AMI
 __|__|__|

https://aws.amazon.com/amazon-linux-2/
Could not chdir to home directory /home/PclusterUser85: No such file or directory
```

Der Fehler beim Erstellen des Home-Verzeichnisses kann durch die `odjjob` im `odjjob-mkhomedir` Cluster-Hauptknoten installierten UND-Pakete verursacht werden.

Ohne ein Home-Verzeichnis und einen SSH-Schlüssel kann der Benutzer keine Jobs oder SSH an die Clusterknoten senden.

Wenn Sie die `odjjob` Pakete in Ihrem System benötigen, stellen Sie sicher, dass der `odjjobd` Dienst ausgeführt wird, und aktualisieren Sie die PAM-Konfigurationsdateien, um sicherzustellen,



dass das Home-Verzeichnis erstellt wurde. Führen Sie dazu die Befehle im Hauptknoten aus, wie im folgenden Beispiel gezeigt.

```
sudo systemctl start oddjobd
sudo authconfig --enablemkhomedir --updateall
```

Wenn Sie die oddjob Pakete in Ihrem System nicht benötigen, deinstallieren Sie sie und aktualisieren Sie die PAM-Konfigurationsdateien, um sicherzustellen, dass das Home-Verzeichnis erstellt wird. Führen Sie dazu die Befehle im Hauptknoten aus, wie im folgenden Beispiel gezeigt.

```
sudo yum remove -y oddjob oddjob-mkhomedir
sudo authconfig --enablemkhomedir --updateall
```

## Behebung von Problemen mit benutzerdefinierten AMIs

Wenn Sie ein benutzerdefiniertes AMI verwenden, werden die folgenden Warnungen angezeigt:

```
"validationMessages": [
 {
 "level": "WARNING",
 "type": "CustomAmiTagValidator",
 "message": "The custom AMI may not have been created by pcluster. You can ignore this warning if the AMI is shared or copied from another pcluster AMI. If the AMI is indeed not created by pcluster, cluster creation will fail. If the cluster creation fails, please go to https://docs.aws.amazon.com/parallelcluster/latest/ug/troubleshooting.html#troubleshooting-stack-creation-failures for troubleshooting."
 },
 {
 "level": "WARNING",
 "type": "AmiOsCompatibleValidator",
 "message": "Could not check node AMI ami-0000012345 OS and cluster OS alinux2 compatibility, please make sure they are compatible before cluster creation and update operations."
 }
]
```

Wenn Sie sicher sind, dass das richtige AMI verwendet wird, können Sie diese Warnungen ignorieren.

Wenn Sie diese Warnungen in future nicht mehr sehen möchten, kennzeichnen Sie das benutzerdefinierte AMI mit den folgenden Tags, wobei *my-os* einer von *linux2*, *ubuntu2204*, *ubuntu2004centos7*, oder *rhel8* und „*3.7.0*“ die verwendete *pcluster* Version ist:

```
$ aws ec2 create-tags \
 --resources ami-yourcustomAmi \
 --tags Key="parallelcluster:version",Value="3.7.0" \
 Key="parallelcluster:os",Value="my-os"
```

## Fehlerbehebung bei einem Timeout für ein Cluster-Update, wenn es nicht läuft **cfn-hup**

Der *cfn-hup* Helper ist ein Daemon, der Änderungen an Ressourcenmetadaten erkennt und benutzerdefinierte Aktionen ausführt, wenn eine Änderung erkannt wird. So nehmen Sie über die UpdateStack API-Aktion Konfigurationsaktualisierungen auf Ihren laufenden Amazon EC2 EC2-Instances vor.

Derzeit wird der *cfn-hup* Daemon von der gestartet. *supervisord* Nach dem Start entzieht sich der *cfn-hup* Prozess jedoch der *supervisord* Kontrolle. Wenn der *cfn-hup* Dämon von einem externen Akteur getötet wird, wird er nicht automatisch neu gestartet. Wenn es *cfn-hup* nicht läuft, startet der CloudFormation Stack während eines Cluster-Updates den Aktualisierungsvorgang wie erwartet, aber das Aktualisierungsverfahren ist auf dem Hauptknoten nicht aktiviert und der Stack erreicht irgendwann ein Timeout. Den Clusterprotokollen können Sie entnehmen `/var/log/chef-client`, dass das Aktualisierungsrezept nie aufgerufen wird.

Überprüfen Sie die Prüfung und starten Sie **cfn-hup** sie neu, falls Fehler auftreten

1. Überprüfen Sie auf dem Hauptknoten, ob *cfn-hup* Folgendes läuft:

```
$ ps aux | grep cfn-hup
```

2. Überprüfen Sie das *cfn-hup* Protokoll `/var/log/cfn-hup.log` und `/var/log/supervisord.log` auf dem Hauptknoten.

3. Wenn es *cfn-hup* nicht läuft, versuchen Sie es neu zu starten, indem Sie Folgendes ausführen:

```
$ sudo /opt/parallelcluster/pyenv/versions/cookbook_virtualenv/bin/supervisorctl \
start cfn-hup
```

## Fehlerbehebung im Netzwerk

### Probleme mit Clustern in einem einzelnen öffentlichen Subnetz

Überprüfen Sie das `cloud-init-output.log` von einem der Rechenknoten aus. Wenn Sie etwas wie das Folgende finden, das darauf hindeutet, dass der Knoten bei der Slurm Initialisierung feststeckt, liegt das höchstwahrscheinlich an einem fehlenden DynamoDB-VPC-Endpunkt. Fügen Sie den DynamoDB-Endpunkt hinzu. Weitere Informationen finden Sie unter [AWS ParallelCluster in einem einzelnen Subnetz ohne Internetzugang ohne Internetzugang ohne Internetzugang ohne Internetzugang](#).

