



Guide de mise en route

Amazon Redshift



Amazon Redshift: Guide de mise en route

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

Les marques et la présentation commerciale d'Amazon ne peuvent être utilisées en relation avec un produit ou un service qui n'est pas d'Amazon, d'une manière susceptible de créer une confusion parmi les clients, ou d'une manière qui dénigre ou discrédite Amazon. Toutes les autres marques commerciales qui ne sont pas la propriété d'Amazon appartiennent à leurs propriétaires respectifs, qui peuvent ou non être affiliés ou connectés à Amazon, ou sponsorisés par Amazon.

Table of Contents

Amazon Redshift sans serveur	1
S'inscrire à AWS	1
Création d'un entrepôt des données avec Amazon Redshift sans serveur	1
Chargement d'exemples de données	4
Exécution d'exemples de requêtes	8
Chargement de données depuis Amazon S3	10
Interroger des sources de données externes	18
Interroger les lacs de données	18
Interroger des sources de données distantes	19
Accéder à des données dans d'autres clusters	19
Entraîner des modèles ML avec des données Redshift	20
Clusters Amazon Redshift provisionnés	22
Console pour clusters provisionnés Amazon Redshift	22
S'inscrire à AWS	22
Déterminer les règles de pare-feu	23
Connexion à Amazon Redshift	25
Clusters Amazon Redshift et chargement de données	26
Avec un exemple de jeu de données	26
Importation de vos propres données dans Amazon Redshift	31
Tâches courantes liées aux bases de données	46
Tâche 1 : Créer une base de données	47
Tâche 2 : Créer un utilisateur	48
Tâche 3 : Créer un schéma	48
Tâche 4 : Créer une table	50
Tâche 5 : Charger des exemples de données	52
Tâche 6 : Interroger les tables système	52
Tâche 7 : Annuler une requête	56
Tâche 8 : Nettoyer vos ressources	59
Présentation conceptuelle d'Amazon Redshift	61
Ressources supplémentaires	65
Historique du document	67
.....	lxix

Amazon Redshift sans serveur

Si vous utilisez Amazon Redshift Serverless pour la première fois, nous vous recommandons de lire les sections suivantes pour vous aider à faire vos premiers pas avec Amazon Redshift Serverless. Le flux de base d'Amazon Redshift sans serveur consiste à créer des ressources sans serveur, à se connecter à Amazon Redshift sans serveur, à charger des exemples de données, puis à exécuter des requêtes sur les données. Dans ce guide, vous pouvez choisir de charger des exemples de données à partir d'Amazon Redshift sans serveur ou d'un compartiment Amazon S3.

- [the section called “S'inscrire à AWS”](#)
- [the section called “Création d'un entrepôt des données avec Amazon Redshift sans serveur”](#)
- [the section called “Chargement de données depuis Amazon S3”](#)

S'inscrire à AWS

Si vous n'avez pas encore de AWS compte, créez-en un. Si vous disposez déjà d'un compte, vous pouvez ignorer ce prérequis et utiliser votre compte existant.

1. Ouvrez <https://portal.aws.amazon.com/billing/signup>.
2. Suivez les instructions en ligne.

Lorsque vous créez un AWS compte, un AWS utilisateur root est créé. L'utilisateur root a accès à tous les AWS services et ressources du compte. En tant que bonne pratique de sécurité, [attribuer un accès administratif à un utilisateur administratif](#), et utilisez uniquement l'utilisateur root pour effectuer [tâches nécessitant un accès utilisateur root](#).

Création d'un entrepôt des données avec Amazon Redshift sans serveur

La première fois que vous vous connectez à la console Amazon Redshift sans serveur, vous êtes invité à accéder à l'expérience de démarrage, que vous pouvez utiliser pour créer et gérer des ressources sans serveur. Dans ce guide, vous allez créer des ressources sans serveur en utilisant les paramètres par défaut d'Amazon Redshift sans serveur.

Pour un contrôle plus précis de votre configuration, choisissez Personnaliser les paramètres.

Pour configurer avec les paramètres par défaut :

1. [Connectez-vous à la console Amazon Redshift AWS Management Console et ouvrez-la à l'adresse https://console.aws.amazon.com/redshiftv2/.](https://console.aws.amazon.com/redshiftv2/)

Choisissez Essayer Amazon Redshift sans serveur..

2. Sous Configuration, choisissez Utiliser les paramètres par défaut. Amazon Redshift sans serveur crée un espace de noms par défaut avec un groupe de travail par défaut associé à cet espace de noms. Choisissez Save configuration.

La capture d'écran suivante présente les paramètres par défaut d'Amazon Redshift sans serveur.

Get started with Amazon Redshift Serverless

To start using Amazon Redshift Serverless, set up your serverless data warehouse and create a database.

You will receive \$0 credit towards your Redshift Serverless usage in this account.

Configuration [Info](#)

Use default settings

Default settings have been defined to help you get started. You can change them at any time later.

Customize settings

Customize your settings for your specific needs.

▼ How it works



Using the default settings

Amazon Redshift Serverless creates a default namespace and workgroup. This configuration uses the default settings and becomes active when you associate the default workgroup to the default namespace.



Customizing the settings

Amazon Redshift Serverless creates a default namespace and workgroup. This configuration becomes active when you associate the default workgroup to the default namespace.

Namespace [Info](#)

Namespace is a collection of database objects and users. Data properties include database name and password, permissions, and encryption and security.

⚠ Your data is encrypted by default with an AWS owned key. To choose a different key, choose **Customize settings**.

Namespace name

default

Database name and password

Database name

dev

Admin user credentials

Created based on IAM credentials

Permissions

Default IAM role

<arn:aws:iam::694237154968:role/service-role/AmazonRedshift-CommandsAccessRole-20220809T211900> [↗](#)

- Une fois la configuration terminée, cliquez sur Continue (Continuer) pour accéder à votre Serverless dashboard (Tableau de bord sans serveur). Vous pouvez constater que le groupe de travail et l'espace de noms sans serveur sont disponibles.

Serverless dashboard [Info](#)

Namespace overview [Info](#)

Namespace data from your account

Total snapshots	Datashares in my account	Datashares requiring authorization	Datashares fr
0	0	0	0

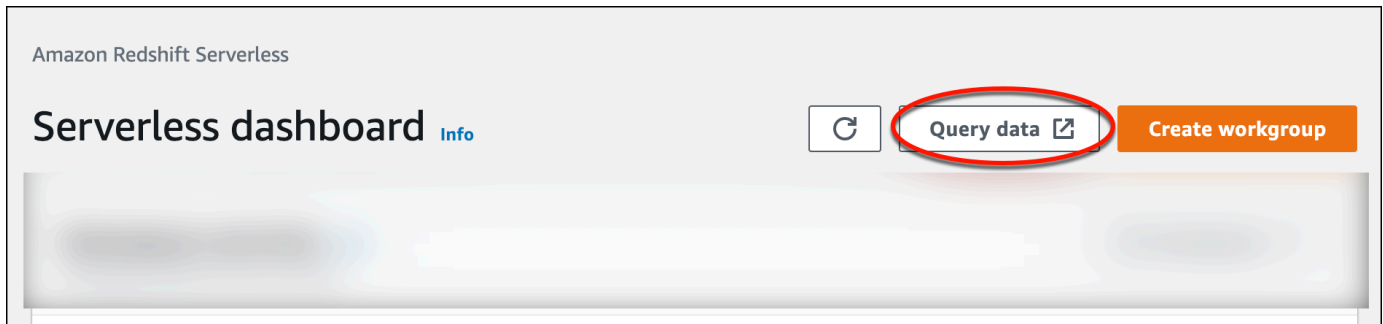
Namespaces / Workgroups [Info](#)

Namespace	Status	Workgroup	Status
default	✔ Available	default	✔ Available

Chargement d'exemples de données

Maintenant que vous avez configuré votre entrepôt des données avec Amazon Redshift sans serveur, vous pouvez utiliser l'éditeur de requête d'Amazon Redshift v2 pour charger des exemples de données.

- Pour lancer l'éditeur de requête v2 à partir de la console Amazon Redshift sans serveur, choisissez Interroger les données. Lorsque vous appelez l'éditeur de requête v2 à partir de la console Amazon Redshift Serverless, un nouvel onglet du navigateur s'ouvre avec l'éditeur de requête. L'éditeur de requête v2 se connecte depuis votre ordinateur client à l'environnement Amazon Redshift sans serveur.



2. Si vous lancez l'éditeur de requêtes v2 pour la première fois, vous devez configurer AWS KMS le chiffrement avant de continuer. En option, vous pouvez également spécifier l'URI d'un compartiment S3 pour le chargement ultérieur des données. Sélectionnez ensuite Configurer le compte.

Configure account

To start using Redshift query editor v2, configure the account.

AWS KMS encryption

Your data is encrypted by default with a key that AWS owns and manages for you. To choose a different key, customize your encryption settings.

You can't change the AWS KMS encryption settings after configuring your account.

Customize encryption settings (advanced)

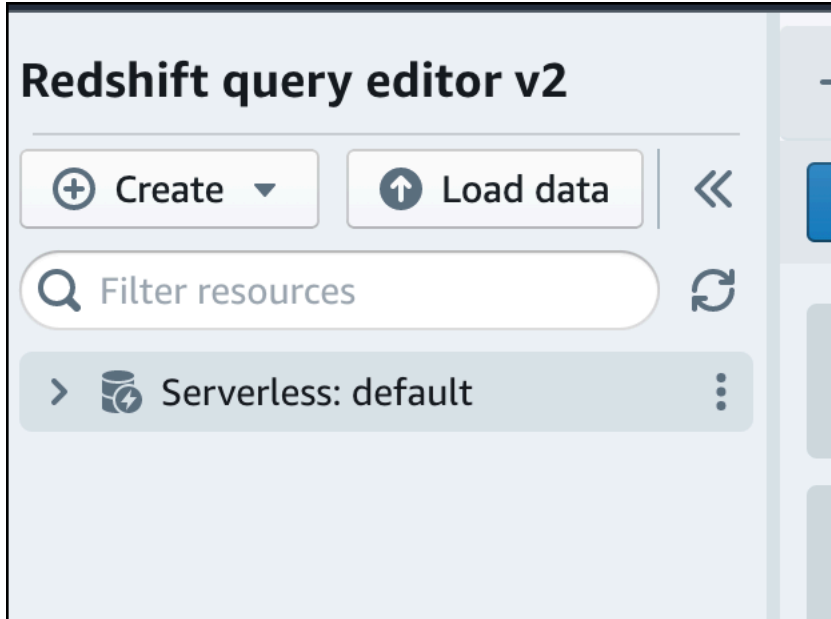
S3 bucket - optional

Enter an Amazon S3 path for this account. This path can be used for features such as load data from a local file store. Account users and the S3 bucket must be configured with the proper permissions to secure user's data. [Learn more](#)

S3 URI

Pour en savoir plus sur la configuration de l'éditeur de requête v2, notamment sur les autorisations nécessaires, voir [Configuration de votre Compte AWS](#) dans le Guide de gestion Amazon Redshift.

3. Pour vous connecter à un groupe de travail, sélectionnez le nom du groupe de travail dans le panneau d'arborescence.



4. Lorsque vous vous connectez à un nouveau groupe de travail pour la première fois dans l'éditeur de requête v2, vous devez sélectionner le type d'authentification à utiliser pour vous connecter au groupe de travail. Pour ce guide, laissez l'option Utilisateur fédéré sélectionnée et choisissez Créer une connexion.

Connect to default ✕

Authentication [Learn more](#)

Federated user
The query editor v2 generates a temporary password to connect to the database.