```
ruby_block[retrieve compute node info] action run[2022-03-11T17:47:11+00:00] INFO:
 Processing ruby_block[retrieve compute node info] action run (aws-parallelcluster-
slurm::init line 31)
```

### Das Cluster-Update ist bei der benutzerdefinierten Aktion **onNodeUpdated** fehlgeschlagen

Wenn ein [HeadNode/CustomActions/OnNodeUpdated](#)-Skript fehlschlägt, schlägt das Update fehl und das Skript wird beim Rollback nicht ausgeführt. Es liegt in Ihrer Verantwortung, die nach Abschluss des Rollbacks erforderlichen Bereinigungen manuell durchzuführen. Wenn das `OnNodeUpdated` Skript beispielsweise den Status eines Felds in einer Konfigurationsdatei ändert (z. B. von `true` bis `false`) und dann fehlschlägt, müssen Sie den Feldwert manuell auf den Zustand vor der Aktualisierung zurücksetzen (z. B. `false` auf `true`). Weitere Informationen finden Sie unter [Benutzerdefinierte Bootstrap-Aktionen](#).

### Fehler bei der benutzerdefinierten Slurm Konfiguration werden angezeigt

Ab AWS ParallelCluster Version 3.6.0 können Sie einzelne `prolog` oder `epilog` Skripte nicht mehr gezielt verwenden, indem Sie sie in eine benutzerdefinierte Slurm Konfiguration aufnehmen. In AWS ParallelCluster Version 3.6.0 und späteren Versionen müssen Sie in den entsprechenden Prolog Ordnern `prolog` und nach benutzerdefinierten `epilog` Skripten suchen. Epilog Diese Ordner sind standardmäßig so konfiguriert, dass sie auf Folgendes verweisen:

- Prolog zeigt auf `/opt/slurm/etc/scripts/prolog.d/`.

- Epilog zeigt auf `opt/slurm/etc/scripts/epilog.d/`.

Wir empfehlen, das `90_plcluster_health_check_manager` Prolog-Skript und das `90_plcluster_noop` Epilog-Skript beizubehalten.

Slurm führt die Skripte in umgekehrter alphabetischer Reihenfolge aus. Prolog Sowohl der Epilog Ordner als auch müssen mindestens eine Datei enthalten. Weitere Informationen finden Sie unter [Slurm prolog und epilog](#) und [Slurm Anpassung der Konfiguration](#).

## Cluster-Alarme

Die Überwachung des Clusterzustands ist für die Sicherstellung einer optimalen Leistung unerlässlich. AWS ParallelCluster ermöglicht es Ihnen, mehrere CloudWatch Alarme für den Hauptknoten des Clusters zu überwachen.

In diesem Abschnitt finden Sie Einzelheiten zu den einzelnen Alarmen im Headknoten-Cluster, einschließlich der Benennungskonventionen, der spezifischen Bedingungen, die Alarme auslösen, und der empfohlenen Schritte zur Fehlerbehebung.

Die Benennungskonvention für Cluster-Alarme lautet `CLUSTER_NAME-COMPONENT-METRIC` z. `mycluster-HeadNode-Cpu` B.

- `CLUSTER_NAME-HeadNode`: signalisiert den Gesamtstatus des Kopfknotens. Es ist rot, wenn mindestens einer der unten aufgeführten Alarme aktiviert ist.
- `CLUSTER_NAME-HeadNode-Health`: rot, wenn mindestens ein EC2 Health Check-Fehler aufgetreten ist. Im Alarmfall empfehlen wir, einen Blick auf die [Problembehandlung bei Instanzen mit fehlgeschlagenen Statusprüfungen](#) zu werfen.
- `CLUSTER_NAME-HeadNode-Cpu`: rot, wenn die CPU-Auslastung mehr als 90% beträgt. Überprüfen Sie im Alarmfall die Prozesse, die die CPU am meisten verbrauchen `ps -aux --sort=-%cpu | head -n 10`.
- `CLUSTER_NAME-HeadNode-Mem`: rot, wenn die Speicherauslastung mehr als 90% beträgt. Überprüfen Sie im Alarmfall die Prozesse, die den Speicher am meisten verbrauchen `ps -aux --sort=-%mem | head -n 10`.
- `CLUSTER_NAME-HeadNode-Disk`: rot, wenn der belegte Festplattenspeicher auf dem Pfad `/` mehr als 90% beträgt. Überprüfen Sie im Alarmfall die Ordner, die den größten Teil des Speicherplatzes mit `du -h --max-depth=2 / 2> /dev/null | sort -hr` belegen.

## Zusätzliche Unterstützung

Eine Liste der bekannten Probleme finden Sie auf der [GitHub Wiki-Hauptseite](#) oder auf der [Problemseite](#).

Bei dringenderen Problemen wenden Sie sich an AWS Support oder öffnen Sie ein [neues GitHub Problem](#).

# AWS ParallelCluster Unterstützungspolitik

AWS ParallelCluster unterstützt mehrere Versionen gleichzeitig. Jede AWS ParallelCluster Version hat ein geplantes Ende der Support-Laufzeit (EOSL). Nach dem EOSL-Datum wird für diese Version kein weiterer Support oder keine Wartung mehr angeboten.

AWS ParallelCluster verwendet ein `major.minor.patch` Versionsschema. Neue Funktionen, Leistungsverbesserungen, Sicherheitsupdates und Bugfixes sind in den neuen Nebenversionsversionen der neuesten Hauptversion enthalten. Nebenversionen sind innerhalb einer Hauptversion abwärtskompatibel. AWS Stellt für kritische Probleme Korrekturen durch Patch-Versionen bereit, jedoch nur für die neuesten Nebenversionen von Releases, die EOSL noch nicht erreicht haben. Wenn Sie die Updates einer neuen Version verwenden möchten, müssen Sie ein Upgrade auf die neue Neben- oder Patch-Version durchführen.

| AWS ParallelCluster Versionen | Datum des Endes der unterstützten Nutzungsdauer (EOSL) |
|-------------------------------|--------------------------------------------------------|
| 3.0. <i>x</i>                 | 31.3.2023                                              |
| 3.1. <i>x</i>                 | 31.8.2023                                              |
| 3.2. <i>x</i>                 | 31.1.2024                                              |
| 3.3. <i>x</i>                 | 31.5.2024                                              |
| 3.4. <i>x</i>                 | 28.06.2024                                             |
| 3.5. <i>x</i>                 | 31.8.2024                                              |
| 3.6. <i>x</i>                 | 30.11.2024                                             |
| 3.7. <i>x</i>                 | 28.02.2025                                             |
| 3.8. <i>x</i>                 | 30.06.2025                                             |
| 3.9. <i>x</i>                 | 09.05.2025                                             |

# Sicherheit in AWS ParallelCluster

Die Sicherheit in der Cloud hat bei AWS höchste Priorität. Als AWS-Kunde profitieren Sie von einer Rechenzentrums- und Netzwerkarchitektur, die zur Erfüllung der Anforderungen von Organisationen entwickelt wurden, für die Sicherheit eine kritische Bedeutung hat.

Sicherheit ist eine übergreifende Verantwortlichkeit zwischen AWS und Ihnen. Das [Modell der geteilten Verantwortung](#) beschreibt dies als Sicherheit der Cloud und Sicherheit in der Cloud.

- Sicherheit der Cloud selbst – AWS ist dafür verantwortlich, die Infrastruktur zu schützen, mit der AWS-Services in der AWS Cloud ausgeführt werden. AWS stellt Ihnen außerdem Services bereit, die Sie sicher nutzen können. Auditoren von Drittanbietern testen und überprüfen die Effektivität unserer Sicherheitsmaßnahmen im Rahmen der [AWS-Compliance-Programme](#) regelmäßig. Informationen zu den Compliance-Programmen, die für AWS ParallelCluster gelten, finden Sie unter [Im Rahmen des Compliance-Programms zugelassene AWS-Services](#).
- Sicherheit in der Cloud — Ihre Verantwortung richtet sich nach dem oder den spezifischen AWS Diensten, die Sie nutzen. Sie sind auch für verschiedene andere damit verbundene Faktoren verantwortlich, einschließlich der Sensibilität Ihrer Daten, der Anforderungen Ihres Unternehmens und der geltenden Gesetze und Vorschriften.

In dieser Dokumentation wird beschrieben, wie Sie das Modell der geteilten Verantwortung anwenden sollten, wenn Sie es verwenden AWS ParallelCluster. Die folgenden Themen veranschaulichen, wie Sie AWS ParallelCluster zur Erfüllung Ihrer Sicherheits- und Compliance-Ziele konfigurieren können. Sie lernen auch, wie Sie es so verwenden AWS ParallelCluster, dass Sie Ihre AWS Ressourcen überwachen und sichern können.

## Themen

- [Sicherheitsinformationen für Dienste, die von AWS ParallelCluster](#)
- [Datenschutz in AWS ParallelCluster](#)
- [Identity and Access Management für AWS ParallelCluster](#)
- [Compliance-Validierung für AWS ParallelCluster](#)
- [Erzwingen einer Mindestversion von TLS 1.2](#)

# Sicherheitsinformationen für Dienste, die von AWS ParallelCluster

- [Sicherheit in Amazon EC2](#)
- [Sicherheit im Amazon API Gateway](#)
- [Sicherheit in AWS Batch](#)
- [Sicherheit in AWS CloudFormation](#)
- [Sicherheit bei Amazon CloudWatch](#)
- [Sicherheit in AWS CodeBuild](#)
- [Sicherheit in Amazon DynamoDB](#)
- [Sicherheit in Amazon ECR](#)
- [Sicherheit in Amazon ECS](#)
- [Sicherheit in Amazon EFS](#)
- [Sicherheit in FSx for Lustre](#)
- [Sicherheit in AWS Identity and Access Management \(IAM\)](#)
- [Sicherheit in EC2 Image Builder](#)
- [Sicherheit in AWS Lambda](#)
- [Sicherheit in Amazon Route 53](#)
- [Sicherheit in Amazon SNS](#)
- [Sicherheit in Amazon SQS \(Für AWS ParallelCluster Version 2.x.\)](#)
- [Sicherheit in Amazon S3](#)
- [Sicherheit in Amazon VPC](#)

## Datenschutz in AWS ParallelCluster

Das [Modell der geteilten Verantwortung](#) von AWS gilt für den Datenschutz in AWS ParallelCluster. Wie in diesem Modell beschrieben, ist AWS verantwortlich für den Schutz der globalen Infrastruktur, in der die gesamte AWS Cloud ausgeführt wird. Sie sind dafür verantwortlich, die Kontrolle über Ihre in dieser Infrastruktur gehosteten Inhalte zu behalten. Sie sind auch für die Sicherheitskonfiguration und AWS-Services die Verwaltungsaufgaben der von Ihnen verwendeten Geräte verantwortlich. Weitere Informationen zum Datenschutz finden Sie unter [Häufig gestellte Fragen zum Datenschutz](#). Informationen zum Datenschutz in Europa finden Sie im Blog-Beitrag [AWS-Modell der geteilten Verantwortung und in der DSGVO](#) im AWS-Sicherheitsblog.



Aus Datenschutzgründen empfehlen wir, AWS-Konto-Anmeldeinformationen zu schützen und einzelne Benutzer mit AWS IAM Identity Center oder AWS Identity and Access Management (IAM) einzurichten. So erhält jeder Benutzer nur die Berechtigungen, die zum Durchführen seiner Aufgaben erforderlich sind. Außerdem empfehlen wir, die Daten mit folgenden Methoden zu schützen:

- Verwenden Sie für jedes Konto die Multi-Faktor Authentifizierung (MFA).
- Verwenden Sie SSL/TLS für die Kommunikation mit AWS-Ressourcen. Wir benötigen TLS 1.2 und empfehlen TLS 1.3.
- Richten Sie die API und die Protokollierung von Benutzeraktivitäten mit AWS CloudTrail ein.
- Verwenden Sie AWS-Verschlüsselungslösungen zusammen mit allen Standardsicherheitskontrollen in AWS-Services.
- Verwenden Sie erweiterte verwaltete Sicherheitsservices wie Amazon Macie, die dabei helfen, in Amazon S3 gespeicherte persönliche Daten zu erkennen und zu schützen.
- Wenn Sie für den Zugriff auf AWS über eine Befehlszeilenschnittstelle oder über eine API FIPS 140-2-validierte kryptografische Module benötigen, verwenden Sie einen FIPS-Endpunkt. Weitere Informationen über verfügbare FIPS-Endpunkte finden Sie unter [Federal Information Processing Standard \(FIPS\) 140-2](#).

Wir empfehlen dringend, in Freitextfeldern, z. B. im Feld Name, keine vertraulichen oder sensiblen Informationen wie die E-Mail-Adressen Ihrer Kunden einzugeben. Dies gilt auch, wenn Sie unter Verwendung der Konsole, der API, AWS CLI oder AWS SDKs mit AWS ParallelCluster oder anderen AWS-Services arbeiten. Alle Daten, die Sie in Tags oder Freitextfelder eingeben, die für Namen verwendet werden, können für Abrechnungs- oder Diagnoseprotokolle verwendet werden. Wenn Sie eine URL für einen externen Server bereitstellen, empfehlen wir dringend, keine Anmeldeinformationen zur Validierung Ihrer Anforderung an den betreffenden Server in die URL einzuschließen.

## Datenverschlüsselung

Ein wesentliches Merkmal eines sicheren Service ist, dass Informationen verschlüsselt werden, wenn sie nicht aktiv verwendet werden.

### Verschlüsselung im Ruhezustand

AWS ParallelCluster speichert selbst keine anderen Kundendaten als die Anmeldeinformationen, die es benötigt, um im Namen des Benutzers mit den AWS Diensten zu interagieren.

Daten auf den Knoten im Cluster können im Ruhezustand verschlüsselt werden.

Für Amazon EBS-Volumes wird die Verschlüsselung mithilfe der KmsKeyId Einstellungen [EbsSettings/Encrypted](#) und [EbsSettings/im EbsSettings](#) Abschnitt konfiguriert. Weitere Informationen finden Sie unter [Amazon EBS-Verschlüsselung](#) im Amazon EC2 EC2-Benutzerhandbuch für Linux-Instances.

Für Amazon EFS-Volumes wird die Verschlüsselung mithilfe der KmsKeyId Einstellungen [EfsSettingsEfsSettings/Encrypted](#) und [im EfsSettings](#) Abschnitt konfiguriert. Weitere Informationen finden Sie unter [So funktioniert Verschlüsselung im Ruhezustand](#) im Amazon Elastic File System-Benutzerhandbuch.

Für FSx for Lustre-Dateisysteme wird die Verschlüsselung ruhender Daten automatisch aktiviert, wenn ein Amazon FSx-Dateisystem erstellt wird. Weitere Informationen finden Sie unter [Verschlüsseln ruhender Daten im](#) Amazon FSx for Lustre-Benutzerhandbuch.

Bei Instance-Typen mit NVMe-Volumes werden die Daten auf NVMe-Instance-Speicher-Volumes mit einer XTS-AES-256-Verschlüsselung verschlüsselt, die auf einem Hardwaremodul auf der Instance implementiert ist. Die Verschlüsselungsschlüssel werden mithilfe des Hardwaremoduls erstellt und sind für jedes NVMe-Instance-Speichergerät eindeutig. Alle Verschlüsselungsschlüssel werden zerstört, wenn die Instance angehalten oder beendet wird, und können nicht wiederhergestellt werden. Sie können diese Verschlüsselung nicht deaktivieren und keine eigenen Verschlüsselungsschlüssel bereitstellen. Weitere Informationen finden Sie unter [Verschlüsselung im Ruhezustand](#) im Amazon EC2 EC2-Benutzerhandbuch für Linux-Instances.

Wenn Sie AWS ParallelCluster einen AWS Service aufrufen, der Kundendaten zur Speicherung auf Ihren lokalen Computer überträgt, finden Sie im Kapitel Sicherheit und Compliance im Benutzerhandbuch dieses Services weitere Informationen darüber, wie diese Daten gespeichert, geschützt und verschlüsselt werden.

## Verschlüsselung während der Übertragung

Standardmäßig werden alle Daten, die von den laufenden Client-Computern AWS ParallelCluster und den AWS Dienstendpunkten übertragen werden, verschlüsselt, indem alles über eine HTTPS/TLS-Verbindung gesendet wird. Der Datenverkehr zwischen den Knoten im Cluster kann je nach den ausgewählten Instanztypen automatisch verschlüsselt werden. Weitere Informationen finden Sie unter [Verschlüsselung bei der Übertragung](#) im Amazon-EC2-Benutzerhandbuch für Linux-Instances.

## Weitere Informationen finden Sie auch unter

- [Datenschutz in Amazon EC2](#)
- [Datenschutz in EC2 Image Builder](#)
- [Datenschutz in AWS CloudFormation](#)
- [Datenschutz in Amazon EFS](#)
- [Datenschutz in Amazon S3](#)
- [Datenschutz in FSx for Lustre](#)

## Identity and Access Management für AWS ParallelCluster

AWS ParallelCluster verwendet Rollen, um auf Ihre AWS Ressourcen und deren Dienste zuzugreifen. Die Instanz- und Benutzerrichtlinien, die zur Erteilung von Berechtigungen AWS ParallelCluster verwendet werden, sind unter [dokumentiert AWS Identity and Access Management Berechtigungen in AWS ParallelCluster](#).

Der einzige wesentliche Unterschied besteht darin, wie Sie sich bei der Verwendung eines Standard-Benutzers und Langzeit-Anmeldeinformationen authentifizieren. Obwohl ein Benutzer ein Passwort benötigt, um auf die Konsole eines AWS Dienstes zuzugreifen, benötigt derselbe Benutzer ein Zugriffsschlüsselpaar, um dieselben Operationen ausführen zu können AWS ParallelCluster. Alle anderen Kurzzeit-Anmeldeinformationen werden auf die gleiche Weise verwendet, wie sie mit der Konsole verwendet werden.

Die von verwendeten Anmeldeinformationen AWS ParallelCluster werden in Klartextdateien gespeichert und sind nicht verschlüsselt.

- Die `$HOME/.aws/credentials`-Datei speichert Langzeit-Anmeldeinformationen, die für den Zugriff auf Ihre AWS-Ressourcen erforderlich sind. Diese beinhalten Ihre Zugriffsschlüssel-ID und Ihren geheimen Zugriffsschlüssel.
- Kurzfristige Anmeldeinformationen, wie z. B. für angenommene Rollen oder für AWS IAM Identity Center-Services, werden ebenfalls in den Ordnern `$HOME/.aws/cli/cache` und `$HOME/.aws/sso/cache` gespeichert.

### Risikominderung

- Wir empfehlen dringend, die Dateisystemberechtigungen für den Ordner \$HOME/.aws und seine untergeordneten Ordner und Dateien so zu konfigurieren, dass der Zugriff ausschließlich auf autorisierte Benutzer beschränkt wird.
- Verwenden Sie möglichst Rollen mit temporären Anmeldeinformationen, um bei einer Gefährdung der Sicherheit von Anmeldeinformationen die Möglichkeit für Schäden zu reduzieren. Verwenden Sie Langzeit-Anmeldeinformationen nur zum Abfragen und Aktualisieren der Anmeldeinformationen von Kurzzeit-Rollen.

## Compliance-Validierung für AWS ParallelCluster

Externe Auditoren bewerten im Rahmen verschiedener AWS-Compliance-Programme die Sicherheit und Compliance der AWS-Services. Die Verwendung AWS ParallelCluster für den Zugriff auf einen Dienst ändert nichts an der Konformität dieses Dienstes.

Eine Liste der AWS Services, die in den Anwendungsbereich bestimmter Compliance-Programme fallen, finden Sie unter [AWS Services im Anwendungsbereich nach Compliance-Programm](#). Allgemeine Informationen finden Sie unter [AWS-Compliance-Programme](#).

Die Auditberichte von externen Auditoren lassen sich mit AWS Artifact herunterladen. Weitere Informationen finden Sie unter [Herunterladen von Berichten in AWS Artifact](#).

Ihre Compliance-Verantwortung bei der Verwendung von AWS ParallelCluster ist von der Sensibilität Ihrer Daten, den Compliance-Zielen Ihres Unternehmens und den geltenden Gesetzen und Vorschriften abhängig. AWS stellt die folgenden Ressourcen zur Unterstützung der Compliance bereit:

- [Kurzanleitungen für Sicherheit und Compliance](#) – In diesen Bereitstellungsleitfäden finden Sie wichtige Überlegungen zur Architektur sowie die einzelnen Schritte zur Bereitstellung von sicherheits- und Compliance-orientierten Basisumgebungen in AWS.
- [AWS Whitepaper zur Erstellung einer Architektur mit HIPAA-konformer Sicherheit und Compliance](#) — In diesem Whitepaper wird beschrieben, wie Unternehmen mithilfe AWS von HIPAA-konforme Anwendungen erstellen können.
- [AWS Ressourcen zur Einhaltung von Vorschriften](#) zur Einhaltung von Vorschriften - Diese Sammlung von Arbeitsbüchern und Leitfäden könnte für Ihre Branche und Ihren Standort zutreffen.
- [Bewertung von Ressourcen](#) mit Regeln im AWS Config Developer Guide - Das AWS Config Service bewertet, wie gut Ihre Ressourcenkonfigurationen mit internen Praktiken, Branchenrichtlinien und Vorschriften übereinstimmen.

- [AWS Security Hub](#) – Dieser AWS-Service liefert einen umfassenden Überblick über den Sicherheitsstatus in AWS. So können Sie die Compliance mit den Sicherheitsstandards in der Branche und den bewährten Methoden abgleichen.

## Erzwingen einer Mindestversion von TLS 1.2

Um die Sicherheit bei der Kommunikation mit AWS -Services zusätzlich AWS ParallelCluster zu erhöhen, sollten Sie Ihre so konfigurieren, dass Sie TLS 1.2 oder höher verwenden. Wenn Sie Python verwenden AWS ParallelCluster, wird Python verwendet, um die TLS-Version festzulegen.

Um sicherzustellen, dass keine TLS-Version vor Version TLS 1.2 AWS ParallelCluster verwendet, müssen Sie möglicherweise OpenSSL neu kompilieren, um das Verwenden dieser Mindestversion zu erzwingen, und dann Python neu kompilieren, um das neu erstellte OpenSSL zu verwenden.

### Ermitteln Ihrer derzeit unterstützten Protokolle

Erstellen Sie zunächst mit OpenSSL ein selbstsigniertes Zertifikat, das für den Testserver und das Python-SDK verwendet werden soll.

```
$ openssl req -subj '/CN=localhost' -x509 -newkey rsa:4096 -nodes -keyout key.pem -out cert.pem -days 365
```

Starten Sie dann einen Testserver mit OpenSSL.

```
$ openssl s_server -key key.pem -cert cert.pem -www
```

Erstellen Sie in einem neuen Terminalfenster eine virtuelle Umgebung und installieren Sie das Python-SDK.

```
$ python3 -m venv test-env
source test-env/bin/activate
pip install botocore
```

Erstellen Sie ein neues Python-Skript namens `check.py`, das die dem SDK zugrunde liegende HTTP-Bibliothek verwendet.

```
$ import urllib3
URL = 'https://localhost:4433/'
```

```
http = urllib3.PoolManager(
 ca_certs='cert.pem',
 cert_reqs='CERT_REQUIRED',
)
r = http.request('GET', URL)
print(r.data.decode('utf-8'))
```

Führen Sie Ihr neues Skript aus.

```
$ python check.py
```

Damit werden Details über die hergestellte Verbindung angezeigt. Suchen Sie in der Ausgabe nach „Protokoll:“. Wenn die Ausgabe „TLSv1.2“ oder höher lautet, ist das SDK auf TLS v1.2 oder höher voreingestellt. Wenn es sich um eine frühere Version handelt, müssen Sie OpenSSL und Python neu kompilieren.

Auch wenn die Installation von Python auf TLS v1.2 oder höher voreingestellt ist, ist es jedoch weiterhin möglich, dass Python neu auf eine frühere Version als TLS v1.2 verhandelt, wenn der Server TLS v1.2 oder höher nicht unterstützt. Um zu überprüfen, ob Python nicht automatisch auf frühere Versionen neu verhandelt, starten Sie den Testserver erneut mit Folgendem.