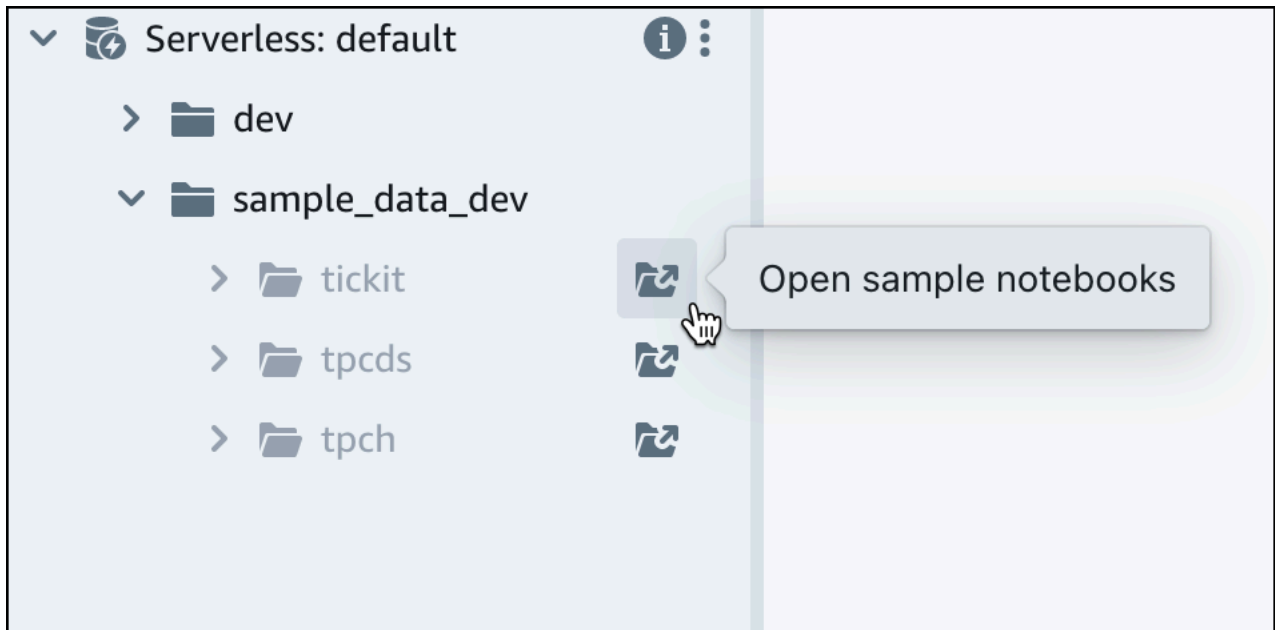
Database user name and password
Provide a database user and password for the database that you are connecting to. The query editor v2 stores your credentials in AWS Secrets Manager on your behalf.

Database

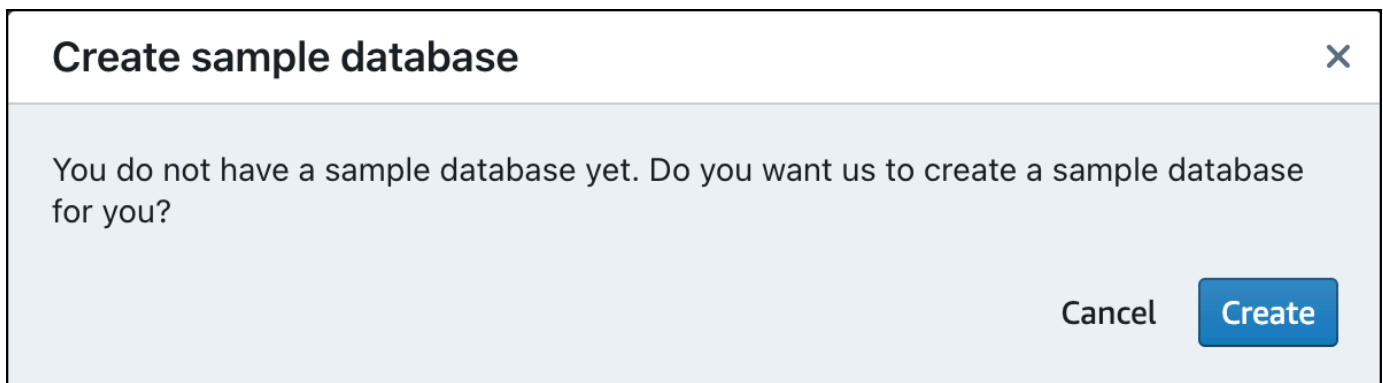
The database name must be 1-64 characters. Valid characters are lowercase alphanumeric characters.

Une fois connecté, vous pouvez choisir de charger des exemples de données à partir d'Amazon Redshift sans serveur ou d'un compartiment Amazon S3.

5. Dans le groupe de travail par défaut Amazon Redshift sans serveur, développez la base de données `sample_data_dev`. Il existe trois schémas types correspondant à trois jeux de données types que vous pouvez charger dans la base de données Amazon Redshift sans serveur. Choisissez l'exemple de jeu de données que vous souhaitez charger, puis sélectionnez Ouvrir les exemples de blocs-notes.



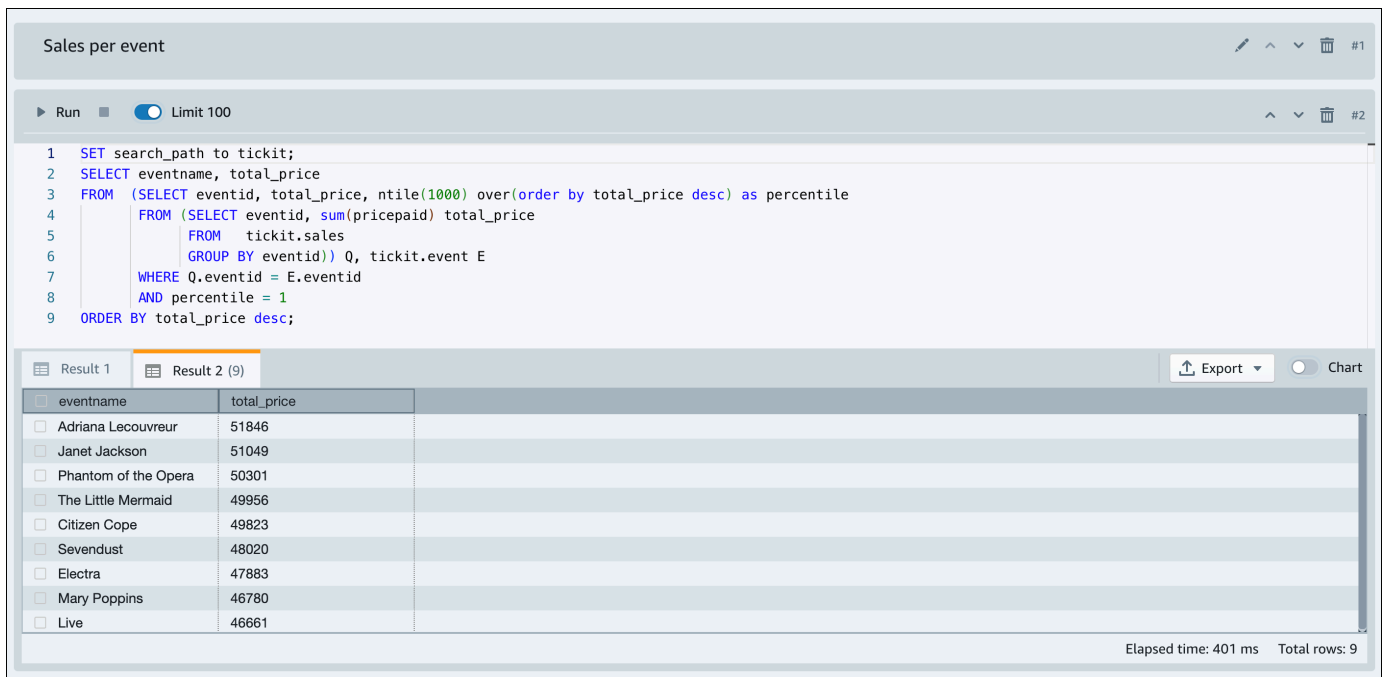
6. Lorsque vous chargez des données pour la première fois, l'éditeur de requête v2 vous invite à créer une base de données d'exemple. Choisissez Créer.



Exécution d'exemples de requêtes

Après avoir configuré Amazon Redshift sans serveur, vous pouvez commencer à utiliser un exemple de jeu de données dans Amazon Redshift sans serveur. Amazon Redshift sans serveur charge automatiquement le jeu de données d'exemple, tel que le jeu de données tickit, et vous pouvez immédiatement interroger les données.

- Une fois qu'Amazon Redshift sans serveur a fini de charger les exemples de données, tous les exemples de requêtes sont chargés dans l'éditeur. Vous pouvez choisir Exécuter tout pour exécuter toutes les requêtes des exemples de blocs-notes.



Sales per event

```

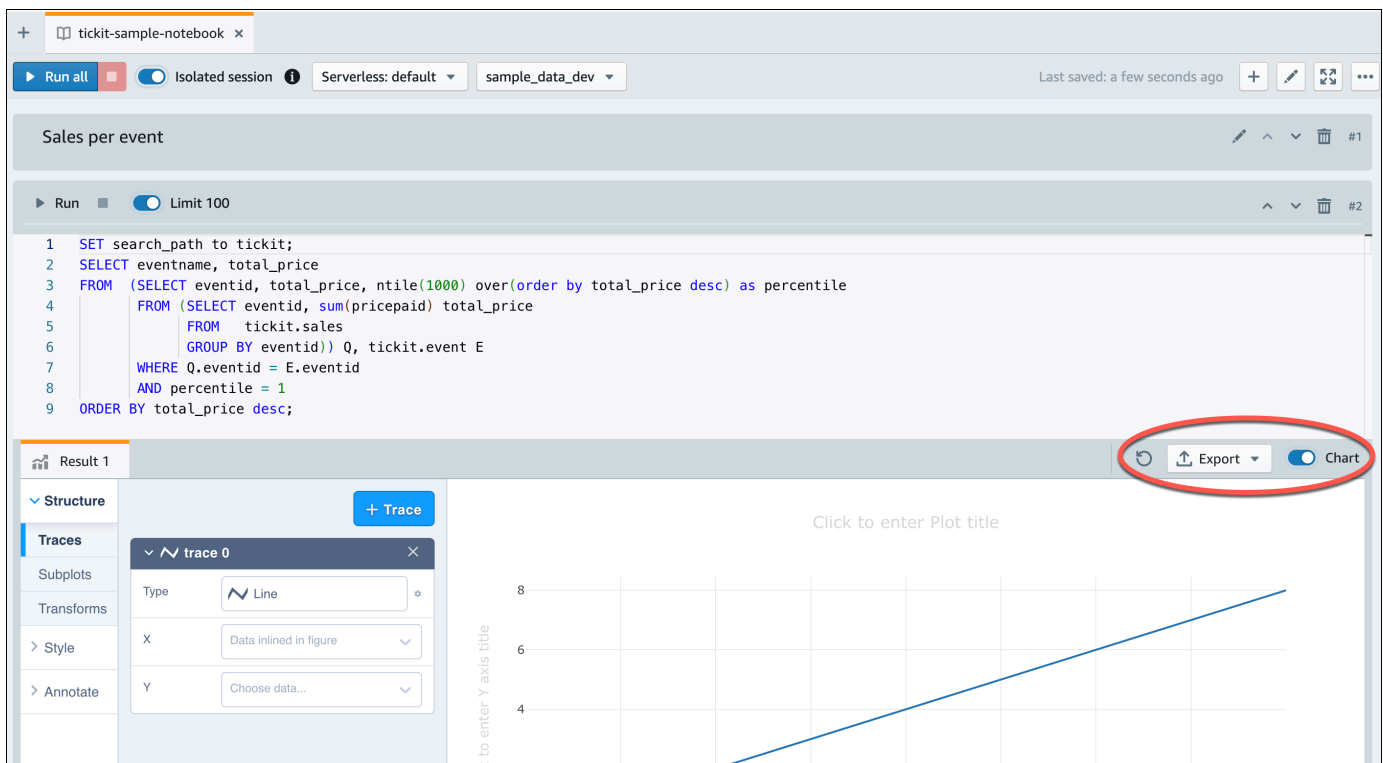
1 SET search_path to tickit;
2 SELECT eventname, total_price
3 FROM (SELECT eventid, total_price, ntile(1000) over(order by total_price desc) as percentile
4       FROM (SELECT eventid, sum(pricepaid) total_price
5             FROM tickit.sales
6             GROUP BY eventid)) Q, tickit.event E
7 WHERE Q.eventid = E.eventid
8      AND percentile = 1
9 ORDER BY total_price desc;

```

eventname	total_price
Adriana Lecouvreur	51846
Janet Jackson	51049
Phantom of the Opera	50301
The Little Mermaid	49956
Citizen Cope	49823
Sevendust	48020
Electra	47883
Mary Poppins	46780
Live	46661

Elapsed time: 401 ms Total rows: 9

Vous pouvez également exporter les résultats dans un fichier JSON ou CSV ou les afficher dans un graphique.



Sales per event

```

1 SET search_path to tickit;
2 SELECT eventname, total_price
3 FROM (SELECT eventid, total_price, ntile(1000) over(order by total_price desc) as percentile
4       FROM (SELECT eventid, sum(pricepaid) total_price
5             FROM tickit.sales
6             GROUP BY eventid)) Q, tickit.event E
7 WHERE Q.eventid = E.eventid
8      AND percentile = 1
9 ORDER BY total_price desc;

```

Export Chart

Click to enter Plot title

8
6
4
2

Click to enter Y axis title

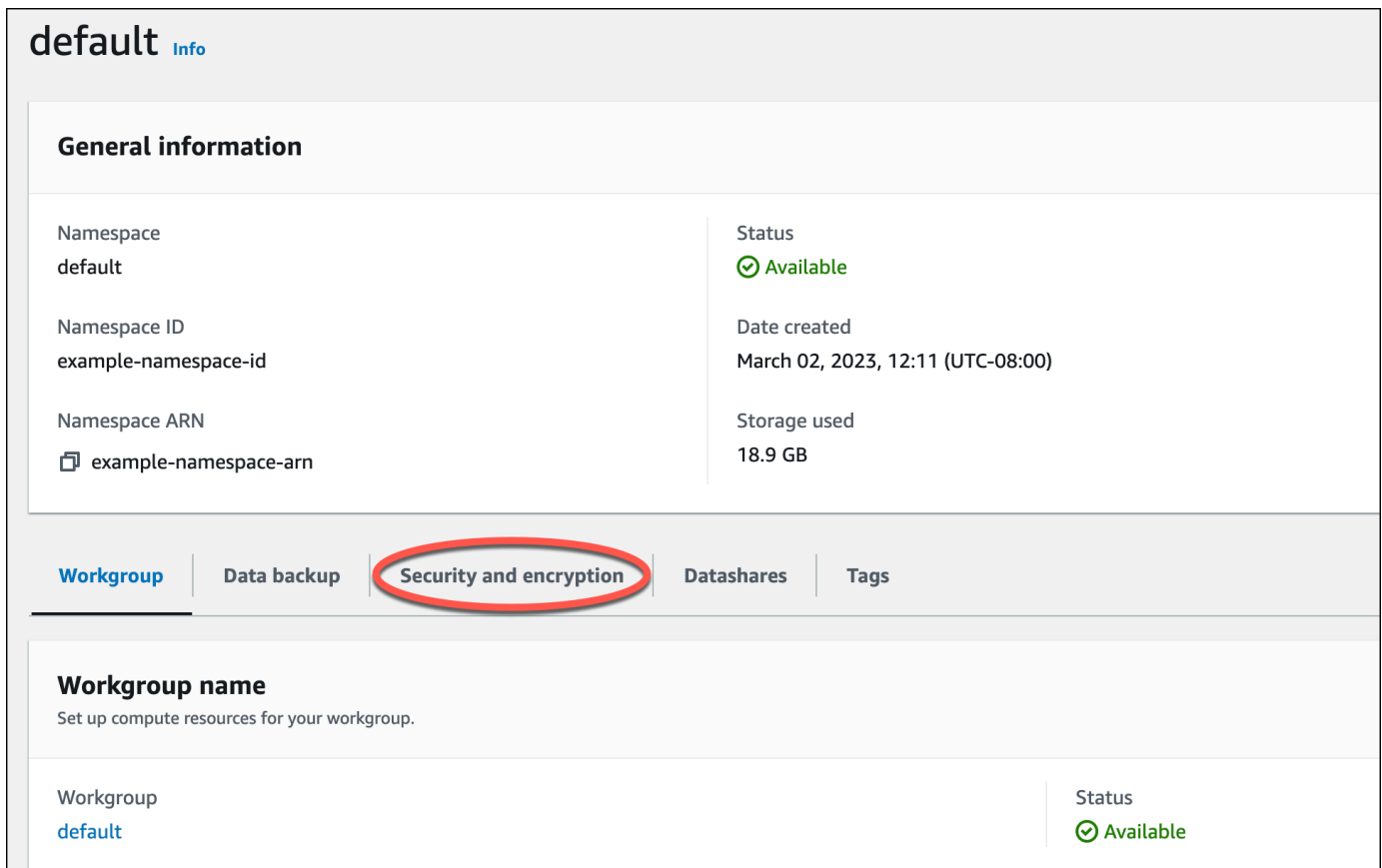
Vous pouvez également charger des données à partir d'un compartiment Amazon S3. Pour en savoir plus, veuillez consulter [the section called "Chargement de données depuis Amazon S3"](#).

Chargement de données depuis Amazon S3

Après avoir créé votre entrepôt des données, vous pouvez charger des données depuis Amazon S3.

À ce stade, vous disposez d'une base de données nommée dev. Ensuite, créez des tables dans la base de données, chargez des données dans les tables et essayez d'exécuter une requête. Pour plus de commodité, les exemples de données que vous chargez sont disponibles dans un compartiment Amazon S3.

1. Avant de pouvoir charger des données depuis Amazon S3, vous devez d'abord créer un rôle IAM avec les autorisations nécessaires et l'attacher à votre espace de noms sans serveur. Pour ce faire, choisissez Configuration d'espace de noms dans le menu de navigation, choisissez Sécurité et chiffrement. Choisissez ensuite Gérer les rôles IAM.



The screenshot displays the configuration page for the 'default' namespace in Amazon Redshift. The page is titled 'default' with an 'Info' link. It features a 'General information' section with the following details:

Namespace	default	Status	Available
Namespace ID	example-namespace-id	Date created	March 02, 2023, 12:11 (UTC-08:00)
Namespace ARN	example-namespace-arn	Storage used	18.9 GB

Below the general information, there is a navigation bar with tabs for 'Workgroup', 'Data backup', 'Security and encryption', 'Datashares', and 'Tags'. The 'Security and encryption' tab is highlighted with a red circle. Underneath, the 'Workgroup name' section is visible, showing the workgroup name as 'default' and its status as 'Available'.

2. Développez le menu Gérer les rôles IAM et choisissez Créer un rôle IAM.

Manage IAM roles

Permissions

i Associate an IAM role so that your serverless endpoint can LOAD and UNLOAD data. You can create an IAM role as the default for this configuration that has the [AmazonRedshiftAllCommandsFullAccess](#) policy attached. This policy includes permissions to run SQL commands to COPY, UNLOAD, and query data with Amazon Redshift Serverless. This policy also grants permissions to run SELECT statements for related services, such as Amazon S3, Amazon CloudWatch logs, Amazon SageMaker, and AWS Glue. You won't be able to run these SQL commands without an IAM role attached to your namespace.

Associated IAM roles (1)

Create, associate, or remove an IAM role. You can associate up to 50 IAM roles. You can also choose an IAM role and set it as the default.

Set default ▼ **Manage IAM roles** ▲

🔍 Search for associated IAM roles or role type

Associate IAM roles
Create IAM role
Remove IAM roles

< 1 >

IAM roles [🔗](#) ▼ **Status** ▼ **Role type** ▼

3. Choisissez le niveau d'accès au compartiment S3 que vous souhaitez accorder à ce rôle, et sélectionnez Créer un rôle IAM par défaut.

Create the default IAM role ✕

i Associate an IAM role so that your serverless endpoint can LOAD and UNLOAD data. You can create an IAM role as the default for this configuration that has the [AmazonRedshiftAllCommandsFullAccess](#) policy attached. This policy includes permissions to run SQL commands to COPY, UNLOAD, and query data with Amazon Redshift Serverless. This policy also grants permissions to run SELECT statements for related services, such as Amazon S3, Amazon CloudWatch logs, Amazon SageMaker, and AWS Glue. You won't be able to run these SQL commands without an IAM role attached to your namespace.

Specify an S3 bucket for the IAM role to access
To create a new bucket, [visit S3](#)

No additional S3 bucket
Create the IAM role without specifying S3 buckets.

Any S3 bucket
Allow users that have access to your Redshift Serverless data to also access any S3 bucket and its contents in your AWS account.

Specific S3 buckets
Specify one or more S3 buckets that the IAM role being created has permission to access.

4. Sélectionnez Enregistrer les modifications. Vous pouvez désormais charger des exemples de données à partir d'Amazon S3.

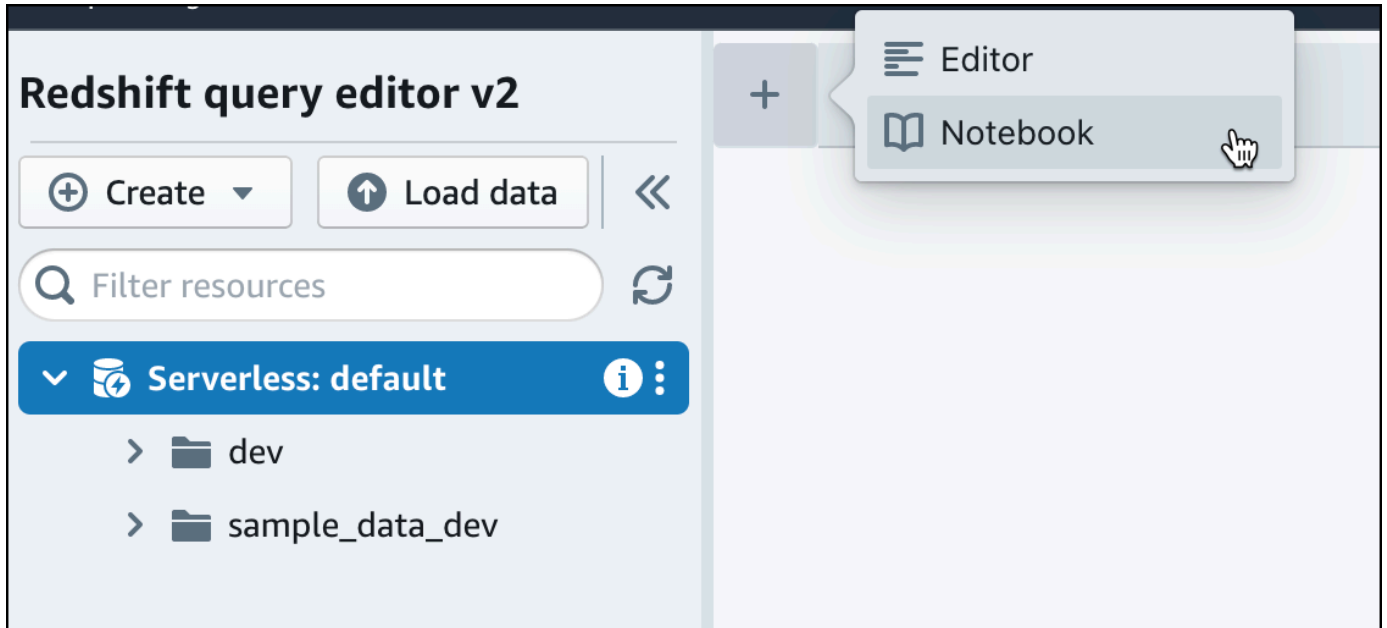
Les étapes suivantes utilisent des données au sein d'un compartiment S3 public d'Amazon Redshift, mais vous pouvez reproduire les mêmes étapes en utilisant votre propre compartiment S3 et vos commandes SQL.