```
$ openssl s_server -key key.pem -cert cert.pem -no_tls1_3 -no_tls1_2 -www
```

Wenn Sie eine frühere Version von OpenSSL verwenden, steht Ihnen die `-no_tls_3`-Flag möglicherweise nicht zur Verfügung. Wenn dies der Fall ist, entfernen Sie das Flag, da die von Ihnen verwendete Version von OpenSSL TLS v1.3 nicht unterstützt. Führen Sie dann das Python-Skript aus.

```
$ python check.py
```

Wenn die Installation von Python korrekterweise für Versionen vor TLS 1.2 nicht neu verhandelt, sollten Sie einen SSL-Fehler erhalten.

```
$ urllib3.exceptions.MaxRetryError: HTTPSConnectionPool(host='localhost',
 port=4433): Max retries exceeded with url: / (Caused by SSLError(SSLError(1, '[SSL:
 UNSUPPORTED_PROTOCOL] unsupported protocol (_ssl.c:1108)')))
```

Wenn Sie eine Verbindung herstellen können, müssen Sie OpenSSL und Python neu kompilieren, um das Aushandeln von Protokollen vor TLS v1.2 zu deaktivieren.

## Kompilieren von OpenSSL und Python

Um sicherzustellen, dass AWS ParallelCluster das nichts vor TLS 1.2 aushandelt, müssen Sie OpenSSL und Python neu kompilieren. Kopieren Sie dazu den folgenden Inhalt, um ein Skript zu erstellen und auszuführen.

```
#!/usr/bin/env bash
set -e

OPENSSL_VERSION="1.1.1d"
OPENSSL_PREFIX="/opt/openssl-with-min-tls1_2"
PYTHON_VERSION="3.8.1"
PYTHON_PREFIX="/opt/python-with-min-tls1_2"

curl -O "https://www.openssl.org/source/openssl-$OPENSSL_VERSION.tar.gz"
tar -xzf "openssl-$OPENSSL_VERSION.tar.gz"
cd openssl-$OPENSSL_VERSION
./config --prefix=$OPENSSL_PREFIX no-ssl3 no-tls1 no-tls1_1 no-shared
make > /dev/null
sudo make install_sw > /dev/null

cd /tmp
curl -O "https://www.python.org/ftp/python/$PYTHON_VERSION/Python-$PYTHON_VERSION.tgz"
tar -xzf "Python-$PYTHON_VERSION.tgz"
cd Python-$PYTHON_VERSION
./configure --prefix=$PYTHON_PREFIX --with-openssl=$OPENSSL_PREFIX --disable-shared > /dev/null
make > /dev/null
sudo make install > /dev/null
```

Dadurch wird eine Version von Python kompiliert, die über ein statisch verknüpftes OpenSSL verfügt, das nicht automatisch eine frühere Version als TLS 1.2 aushandelt. Dadurch werden auch OpenSSL im `/opt/openssl-with-min-tls1_2`-Verzeichnis und Python im `/opt/python-with-min-tls1_2`-Verzeichnis installiert. Nachdem Sie dieses Skript ausgeführt haben, bestätigen Sie die Installation der neuen Version von Python.