Charger un exemple de données à partir d'Amazon S3

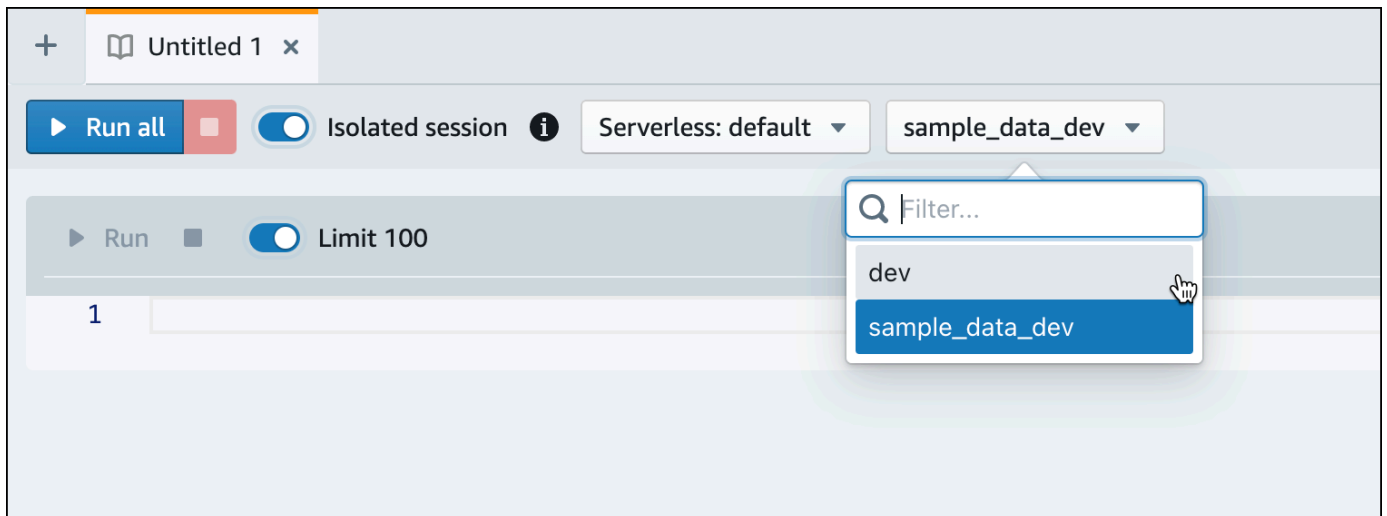
1. Dans l'éditeur de requête v2, choisissez



Ajouter, puis Bloc-notes pour créer un nouveau bloc-notes SQL.



2. Passez à la base de données dev.



3. Créez des tables.

Si vous utilisez l'éditeur de requête v2, copiez et exécutez les instructions create table suivantes pour créer des tables dans la base de données dev. Pour plus d'informations sur la syntaxe,

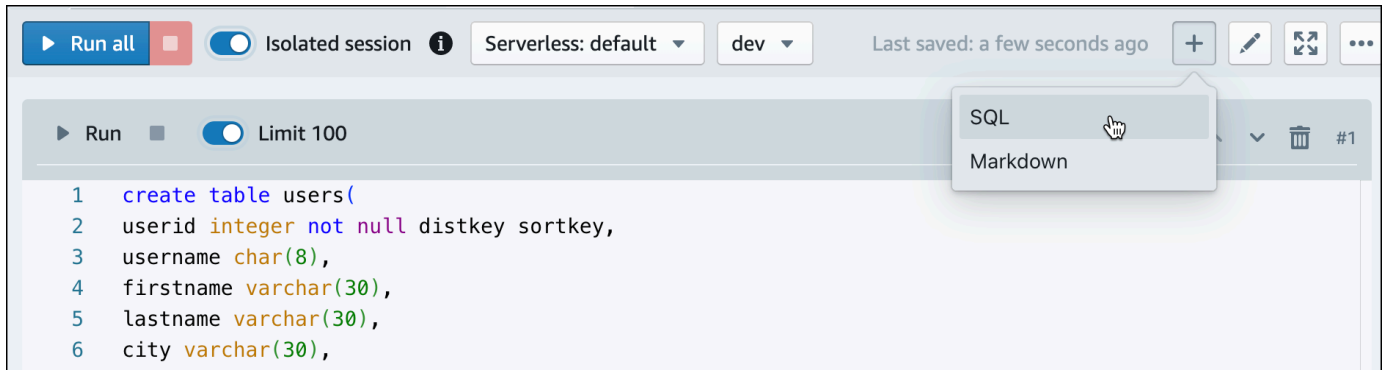
consultez [CREATE TABLE](#) dans le Guide du développeur de base de données Amazon Redshift.

```
create table users(  
userid integer not null distkey sortkey,  
username char(8),  
firstname varchar(30),  
lastname varchar(30),  
city varchar(30),  
state char(2),  
email varchar(100),  
phone char(14),  
likesports boolean,  
liketheatre boolean,  
likeconcerts boolean,  
likejazz boolean,  
likeclassical boolean,  
likeopera boolean,  
likerock boolean,  
likevegas boolean,  
likebroadway boolean,  
likemusicals boolean);
```

```
create table event(  
eventid integer not null distkey,  
venueid smallint not null,  
catid smallint not null,  
dateid smallint not null sortkey,  
eventname varchar(200),  
starttime timestamp);
```

```
create table sales(  
salesid integer not null,  
listid integer not null distkey,  
sellerid integer not null,  
buyerid integer not null,  
eventid integer not null,  
dateid smallint not null sortkey,  
qtysold smallint not null,  
pricepaid decimal(8,2),  
commission decimal(8,2),  
saletime timestamp);
```

4. Dans l'éditeur de requête v2, créez une nouvelle cellule SQL dans votre bloc-notes.



5. Utilisez maintenant la commande COPY dans l'éditeur de requête v2 pour charger de grands jeux de données depuis Amazon S3 ou Amazon DynamoDB dans Amazon Redshift. Pour plus d'informations sur la syntaxe COPY, consultez [COPY](#) dans le Guide du développeur de la base de données Amazon Redshift.

Vous pouvez exécuter la commande COPY avec quelques exemples de données disponibles dans un compartiment S3 public. Exécutez les commandes SQL suivantes dans l'éditeur de requête v2.

```
COPY users
FROM 's3://redshift-downloads/ticket/allusers_pipe.txt'
DELIMITER '|'
TIMEFORMAT 'YYYY-MM-DD HH:MI:SS'
IGNOREHEADER 1
REGION 'us-east-1'
IAM_ROLE default;

COPY event
FROM 's3://redshift-downloads/ticket/allevnts_pipe.txt'
DELIMITER '|'
TIMEFORMAT 'YYYY-MM-DD HH:MI:SS'
IGNOREHEADER 1
REGION 'us-east-1'
IAM_ROLE default;

COPY sales
FROM 's3://redshift-downloads/ticket/sales_tab.txt'
DELIMITER '\t'
TIMEFORMAT 'MM/DD/YYYY HH:MI:SS'
IGNOREHEADER 1
REGION 'us-east-1'
```

```
IAM_ROLE default;
```

6. Après avoir chargé les données, créez une autre cellule SQL dans votre bloc-notes et essayez quelques exemples de requêtes. Pour plus d'informations sur l'utilisation de l'instruction `SELECT`, consultez [SELECT](#) dans le Guide du développeur Amazon Redshift. Pour comprendre la structure et les schémas des données de l'échantillon, explorez l'éditeur de requête v2.

```
-- Find top 10 buyers by quantity.
SELECT firstname, lastname, total_quantity
FROM (SELECT buyerid, sum(qtysold) total_quantity
      FROM sales
      GROUP BY buyerid
      ORDER BY total_quantity desc limit 10) Q, users
WHERE Q.buyerid = userid
ORDER BY Q.total_quantity desc;

-- Find events in the 99.9 percentile in terms of all time gross sales.
SELECT eventname, total_price
FROM (SELECT eventid, total_price, ntile(1000) over(order by total_price desc) as
      percentile
      FROM (SELECT eventid, sum(pricepaid) total_price
            FROM sales
            GROUP BY eventid)) Q, event E
WHERE Q.eventid = E.eventid
      AND percentile = 1
ORDER BY total_price desc;
```

Maintenant que vous avez chargé des données et exécuté quelques exemples de requêtes, vous pouvez explorer d'autres domaines d'Amazon Redshift sans serveur. Consultez la liste suivante pour en savoir plus sur la manière dont vous pouvez utiliser Amazon Redshift sans serveur.

- Vous pouvez charger des données à partir d'un compartiment Amazon S3. Pour plus d'informations, consultez [Chargement des données](#) à partir d'Amazon S3.
- Vous pouvez utiliser l'éditeur de requête v2 pour charger des données à partir d'un fichier local séparé par des caractères et d'une taille inférieure à 5 Mo. Pour plus d'informations, consultez [Chargement de données à partir d'un fichier local](#).
- Vous pouvez vous connecter à Amazon Redshift sans serveur avec des outils SQL tiers grâce au pilote JDBC et ODBC. Pour plus d'informations, consultez [Connexion à Amazon Redshift sans serveur](#).

- Vous pouvez également utiliser l'API de données Amazon Redshift pour vous connecter à Amazon Redshift sans serveur. Pour plus d'informations, consultez [Utilisation de l'API de données Amazon Redshift](#).
- Vous pouvez utiliser vos données dans Amazon Redshift sans serveur avec Redshift ML pour créer des modèles de machine learning avec la commande CREATE MODEL. Consultez [Tutoriel : création de modèles de désabonnement des clients](#) pour apprendre à élaborer un modèle ML Redshift.
- Vous pouvez interroger les données d'un lac de données Amazon S3 sans charger de données dans Amazon Redshift sans serveur. Pour plus d'informations, consultez [Interrogation d'un lac de données](#).

Interroger des sources de données en dehors de votre base de données Amazon Redshift.

Vous trouverez ci-dessous des informations relatives la façon de commencer à interroger des données sur des sources distantes, y compris des clusters Amazon Redshift distants. Vous trouverez également des informations relatives à l'entraînement des modèles de machine learning (ML) à l'aide d'Amazon Redshift.

Rubriques

- [Interroger votre lac de données](#)
- [Interroger des données sur des sources de données distantes](#)
- [Accéder aux données dans d'autres clusters Amazon Redshift](#)
- [Entraîner des modèles de machine learning avec les données d'Amazon Redshift](#)

Interroger votre lac de données

Vous pouvez utiliser Amazon Redshift Spectrum pour interroger les données dans les fichiers Amazon S3 sans avoir à les charger dans les tables Amazon Redshift. Amazon Redshift fournit une fonctionnalité SQL conçue pour le traitement analytique en ligne rapide (OLAP) de très grands jeux de données stockés dans des clusters Amazon Redshift et des lacs de données Amazon S3. Vous pouvez interroger des données dans de nombreux formats, notamment Parquet, ORC, RCFile, TextFile, SequenceFile, RegexSerde, OpenCSV et AVRO. Vous créez des schémas et des tables externes pour définir la structure des fichiers dans Amazon S3. Ensuite, utilisez un catalogue de données externe tel que AWS Glue ou votre métastore Apache Hive. Les modifications apportées à un type de catalogue de données externe sont instantanément répercutées dans vos clusters Amazon Redshift.

Une fois vos données enregistrées dans un catalogue de données AWS Glue et activées avec AWS Lake Formation, vous pouvez les interroger à l'aide de Redshift Spectrum.

Redshift Spectrum réside sur des serveurs Amazon Redshift dédiés indépendants de votre cluster. Il transmet à la couche Redshift Spectrum de nombreuses tâches nécessitant une importante capacité de calcul, telles que le regroupement et le filtrage des prédicats. Redshift Spectrum évolue également intelligemment pour tirer parti du traitement massivement parallèle.

Vous pouvez également partitionner les tables externes en une ou plusieurs colonnes pour optimiser les performances de la requête en éliminant les partitions. Vous pouvez interroger les tables externes et les joindre avec les tables Amazon Redshift. Vous pouvez ajouter des tables Redshift Spectrum à plusieurs clusters Amazon Redshift et interroger les mêmes données dans Amazon S3 à partir de n'importe quel cluster de la même région AWS. Lors de la mise à jour des fichiers de données Amazon S3, les données peuvent être interrogées immédiatement depuis n'importe lequel de vos clusters Amazon Redshift.

Pour plus d'informations sur Redshift Spectrum, y compris sur l'utilisation de Redshift Spectrum et des lacs de données, consultez la section [Premiers pas avec Amazon Redshift Spectrum](#) dans le Manuel du développeur de bases de données Amazon Redshift.

Interroger des données sur des sources de données distantes

Vous pouvez joindre des données à partir d'une base de données Amazon RDS, d'une base de données Amazon Aurora ou d'Amazon S3 avec les données de votre base de données Amazon Redshift à l'aide d'une requête fédérée. Vous pouvez utiliser Amazon Redshift pour interroger directement les données opérationnelles (sans les déplacer), appliquer des transformations et insérer des données dans vos tables Redshift. Une partie du calcul des requêtes fédérées est distribuée aux sources de données distantes.

Pour exécuter des requêtes fédérées, Amazon Redshift établit d'abord une connexion à la source de données distante. Amazon Redshift récupère ensuite les métadonnées sur les tables dans la source de données distante, émet des requêtes, puis récupère les lignes de résultats. Amazon Redshift distribue ensuite les lignes de résultats aux nœuds de calcul Amazon Redshift pour un traitement ultérieur.