```
$ /opt/python-with-min-tls1_2/bin/python3 --version
```

Dadurch sollte Folgendes ausgedruckt werden.

**Python 3.8.1**

Um zu bestätigen, dass diese neue Version von Python keine frühere Version als TLS 1.2 aushandelt, führen Sie die Schritte unter [Ermitteln Ihrer derzeit unterstützten Protokolle](#) mit der neu installierten Python-Version (also `/opt/python-with-min-tls1_2/bin/python3`) erneut aus.



## Versionshinweise und Dokumentenverlauf

Die folgende Tabelle beschreibt die wichtigen Updates und neuen Funktionen für das AWS ParallelCluster -Benutzerhandbuch. Wir aktualisieren die Dokumentation regelmäßig, um das Feedback, das Sie uns senden, einzuarbeiten.

| Änderung                                                                | Beschreibung                                                                                                                                                                                                                                                                                                                                                                                                                                                   | Datum          |
|-------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------|
| <a href="#">AWS ParallelCluster UI-Version 2024.04.0 veröffentlicht</a> | <p>AWS ParallelCluster UI-Version 2024.04.0 veröffentlicht.</p> <p>Features:</p> <ul style="list-style-type: none"> <li>• Unterstützung für Version 3.9.1 hinzugefügt AWS ParallelCluster</li> <li>• <a href="#">Vollständiges Changelog</a></li> </ul>                                                                                                                                                                                                        | 17. April 2024 |
| <a href="#">AWS ParallelCluster Version 3.9.1 veröffentlicht</a>        | <p>Wir freuen uns, die Veröffentlichung von AWS ParallelCluster 3.9.1 bekannt zu geben</p> <p>Um ein Upgrade durchzuführen, geben Sie Folgendes ein: <code>sudo pip install --upgrade aws-parallelcluster</code></p> <p>Fehlerbehebungen</p> <ul style="list-style-type: none"> <li>• Entfernt das rekursive Löschen des gemeinsam genutzten Speichers <code>mountdir</code> beim Aushängen von Dateisystemen als Teil des Update-Cluster-Vorgangs.</li> </ul> | 11. April 2024 |

## [AWS ParallelCluster Version 3.9.1 veröffentlicht](#)

Wir freuen uns, die Veröffentlichung von AWS ParallelCluster 3.9.1 bekannt zu geben

11. April 2024

Um ein Upgrade durchzuführen, geben Sie Folgendes ein: `sudo pip install --upgrade aws-parallelcluster`

### Fehlerbehebungen

- Entfernt das rekursive Löschen des gemeinsam genutzten Speichers `mountdir` beim Aushängen von Dateisystemen als Teil des Update-Cluster-Vorgangs.

## [AWS ParallelCluster UI-Version 2024.03.0 veröffentlicht](#)

AWS ParallelCluster UI-Version 2024.03.0 veröffentlicht.

12. März 2024

### Features:

- Unterstützung für Version 3.9.0 hinzugefügt AWS ParallelCluster
- Unterstützung für Ubuntu 22.04 und Red Hat Enterprise Linux 9 hinzugefügt
- Veraltetes Ubuntu 18.04

### Fehlerkorrekturen

- Es wurde ein Problem behoben, das dazu führte, dass einige Cluster nicht angezeigt wurden, wenn viele Cluster verwendet wurden

Einzelheiten zu den Änderungen finden Sie in den CHANGELOG Dateien für das [aws-parallelcluster-ui](#) Paket unter GitHub.

## [AWS ParallelCluster Version 3.9.0 veröffentlicht](#)

Wir freuen uns, die Veröffentlichung von AWS ParallelCluster 3.9.0 bekannt zu geben

5. März 2024

Um ein Upgrade durchzuführen, geben Sie Folgendes ein: `sudo pip install --upgrade aws-parallelcluster`

Verbesserungen:

- Fügen Sie den Konfigurationsparameter `DeploymentSettings/DefaultUserHome` hinzu, damit Benutzer das Home-Verzeichnis des Standardbenutzers `/local/home` anstelle von `/home` (Standard) verschieben können.
- Ermöglicht die Aktualisierung von `Queue` und `ComputeResource` Konfigurationsparametern `MinCountMaxCount`, ohne dass die Rechenflotte gestoppt werden muss. Es ist jetzt möglich, sie zu aktualisieren, indem Sie sie `Scheduling/SlurmSettings/QueueUpdateStrategy` auf `TERMINATE` setzen. AWS ParallelCluster beendet nur die Knoten, die

bei einer Größenänderung der Clusterkapazität durch ein Cluster-Update entfernt wurden.

- Erlaubt die Aktualisierung des externen gemeinsam genutzten Speichers vom Typ Efs,, FsxLustre,, FsxOntap, FileCache ohne die Rechen FsxOpenZfs - und Anmeldeflotte zu ersetzen.
- Unterstützung für RHEL9 hinzugefügt.
- Unterstützung für Rocky Linux 9 hinzugefügt, wie es durch build-image den Prozess CustomAmi erstellt wurde. Derzeit ist kein offizielles AWS ParallelCluster Rocky9 Linux AMI verfügbar.
- CommunicationParameters Aus der Sperrliste für benutzerdefinierte Slurm Einstellungen entfernen.
- DeploymentSettings /DisableSudoAccessForDefaultUser Parameter hinzufügen, um den Sudo-Zugriff des Standardbenutzers in unterstützten Betriebssystemen zu deaktivieren.

- Änderungen an FSx for Lustre-Dateisystemen erstellt von ParallelCluster: Ändern Sie die Lustre-Serverversion in 2.15.
- Fügen Sie die Möglichkeit hinzu, beim Erstellen eines AMI über das ['cluster']['nvidia']['kernel\_open'] Cookbook-Knotenattribut zwischen Open- und Closed Source-Nvidia-Treibern zu wählen.
- \* Fügen Sie eine clustermgtd-Konfigurationsoption hinzu, um eine konfigurierbare Anzahl von Wiederholungsversuchen `ec2_instance_missing_max_count` zu ermöglichen, damit die Konsistenz der EC2-Describe-Instances mit den ausgeführten Instances gewährleistet ist.

## Änderungen

- Führen Sie ein Upgrade auf 23.11.4 Slurm (von 23.02.7) durch.
- Aktualisieren Sie den NVIDIA-Treiber auf Version 535.154.05.
- Unterstützung für Python 3.11, 3.12 in pcluster CLI

hinzugefügt und. aws-parallelcluster-batch-cli

- Erstellen Sie Netzwerkschnittstellen mithilfe des Netzwerkkartenindex aus der NetworkCardIndex Liste der DescribeInstances EC2-Antworten, anstatt eine Schleife über den Bereich zu laufen. MaximumNetworkCards
- Schlägt die Clustererstellung fehl, wenn Sie die Instance-Typen P3, G3, P2 und G2 verwenden, da deren GPU-Architektur nicht mit den in Version 3.8.0 eingeführten Open-Source-Nvidia-Treibern (OpenRM) kompatibel ist.
- Aktualisieren Sie die Abhängigkeiten von Drittanbieter-Kochbüchern: nfs-5.1.2 (von nfs-5.0.0)
- Aktualisieren Sie das EFA-Installationsprogramm auf 1.30.0.
  - EFA-Treiber: efa-2.6.0-1
  - EFA-Konfiguration: efa-config-1.15-1
  - EFA-Profil: efa-profile-1.6-1

- libFabric-aws:  
libfabric-aws-1.19.0
- RDMA-Kern: rdma-core-46.0-1
- MPI öffnen: und  
openmpi40-aws-4.1.6-2 openmpi50-aws-5.0.0-11
- NICE DCV auf Version aktualisieren 2023.1-16388.
  - Server: 2023.1.16388-1
  - xdcv: 2023.1.565-1
  - gl: 2023.1.1047-1
  - Webviewer: 2023.1.16388-1

## Fehlerbehebungen

- Behebung eines Problems, das dazu führte, dass der Job fehlschlug, wenn er als Active Directory-Benutzer von Anmeldeknoten aus eingereicht wurde. Das Problem wurde durch eine unvollständige Konfiguration der Integration mit dem externen Active Directory auf dem Hauptknoten verursacht.
- Refaktorisieren Sie die in der CloudFormation Vorlage



parallelclutser-policies.yaml definierten IAM-Richtlinien, um zu verhindern, dass API-Bereitstellungsfehler aufgrund von Richtlinien, die die IAM-Grenzwerte überschreiten, verursacht werden. ParallelCluster

- Behebung eines Problems, das dazu führte, dass Anmeldeknoten nicht starteten, wenn der Hauptknoten mehr Zeit als erwartet für das Schreiben von Schlüsseln benötigt.

Einzelheiten zu den Änderungen finden Sie in den CHANGELOG Dateien für das [aws-parallelcluster-ui](#) Paket unter GitHub.

[AWS ParallelCluster UI-Version 2024.02.0 veröffentlicht](#)

AWS ParallelCluster UI-Version 2024.02.0 veröffentlicht

8. Februar 2024

Änderungen:

- Die Lambda-Laufzeitumgebung wurde auf Python v3.9 aktualisiert

Einzelheiten zu den Änderungen finden Sie in den CHANGELOG Dateien für das [aws-parallelcluster-ui](#) Paket unter GitHub

[AWS ParallelCluster UI-Version 2023.12.0 veröffentlicht](#)

AWS ParallelCluster UI-Version 2023.12.0 veröffentlicht.

21. Dezember 2023

Features:

- Unterstützung für die PCI-Bereitstellung mit privaten Netzwerken hinzugefügt.
- Es wurde die Möglichkeit hinzugefügt, optional eine Berechtigungsgrenze auf jede IAM-Rolle anzuwenden, die von den PCUI- und PCAPI-Infrastrukturen erstellt wurde
- Es wurde die Möglichkeit hinzugefügt, optional ein Präfix auf jede IAM-Rolle und -Richtlinie anzuwenden, die von der PCUI- und PCAPI-Infrastruktur erstellt wurden.
- Unterstützung für ParallelCluster Version 3.8.0 ohne Funktionsparität im Assistenten hinzugefügt.

Einzelheiten zu den Änderungen finden Sie in den CHANGELOG Dateien für das [aws-parallelcluster-ui](#) Paket unter [GitHub](#)

[AWS ParallelCluster Version  
3.8.0 veröffentlicht](#)

AWS ParallelCluster Version  
3.8.0 veröffentlicht.

19. Dezember 2023

Verbesserungen:

- Unterstützung für EC2 Capacity Blocks für ML hinzugefügt.
- Unterstützung für Rocky Linux 8 hinzugefügt, wie es durch den `build-image` Prozess `CustomAmi` erstellt wurde. Derzeit ist kein offizielles AWS ParallelCluster Rocky8 Linux AMI verfügbar.
- Fügen Sie einen `Scheduling/Scaling Strategy` Parameter hinzu, um die Cluster-Skalierungsstrategie zu steuern, die beim Starten von EC2-Instances für Slurm-Rechenknoten verwendet werden soll. Mögliche Werte sind `all-or-nothing` ,, `greedy-all-or-nothing` `best-effort` , `all-or-nothing` wobei dies die Standardeinstellung ist.
- Fügen Sie einen `HeadNode/SharedStorageType` Parameter

hinzu, um EFS-Speicher anstelle von NFS-Exporten aus dem Stammvolumen des Hauptknotens für gemeinsam genutzte Dateisystemressourcen innerhalb des Clusters zu verwenden: `ParallelCluster`, `Intel`, `Slurm` und `Datan`. `/home` Diese Erweiterung reduziert die Belastung des Headnode-Netzwerks.

- Erlaubt das `/home` Mounten als externer EFS- oder FSx-gemeinsam genutzter Speicher über den `SharedStorage` Abschnitt der Konfigurationsdatei.
- Fügen Sie einen neuen Parameter `hinzuSlurmSettings/MungeKeySecretArn`, um die Verwendung eines externen benutzerdefinierten MUNGE-Schlüssels aus AWS Secrets Manager zu ermöglichen.
- Fügen Sie einen `Monitoring/Alarms/Enabled` Parameter hinzu, um Amazon CloudWatch Alarms für den Cluster umzuschalten.
- Fügen Sie Head-Node -Alarmer hinzu, um die EC2-Zustandsprüfungen,

die CPU-Auslastung und den Gesamtstatus des Hauptknotens zu überwachen, und fügen Sie sie dem CloudWatch Dashboard hinzu, das mit dem Cluster erstellt wurde.

- Fügen Sie Unterstützung für Data Repository Associations hinzu, wenn Sie `PERSISTENT_2` als `DeploymentType` für einen verwalteten FSx for Lustre verwenden.
- Fügen Sie einen `Scheduling/SlurmSettings/Database/DatabaseName` Parameter hinzu, damit Benutzer einen benutzerdefinierten Namen für die Datenbank auf dem Datenbankserver angeben können, der für die Slurm-Buchhaltung verwendet werden soll.
- Geben Sie `InstanceType` bei der Konfiguration `CapacityReservationTarget/CapacityReservationId` in der Rechenressource einen optionalen Konfigurationsparameter an.
- Fügen Sie die Möglichkeit hinzu, ein Präfix für IAM-Rollen und -Richtlinien

anzugeben, die von der AWS ParallelCluster API erstellt wurden.

- Fügt die Möglichkeit hinzu, eine Berechtigungsgrenze anzugeben, die für von AWS ParallelCluster der API erstellte IAM-Rollen und -Richtlinien angewendet werden soll.

## Änderungen

- Aktualisieren Sie Slurm auf 23.02.7 (von 23.02.6).
- Aktualisieren Sie den NVIDIA-Treiber auf Version 535.129.03.
- Aktualisieren Sie das CUDA Toolkit auf Version 12.2.2.
- Verwenden Sie Open-Source-NVIDIA-GPU-Treiber (OpenRM) als NVIDIA-Kernelmodul für Linux anstelle des NVIDIA-ClosedSource-Moduls.
- Entfernen Sie die Unterstützung der `all_or_nothing_batch` Konfigurationsparameter im Slurm-Resume-Programm zugunsten der neuen Scheduling/Scaling Strategy Cluster-Konfiguration.

- Die Benennungskonvention für Cluster-Alarmer wurde in '[Clustername] - [Komponentenname] - [Metrik]' geändert.
- Ändern Sie die standardmäßigen EBS-Volumentypen in ADC-Regionen von gp2 auf gp3, sowohl für das Stammvolumen als auch für zusätzliche Volumes.
- Die optionale Berechtigungsgrenze für die AWS ParallelCluster API wird jetzt auf jede IAM-Rolle angewendet, die von der API-Infrastruktur erstellt wurde.
  - Aktualisieren Sie das EFA-Installationsprogramm auf 1.29.1
  - EFA-Treiber: efa-2.6.0-1
  - EFA-Konfiguration: efa-config-1.15-1
  - EFA-Profil: efa-profile-1.5-1
  - libFabric-aws: libfabric-aws-1.19.0-1
  - RDMA-Kern: rdma-core-46.0-1
  - MPI öffnen: openmpi40-aws-4.1.6-1

- Aktualisieren Sie GdrCopy auf Version 2.4 in allen unterstützten Betriebssystemen, außer Centos 7, wo Version 2.3.1 verwendet wird.
- Führen Sie ein Upgrade auf Version 2.0-28 durch. `aws-cfn-bootstrap`
- Unterstützung für Python 3.10 in `aws-parallelcluster-batch-cli` hinzugefügt.

### Fehlerbehebungen

- Korrigiert die inkonsistente Skalierungskonfiguration nach dem Rollback des Cluster-Updates, wenn die Liste der in den Compute-Ressourcen deklarierten Instanztypen geändert wurde.
- Behebung der Generierung von SSH-Schlüsseln für Benutzer beim Wechseln von Benutzern ohne Root-Rechte in Clustern, die über Cluster-Konfigurationsdateien in einen externen LDAP-Server integriert sind.
- Das Deaktivieren des Slurm-Stromsparmodes bei der Einstellung wurde behoben. `ScaledownIdleTime = -1`



- Korrigiert den fest codierten Pfad zum Slurm-Installation sverzeichnis im update\_slurm\_database\_password.sh Skript für Slurm Accounting.

[AWS ParallelCluster Version 3.7.2 veröffentlicht](#)

AWS ParallelCluster Version 3.7.2 veröffentlicht.

25. Oktober 2023

Änderungen:

- Aktualisiere Slurm auf 23.02.6.

## [AWS ParallelCluster UI-Version 2023.10.0 veröffentlicht](#)

AWS ParallelCluster UI-Version 2023.10.0 veröffentlicht.

20. Oktober 2023

### Features:

- Unterstützung für ParallelCluster 3.7.2 hinzugefügt, wobei die Funktionsparität im Assistenten auf FSx File Cache beschränkt ist, und speicherbasierte Scheduling-Kompatibilität mit mehreren Instanztypen.

### Fehlerbehebungen:

- Es wurde ein Problem behoben, das zu Benutzeroberflächenfehlern führte, wenn PCUI keine Berechtigungen für die Interaktion mit dem Cost Explorer hatte.

### Verbesserungen

- Die Sicherheit wurde verbessert, indem die TTL des Zugriffstokens von 10 Minuten auf 5 Minuten reduziert wurde.

Einzelheiten zu den Änderungen finden Sie in den CHANGELOG Dateien für das

[aws-parallelcluster-ui](#) Paket

unter GitHub.

[AWS ParallelCluster Version 3.7.1 veröffentlicht](#)

AWS ParallelCluster Version 3.7.1 veröffentlicht.

22. September 2023

### Änderungen:

- Upgraden Sie Slurm auf 23.02.5 (von 23.02.4).
  - Aktualisieren Sie Pmix auf 4.2.6 (von 3.2.3).
  - Aktualisieren Sie libjwt auf 1.15.3 (von 1.12.0).
- Aktualisieren Sie das EFA-Installationsprogramm auf und beheben Sie damit das RDMA-Schreibdatenproblem in 1.26.1 P5.
  - efa-2.5.0-1 EFA-Treiber:.
  - EFA-Konfiguration: efa-config-1.15-1
  - EFA-Profil: efa-profile-1.5-1
  - libFabric-aws: libfabric-aws-1.18.2-1
  - Erdma-Kern: rdma-core-46.0-1
  - Öffnen Sie MPI: openmpi40-aws-4.1.5-4

[AWS ParallelCluster Version  
3.7.0 veröffentlicht](#)

AWS ParallelCluster Version  
3.7.0 veröffentlicht.

30. August 2023

Verbesserungen:

- Support die Konfiguration statischer und dynamischer Knotenprioritäten in Rechenressourcen mithilfe einer AWS ParallelCluster YAML-Konfigurationsdatei.
- Unterstützung für Ubuntu 22 hinzufügen. RSA-Schlüssel werden standardmäßig nicht unterstützt.
- Fügen Sie die Einstellung für die Warteschlangenkonfiguration hinzu `JobExclusiveAllocation`, um Knoten in einer Partition zu einem bestimmten Zeitpunkt ausschließlich einem einzelnen Job zuzuweisen.
- Lassen Sie `aws-parallelcluster-node` das Override-Paket bei der Clustererstellung und beim Cluster-Update zu. Für den Hauptknoten gilt dies für das Cluster-Update. Nur für Entwicklungszwecke nützlich.

- Vermeiden Sie den Start von NFS-Servern auf Rechenknoten.
- Unterstützung für Anmeldeknoten hinzufügen.
- Erlaubt die speicherbasierte Planung, wenn mehrere Instanztypen für eine Slurm-Rechenressource angegeben sind.
- Unterstützung hinzugefügt, um vorhandenen Amazon File Cache als gemeinsam genutzten Speicher zu mounten.

#### Änderungen:

- Weisen Sie Slurm dynamischen Knoten standardmäßig eine Priorität (Gewichtung) von 1000 zu. Auf diese Weise Slurm können statische Knoten im Leerlauf gegenüber dynamischen Knoten im Leerlauf priorisiert werden.
- Sorgen Sie dafür, dass `aws-parallelcluster-node` Daemons nur verwaltete Partitionen verwalten AWS ParallelCluster. Slurm
- Erhöhen Sie das EFS-`utils` Watchdog-

Abfrageintervall auf 10 Sekunden. Diese Änderung gilt `true`, wenn sie auf `gesetzt` `EncryptionInTransit` ist. Dies ist die einzige Bedingung, unter der der Watchdog ausgeführt wird.

- Aktualisieren Sie das EFA-Installationsprogramm auf 1.25.1
  - EFA-Treiber: `efa-2.5.0-1` (von) `efa-2.1.1g`
  - EFA-Config: (von) `efa-config-1.15-1` `efa-config-1.13-1`
  - EFA-Profil: `efa-profile-1.5-1` (keine Änderung)
  - `libFabric-aws`: (von) `libfabric-aws-1.18.1-0` `libfabric-aws-1.17.1-1`
  - RDMA-Core: (von) `rdma-core-46.0-1` `rdma-core-43.0-1`
  - MPI öffnen: `openmpi40-aws-4.1.5-4` (von) `openmpi40-aws-4.1.5-1`
- Führen Sie ein Upgrade Slurm auf Version 23.02.4 durch.

- Ändern Sie den Standardwert von `Imds/ImdsSupport` von v1.0 auf v2.0.
- Verwerfen Sie Ubuntu 18.
- Aktualisieren Sie die Standardgröße des Root-Volumes auf 40 GB, um die Beschränkungen auf Centos 7 zu berücksichtigen.
- Beschränken Sie die Zugriffsrechte für die Datei `/tmp/wait_condition_handle.txt` innerhalb des Hauptknotens, sodass nur Root-Benutzer sie lesen können.
- Erstellen Sie eine JSON-Datei mit einer Slurm-Partition-Nodelist-Zuordnung, die von den Node-Paket-Daemons verwendet werden soll, um vom PC verwaltete Slurm-Partitionen und -Knotenlisten zu erkennen.
- Aktualisieren Sie den NVIDIA-Treiber auf Version 535.54.03.
- Aktualisieren Sie die CUDA-Bibliothek auf Version 12.2.0.
- Aktualisieren Sie den NVIDIA Fabric Manager auf `nvidia-fabricmanager-535`.
- Aktualisieren Sie ARM PL auf Version 23.04.1 nur für Ubuntu 22.04.

- Aktualisieren Sie NICE DCV auf Version 2023.0-15487 .
  - Server: 2023.0.15487-1
  - xdcv: 2023.0.551-1
  - gl: 2023.0.1039-1
  - Webviewer: 2023.0.15487-1

#### Fehlerbehebungen:

- Fügen Sie dem Scaledown IdleTime Wert eine Validierung hinzu, um zu verhindern, dass ein Wert niedriger als -1 gesetzt wird.
- Behebung eines Fehlers bei der Clustererstellung mit Ubuntu Deep Learning AMI auf GPU-Instanzen mit aktiviertem DCV.
- Problem behoben, das dazu führte, dass beim Erstellen eines ParallelCluster CloudFormation benutzerdefinierten Ressourcenanbieters mit fehlerhafte IAM-Richtlinien erstellt wurden. CustomLambdaRole
- Behebt ein Problem, das zu einer falschen Ausrichtung des DNS-Namens von Rechenknoten auf Instances



mit mehreren Netzwerkschnittstellen führte, wenn Equals to verwendet wurde SlurmSettings/Dns/UseEc2Hostnames True

Einzelheiten zu den Änderungen finden Sie in den CHANGELOG Dateien für [aws-parallelcluster](#), und packages on. [aws-parallelcluster-cookbookaws-parallelcluster-node](#) GitHub

[Veröffentlichung nur zur Dokumentation](#)

AWS ParallelCluster Spezifisches Benutzerhandbuch für Version 3 veröffentlicht. 17. Juli 2023

Version nur zur Dokumentation:

- AWS ParallelCluster Version 3 hat ein eigenes separates Benutzerhandbuch.

[AWS ParallelCluster Version  
3.6.1 veröffentlicht](#)

AWS ParallelCluster Version  
3.6.1 veröffentlicht.

5. Juli 2023

Änderungen:

- Vermeiden Sie die doppelte Anzahl von Knoten, `clustermgtd` wenn Rechenknoten zu mehreren Slurm Partitionen hinzugefügt werden.

Fehlerbehebungen:

- Entfernen Sie die Hardcodierung des Root-Volume-Gerätenamens (`/dev/sda1` und `/dev/xvda`) und rufen Sie ihn aus den AMI (s) ab, die während verwendet wurden `create-cluster`.
- Behebung eines Fehlers bei der Clustererstellung bei Verwendung einer CloudFormation benutzerdefinierten Ressource mit `ElasticIp` Einstellung auf `True`.
- Behebt Fehler beim Erstellen und Aktualisieren von Clustern, wenn eine AWS CloudFormation benutzerdefinierte Ressource mit großen

Konfigurationsdateien verwendet wurde.

- Behebt ein Problem, durch das der `ptrace` Schutz auf Ubuntu nicht deaktiviert werden konnte und das Cross Memory Attach (CMA) in `libfabric` nicht zuließ.
- Behebt die schnelle Failover-Logik mit unzureichender Kapazität, wenn mehrere Instanztypen verwendet werden und keine Instanzen zurückgegeben werden.

Einzelheiten zu den Änderungen finden Sie in den `CHANGELOG` Dateien für den [aws-parallelcluster](#) und in den Paketen unter [aws-parallelcluster-cookbookaws-parallelcluster-node](#) [GitHub](#)

[AWS ParallelCluster UI-Version 2023.06.0 veröffentlicht](#)

AWS ParallelCluster UI-Version 2023.06.0 veröffentlicht. 7. Juni 2023

Änderungen:

- Die AWS ParallelCluster Standard-API-Version wurde auf 3.6.0 aktualisiert.

Fehlerbehebungen:

- Eine fehlerhafte Bereitstellung für die Region AWS GovCloud (US-West) wurde behoben.
- Das geteilte Panel lädt die Cluster-Details jetzt korrekt, nachdem die Erstellung gestartet wurde.

Hinweise:

- Die Funktion zur Kostenüberwachung ist in nicht verfügbar AWS GovCloud (US) Regions.

Einzelheiten zu den Änderungen finden Sie in den CHANGELOG Dateien für das [aws-parallelcluster-ui](#) Paket unter GitHub.

[AWS ParallelCluster Version  
3.6.0 veröffentlicht](#)

AWS ParallelCluster Version  
3.6.0 veröffentlicht.

22. Mai 2023

Dokumentation:

- Dokumentation für die [AWS ParallelClusterPython-Bibliothek-API](#) hinzufügen.

Verbesserungen:

- Unterstützung für RHEL8 hinzufügen.
- Fügen Sie eine [AWS CloudFormation benutzerdefinierte Ressource](#) zum Erstellen und Verwalten von Clustern mit hinzu. CloudFormation
- Fügen Sie Unterstützung für [das Anpassen der Slurm Clusterkonfiguration](#) in der AWS ParallelCluster YAML-Konfigurationsdatei hinzu.
- Erstellen Sie Slurm mit Unterstützung für LUA.
- Erhöhen Sie das Limit für die maximale Anzahl von Warteschlangen pro Cluster von 10 auf 50. Jede Warteschlange kann bis zu 50 Rechenressourcen enthalten. Jeder Cluster

kann bis zu 50 Rechenresourcen haben.

- Fügen Sie Unterstützung für die Angabe einer Sequenz von mehreren [benutzerdefinierten Aktionsskripten](#) für ein Ereignis hinzu, das in den OnNodeUpdated Parametern OnNodeStart OnNodeConfigured , und konfiguriert ist.
- Fügen Sie den neuen Konfigurationsabschnitt healthChecks / hinzuGpu, um GPU-Zustandsprüfungen auf einen Rechenknoten anzuwenden, bevor ein Job ausgeführt wird.
- Fügen Sie Unterstützung für Tags in der ComputeResources Konfiguration SlurmQueues undSlurmQueues /hinzu.
- Fügen Sie [DetailedMonitoring](#) in der Monitoring Konfiguration Unterstützung für hinzu.
- Fügen Sie mem\_used\_percent im AWS ParallelCluster [CloudWatch Dashboard disk\\_used\\_percent](#) Metriken für den Arbeitsspeicher und die Festplattenauslastung

des Hauptknotens hinzu und richten Sie Alarme zur Überwachung dieser Messwerte ein.

- Fügen Sie Unterstützung für die [Protokollrotation](#) für AWS ParallelCluster verwaltete Protokolle hinzu.
- Verfolgen Sie häufige Rechenknotenfehler und die längste Leerlaufzeit dynamischer Knoten im [CloudWatch Dashboard](#).
- Erzwingen Sie, dass der DCV Authenticator Server bei der Erstellung des SSL-Sockets mindestens TLS-1.2 das Protokoll verwendet.
- Installieren Sie das [NVIDIA Data Center GPU Manager \(DCGM\)](#) -Paket auf allen unterstützten Betriebssystemen außer und.  
aarch64 centos7  
alinux2
- Laden Sie standardmäßig das Kernelmodul [nvidia-uvm](#), um dem CUDA-Treiber Unified Virtual Memory (UVM) -Funktionalität zur Verfügung zu stellen.
- Installieren Sie den [NVIDIA Persistence Daemon](#) als Systemdienst.

## Änderungen:

- Upgrade Slurm auf Version 23.02.2 (von Version 22.05.8).
- Aktualisieren Sie munge auf Version 0.5.15 (von Version 0.5.14).
- Stellen Sie den Slurm `TreeWidth` auf 30 ein.
- Stellen Sie die epilog Konfigurationen Slurm `prolog /opt/slurm/etc/scripts/prolog.d/` und `/opt/slurm/etc/scripts/epilog.d/` jeweils auf das Zielverzeichnis ein.
- Stellen Sie Slurm `BatchStartTimeout` den Wert auf maximal 3 Minuten für die Ausführung von Prolog Skripten während der Registrierung des Rechenknotens ein.
- Erhöhen Sie die Standardinstellung für CloudWatch Protokolle `RetentionInDays` von 14 auf 180 Tage.
- Aktualisieren Sie das EFA-Installationsprogramm auf 1.22.1.
  - Dkms: 2.8.3-2



- EFA-Treiber: efa-2.1.1g (keine Änderung)
- EFA-Config: efa-config-1.13-1 (keine Änderung)
- EFA-Profil: efa-profile-1.5-1 (keine Änderung)
- libFabric-aws: (von) libfabric-aws-1.17.1-1 libfabric-aws-1.17.0-1
- RDMA-Core: (keine Änderung) rdma-core-43.0-1
- MPI öffnen: openmpi40-aws-4.1.5-1 (keine Änderung)
- Führen Sie ein Upgrade der Lustre-Client-Version 2.12 auf Amazon Linux 2 durch. Der Lustre-Client 2.12 wurde auf Ubuntu 20.04, 18.04 und CentOS >= 7.7 installiert.
- Aktualisieren Sie die Lustre-Client-Version 2.10.8 auf CentOS 7.6.
- Aktualisieren Sie den NVIDIA-Treiber auf Version 470.182.03 (von Version 470.141.03 ).
- Aktualisieren Sie den NVIDIA Fabric Manager auf

Version 470.182.03 (von Version 470.141.03 ).

- Aktualisieren Sie das NVIDIA CUDA Toolkit auf Version 11.8.0 (von Version 11.7.1).
- Aktualisieren Sie das NVIDIA CUDA-Beispiel auf Version. 11.8.0
- Aktualisieren Sie die Intel MPI-Bibliothek auf Version 2021 Update 9 (von Version 2021 Update 6). Weitere Informationen finden Sie unter [Intel® MPI Library 2021 Update 9](#).
- Aktualisieren Sie NICE DCV auf Version 2023.0-15022 (von Version 2022.2-14521 ).
  - Server: 2023.0.15022-1 (von Version 2022.2-14521-1 ).
  - xdcv: 2023.0.547-1 (aus Version 2022.2.519-1 ).
  - gl: 2023.0.1027-1 (aus Version 2022.2.1012-1 ).
  - web\_viewer: 2023.0.15022-1 (aus Version 2022.2.14521-1 ).

- Auf Version aktualisieren  
`erenaws-cfn-bootstrap .2.0-24`
- Aktualisieren Sie das Image, das von der CodeBuild Umgebung beim Erstellen von Container-Images für AWS Batch Cluster verwendet wird:
  - `aws/codebuild/amazonlinux2-x86_64-standard:4.0` (von `aws/codebuild/amazonlinux2-x86_64-standard:3.0` ).
  - `aws/codebuild/amazonlinux2-aarch64-standard:2.0` (von `aws/codebuild/amazonlinux2-aarch64-standard:1.0` ).

#### Fehlerbehebungen:

- Korrigieren Sie die Netzwerksicherheit sgruppensvalidatoren von Amazon EFS und Amazon FSx, um zu vermeiden, dass falsche Fehler gemeldet werden.
- Korrigiert das fehlende Tagging von Ressourcen, die von Image Builder

während des `build-image` Vorgangs erstellt wurden.

- Korrigieren Sie die Aktualisierungsrichtlinie `MaxCount`, um immer numerische Vergleiche für die `MaxCount` Immobilie durchzuführen.
- Korrigieren Sie die IP-Ausrichtung auf Rechenknoteninstanzen mit mehreren Netzwerkkarten.
- Korrigiert die Ersetzung von `StoragePass` `inslurm_parallelcluster_slurmdbd.conf`, wenn eine Aktualisierung der Warteschlangenparameter durchgeführt wird und die Slurm Konfigurationen nicht aktualisiert werden.
- Behebt ein Problem, das dazu führt, dass beim Erstellen eines Clusters mit einem vorhandenen EFS-Dateisystem fehlerhafte Sicherheitsgruppen erstellt werden.
- Behebt das Problem, dass der `cfn-hup` Daemon beim Neustart fehlschlägt.
- Betrachten Sie dynamische Knoten, die als Bootstrap-

Fehler gekennzeichnet sind, für den Slurm geschützt en Modus. INVALID\_REGION Statische Knoten, bei denen die Slurm Registrierung fehlschlägt, werden nach dem bereits als Bootstrap-Fehler behandelt . node\_replacement\_timeout