Pour plus d'informations sur la configuration de votre environnement pour les requêtes fédérées, consultez l'une des rubriques suivantes : Manuel du développeur de base de données Amazon Redshift :

- [Commencer à utiliser les requêtes fédérées vers PostgreSQL](#)
- [Commencer à utiliser des requêtes fédérées vers MySQL](#)

Accéder aux données dans d'autres clusters Amazon Redshift

Grâce au partage de données Amazon Redshift, vous pouvez partager en toute sécurité et facilement des données en direct sur des clusters Amazon Redshift ou des comptes AWS à des fins de

lecture. Vous pouvez bénéficier d'un accès instantané, granulaire et performant aux données sur l'ensemble des clusters Amazon Redshift, sans avoir à les copier ou à les déplacer manuellement. Vos utilisateurs peuvent voir les dernières informations mises à jour telles quelles dans les clusters Amazon Redshift. Vous pouvez partager des données à différents niveaux, comme les bases de données, les schémas, les tables, les vues (y compris les vues standard, à liaison tardive et matérialisées) et les fonctions SQL définies par l'utilisateur (UDF).

Le partage de données Amazon Redshift est particulièrement utile dans les situations suivantes :

- Centralisation des applications essentielles à vos activités : utilisez un cluster central Extract-transform-load (ETL) qui partage des données avec plusieurs clusters de Business Intelligence (BI) ou d'analytique. Cette approche fournit un isolement de l'application en lecture et une rétrofacturation pour les applications individuelles.
- Partage de données entre environnements : partagez des données entre les environnements de développement, de test et de production. Vous pouvez améliorer l'agilité des équipes en partageant des données selon différents niveaux de précision.

Pour plus d'informations sur le partage de données, consultez [Mise en route du partage de données](#) dans le Manuel du développeur de bases de données Amazon Redshift.

Entraîner des modèles de machine learning avec les données d'Amazon Redshift

À l'aide d'Amazon Redshift ML, vous pouvez entraîner un modèle en fournissant les données à Amazon Redshift. Amazon Redshift ML crée ensuite des modèles qui capturent les modèles dans les données entrées. Vous pouvez ensuite utiliser ces modèles pour générer des prédictions pour de nouvelles données entrées sans encourir de coûts supplémentaires. En utilisant Amazon Redshift ML, vous pouvez entraîner des modèles de machine learning en utilisant des instructions SQL et les invoquer dans des requêtes SQL pour la prédiction. Vous pouvez continuer à améliorer la précision des prédictions en changeant itérativement les paramètres et en améliorant vos données d'entraînement.

Amazon Redshift ML permet aux utilisateurs de SQL de créer, d'entraîner et de déployer facilement des modèles de machine learning à l'aide d'instructions SQL familières. En utilisant Amazon Redshift ML, vous pouvez utiliser vos données dans les clusters Amazon Redshift pour entraîner des modèles avec Amazon SageMaker Autopilot et obtenir automatiquement le meilleur modèle. Vous pouvez ensuite repérer les modèles et faire des prédictions à partir d'une base de données Amazon Redshift.

Pour plus d'informations sur Amazon Redshift ML, consultez [Démarrer avec Amazon Redshift ML](#) du guide du développeur de bases de données Amazon Redshift.

Clusters Amazon Redshift provisionnés

Si vous utilisez Amazon Redshift pour la première fois, nous vous recommandons de lire les sections suivantes pour vous aider à faire vos premiers pas avec Amazon Redshift.

Rubriques

- [Console pour clusters provisionnés Amazon Redshift](#)
- [Connexion aux clusters provisionnés d'Amazon Redshift](#)
- [Clusters Amazon Redshift et chargement de données](#)
- [Tâches courantes liées aux bases de données](#)

Console pour clusters provisionnés Amazon Redshift

Avant de commencer à configurer un cluster Amazon Redshift, vérifiez que vous remplissez les prérequis suivants dans cette section :

- [S'inscrire à AWS](#)
- [Déterminer les règles de pare-feu](#)

S'inscrire à AWS

Si vous n'en avez pas encore Compte AWS, inscrivez-vous. Si vous disposez déjà d'un compte, vous pouvez ignorer ce prérequis et utiliser votre compte existant.

1. Ouvrez <https://portal.aws.amazon.com/billing/signup>.
2. Suivez les instructions en ligne.

Dans le cadre de la procédure d'inscription, vous recevrez un appel téléphonique et vous saisirez un code de vérification en utilisant le clavier numérique du téléphone.

Lorsque vous vous inscrivez à un Compte AWS, un Utilisateur racine d'un compte AWS est créé. Par défaut, seul l'utilisateur root a accès à l'ensemble des Services AWS et des ressources de ce compte. La meilleure pratique de sécurité consiste à [attribuer un accès administratif à un utilisateur administratif](#), et à uniquement utiliser l'utilisateur root pour effectuer [tâches nécessitant un accès utilisateur root](#).

Déterminer les règles de pare-feu

Dans le cadre de ce didacticiel, vous devez spécifier un port lorsque vous lancez votre cluster Amazon Redshift. Vous créez également une règle de trafic entrant dans un groupe de sécurité afin de permettre l'accès à votre cluster via le port.

Si votre ordinateur client se trouve derrière un pare-feu, assurez-vous de connaître un port ouvert que vous pouvez utiliser. Ce port ouvert vous permet de vous connecter au cluster à partir d'un outil client SQL et d'exécuter les requêtes. Si vous ne connaissez pas de port ouvert, travaillez avec quelqu'un qui comprend vos règles de pare-feu réseau afin de déterminer un port ouvert dans votre pare-feu.

Bien qu'Amazon Redshift utilise le port 5439 par défaut, la connexion ne fonctionne pas si ce port n'est pas ouvert dans votre pare-feu. Vous ne pouvez pas modifier le numéro de port de votre cluster Amazon Redshift après sa création. Par conséquent, assurez-vous que vous spécifiez un port ouvert qui fonctionne dans votre environnement au cours de la procédure de lancement.

Ce prérequis est valable uniquement si vous ajoutez vos données à Amazon Redshift. Pour plus d'informations, consultez [Importation de vos propres données dans Amazon Redshift](#).

Une fois connecté à la console Amazon Redshift, vous pouvez créer et gérer tous les objets Amazon Redshift, y compris des clusters, des bases de données et des nœuds. Vous pouvez également afficher des requêtes, exécuter des requêtes et effectuer d'autres opérations de langage de définition de données (DDL) et de langage de manipulation de données (DML).

Si vous utilisez Amazon Redshift pour la première fois, nous vous recommandons de commencer en accédant aux pages Dashboard (Tableau de bord), Clusters et Query editor v2 (Éditeur de requête v2) pour faire vos premiers pas avec la console.

Pour démarrer avec la console Amazon Redshift, regardez la vidéo suivante : [Getting started with Amazon Redshift](#).

Vous trouverez ci-dessous des descriptions des éléments du panneau de navigation de la console Amazon Redshift :

- Amazon Redshift sans serveur : accédez aux données et analysez-les sans avoir besoin de configurer, de régler et de gérer des clusters alloués Amazon Redshift.
- Provisioned clusters dashboard (Tableau de bord de clusters alloués) : cochez Cluster metrics (Métriques du cluster) et Query overview (Présentation des requêtes) pour obtenir des informations sur les données de métriques (telles que l'utilisation du processeur) et les informations de requête.

Leur utilisation peut vous aider à déterminer si vos données de performances sont anormales sur une plage de temps spécifiée.

- **Clusters** : consultez la liste des clusters de votre AWS compte, choisissez un cluster pour commencer à interroger ou effectuez des actions liées au cluster. Vous pouvez également créer un cluster à partir de cette page.
- **Query editor (Éditeur de requête)** : exécutez des requêtes sur des bases de données hébergées sur votre cluster Amazon Redshift, enregistrez les requêtes en vue d'une réutilisation ou planifiez leur exécution ultérieurement (dans l'éditeur de requêtes uniquement).
- **Query editor v2** : Query editor v2 est une application client SQL web distincte qui permet de créer et d'exécuter des requêtes sur votre entrepôt des données Amazon Redshift. Vous pouvez visualiser vos résultats dans des diagrammes et collaborer en partageant vos requêtes avec d'autres membres de votre équipe.
- **Queries and loads (Requêtes et charges)** : obtenez des informations à des fins de référence ou de dépannage, telles qu'une liste de requêtes récentes et le texte SQL de chaque requête.
- **Unités de partage des données** : un administrateur de compte producteur peut autoriser des comptes consommateur à accéder à des unités de partage des données ou choisir de ne pas autoriser l'accès. Pour utiliser un partage de données autorisé, un administrateur de compte client peut associer le partage de données à un compte complet ou à des espaces de noms de cluster spécifiques d'un AWS compte. Un administrateur peut également refuser une unité de partage des données.
- **Configurations** : connectez-vous aux clusters Amazon Redshift depuis les outils client SQL via des connexions Java Database Connectivity (JDBC) et Open Database Connectivity (ODBC). Vous pouvez également configurer un point de terminaison de cloud privé virtuel (VPC) géré par Amazon Redshift. Cela permet d'établir une connexion privée entre un VPC basé sur le service Amazon VPC qui contient un cluster et un autre VPC exécutant un outil client.
- **Advisor (Conseiller)** : obtenez des recommandations spécifiques concernant les modifications que vous pouvez apporter à votre cluster Amazon Redshift afin de prioriser vos optimisations.
- **AWS Marketplace** : obtenez des informations sur les autres outils ou AWS services compatibles avec Amazon Redshift.
- **Alarms (Alarmes)** : créez des alarmes sur les métriques de cluster pour afficher les données de performance et suivre les métriques sur une période que vous spécifiez.
- **Events (Événements)** : suivez les événements et obtenez des rapports sur des informations telles que la date à laquelle l'événement s'est produit, une description ou la source de l'événement.

- What's new (Nouveautés) : consultez les nouvelles fonctions et les mises à jour des produits Amazon Redshift.

Connexion aux clusters provisionnés d'Amazon Redshift

Pour se connecter à des clusters Amazon Redshift, à partir de la page Clusters, développez Connect to Amazon Redshift clusters (Se connecter aux clusters Amazon Redshift) et effectuez l'une des actions suivantes :

- Utilisez l'éditeur de requêtes v2 pour exécuter des requêtes sur les bases de données hébergées par votre cluster Amazon Redshift. Après avoir créé votre cluster, vous pouvez exécuter immédiatement des requêtes en utilisant l'éditeur de requêtes v2.

Pour plus d'informations, consultez [Interrogation d'une base de données à l'aide de l'éditeur de requêtes Amazon Redshift v2](#).

- Connectez-vous à Amazon Redshift depuis vos outils client à l'aide de pilotes JDBC ou ODBC en copiant l'URL du pilote JDBC ou ODBC.

Pour utiliser les données de votre cluster, vous avez besoin de pilotes ODBC ou JDBC pour la connectivité dans votre ordinateur ou instance client. Codez vos applications afin d'utiliser des opérations d'API d'accès aux données ODBC ou JDBC ou utilisez des outils clients SQL qui prennent en charge JDBC ou ODBC.

Pour plus d'informations sur la façon de trouver la chaîne de connexion de votre cluster, consultez [Recherche de votre chaîne de connexion au cluster](#).

- Si votre outil client SQL nécessite un pilote, vous pouvez télécharger un pilote spécifique au système d'exploitation pour vous connecter à Amazon Redshift à partir de vos outils client.

Pour plus d'informations sur l'installation du pilote approprié pour votre client SQL, consultez [Configuration d'une connexion au pilote JDBC version 2.0](#).

Pour plus d'informations sur la façon de configurer une connexion ODBC, consultez [Configuration d'une connexion ODBC](#).

Clusters Amazon Redshift et chargement de données

Dans cette section, vous trouverez deux didacticiels qui vous guideront dans le processus de création d'un cluster Amazon Redshift. Dans l'un, vous utilisez un exemple de jeu de données, et dans l'autre, vous apportez votre propre jeu de données.

Rubriques

- [Avec un exemple de jeu de données](#)
- [Importation de vos propres données dans Amazon Redshift](#)

Avec un exemple de jeu de données

Ce didacticiel vous guidera dans le processus de création d'un cluster Amazon Redshift en utilisant un exemple de jeu de données. Amazon Redshift charge automatiquement l'exemple de jeu de données lorsque vous créez un cluster. Vous pouvez immédiatement effectuer des requêtes sur les données après que le cluster est créé.

Avant de commencer à configurer un cluster Amazon Redshift, assurez-vous d'avoir effectué les étapes [S'inscrire à AWS](#) et [Déterminer les règles de pare-feu](#).

Dans ce didacticiel, vous exécuterez les étapes suivantes :



Step 1:
Create cluster



Step 2:
Try sample queries

Rubriques

- [Étape 1 : Créer un exemple de cluster Amazon Redshift](#)
- [Étape 2 : Essayer des exemples de requêtes à l'aide des éditeurs de requête](#)

Important

L'exemple de cluster que vous créez s'exécute dans un environnement en direct. Le tarif à la demande est de 0,25 USD par heure pour l'utilisation de l'exemple de cluster conçu dans ce didacticiel jusqu'à ce que vous le supprimiez. Pour plus d'informations sur la tarification, consultez [Tarification Amazon Redshift](#). Si vous avez des questions ou que vous êtes

bloqué, vous pouvez contacter l'équipe Amazon Redshift en écrivant dans notre [Forum de discussion](#).

Ce didacticiel n'est pas destiné aux environnements de production et ne traite pas des options en profondeur. Une fois que vous avez terminé les étapes de ce didacticiel, vous pouvez utiliser la section [Ressources supplémentaires](#) pour trouver des informations plus complètes. Ces informations peuvent vous aider à planifier, déployer et gérer vos clusters, et à utiliser les données dans votre entrepôt des données.

Étape 1 : Créer un exemple de cluster Amazon Redshift

Une fois les conditions préalables remplies, vous pouvez commencer à créer votre cluster Amazon Redshift, sur la base d'un exemple de jeu de données.

Pour créer un cluster Amazon Redshift basé sur un exemple de jeu de données :

1. [Connectez-vous à la console Amazon Redshift AWS Management Console et ouvrez-la à l'adresse https://console.aws.amazon.com/redshiftv2/](https://console.aws.amazon.com/redshiftv2/).
2. Pour créer un cluster, procédez comme suit :
 - Sur la page du service Amazon Redshift, choisissez Create cluster (Créer un cluster). La page Créer un cluster s'affiche.
 - Sur le [site https://console.aws.amazon.com/redshiftv2/](https://console.aws.amazon.com/redshiftv2/), choisissez le tableau de bord des clusters provisionnés, puis sélectionnez Create cluster.
 - Sur le [site https://console.aws.amazon.com/redshiftv2/](https://console.aws.amazon.com/redshiftv2/), choisissez Clusters, puis Create cluster.
3. Dans la section Cluster configuration (Configuration de cluster), spécifiez un Cluster identifier (Identifiant du cluster). Cet identifiant doit être unique. Il doit comporter de 1 à 63 caractères et doit être composé des caractères valides suivants : a-z (minuscules uniquement) et - (trait d'union).

Dans le cadre de ce didacticiel, saisissez **exemplecluster**.

4. Si votre organisation est éligible et que votre cluster est créé dans un environnement Région AWS où Amazon Redshift Serverless n'est pas disponible, vous pouvez peut-être créer un cluster dans le cadre du programme d'essai gratuit d'Amazon Redshift. Choisissez Production ou Essai gratuit pour répondre à la question Quelle est l'utilisation prévue de ce cluster ? Lorsque vous choisissez Essai gratuit, vous créez une configuration avec le type de nœud dc2.large.

Pour plus d'informations sur le choix d'un essai gratuit, accédez à [Essai gratuit d'Amazon Redshift](#). Pour obtenir la liste des Régions AWS endroits où Amazon Redshift Serverless est disponible, consultez les points de terminaison répertoriés pour l'API [Redshift](#) Serverless dans le. Référence générale d'Amazon Web Services

Après avoir choisi votre type de nœud, effectuez l'une des opérations suivantes :

- Dans Sample data (Exemples de données), choisissez Load sample data (Charger les exemples de données) pour charger l'exemple de jeu de données dans votre cluster Amazon Redshift. Amazon Redshift charge l'exemple de jeu de données Tackit dans la base de données dev et le schéma public par défaut. Vous pouvez commencer à utiliser l'éditeur de requêtes v2 pour interroger des données.
- Pour apporter vos propres données à votre cluster Amazon Redshift, choisissez Production. Puis, dans Sample data (Exemples de données), choisissez Load sample data (Charger les exemples de données). Pour plus d'informations sur l'ajout de vos propres données, consultez [Importation de vos propres données dans Amazon Redshift](#).

Amazon Redshift charge automatiquement l'exemple de jeu de données dans votre exemple de cluster Amazon Redshift.

5. Dans la section Configuration de la base de données, spécifiez une valeur pour Nom de l'utilisateur administrateur. Pour Mot de passe administrateur, choisissez l'une des options suivantes :

- Générez un mot de passe : utilisez un mot de passe généré par Amazon Redshift.
- Ajouter manuellement un mot de passe d'administrateur : utilisez votre propre mot de passe.
- Gérez les informations d'identification d'administrateur dans AWS Secrets Manager : Amazon Redshift utilise AWS Secrets Manager pour générer et gérer votre mot de passe d'administrateur. L'utilisation AWS Secrets Manager pour générer et gérer le secret de votre mot de passe entraîne des frais. Pour en savoir plus sur la tarification AWS Secrets Manager , consultez [Tarification d'AWS Secrets Manager](#).

Dans le cadre de ce didacticiel, utilisez les valeurs suivantes :

- Admin user name (Nom de l'utilisateur Admin) : saisissez **awsuser**.
- Admin user password (Mot de passe d'utilisateur de l'administrateur) : saisissez une valeur pour le mot de passe.

6. Choisissez Créer un cluster.

Ce didacticiel utilise l'éditeur de requêtes Amazon Redshift v2. Vous pouvez utiliser cet éditeur pour interroger des données immédiatement, une fois qu'Amazon Redshift a fini de créer le cluster.

Vous pouvez également choisir d'autres outils clients SQL qui prennent en charge les pilotes JDBC ou ODBC pour travailler avec les données de votre cluster. Pour plus d'informations, consultez [Connexion à un cluster Amazon Redshift à l'aide des outils client SQL](#) dans le Guide de gestion Amazon Redshift.

Étape 2 : Essayer des exemples de requêtes à l'aide des éditeurs de requête

Lorsqu'Amazon Redshift crée votre cluster Amazon Redshift, il télécharge automatiquement l'exemple de jeu de données Tickit. La création du cluster peut prendre quelques minutes. Une fois la création terminée, l'état du cluster devient ACTIVE. Vous pouvez afficher les exemples de tables Tickit à partir de l'exemple de jeu de données.

Utilisation de l'éditeur de requêtes

Vous pouvez afficher les exemples de tables Tickit dans l'éditeur de requêtes v2 en choisissant le cluster, la base de données dev et le schéma public.

Une fois le cluster Amazon Redshift créé, dans Connect to Amazon Redshift cluster (Connexion aux clusters Amazon Redshift), choisissez Query data (Requête de données).

Dans l'éditeur de requêtes v2, connectez-vous à une base de données et choisissez le nom du cluster dans le panneau d'arborescence. Si vous y êtes invité, saisissez les paramètres de connexion.

Lorsque vous vous connectez à un cluster et à ses bases de données, vous fournissez un nom de base de données et un nom d'utilisateur. Vous fournissez également les paramètres requis pour l'une des méthodes d'authentification suivantes :

Nom d'utilisateur et mot de passe de la base de données

Avec cette méthode, fournissez également un mot de passe pour la base de données à laquelle vous vous connectez.

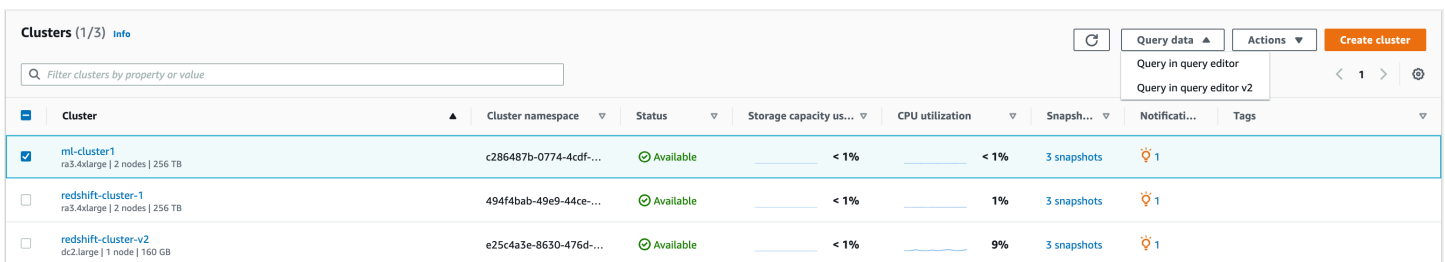
Informations d'identification temporaires

Avec cette méthode, l'éditeur de requêtes v2 génère un mot de passe temporaire pour se connecter à la base de données.

Lorsque vous sélectionnez un cluster avec l'éditeur de requêtes v2, selon le contexte, vous pouvez créer, modifier et supprimer des connexions à l'aide du menu contextuel (clic droit).

Par défaut, Amazon Redshift crée une base de données par défaut nommée `dev` et un schéma par défaut nommé `public`. Pour afficher les fichiers de données individuels de l'exemple de jeu de données, choisissez un cluster, allez dans l'éditeur de requêtes v2 et choisissez la base de données `dev`, le schéma `public`, puis `Tables`.

Sinon, dans le panneau de navigation, sélectionnez `Clusters`, puis le cluster sur lequel vous souhaitez interroger des données. Ensuite, sous `Interroger des données`, choisissez `Interroger` dans l'éditeur de requêtes ou `Interroger dans l'éditeur de requêtes v2` pour interroger des données dans votre éditeur de requêtes spécifié.



Cluster	Cluster namespace	Status	Storage capacity us...	CPU utilization	Snapsh...	Notificati...	Tags
<input checked="" type="checkbox"/> ml-cluster1 ra3.4xlarge 2 nodes 256 TB	c286487b-0774-4cdf-...	Available	< 1%	< 1%	3 snapshots	1	
<input type="checkbox"/> redshift-cluster-1 ra3.4xlarge 2 nodes 256 TB	494f4bab-49e9-44ce-...	Available	< 1%	1%	3 snapshots	1	
<input type="checkbox"/> redshift-cluster-v2 dc2.large 1 node 160 GB	e25c4a3e-8630-476d-...	Available	< 1%	9%	3 snapshots	1	

Essayer des exemples de requêtes

Essayez quelques exemples de requêtes dans l'un des éditeurs de requête, comme indiqué ci-dessous. Pour plus d'informations sur l'utilisation de l'instruction `SELECT`, consultez [SELECT](#) du Manuel du développeur de base de données Amazon Redshift.

```
-- Find total sales on a given calendar date.
SELECT sum(qtysold)
FROM   sales, date
WHERE  sales.dateid = date.dateid
AND    caldate = '2008-01-05';

-- Find top 10 buyers by quantity.
SELECT firstname, lastname, total_quantity
FROM   (SELECT buyerid, sum(qtysold) total_quantity
        FROM   sales
        GROUP BY buyerid
        ORDER BY total_quantity desc limit 10) Q, users
WHERE  Q.buyerid = userid
ORDER BY Q.total_quantity desc;

-- Find events in the 99.9 percentile in terms of all time gross sales.
```

```
SELECT eventname, total_price
FROM (SELECT eventid, total_price, ntile(1000) over(order by total_price desc) as
percentile
      FROM (SELECT eventid, sum(pricepaid) total_price
            FROM sales
            GROUP BY eventid)) Q, event E
WHERE Q.eventid = E.eventid
AND percentile = 1
ORDER BY total_price desc;
```

Une fois que vous avez terminé les étapes de ce didacticiel, vous pouvez utiliser la section [Ressources supplémentaires](#) pour trouver des informations plus complètes. Ces informations peuvent vous aider à planifier, déployer et gérer vos clusters, et à utiliser les données dans votre entrepôt des données.