Einzelheiten zu den Änderungen finden Sie in den CHANGELOG Dateien für [aws-parallelcluster](#), [aws-parallelcluster-cookbook](#) und packages on. [aws-parallelcluster-node](#) GitHub

[AWS ParallelCluster UI-Version 2023.05.0 veröffentlicht](#)

AWS ParallelCluster UI-Version 2023.05.0 veröffentlicht.

16. Mai 2023

Verbesserungen:

- Fügen Sie ab AWS ParallelCluster Version 3.6.0 Unterstützung für RHEL 8 hinzu.
- Fügen Sie die Überwachung der Cluster-Kosten hinzu.
- Erhöhen Sie ab AWS ParallelCluster Version 3.6.0 die Quoten für Warteschlangen und Rechenressourcen.

Änderungen:

- Die Benutzeroberfläche des Assistenten zur Clustererstellung wurde verbessert.
- Die Geschwindigkeit der Bereitstellung der AWS ParallelCluster Benutzeroberfläche wurde erhöht.
- Die Oberfläche zum Hinzufügen eines neuen Benutzers wurde verbessert.
- Warteschlangen befinden sich standardmäßig im Hauptknoten-Subnetz.

Fehlerbehebungen:

- Wechseln Sie nach Abschluss der Clustererstellung zur richtigen Region.
- Korrigieren Sie die Anzeige der Ladeanzeige in der Funktion „Cluster bearbeiten“.
- Korrigiert die Clustererstellung, wenn die SnapshotId EBS-Eigenschaft entfernt wird.

Einzelheiten zu den Änderungen finden Sie in den CHANGELOG Dateien für das [aws-parallelcluster-ui](#) Paket unter GitHub.

[AWS ParallelCluster UI-Version 2023.04.0 veröffentlicht](#)

AWS ParallelCluster UI-Version 2023.04.0 veröffentlicht.

17. April 2023

Verbesserungen:

- Neugestaltung des Assistenten zur Clustererstellung.
- Neugestaltung der Seite mit Cluster-Protokollen.
- Fügen Sie eine benutzerdefinierte Namenseinstellung für gemeinsam genutzten Speicher hinzu.
- Fügen Sie beim Hinzufügen von Speicher zu einem Cluster mehrere Speicheroptionen hinzu.
- `DeletionPolicy` Unterstützung für Amazon EFS und FSx for Lustre hinzugefügt.
- `ImdsSupport` Einstellung zur Cluster-Konfiguration hinzufügen.
- Unterstützung für C7-Instanztypen hinzufügen.
- Tutorial [Zu einer früheren AWS Systems Manager-Dokumentversion zurückkehren](#) hinzugefügt.

Änderungen:



- Cluster-Konfiguration YAML mit einer Größe von bis zu 1 MB.
- Der Benutzer wurde aufgrund einer Autorisierung mit temporären Boto3 IAM-Anmeldeinformationen nicht abgemeldet.
- Multi-Threading-Optionen wurden deaktiviert, wenn eine HPC-Instanz ausgewählt wurde.
- Die Option „Rollback deaktivieren“ wurde auf der Cluster-Erstellungsseite entfernt.
- Der Benutzer kann die AWS ParallelCluster Benutzeroberfläche erst verwenden, wenn die erforderlichen Informationen bereitgestellt wurden.
- Es können bis zu 10 Warteschlangen hinzugefügt werden.
- Das SSM-SessionManagerRunShell Dokument wird bei der Installation der AWS ParallelCluster Benutzeroberfläche nicht überschrieben.

Fehlerbehebungen:

- Korrigiert den defekten Link zum Zurücksetzen des Passworts
- Fehler beheben, der `delete stack` dadurch verursacht wurde, dass er `EcrPrivateRepository` nicht leer war
- Das Initialisierungsproblem des Kästchens „SSH-Schlüssel generieren“ im Abschnitt „Eigenschaften für die Verwaltung mehrerer Benutzer“ wurde behoben.
- Absturz behoben, der durch einen Job mit undefinierten Eigenschaften verursacht wurde.
- Die SCRATCH FSx-Einstellungen wurden korrigiert.
- Die Schaltflächen zum Starten und Stoppen von Instanzen wurden behoben, die auch nach einmaligem Klicken immer noch aktiviert waren.

Einzelheiten zu den Änderungen finden Sie in den `CHANGELOG` Dateien für das [aws-parallelcluster-ui](#) Paket unter GitHub.

[AWS ParallelCluster Version  
3.5.1 veröffentlicht](#)

AWS ParallelCluster Version  
3.5.1 veröffentlicht.

29. März 2023

Verbesserungen:

- Fügen Sie eine eigenständige [ausführbare pcluster CLI-Installationsdatei](#) hinzu.

Änderungen:

- Aktualisieren Sie das EFA-Installationsprogramm auf 1.22.0.
  - EFA-Treiber: efa-2.1.1g (von efa-2.1.1-1)
  - EFA-Config: (aus efa-config-1.12-1 efa-config-1.13-1 )
  - efa-profile-1.5-1 EFA-Profil: (keine Änderung)
  - libFabric-aws: (von libfabric-aws-1.17.0-1 libfabric-aws-1.16.1amzn3.0-1)
  - RDMA-Core: (keine Änderung) rdma-core-43.0-1
  - MPI öffnen: openmpi40-aws-4.1.5-1 (von openmpi40-aws-4.1.4-3)

Aktualisieren Sie NICE DCV auf Version 2022.2-14521 .

- Server: 2022.2.14521-1
- xdcv: 2022.2.519-1
- gl: 2022.2.1012-1
- Webviewer: 2022.2.14521-1

Fehlerbehebungen:

- Beheben Sie potenzielle Fehler beim Starten von Knoten, die durch Musterabgleiche zwischen `MountDir` und `/etc/exports` beim Entfernen gemeinsam genutzter Amazon EBS-Volumes im Rahmen eines Cluster-Updates verursacht wurden.
- Korrektur, um zu verhindern, dass `compute_console_output` Protokolldateien bei jeder `clustermgmt` Iteration gekürzt werden.

Einzelheiten zu den Änderungen finden Sie in den `CHANGELOG` Dateien für [aws-parallelcluster](#), und [packages on. aws-parallelcluster](#)

[lelcluster-cookbookaws-para](#)

[lelcluster-node](#) GitHub

[AWS ParallelCluster Version  
3.5.0 veröffentlicht](#)

AWS ParallelCluster Version  
3.5.0 veröffentlicht.

20. Februar 2023

Verbesserungen:

- Greifen Sie über die [AWS ParallelCluster Benutzeroberfläche](#) auf Cluster zu und verwalten Sie sie.
- Fügen Sie versionierte AWS ParallelCluster Richtlinien zu einer CloudFormation Vorlage hinzu, auf die Sie in Ihren Workloads verweisen können.
- Fügen Sie eine AWS ParallelCluster Python-Bibliothek hinzu, die Sie mit Ihrem eigenen Code verwenden können.
- Fügen Sie Amazon die Protokollierung der Ausgabe der Compute-Knoten-Konsole bei CloudWatch einem Bootstrap-Fehler des Compute-Knotens hinzu.
- Fügen Sie ein Feld für Fehler hinzu, das den Fehlercode und den Grund für die `describe-cluster` Ausgabe enthält, wenn die Clustererstellung fehlschlägt.

- Fügen Sie Validatoren hinzu, um die böswillige Eingabe von Zeichenketten beim Aufrufen des Unterprozessmoduls zu verhindern.
- Schlägt die Clustererstellung fehl, wenn sich der Clusterstatus `PROTECTED` während der Bereitstellung statischer Knoten auf ändert.

#### Änderungen:

- Upgrade auf Slurm Version `22.05.8` (von Version `22.05.7`)
- Aktualisieren Sie das EFA-Installationsprogramm auf `1.21.0`.
  - EFA-Treiber: `efa-2.1.1-1` (von `efa-2.1`)
  - EFA-Config: (aus `efa-config-1.11-1` `efa-config-1.12-1` )
  - `efa-profile-1.5-1` EFA-Profil: (keine Änderung)
  - `libfabric-aws`: (von) `libfabric-aws-1.16.1amzn3.0-1` `libfabric-aws-1.16.1`

- RDMA-Core: (von) `rdma-core-43.0-1` `rdma-core-43.0-2`
- MPI öffnen: `openmpi40-aws-4.1.4-3` (keine Änderung)
- Machen Sie die Slurm Controller-Protokolle ausführlicher und aktivieren Sie zusätzliche Protokollierung für das Slurm Energiespar-Plugin.

#### Fehlerbehebungen:

- Korrigieren Sie die Erstellung von Cluster-Datenbanken, indem Sie sicherstellen, dass der Clustername nicht länger als 40 Zeichen ist, wenn Slurm Accounting aktiviert ist.
- Behebt ein Problem, das dazu führte, `clustermgtd` dass Rechenknoten, die neu gestartet wurden, ersetzt wurdenSlurm, wenn die EC2-Instanzstatusprüfungen fehlschlagen.
- Behebung eines Problems, das aufgrund einer falschen IAM-Richtlinie auf dem Hauptknoten verhinderte, dass Rechenknoten, deren Kapazitätsreserve



rungen von anderen Konten  
gemeinsam genutzt wurden,  
gestartet werden konnten.

Einzelheiten zu den  
Änderungen finden Sie in  
den CHANGELOG Dateien für  
[aws-parallelcluster](#), [aws-parallelcluster-cookbook](#), [aws-parallelcluster-node](#) und packages  
auf [aws-parallelcluster-ui](#)  
GitHub

[AWS ParallelCluster Version  
3.4.1 veröffentlicht](#)

AWS ParallelCluster Version  
3.4.1 veröffentlicht.

13. Januar 2023

Fehlerbehebungen:

- Behebt ein Slurm Scheduler -Problem, das zur falschen Anwendung von Updates in der internen Registrierung der Rechenknoten führen konnte. Als Folge dieses Problems könnten EC2-Instances nicht mehr verfügbar sein oder durch einen falschen Instance-Typ gesichert werden.

Einzelheiten zu den Änderungen finden Sie in den CHANGELOG Dateien für den [aws-parallelcluster](#) und in den Paketen unter. [aws-parallelcluster-cookbookaws-parallelcluster-node](#) [GitHub](#)

[AWS ParallelCluster Version  
3.4.0 veröffentlicht](#)

AWS ParallelCluster Version  
3.4.0 veröffentlicht.

22. Dezember 2022

Verbesserungen:

- Fügen Sie Unterstützung für das Starten von Knoten in mehreren Verfügbarkeitszonen hinzu, um die Kapazitätsverfügbarkeit zu erhöhen.
- Fügen Sie Unterstützung für die Angabe mehrerer Subnetze für jede Warteschlange hinzu, um die Kapazitätsverfügbarkeit zu erhöhen.
- Fügen Sie einen neuen Konfigurationsparameter in [Iam](#)/hinzu [ResourcePrefix](#) , um ein Präfix für den Pfad und den Namen der IAM-Ressourcen anzugeben, die von erstellt wurden. AWS ParallelCluster
- Fügen Sie den neuen Konfigurationsabschnitt [DeploymentSettings](#) /hinzu, [LambdaFunctionsVpcConfig](#) um die von AWS ParallelCluster Lambda-Funktionen

verwendete VPC-Konfiguration anzugeben.

- Fügen Sie die Möglichkeit hinzu, ein benutzerdefiniertes Skript anzugeben, das während eines Cluster-Updates im Hauptknoten ausgeführt werden soll. Das Skript kann mit [HeadNode/CustomActions](#) angegeben werden, wenn es Slurm als Scheduler verwendet wird.

#### Änderungen:

- Entfernen Sie die Erstellung von Amazon EFS-Mount-Zielen für bestehende Dateisysteme.
- Mounten Sie EFS-Dateisysteme mit `amazon-efs-utils`. EFS-Dateisysteme können mithilfe von Übertragungsverschlüsselung und einem autorisierten IAM-Benutzer bereitgestellt werden.
- Installieren Sie `stunnel` 5.67 auf CentOS7 und Ubuntu, um die EFS-Verschlüsselung bei der Übertragung zu unterstützen.

- Aktualisieren Sie das EFA-Installationsprogramm auf (von) 1.20.0 1.18.0
  - EFA-Treiber: efa-2.1 (von) efa-1.16.0-1
  - EFA-Config: efa-config-1.11-1 (keine Änderung)
  - EFA-Profil: efa-profile-1.5-1 (keine Änderung)
  - libFabric-aws: (von) libfabric-aws-1.16.1 libfabric-aws-1.16.0~amzn4.0-1
  - RDMA-Kern: von () rdma-core-43.0-2 rdma-core-41.0-2
  - Öffne MPI: openmpi40-aws-4.1.4-3 von () openmpi40-aws-4.1.4-2
- Führen Sie ein Slurm Upgrade auf Version 22.05.7 (von 22.05.5) durch.
- Aktualisieren Sie Python auf 3.9.16 und 3.7.16. (von 3.9.15 und 3.7.13).
- Mit Slurm 22.05.7 werden dynamische Knoten im IDLE+CLOUD+COMPLETING+POWER\_DOWN

+NOT\_RESPONDING  
Status nicht als fehlerhaft  
angesehen.

Einzelheiten zu den  
Änderungen finden Sie in den  
CHANGELOG Dateien für den  
[aws-parallelcluster](#) und in  
den Paketen unter [aws-parallelcluster-cookbook](#)  
[aws-parallelcluster-node](#) GitHub

[AWS ParallelCluster Version  
3.3.1 veröffentlicht](#)

AWS ParallelCluster Version  
3.3.1 veröffentlicht.

2. Dezember 2022

Änderungen:

- Offizielle AWS ParallelCluster Produkt-AMIs sind jetzt verfügbar, nachdem Amazon EC2 nach zwei Jahren nicht mehr unterstützt wurde.
- Erhöhen Sie die Speichergrenze der AWS ParallelCluster API Lambda auf 2048, um Kaltstart-Strafen zu reduzieren und Timeouts zu vermeiden.

Fehlerbehebungen:

- Vermeiden Sie den Austausch von verwalteten FSx for Lustre-Dateisystemen und den Verlust von Daten bei Cluster-Updates, die Änderungen an der Subnetz-ID der Compute-Flotte beinhalten.
- [SharedStorage DeletionPolicy](#) gilt für Cluster-Aktualisierungsaktionen.

Einzelheiten zu den Änderungen finden Sie in der CHANGELOG Datei für das [aws-parallelcluster-Paket](#) unter. GitHub

[AWS ParallelCluster nur Dokumentation, hpc6id, Hinweis](#)

AWS ParallelCluster Update nur zur Dokumentation

2. Dezember 2022

- AWS ParallelCluster unterstützt den Instanztyp hpc6id für die Einstellung/nicht. [HeadNodeInstanceType](#)



[AWS ParallelCluster Version  
3.1.5 veröffentlicht](#)

AWS ParallelCluster Version  
3.1.5 veröffentlicht.

16. November 2022

Verbesserungen:

- Behebt das Slurm-Problem, das die Kündigung inaktiver Knoten verhindert.
- Aktualisieren Sie das EFA-Installationsprogramm auf 1.18.0
  - EFA-Treiber: efa-1.16.0-1
  - EFA-Config: (von) efa-config-1.11-1 efa-config-1.9-1
  - EFA-Profil: efa-profile-1.5-1 (keine Änderung)
  - libFabric-aws: (von) libfabric-aws-1.16.0~amzn4.0-1 libfabric-1.13.2
  - RDMA-Core: (von) rdma-core-41.0-2 rdma-core-37.0
  - MPI öffnen: openmpi40-aws-4.1.4-2 (von) openmpi40-aws-4.1.1-2

Änderungen:

- `lambda:ListTags` Fügt und `lambda:UntagResource` zu dem von der AWS ParallelCluster API `ParallelClusterUserRole` verwendeten Stack für ein Cluster-Update hinzu.
- Aktualisieren Sie die Intel MPI Library auf Version 2021 Update 6 (von Version 2021 Update 4). Weitere Informationen finden Sie unter [Intel® MPI Library 2021 Update 6](#).
- Aktualisieren Sie den NVIDIA-Treiber auf Version 470.141.03 (von 470.103.01).
- Aktualisieren Sie NVIDIA Fabric Manager auf Version 470.141.03 (von 470.103.01).

[Einzelheiten zu den Änderungen finden Sie in den CHANGELOG Dateien für den `aws-parallelcluster` und in den Paketen unter `aws-parallelcluster-cookbookaws-parallelcluster-node` GitHub](#)

[AWS ParallelCluster Version  
3.3.0 veröffentlicht](#)

AWS ParallelCluster Version  
3.3.0 veröffentlicht.

02. November 2022

Verbesserungen:

- Unterstützung für die Konfiguration der Zuweisung mehrerer Instanzen für eine Rechenressource bei Verwendung Slurm als Scheduler hinzugefügt. Weitere Informationen finden Sie unter [Zuweisung mehrerer Instanztypen mit Slurm](#).
- Unterstützung für das Hinzufügen und Entfernen [SharedStorage](#) mit einem Cluster-Update unter Verwendung einer aktualisierten Konfiguration hinzugefügt. Weitere Informationen finden Sie unter [Gemeinsamer Speicher](#).
- Fügen Sie neue Konfigurationsparameter [Efs](#) und Einstellungen [DeletionPolicy](#) für [FsxLustre](#) [\\_gemeinsam genutzten Speicher](#) hinzu, um die Aufbewahrung von Speicherplatz zu unterstützen.

- Unterstützung für Slurm Accounting mit dem neuen Konfigurationsparameter [Scheduling /SlurmSettings](#) /hinzugefügt [Database](#). Weitere Informationen finden Sie unter [SlurmAbrechnung mit AWS ParallelCluster](#).
- Unterstützung für Kapazitätsreservierungen auf Abruf und Ressourcengruppen für Kapazitätsreservierungen hinzugefügt. Weitere Informationen finden Sie unter [Erstellen Sie Instances mit ODCR Starten Sie Instances](#).
- Fügen Sie einen neuen Konfigurationsparameter hinzu, um die IMDS-Version anzugeben, die in einem Cluster oder einer Build-Image-Infrastruktur in den Cluster-Konfigurationen [Imds/ImdsSupport](#) und Build [ImdsSupport](#) , [Imds](#)/unterstützt werden soll.
- Fügen Sie Unterstützung für [Networking /Placement Group](#) im [ComputeResources](#) Abschnitt [SlurmQueues](#) /hinzu.
- Fügen Sie Unterstützung für Instances mit mehreren

Netzwerkschnittstellen hinzu, die auf nur eine ENI pro Gerät beschränkt sind.

- Verbessern Sie die Netzwerkvalidierung für externe Amazon EFS-Dateisysteme, indem Sie den CIDR-Block in der angehängten Sicherheitsgruppe überprüfen.
- Fügen Sie einen Validator hinzu, um zu überprüfen, ob konfigurierte Instance-Typen Platzierungsgruppen unterstützen.
- Konfigurieren Sie NFS-Threads auf  $\min(256, \max(8, \text{num\_cores} * 4))$ , um eine bessere Stabilität und Leistung zu gewährleisten.
- Verschieben Sie die NFS-Installation zur Build-Zeit, um die Konfigurationszeit zu reduzieren.
- Aktivieren Sie die serverseitige Verschlüsselung für das EcrImageBuilder SNS-Thema, das bei der Bereitstellung der AWS ParallelCluster API erstellt wird und zur Benachrichtigung bei Docker-Image-Build-Ereignissen verwendet wird.

## Änderungen:

- Ändert das Verhalten von [SlurmQueues /Networking /PlacementGroup /Enabled](#). Es erstellt jetzt eine eindeutige verwaltete Platzierungsgruppe für jede Rechenressource anstelle einer einzigen verwalteten Platzierungsgruppe für alle Rechenressourcen.
- Fügen Sie Unterstützung für [SlurmQueues /Networking /PlacementGroup /Nameals](#) bevorzugte Benennungsmethode hinzu.
- Verschieben Sie die Kopfknoten-Tags von Launch Template in die Instanzdefinition, um zu vermeiden, dass Head-Knoten-Tags bei Tag-Updates ersetzt werden.
- Deaktivieren Sie Multithreading über ein Skript, das von der Startvorlage ausgeführt wird `cloud-init` und nicht über das in der Startvorlage `CpuOptions` festgelegte Skript.

- Aktualisieren Sie Python auf Version 3.9 und NodeJS auf Version 16 in der API-Infrastruktur, dem API-Docker-Container und den Cluster-Lambda-Ressourcen.
- Entfernen Sie die Unterstützung für Python 3.6 in `aws-parallelcluster-batch-cli`.
- Führen Sie ein Slurm Upgrade auf Version 22.05.5 (von 21.08.8-2) durch.
- Aktualisieren Sie den NVIDIA-Treiber auf Version 470.141.03 (von 470.129.06).
- Aktualisieren Sie NVIDIA Fabric Manager auf Version 470.141.03 (von 470.129.06).
- Aktualisieren Sie das NVIDIA CUDA Toolkit auf Version 11.7.1 (von 11.4.4).
- Aktualisieren Sie Python, das in AWS ParallelCluster Virtualenvs verwendet wird, von 3.7.13 auf 3.9.15.
- Aktualisieren Sie das EFA-Installationsprogramm auf Version 1.18.0.

- EFA-Treiber: (keine Änderung) efa-1.16.0-1
- EFA-Konfiguration: () efa-config-1.11-1 from efa-config-1.10-1
- EFA-Profil: efa-profile-1.5-1 (keine Änderung)
- libFabric-aws: (von) libfabric-aws-1.16.0~amzn4.0-1 libfabric-aws-1.16.0~amzn2.0-1
- RDMA-Core: (von) rdma-core-41.0-2 rdma-core-37.0
- MPI öffnen: openmpi40-aws-4.1.4-2 (von) openmpi40-aws-4.1.1-2
- Aktualisieren Sie NICE DCV auf Version 2022.1-13300 (von 2022.0-12760 ).
- Aktiviert die Unterdrückung von SingleSubnetValidator für Queues.
- Ersetzen Sie keine DRAIN Knoten, wenn sie sich im COMPLETING Status



befinden, da Epilog  
möglicherweise noch läuft.

#### Fehlerbehebungen:

- Korrigiert, dass die Überprüfung des Filterparameters im AWS ParallelCluster `ListClusterLogStreams` Befehl fehlschlägt, wenn falsche Filter übergeben werden.
- Korrigiert die Validierung des Parameters [SharedStorage /EfsSettings](#), sodass die Validierung fehlschlägt, wenn `FileSystemId` er zusammen mit anderen [SharedStorage EfsSettings](#)/-Parametern angegeben wurde. `FileSystemId` War bisher nicht enthalten.
- Das Cluster-Update wurde behoben, wenn die Reihenfolge [SharedStorage](#) zusammen mit anderen Änderungen in der Konfiguration geändert wurde.
- Fehler `UpdateParallelClusterLambdaRole` in der AWS ParallelCluster API behoben, in die Logs hochgeladen werden sollen CloudWatch.

- Fix, dass Cinc bei der Installation von Paketen vor der Ausführung von Kochbüchern nicht das lokale CA-Zertifikatspaket verwendet.
- Behebt ein Problem beim Upgraden von Ubuntu mit dem `pcluster build-image` eingestellten `ZeitpunktBuild:UpdateOsPackages:Enabled:true`.
- Das Parsen der YAML-Cluster-Konfiguration wurde behoben, indem es bei doppelten Schlüsseln fehlschlug.

Einzelheiten zu den Änderungen finden Sie in den `CHANGELOG` Dateien für den [aws-parallelcluster](#) und in den Paketen unter [aws-parallelcluster-cookbookaws-parallelcluster-node](#) [GitHub](#)

[AWS ParallelCluster Nur Dokumentation, API-Referenz hinzugefügt.](#)

AWS ParallelCluster Update nur zur Dokumentation

27. Oktober 2022

- Version 3 [AWS ParallelCluster-API-Referenz](#) wurde der Dokumentation hinzugefügt.

[AWS ParallelCluster Version  
3.2.1 veröffentlicht](#)

AWS ParallelCluster Version  
3.2.1 veröffentlicht.

3. Oktober 2022

Verbesserungen:

- Verbessern Sie die Logik, um die Host-Routing-Tabellen den verschiedenen Netzwerkkarten zuzuordnen, um EC2-Instances mit mehreren NICs besser zu unterstützen.

Änderungen:

- Aktualisieren Sie den NVIDIA-Treiber auf Version 470.141.03.
- Aktualisieren Sie NVIDIA Fabric Manager auf Version 470.141.03.
- Deaktivieren Sie die cron-Jobaufgaben `man-db` und `mdmlocate`, was sich negativ auf die Knotenleistung auswirken kann.
- Aktualisieren Sie die Intel MPI Library auf 2021.6.0.602.
- Führen Sie als Reaktion auf dieses Sicherheitsrisiko ein Upgrade von Python von 3.7.10 auf 3.7.13 durch.

## Fehlerbehebungen:

- Vermeiden Sie Fehler `DescribeCluster` beim Einschalten, wenn die Clusterkonfiguration nicht verfügbar ist.

Einzelheiten zu den Änderungen finden Sie in den CHANGELOG Dateien für den [aws-parallelcluster](#) und den Paketen [aws-parallelcluster-cookbook](#) unter [aws-parallelcluster-node](#) [GitHub](#)

[AWS ParallelCluster Version  
3.2.0 veröffentlicht](#)

AWS ParallelCluster Version  
3.2.0 veröffentlicht.

27. Juli 2022

Verbesserungen:

- Unterstützung für [speicherbasierte Planung](#) in Slurm hinzugefügt.
  - Konfigurieren Sie den realen Speicher der Rechenknoten in der Slurm-Cluster-Konfiguration.
  - Fügen Sie den neuen Konfigurationsparameter [Scheduling /SlurmSettings / EnableMemoryBasedScheduling](#) , um die speicherbasierte Planung in Slurm zu aktivieren.
  - Fügen Sie den neuen Konfigurationsparameter [Scheduling /SlurmQueues /ComputeResources /hinzu SchedulableMemory](#) , um den Standardwert des Speichers zu überschreiben, den der Scheduler auf Rechenknoten erkennt.

- Verbessern Sie die Flexibilität bei Aktualisierungen der Cluster-Konfiguration, um das Stoppen und Starten des gesamten Clusters nach Möglichkeit zu vermeiden. Fügen Sie den neuen Konfigurationsparameter [Scheduling /SlurmSettings /hinzu](#) [QueueUpdateStrategy](#), um die bevorzugte Strategie festzulegen, die verwendet werden soll, wenn Rechenknoten ein Konfigurationsupdate und einen Austausch benötigen.
- Verbessern Sie den Failover-Mechanismus für verfügbare Rechenressourcen, wenn bei EC2-Instances Probleme mit unzureichender Kapazität auftreten. [Deaktivieren Sie Rechenknoten für einen konfigurierbaren Zeitraum, wenn der](#) Start eines Knotens aufgrund unzureichender Kapazität fehlschlägt.
- Unterstützung für das Mounten vorhandener Systeme [FSx for ONTAP](#) und [FSx for OpenZFS](#) Dateisysteme hinzufügen.

- Unterstützung für das Mounten mehrerer Instanzen vorhandener [Amazon Elastic File Systems](#)-, [FSx for Lustre](#)-, [FSx for ONTAP](#)- und [FSx for OpenZFS](#)-Dateisysteme hinzugefügt.
- Fügen Sie Unterstützung für den [Bereitstellungstyp FSx for Lustre Persistent\\_2](#) hinzu, wenn Sie ein neues [Dateisystem](#) erstellen.
- Fordert den Benutzer auf, EFA für unterstützte Instanztypen zu aktivieren, wenn er den Assistenten verwendet. [pcluster configure](#)
- Unterstützung für den Neustart von Rechenknoten mit Slurm hinzugefügt.
- Verbessern Sie den Umgang mit Slurm-Betriebeszuständen, um auch das manuelle Herunterfahren von Knoten zu berücksichtigen.
- Installieren Sie NVIDIA GDRCopy 2.3 in den AMIs des Produkts, um das Kopieren von GPU-Speichern mit niedriger Latenz zu ermöglichen.

## Änderungen:

- Aktualisieren Sie das EFA-Installationsprogramm auf Version 1.17.2.
  - EFA-Treiber: efa-1.16.0-1
  - EFA-Konfiguration: efa-config-1.10-1
  - EFA-Profil: efa-profile-1.5-1
  - Libfabric: libfabric-aws-1.16.0~amzn2.0-1
  - RDMA-Kern: rdma-core-41.0-2
  - MPI öffnen: openmpi40-aws-4.1.4-2
- Aktualisieren Sie NICE DCV auf Version 2022.0-12760.
- Aktualisieren Sie den NVIDIA-Treiber auf Version 470.129.06.
- Aktualisieren Sie NVIDIA Fabric Manager auf Version 470.129.06.
- Ändern Sie die standardmäßigsten EBS-Volumetypen sowohl im Stammvolume als auch in den zusätzlichen Volumes von gp2 auf gp3.
- Änderungen an den Dateisystemen FSx for



Lustre, erstellt von: AWS  
ParallelCluster

- Ändern Sie den Standard-Bereitstellungstyp in. `Scratch_2`
- Ändern Sie die Lustre-Serverversion auf `2.12`.
- Erfordert nicht, dass `PlacementGroup /Enabled` auf `true` gesetzt ist, wenn ein vorhandenes `PlacementGroup /Id` übergeben wird.
- Erlaubt nicht, `PlacementGroup /zu setzenId`, wenn `PlacementGroup /Enabled` explizit auf `false` gesetzt ist.
- Füge allen Ressourcen, die von `parallelcluster:create-cluster-name` erstellt wurden, ein `Tag` hinzu `parallelcluster:cluster-name` AWS ParallelCluster.
- Fügt `lambda:UpdateResource` dem AWS ParallelCluster API-Stack für das Cluster-Update hinzu `lambda:ListTags` und `ParallelClusterUserRole` wird vom API-Stack verwendet.
- Beschränken Sie den IMDS IPv6-Zugriff nur auf Root- und Cluster-Admin-Benutzer, wenn

- der Konfigurationsparameter `meterHeadNode/Imds/aktiviert Secured` ist.
- Verwenden Sie bei einem benutzerdefinierten AMI die Größe des AMI-Root-Volumens anstelle der ParallelCluster Standardgröße von 35 GiB. Der Wert kann in der Cluster-Konfigurationsdatei geändert werden.
  - Automatische Deaktivierung der Rechenflotte, wenn der Konfigurationsparameter `Scheduling //SlurmQueues ComputeResources /` unter dem erforderlichen Mindestpreis für die Erfüllung von Spot-Anfragen `SpotPrice` liegt.
  - Zeigt `requested_value` `current_value` Werte im Änderungssatz an, wenn während einer Aktualisierung ein Abschnitt hinzugefügt oder entfernt wird.
  - Deaktivieren Sie den `aws-ubuntu-eni-helper` Dienst, der in Deep Learning-AMIs verfügbar ist, um Konflikte `configure_nw_interface.sh` bei der Konfigura

tion von Instances mit mehreren Netzwerkkarten zu vermeiden.

- Entfernen Sie die Unterstützung für Python 3.6.
- Setzen Sie MTU für alle Netzwerkschnittstellen auf 9001, wenn Sie Instanzen mit mehreren Netzwerkkarten konfigurieren.
- Entfernen Sie den letzten Punkt, wenn Sie den FQDN des Compute-Knotens konfigurieren.
- Verwalten Sie statische Knoten in. `POWERING_DOWN`
- Ersetzt den dynamischen Knoten nicht, `POWER_DOWN` da Jobs möglicherweise noch ausgeführt werden.
- Neustart `clustermgtd` und `slurmctld` Daemons bei der Cluster-Aktualisierung nur dann, wenn die `Scheduling` Parameter in der Clusterkonfiguration aktualisiert werden.
- Update `slurmctld` - und `slurmd` `systemd` Servicedateien.
- Beschränken Sie den IPv6-Zugriff auf IMDS nur auf Root- und Cluster-Admin-Benutzer, wenn

der Konfigurationsparameter `HeadNode/Imds/aktiviertSecured` ist.

- Stellen Sie die Slurm-Konfiguration so ein `AuthInfo=cred_expire=70`, dass die Zeit in der Warteschlange reduziert wird. Jobs müssen warten, bevor sie erneut gestartet werden, wenn Knoten nicht verfügbar sind.
- Aktualisieren Sie die Abhängigkeiten von Drittanbieter-Kochbüchern:
  - `apt-7.4.2` (von `apt-7.4.0`)
  - `Line-4.5.2` (von `Line-4.0.1`)
  - `openssh-2.10.3` (von `openssh-2.9.1`)
  - `pyenv-3.5.1` (von `pyenv-3.4.2`)
  - `selinux-6.0.4` (von `selinux-3.1.1`)
  - `yum-7.4.0` (von `yum-6.1.1`)
  - `yum-epel-4.5.0` (von `yum-epel-4.1.2`)

Fehlerbehebungen:

- Korrigieren Sie das Standardverhalten, bei dem die AWS ParallelCluster Validierungs- und Testschritte beim Erstellen eines

- benutzerdefinierten AMI übersprungen werden.
- Behebt ein Leck im Datei-Handlecomputemgtd .
  - Behebung einer Race-Bedingung, die sporadisch dazu führte, dass gestartete Instances sofort beendet wurden, weil sie in der DescribeInstances EC2-Antwort noch nicht verfügbar waren.
  - Korrigiert die Unterstützung für den DisableSimultaneousMultithreading Parameter bei Instance-Typen mit ARM-Prozessoren.
  - Behebung eines Fehlers beim AWS ParallelCluster API-Stack-Update beim Upgrade von einer früheren Version. Fügen Sie das für die ListImagePipelineImages Aktion verwendete ResourcePattern in hinzuEcrImageDeletionLambdaRole .
  - Die AWS ParallelCluster API hat fehlende Berechtigungen für den Import oder Export aus Amazon S3 beim Erstellen eines FSx for Lustre-Dateisystems

hinzugefügt. Dieser Fehler wurde behoben.

Einzelheiten zu den Änderungen finden Sie in den CHANGELOG Dateien für den [aws-parallelcluster](#) und in den Paketen unter. [aws-parallelcluster-cookbookaws-parallelcluster-node](#) [GitHub](#)

[AWS ParallelCluster Bisher wurden nur Dokumentationsaktualisierungen in diesem Jahr aktualisiert](#)

AWS ParallelCluster Updates nur für die Dokumentation.

6. Juli 2022

Neue Abschnitte:

- [Best Practices: BudgetwarnungenV3](#)
- [Bewährte Methoden: Verschieben eines Clusters auf eine neueAWS ParallelCluster Minor- oder Patch-VersionV3](#)
- [Arbeiten mit Amazon S3V3](#)
- [Arbeiten mit Spot-InstancesV3](#)
- [SlurmCluster-geschützter ModusV3](#)
- [AWS ParallelClusterRessourcen und TaggingV3](#)
- [CloudWatchAmazon-DashboardV3](#)