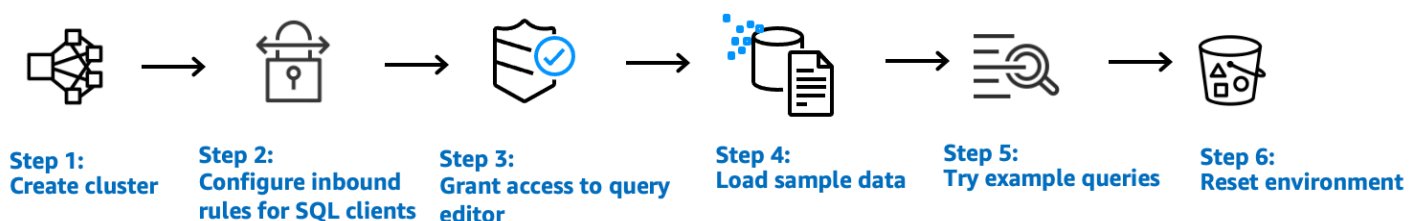
Vous pouvez également essayer le didacticiel [Importation de vos propres données dans Amazon Redshift](#) pour créer un cluster avec votre propre jeu de données.

Importation de vos propres données dans Amazon Redshift

Dans ce didacticiel, vous suivez le processus de création d'un cluster Amazon Redshift en important votre propre jeu de données dans Amazon Redshift. Vous pouvez utiliser cet exemple de cluster pour évaluer le service Amazon Redshift.

Avant de commencer à configurer un cluster Amazon Redshift, assurez-vous d'avoir effectué les étapes [S'inscrire à AWS](#) et [Déterminer les règles de pare-feu](#).

Dans ce didacticiel, vous effectuerez les étapes suivantes.



⚠ Important

L'exemple de cluster que vous créez s'exécute dans un environnement en direct. Le tarif à la demande est de 0,25 USD par heure pour l'utilisation de l'exemple de cluster conçu dans ce didacticiel jusqu'à ce que vous le supprimiez. Pour plus d'informations sur la tarification,

consultez la [Page de tarification Amazon Redshift](#). Si vous avez des questions ou que vous êtes bloqué, vous pouvez contacter l'équipe Amazon Redshift en écrivant dans notre [Forum de discussion](#).

Ce didacticiel n'est pas destiné aux environnements de production et ne traite pas des options en profondeur. Une fois que vous avez terminé les étapes de ce didacticiel, vous pouvez utiliser la section [Ressources supplémentaires](#) pour trouver des informations plus complètes. Ces informations peuvent vous aider à planifier, déployer et gérer vos clusters, et à utiliser les données dans votre entrepôt des données.

Rubriques

- [Étape 1 : Créer un exemple de cluster Amazon Redshift](#)
- [Étape 2 : Configurer les règles entrantes pour les clients SQL](#)
- [Étape 3 : Accorder des privilèges d'accès aux éditeurs de requête et exécuter des requêtes](#)
- [Étape 4 : Charger les données d'Amazon S3 vers Amazon Redshift](#)
- [Étape 5 : Essayer des exemples de requêtes à l'aide de l'éditeur de requêtes](#)
- [Étape 6 : Réinitialiser votre environnement](#)

Étape 1 : Créer un exemple de cluster Amazon Redshift

Pour toute opération qui accède aux données d'une autre AWS ressource, votre cluster doit être autorisé à accéder à la ressource et aux données de la ressource en votre nom. Un exemple est l'utilisation d'une instruction COPY pour charger des données depuis Amazon Simple Storage Service (Amazon S3). Vous fournissez ces autorisations en utilisant AWS Identity and Access Management (IAM). Vous pouvez effectuer cette opération via un rôle IAM attaché à votre cluster. Vous pouvez également fournir la clé AWS d'accès à un utilisateur IAM disposant des autorisations nécessaires. Pour plus d'informations sur les informations d'identification et les autorisations d'accès, consultez [Informations d'identification et autorisations d'accès](#).


Pour protéger au mieux vos données sensibles et vos informations d' AWS accès, nous vous recommandons de créer un rôle IAM et de l'associer à votre cluster. Pour plus d'informations sur l'octroi d'autorisations d'accès, consultez [Autorisations d'accès aux autres ressources AWS](#).

Le cluster que vous êtes sur le point de créer est opérationnel et non pas exécuté dans un environnement de test (sandbox). Des frais d'utilisation sont perçus pour l'utilisation d'Amazon

Redshift pour le cluster jusqu'à ce que vous le supprimiez. Si vous terminez le didacticiel décrit ici en une seule fois et que vous supprimez votre cluster lorsque vous avez terminé, le montant total des frais est minimal.


Pour créer un cluster Amazon Redshift

1. [Connectez-vous à la console Amazon Redshift AWS Management Console et ouvrez-la à l'adresse https://console.aws.amazon.com/redshiftv2/.](https://console.aws.amazon.com/redshiftv2/)

 Important

Si vous utilisez les informations d'identification utilisateur IAM, vérifiez que l'utilisateur dispose des autorisations nécessaires pour exécuter les opérations de cluster. Pour plus d'informations, consultez [Contrôle de l'accès aux utilisateurs IAM](#) dans le Guide de gestion Amazon Redshift.

2. En haut à droite, choisissez l' Région AWS endroit où vous souhaitez créer le cluster.
3. Dans le menu de navigation, choisissez Clusters, puis choisissez Créer un cluster. La page Créer un cluster s'affiche.
4. Dans la section Cluster configuration (Configuration du cluster), spécifiez les valeurs Cluster identifier (Identifiant du cluster), du Node type (Type de nœud) et Nodes (Nœuds) :
 - Cluster identifier (Identifiant du cluster) : saisissez **exemplecluster** pour ce didacticiel. Cet identifiant doit être unique. Il doit comporter de 1 à 63 caractères et doit être composé des caractères valides suivants : a-z (minuscules uniquement) et - (trait d'union).
 - Choisissez l'une des méthodes suivantes pour dimensionner votre cluster :

 Note

L'étape suivante suppose une AWS région qui prend en charge les types de nœuds RA3. Pour obtenir une liste des régions AWS qui prennent en charge les types de nœuds RA3, consultez [Présentation des types de nœuds RA3](#) dans le Guide de la gestion du cluster Amazon Redshift. Pour en savoir plus sur les spécifications de chaque type et taille de nœud, consultez [Détails du type de nœud](#).

- Si votre AWS région prend en charge les types de nœuds RA3, choisissez Production ou Essai gratuit pour répondre à la question Pourquoi prévoyez-vous d'utiliser ce cluster ?

- Si vous ne savez pas quelle taille donner à votre cluster, choisissez Help me choose (Aidez-moi à choisir). Cette opération lance un calculateur de dimensionnement qui vous pose des questions sur la taille et les caractéristiques d'interrogation des données que vous prévoyez de stocker dans votre entrepôt des données.

Si vous connaissez la taille requise de votre cluster (c'est-à-dire le type et le nombre de nœuds), choisissez I'll choose (Je vais choisir). Choisissez ensuite le Node type (Type de nœud) et le nombre de Nodes (Nœuds) pour dimensionner votre cluster pour la preuve de concept.

- Sélectionnez dc2.large pour Node type (Type de nœud) et 2 pour Nodes (Nœuds) pour ce didacticiel.
 - Si vous avez choisi Production pour votre cluster, effectuez l'une des opérations suivantes :
 - Pour utiliser l'exemple de jeu de données fourni par Amazon Redshift, dans Sample data (Exemples de données), choisissez Load sample data (Charger les exemples de données). Amazon Redshift charge l'exemple de jeu de données Tockit dans la base de données dev et le schéma public par défaut.
 - Pour apporter vos propres données à Amazon Redshift, continuez avec le reste du didacticiel.
5. Dans la section Configuration de la base de données, spécifiez une valeur pour Nom de l'utilisateur administrateur. Pour Mot de passe administrateur, choisissez l'une des options suivantes :
- Générez un mot de passe : utilisez un mot de passe généré par Amazon Redshift.
 - Ajouter manuellement un mot de passe d'administrateur : utilisez votre propre mot de passe.
 - Gérez les informations d'identification d'administrateur dans AWS Secrets Manager : Amazon Redshift utilise AWS Secrets Manager pour générer et gérer votre mot de passe d'administrateur. L'utilisation AWS Secrets Manager pour générer et gérer le secret de votre mot de passe entraîne des frais. Pour en savoir plus sur la tarification AWS Secrets Manager , consultez [Tarification d'AWS Secrets Manager](#).

Dans le cadre de ce didacticiel, utilisez les valeurs suivantes :

- Admin user name (Nom de l'utilisateur Admin) : saisissez **awsuser**.
- Admin user password (Mot de passe d'utilisateur de l'administrateur) : saisissez une valeur pour le mot de passe.

6. Pour ce didacticiel, créez un rôle IAM et définissez-le comme rôle par défaut pour votre cluster, comme décrit ci-dessous. Il ne peut y avoir qu'un seul rôle IAM par défaut défini pour un cluster.
 - a. Sous Cluster permissions (Autorisations de cluster), pour Manage IAM roles (Gérer les rôles IAM), choisissez Create IAM role (Créer un rôle IAM).
 - b. Spécifiez un compartiment Amazon S3 auquel le rôle IAM doit accéder à l'aide de l'une des méthodes suivantes :
 - Choisissez Pas de compartiment Amazon S3 supplémentaire pour permettre au rôle IAM créé d'accéder uniquement aux compartiments Amazon S3 nommés en tant que `redshift`.
 - Choisissez N'importe quel compartiment Amazon S3 pour permettre au rôle IAM créé d'accéder à tous les compartiments Amazon S3.
 - Choisissez Specific Amazon S3 buckets (Compartiments Amazon S3 spécifiques) pour spécifier un ou plusieurs compartiments Amazon S3 pour le rôle IAM en cours de création est autorisé à accéder. Choisissez ensuite un ou plusieurs compartiments Amazon S3 dans le tableau.
 - c. Choisissez Create IAM role as default (Créer un rôle IAM par défaut). Amazon Redshift crée et définit automatiquement le rôle IAM comme rôle par défaut pour votre cluster.

Étant donné que vous avez créé votre rôle IAM à partir de la console, il a la politique `AmazonRedshiftAllCommandsFullAccess` attachée. Cela permet à Amazon Redshift de copier, de charger, d'interroger et d'analyser des données provenant des ressources Amazon dans votre compte IAM.

Pour obtenir des informations sur la façon de gérer le rôle IAM par défaut pour un cluster, consultez [Création d'un rôle IAM par défaut pour Amazon Redshift](#).

7. (Facultatif) Dans la section Additional configurations (Configurations supplémentaires), désactivez l'option Use defaults (Utiliser les valeurs par défaut) pour modifier les paramètres Network and security (Réseau et sécurité), Database configuration (Configuration de la base de données), Maintenance, Monitoring (Surveillance) et Backup (Sauvegarde).

Dans certains cas, vous pouvez créer votre cluster avec l'option Load sample data (Charger des exemples de données) et souhaiter activer le routage Amazon VPC amélioré. Si c'est le cas, le cluster de votre cloud privé virtuel (VPC) nécessite l'accès au point de terminaison Amazon S3 pour que les données soient chargées.

Pour rendre le cluster accessible au public, vous pouvez effectuer l'une des deux actions suivantes. Vous pouvez configurer une traduction d'adresses réseau (NAT) dans votre VPC pour que le cluster puisse accéder à Internet. Vous pouvez également configurer un point de terminaison d'un VPC Amazon S3 dans votre VPC. Pour plus d'informations sur le routage Amazon VPC amélioré, consultez [Activation du routage Amazon VPC amélioré](#) dans le Guide de gestion Amazon Redshift.

8. Choisissez Créer un cluster.

Étape 2 : Configurer les règles entrantes pour les clients SQL

Plus loin dans ce didacticiel, vous accéderez à votre cluster à partir d'un cloud privé virtuel (VPC) basé sur le service Amazon VPC. Toutefois, si vous utilisez un client SQL qui est situé en dehors de votre pare-feu pour accéder au cluster, assurez-vous d'accorder un accès entrant.

Vous pouvez ignorer cette étape si vous avez l'intention d'accéder au cluster avec l'éditeur de requêtes Amazon Redshift depuis l'intérieur de votre VPC.

Pour vérifier votre pare-feu et octroyer un accès entrant à votre cluster

1. Vérifiez les règles de votre pare-feu si votre cluster doit faire l'objet d'un accès depuis l'extérieur du pare-feu. Par exemple, votre client peut être une instance Amazon Elastic Compute Cloud (Amazon EC2) ou un ordinateur externe.