- [Integration mit Amazon CloudWatch LogsV3](#)
- [Elastic Fabric AdapterV3](#)
- [AWS ParallelCluster AMI-AnpassungV3](#)
- [Erstellen Sie Instances mit ODCR Starten Sie InstancesV3](#)
- [AMI-Patching und Austausch von EC2-InstancesV3](#)

- [Funktionsweise von AWS ParallelCluster V3](#)
- [Konfiguration der Verschlüsselung von gemeinsam genutztem Speicher mit einem AWS KMS Schlüssel V3](#)
- [Ausführung von Jobs in einem Cluster mit mehreren Warteschlangen V3](#)
- [Verwenden der AWS ParallelCluster-API V3](#)

Aktualisierungen des Abschnitts:

- [Best Practices: Netzwerkeistung V3](#): Es wurden bewährte Methoden für die Verwendung von Elastic Fabric Adaptor hinzugefügt.
- [AWS Identity and Access Management Berechtigungen in AWS ParallelCluster V3](#): Verschiedene Updates und hinzugefügt [Zusätzliche AWS ParallelCluster pcluster Benutzerrichtlinie bei der Verwendung von Amazon FSx for Lustre](#).
- [Fehlerbehebung für AWS ParallelCluster V3](#): Verschiedene Updates.



[AWS ParallelCluster Version  
3.1.4 veröffentlicht](#)

AWS ParallelCluster Version  
3.1.4 veröffentlicht.

16. Mai 2022

Verbesserungen:

- Fügen Sie die Validierung für [Directory Service / PasswordSecretArn](#) hinzu, damit sie fehlschlägt, wenn das Geheimnis nicht existiert.

Unterstützung für die Aktivierung der JWT-Authentifizierung (Slurm) hinzugefügt.

Änderungen:

- Aktualisieren Sie Slurm auf Version 21.08.8-2.
- Erstellen Sie Slurm mit JWT-Unterstützung.
- Erfordert nicht, dass [Placement Group / Enabled](#) auf gesetzt wird, `true` wenn ein PlacementGroup vorhanden/übergeben wird. `Id`
- `ParallelClusterUserRole` Wird vom ParallelCluster API-Stack für die Clustererstellung und

Image-Erstellung verwendet.  
Lambda: TagResource

Fehlerbehebungen:

- Korrigiert die Fähigkeit, die Logs eines Clusters zu exportieren, wenn der `export-cluster-logs` Befehl mit der `--filters` Option verwendet wird.
- Der AWS Batch Docker-Einstiegspunkt wurde so korrigiert, dass er ein `/home` gemeinsames Verzeichnis verwendet, um die parallele Ausführung von Jobs auf mehreren Knoten zu koordinieren.
- Setzen Sie die Knotenadresse zurück, wenn Sie Slurm Unhealthy static node auf down setzen, um zu verhindern, dass statische Knoten, die aufgrund unzureichender Kapazität ausgefallen sind, als Bootstrap-Ausfallknoten behandelt werden.

Einzelheiten zu den Änderungen finden Sie in den CHANGELOG Dateien für den [aws-parallelcluster](#) und in den Paketen unter [aws-paral](#)

[lelcluster-cookbookaws-para](#)

[lelcluster-node](#) GitHub

[AWS ParallelCluster Version  
3.1.3 veröffentlicht](#)

AWS ParallelCluster Version  
3.1.3 veröffentlicht.

20. April 2022

Verbesserungen:

- Führen Sie die SSH-Schlüsselerstellung zusammen mit der Erstellung des HOME-Verzeichnisses aus, z. B. während der SSH-Anmeldung, wenn Sie zu einem anderen Benutzer wechseln und wenn Sie einen Befehl als ein anderer Benutzer ausführen.
- Fügen Sie im Konfigurationsparameter/Unterstützung für FQDN und LDAP Distinguished Names hinzu. [Directory Service DomainName](#)  
Der neue Validator überprüft jetzt beide Syntaxen.
- Das neue `update_directory_service_password.sh` Skript, das auf dem Hauptknoten bereitgestellt wird, unterstützt die manuelle Aktualisierung des Active Directory-Passworts in der SSSD-Konfiguration. Das Passwort wird vom AWS Secrets Manager AS aus

der Clusterkonfiguration abgerufen.

- Fügen Sie Unterstützung für die Bereitstellung der API-Infrastruktur in Umgebungen ohne Standard-VPC hinzu.

#### Änderungen:

- Deaktiviert tiefere C-States in offiziellen x86\_64-AMIs und AMIs, die per `build-image` Befehl erstellt wurden, um eine hohe Leistung und geringe Latenz zu gewährleisten.
- Aktualisierungen von Betriebssystempaketen und Sicherheitskorrekturen.
- Ändern Sie die Amazon Linux 2-Basis-Images, um AMIs mit Kernel 5.10 zu verwenden.

#### Fehlerbehebungen:

- Korrigieren Sie den Build-Image-Stack `DELETE_FAILED` nach erfolgreicher Image-Erstellung aufgrund neuer EC2 Image Builder Builder-Richtlinien.
- Korrigieren Sie die SSSD-Eigenschaft des Konfigurationsparameters [Directory](#)

[Service //DomainAdd](#)  
[r](#) conversion to ldap\_uri,  
wenn sie mehrere  
Domainadressen enthält.

Einzelheiten zu den  
Änderungen finden Sie in den  
CHANGELOG Dateien für den  
[aws-parallelcluster](#) und in  
den Paketen unter. [aws-parallelcluster-cookbook](#) [GitHub](#)

[AWS ParallelCluster Version  
3.1.2 veröffentlicht](#)

AWS ParallelCluster Version  
3.1.2 veröffentlicht.

2. März 2022

Änderungen:

- Aktualisieren Sie Slurm auf Version 21.08.6 (von 21.08.5).

Fehlerbehebungen:

- Korrigieren Sie das Update der `/etc/hosts` Datei auf Rechenknoten, wenn ein Cluster in Subnetzen ohne Internetzugang bereitgestellt wird.
- Korrigieren Sie den Bootstrap der Rechenknoten so, dass sie auf die Initialisierung kurzlebiger Laufwerke warten, bevor sie dem Cluster beitreten.

[Einzelheiten zu den Änderungen finden Sie in den CHANGELOG Dateien für das aws-parallelcluster-Paket unter GitHub](#)

[AWS ParallelCluster Version  
3.1.1 veröffentlicht](#)

AWS ParallelCluster Version  
3.1.1 veröffentlicht.

10. Februar 2022

- Fügen Sie Unterstützung für Clusterumgebungen [mit mehreren Benutzern hinzu, indem Sie Active Directory-Domänen \(AD\) integrieren](#), die über AWS Directory Service verwaltet werden.
- Fügen Sie Unterstützung für [UseEc2Hostnames](#) in der Cluster-Konfigurationsdatei hinzu. Wenn auf true gesetzt, verwenden Sie EC2-Standard-Hostnamen (z. B. ip-1-2-3-4) für Rechenknoten.
- Fügen Sie Unterstützung für die Clustererstellung in [Subnetzen](#) ohne Internetzugang hinzu.
- Fügen Sie Unterstützung für mehrere Compute-Instanztypen pro Warteschlange hinzu.
- Fügen Sie Unterstützung für GPU-Planung mit Slurm auf ARM-Instanzen mit NVIDIA-Karten hinzu.
- Fügen Sie der AWS ParallelCluster CLI abgekürzte Flags für `cluster-name region`



(-n-r), `image-id` (-i), `()` und `cluster-configuration` /`image-configuration` (-c) hinzu.

- Unterstützung für die `NEW_CHANGED_DELETED` Option für den Parameter FSx for Lustre [AutoImportPolicy](#) hinzugefügt.
- Fügen Sie den `LaunchTemplates` EC2-Ressourcen, die von Rechenknoten verwendet werden, ein `parallelcluster:compute-resource-name` Tag hinzu.
- Verbessern Sie die innerhalb des Clusters erstellten Sicherheitsgruppen, um eingehende Verbindungen von benutzerdefinierten Sicherheitsgruppen zu ermöglichen, wenn `SecurityGroups` Parameter für einige Hauptknoten und/oder Warteschlangen angegeben werden.
- Installieren Sie die NVIDIA-Treiber und die CUDA-Bibliothek für ARM.

Änderungen:

- Aktualisieren Sie Slurm auf Version 21.08.5 (von 20.11.8).
- Aktualisieren Sie das Slurm-Plugin auf Version 21.08 (von 20.11).
- Aktualisieren Sie NICE DCV auf Version 2021.3-11591 (von 2021.1-10851).
- Aktualisieren Sie den NVIDIA-Treiber auf Version 470.103.01 (von 470.57.02).
- Aktualisieren Sie den NVIDIA Fabric Manager auf Version 470.103.01 (von 470.57.02).
- Aktualisieren Sie CUDA auf Version 11.4.4 (von 11.4.0).
- [Intel MPI](#) wurde auf Version 2021 Update 4 aktualisiert (aktualisiert von Version 2019 Update 8). Weitere Informationen finden Sie unter [Intel® MPI Library 2021 Update 4](#).
- Führen Sie ein Upgrade von PMix auf Version 3.2.3 (von) durch 3.1.5.
- Entfernen Sie das Dumping ausgefallener Rechenknoten nach `/home/logs/compute`. Die Protokoll

dateien für Rechenknoten sind in CloudWatch und in den EC2-Konsolenprotokollen verfügbar.

- Aktivieren Sie die Möglichkeit, Validatoren zu unterdrücken `SlurmQueues` und zu `ComputeResources` verlängern.
- Deaktivieren Sie das Paket-Update beim Start der Instance auf Amazon Linux 2.
- Deaktivieren Sie ImageBuilder erweiterte EC2-Bildmetadaten, wenn Sie AWS ParallelCluster benutzerdefinierte Images erstellen.
- Legen Sie die `cloud-init` Datenquelle explizit auf EC2 fest. Dies spart Startzeit für Ubuntu- und CentOS-Plattformen.
- Verwenden Sie im Namen der Vorlage für den Start der Rechenflotte den Namen der Rechenressource anstelle des Instanztyps.
- Leitet `stderr` und `stdout` zur CLI-Protokolldatei um, um unerwünschten Text in der `pcluster-CLI`-Ausgabe zu verhindern.
- Verschiebe die Rezepte für die Konfiguration und

Installation in separate Kochbücher, die vom Hauptbuch aufgerufen werden. Bestehende Einstiegspunkte werden beibehalten und sind abwärtskompatibel.

- Laden Sie die Abhängigkeiten der Intel HPC-Plattform während der AMI-Build-Zeit herunter, um zu vermeiden, dass während der Clustererstellung eine Verbindung zum Internet hergestellt wird.
- Entfernen Sie bei der Konfiguration - von Slurm-Knoten nicht den Namen der Rechenressource.
- Konfigurieren Sie keine GPUs in Slurm, wenn der NVIDIA-Treiber nicht installiert ist.
- Korrigieren Sie die `ecs:ListContainerInstances` Erlaubnis in `BatchUserRole`
- Korrigiert den Export von Clusterprotokollen, wenn kein Präfix angegeben wurde, das zuvor in ein `None` Präfix exportiert wurde.
- Behebt, dass das Rollback nicht durchgeführt wird,

wenn das Cluster-Update fehlschlägt.

- Korrigieren Sie `ecs:ListContainerInstances` die Berechtigung `inBatchUserRole` .
- Korrigieren Sie das `RootVolume` Schema für, `HeadNode` indem Sie einen Fehler auslösen, wenn ein nicht unterstütztes Objekt angegeben `KmsKeyId` wird.
- Korrigiert die fehlenden Kennzahlen von Amazon FSx, die im CloudWatch Dashboard angezeigt werden sollen.
- Korrigieren `EfaSecurityGroupValidator` .  
Bisher konnte es zu falschen Fehlern kommen, wenn benutzerdefinierte Sicherheitsgruppen bereitgestellt und EFA aktiviert war.

Einzelheiten zu den Änderungen finden Sie in den `CHANGELOG` Dateien für den [aws-parallelcluster](#) und in den Paketen unter. [aws-parallelcluster-cookbookaws-parallelcluster-node](#) [GitHub](#)

[AWS ParallelCluster Version  
3.0.3 veröffentlicht](#)

AWS ParallelCluster Version  
3.0.3 veröffentlicht.

17. Januar 2022

- Deaktivieren Sie `log4j-cve-2021-44228-hotpatch agent` (`Log4jHotPatch` ) auf Amazon Linux 2, um mögliche Leistungseinbußen zu vermeiden. Weitere Informationen finden Sie unter [Amazon Linux Hotpatch-Ankündigung für Apache Log4j](#).

Einzelheiten zu den Änderungen finden Sie in den CHANGELOG Dateien für den [aws-parallelcluster](#) und die [Pakete](#) auf [aws-parallelcluster-cookbook](#) GitHub

[AWS ParallelCluster Version 3.0.2 veröffentlicht](#)

AWS ParallelCluster Version 3.0.2 veröffentlicht.

5. November 2021

Aktualisieren Sie das [Elastic Fabric Adapter](#) Installationsprogramm auf 1.14.1

- EFA-Konfiguration: efa-config-1.9-1 (von) efa-config-1.9
- EFA-Profil: efa-profile-1.5-1 (von) efa-profile-1.5
- EFA-Kernel-Modul: efa-1.14.2 (von) efa-1.13.0
- RDMA-Kern: rdma-core-37.0 (von) rdma-core-35
- Libfabric: libfabric-1.13.2 (von) libfabric-1.13.0
- MPI öffnen: openmpi40-aws-4.1.1-2 (keine Änderung)

GPUDirect RDMA ist immer aktiviert, wenn es vom Instanztyp unterstützt wird. Die [GdrSupport](#)Konfigurationsoption hat keine Auswirkung.

Einzelheiten zu den Änderungen finden Sie in den

CHANGELOG Dateien für den [aws-parallelcluster](#) und in den Paketen unter. [aws-parallelcluster-cookbookaws-parallelcluster-node](#) [GitHub](#)



## [AWS ParallelCluster Version 3.0.1 veröffentlicht](#)

AWS ParallelCluster Version 3.0.1 veröffentlicht.

27. Oktober 2021

Tool zur Migration der Clusterkonfiguration

- Kunden können ihre Clusterkonfigurationen jetzt vom Format der AWS ParallelCluster Version 2 auf das YAML-basierte Format der AWS ParallelCluster Version 3 migrieren. Weitere Informationen finden Sie unter [pcluster3-config-converter](#).

Der Hauptknoten kann gestoppt werden

- Nach dem Stoppen der Rechenflotte kann der Hauptknoten mithilfe der Amazon EC2 EC2-Konsole oder des Befehls [AWS CLI stop-instances](#) gestoppt und später neu gestartet werden.

Standardmäßig aus Datei gelesen AWS-Region  
~/.aws/config

- Wenn für den [pcluster](#) Befehl nicht in der AWS-Region Konfigurationsdatei

i, in der Umgebung oder in der Befehlszeile AWS-Region angegeben ist, wird der in der `region` Einstellung im `[default]` Abschnitt der `~/.aws/config` Datei angegebene Standard verwendet.

Einzelheiten zu den Änderungen finden Sie in den `CHANGELOG` Dateien für den [aws-parallelcluster](#) und in den Paketen unter [aws-parallelcluster-cookbookaws-parallelcluster-node](#) [GitHub](#)

[AWS ParallelCluster Version  
3.0.0 veröffentlicht](#)

AWS ParallelCluster Version  
3.0.0 veröffentlicht.

10 September 2021

Support für Clusterma-  
nagement über Amazon API  
Gateway

- Kunden können jetzt Cluster über HTTP-Endpunkte mit Amazon API Gateway verwalten und bereitstellen. Dies eröffnet neue Möglichkeiten für skriptgesteuerte oder ereignisgesteuerte Workflows.

Die AWS ParallelCluster Befehlszeilenschnittstelle (CLI) wurde ebenfalls aus Gründen der Kompatibilität mit dieser API neu gestaltet und enthält eine neue JSON-Ausgabeoption. Diese neue Funktionalität ermöglicht es Kunden, ähnliche Bausteinfunktionen auch über die CLI zu implementieren.

Verbesserte benutzerdefinierte  
AMI-Erstellung

- Kunden haben jetzt Zugriff auf einen robusteren Prozess für die Erstellung

und Verwaltung benutzerdefinierter AMIs mit EC2 Image Builder. Benutzerdefinierte AMIs können jetzt über eine separate AWS ParallelCluster Konfigurationsdatei verwaltet und mit dem [pcluster build-image](#) Befehl in der AWS ParallelCluster Befehlszeilenschnittstelle erstellt werden.

Einzelheiten zu den Änderungen finden Sie in den CHANGELOG Dateien für den [aws-parallelcluster](#) und in den Paketen unter [aws-parallelcluster-cookbookaws-parallelcluster-node](#) [GitHub](#)

Die vorliegende Übersetzung wurde maschinell erstellt. Im Falle eines Konflikts oder eines Widerspruchs zwischen dieser übersetzten Fassung und der englischen Fassung (einschließlich infolge von Verzögerungen bei der Übersetzung) ist die englische Fassung maßgeblich.