Pour plus d'informations sur les règles de pare-feu, consultez [Règles des groupes de sécurité](#) dans le Guide de l'utilisateur Amazon EC2 pour les instances Linux.

2. Pour accéder au cluster depuis un client externe Amazon EC2, ajoutez une règle d'entrée au groupe de sécurité associé à votre cluster afin d'autoriser le trafic entrant. Vous ajoutez des règles de groupe de sécurité Amazon EC2 dans la console Amazon EC2. Par exemple, un CIDR/IP de 192.0.2.0/24 autorise les clients de cette plage d'adresses IP à se connecter à votre cluster. Cherchez le CIDR/IP de votre environnement.

Étape 3 : Accorder des privilèges d'accès aux éditeurs de requête et exécuter des requêtes

Pour interroger les bases de données hébergées par votre cluster Amazon Redshift, deux options s'offrent à vous :

1. Connectez-vous à votre cluster et exécutez des requêtes à l' AWS Management Console aide de l'un des éditeurs de requêtes.

Si vous utilisez l'un des éditeurs de requête, vous n'avez pas besoin de télécharger et configurer une application client SQL.

2. Connectez-vous à votre cluster par le biais d'un outil client SQL, tel que SQL Workbench/J. Pour plus d'informations sur l'utilisation de SQL Workbench/J, consultez [Se connecter à votre cluster en utilisant SQL Workbench/J](#) dans le Guide de gestion Amazon Redshift.

Utiliser l'un des éditeurs de requête Amazon Redshift est la façon la plus simple d'exécuter des requêtes sur les bases de données hébergées par votre cluster Amazon Redshift. Après avoir créé votre cluster, vous pouvez exécuter immédiatement des requêtes en utilisant la console Amazon Redshift. Pour obtenir des informations sur les éléments à prendre en compte lorsque vous utilisez l'éditeur de requêtes Amazon Redshift, consultez [Interrogation d'une base de données à l'aide de l'éditeur de requêtes](#) dans le Guide de gestion Amazon Redshift.

Accorder des privilèges d'accès à l'éditeur de requêtes v2

La première fois qu'un administrateur configure l'éditeur de requêtes v2 pour votre Compte AWS, il choisit AWS KMS key celui qui est utilisé pour chiffrer les ressources de l'éditeur de requêtes v2. Par défaut, une clé appartenant à AWS est utilisée pour chiffrer les ressources. Un administrateur peut également utiliser une clé gérée par le client en choisissant l'Amazon Resource Name (ARN) pour la clé dans la page de configuration. Une fois que vous avez configuré un compte, les paramètres de AWS KMS chiffrement ne peuvent pas être modifiés. Pour plus d'informations, consultez [Configuration de votre Compte AWS](#).

Pour accéder à l'éditeur de requêtes v2, vous avez besoin d'une autorisation. Un administrateur peut associer l'une des politiques AWS gérées suivantes au rôle ou à l'utilisateur IAM pour accorder des autorisations. Ces politiques AWS gérées sont rédigées avec différentes options qui contrôlent la manière dont le balisage des ressources permet le partage des requêtes. Vous pouvez attacher des politiques IAM à l'aide de la console IAM à l'adresse <https://console.aws.amazon.com/iam/>.

Vous pouvez également créer votre propre politique basée sur les autorisations autorisées et refusées dans les politiques gérées fournies. Si vous utilisez l'éditeur de politique de la console IAM pour créer votre propre politique, choisissez SQL Workbench en tant que service pour lequel vous créez la politique dans l'éditeur visuel. L'éditeur de requêtes v2 utilise le nom de service AWS SQL Workbench dans l'éditeur visuel et dans le simulateur de politique IAM.

Pour plus d'informations, consultez [Accessing the query editor v2](#).

Utilisation de l'éditeur de requêtes v2

Pour utiliser l'éditeur de requêtes v2 pour interroger une base de données, consultez [Utilisation de l'éditeur de requêtes v2](#).

Accorder des privilèges d'accès à l'éditeur de requêtes

Pour utiliser l'éditeur de requêtes Amazon Redshift, vous avez besoin d'une autorisation. Pour activer l'accès, attachez les politiques IAM `AmazonRedshiftQueryEditor` et `AmazonRedshiftReadOnlyAccess` à un rôle IAM, puis assignez ce rôle à l'utilisateur que vous utilisez pour accéder à votre cluster.

Pour attacher les politiques IAM requises pour l'éditeur de requêtes

1. Connectez-vous à la console IAM AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/iam/>.
2. Sélectionnez Roles (Rôles).
3. Choisissez le rôle qui a besoin d'accéder à l'éditeur de requêtes.
4. Choisissez Ajouter des autorisations sous l'onglet Autorisation.
5. Choisissez Attach Policies (Attacher des politiques).
6. Pour Noms des politiques, sélectionnez `AmazonRedshiftQueryEditor` et `AmazonRedshiftReadOnlyAccess`.
7. Choisissez Add permissions (Ajouter des autorisations).


Pour créer un rôle IAM afin d'autoriser votre cluster Amazon Redshift à communiquer avec d'autres services AWS en votre nom, procédez comme suit. Les valeurs utilisées dans cette section sont des exemples. Vous pouvez choisir des valeurs en fonction de vos besoins.

Pour créer un rôle IAM afin de permettre à Amazon Redshift d'accéder aux services AWS

1. Ouvrez la [console IAM](#).
2. Dans le panneau de navigation, choisissez Roles (Rôles).
3. Sélectionnez Create role (Créer un rôle).
4. Choisissez Service AWS , puis Redshift.
5. Sous Sélectionner votre cas d'utilisation, choisissez Redshift - Personnalisable, puis Suivant : Autorisations. La page Attacher une politique d'autorisations s'affiche.
6. Ajoutez les `AmazonRedshiftReadOnlyAccess` autorisations `AmazonRedshiftQueryEditor`.

Choisissez Suivant : Balises.

7. La page Ajouter des balises s'affiche. Si vous le souhaitez, vous pouvez ajouter des balises. Choisissez Suivant : vérification.
8. Pour Nom du rôle, indiquez le nom de votre rôle, par exemple **RedshiftQueryEditorAccess**. Sélectionnez Create role (Créer un rôle).
9. Le nouveau rôle est disponible pour tous les utilisateurs des clusters qui utilisent ce rôle. Pour limiter l'accès à des utilisateurs spécifiques uniquement sur certains clusters, ou sur des clusters dans des régions spécifiques, modifiez la relation d'approbation de ce rôle. Pour plus d'informations, consultez [Restriction de l'accès aux rôles IAMM](#) dans le Guide de gestion Amazon Redshift.
10. Associez le rôle à votre cluster. Vous pouvez associer un rôle IAM à un cluster lors de la création du cluster, ou vous pouvez ajouter le rôle à un cluster existant. Pour plus d'informations, consultez [Association des rôles IAM aux clusters](#) dans le Guide de gestion Amazon Redshift.

 Note

Pour restreindre l'accès à des données spécifiques, utilisez un rôle IAM qui accorde les moindres privilèges requis.

Utilisation de l'éditeur de requêtes

Pour utiliser l'éditeur de requêtes pour interroger une base de données, consultez [Interrogation d'une base de données à l'aide de l'éditeur de requêtes](#).

Étape 4 : Charger les données d'Amazon S3 vers Amazon Redshift

L'utilisation de l'un des éditeurs de requête Amazon Redshift est le moyen le plus simple de charger des données vers des tables. Après avoir créé votre cluster, vous pouvez charger des données d'Amazon S3 vers votre cluster à l'aide de la console Amazon Redshift.

Utiliser l'éditeur de requêtes v2 simplifie le chargement des données à l'aide de l'assistant de chargement des données. Pour plus d'informations, consultez [Chargement de vos propres données d'Amazon S3 vers Amazon Redshift à l'aide de l'éditeur de requêtes v2](#). Vous pouvez également utiliser l'éditeur de requêtes v2 pour créer des tables et charger vos données. Pour plus d'informations, consultez [Chargement d'exemples de données depuis Amazon S3 à l'aide de l'éditeur de requêtes](#).

Chargement de vos propres données d'Amazon S3 vers Amazon Redshift à l'aide de l'éditeur de requêtes v2

Pour charger vos propres données depuis Amazon S3 vers Amazon Redshift, Amazon Redshift nécessite un rôle IAM disposant des privilèges requis pour charger des données à partir du compartiment Amazon S3 spécifié.

Connectez-vous d'abord à une base de données. Ensuite, créez des tables dans la base de données. Ensuite, chargez vos propres données d'Amazon S3 vers Amazon Redshift. Pour plus d'informations sur l'utilisation de l'éditeur de requêtes v2, consultez [Working with query editor v2](#) dans le Guide de gestion Amazon Redshift.

La commande COPY générée et utilisée dans l'assistant de chargement de données de l'éditeur de requêtes v2 prend en charge tous les paramètres disponibles pour la syntaxe de la commande COPY afin de charger des données depuis Amazon S3. Pour plus d'informations sur la commande COPY et ses options utilisées pour copier une charge à partir d'Amazon S3, consultez [Commande COPY depuis Amazon Simple Storage Service](#) dans le Guide du développeur de base de données Amazon Redshift.

Chargement d'exemples de données depuis Amazon S3 à l'aide de l'éditeur de requêtes

À ce stade, vous disposez d'une base de données appelée dev à laquelle vous êtes connecté. Ensuite, vous créez des tables dans la base de données, chargez des données dans les tables et essayez d'exécuter une requête. Pour plus de commodité, les exemples de données que vous chargez sont disponibles dans un compartiment Amazon S3.

Note

Si vous utilisez un outil client SQL, veillez à ce que votre client SQL soit connecté au cluster.

Une fois que vous avez terminé cette étape, procédez comme suit :

- Essayez des exemples de requêtes sur [Étape 5 : Essayer des exemples de requêtes à l'aide de l'éditeur de requêtes](#).
- Réinitialisez votre environnement à [Étape 6 : Réinitialiser votre environnement](#).
- Vous trouverez plus d'informations sur Amazon Redshift en consultant [Ressources supplémentaires](#).

Note

Pour essayer d'interroger des données dans l'éditeur de requêtes sans charger vos propres données, sélectionnez Load sample data (Charger des exemples de données) dans Sample data (Exemples de données). Si vous le faites, Amazon Redshift charge son exemple de jeu de données dans votre cluster Amazon Redshift automatiquement pendant la création du cluster.

Pour charger des exemples de données à partir d'Amazon S3

1. Créez des tables.

Si vous utilisez l'éditeur de requêtes Amazon Redshift, copiez et exécutez individuellement les instructions de création de table suivantes pour créer des tables dans la base de données dev. Pour plus d'informations sur la syntaxe, consultez [CREATE TABLE](#) dans le Guide du développeur de base de données Amazon Redshift.

```
create table users(  
  userid integer not null distkey sortkey,  
  username char(8),  
  firstname varchar(30),  
  lastname varchar(30),  
  city varchar(30),  
  state char(2),  
  email varchar(100),  
  phone char(14),  
  likesports boolean,  
  liketheatre boolean,  
  likeconcerts boolean,  
  likejazz boolean,  
  likeclassical boolean,  
  likeopera boolean,  
  likerock boolean,  
  likevegas boolean,  
  likebroadway boolean,  
  likemusicals boolean);
```

```
create table venue(  

```

```
venueid smallint not null distkey sortkey,  
venue name varchar(100),  
venue city varchar(30),  
venue state char(2),  
venue seats integer);
```

```
create table category(  
catid smallint not null distkey sortkey,  
catgroup varchar(10),  
catname varchar(10),  
catdesc varchar(50));
```

```
create table date(  
dateid smallint not null distkey sortkey,  
caldate date not null,  
day character(3) not null,  
week smallint not null,  
month character(5) not null,  
qtr character(5) not null,  
year smallint not null,  
holiday boolean default('N'));
```

```
create table event(  
eventid integer not null distkey,  
venueid smallint not null,  
catid smallint not null,  
dateid smallint not null sortkey,  
eventname varchar(200),  
starttime timestamp);
```

```
create table listing(  
listid integer not null distkey,  
sellerid integer not null,  
eventid integer not null,  
dateid smallint not null sortkey,  
numtickets smallint not null,  
priceperticket decimal(8,2),  
totalprice decimal(8,2),  
listtime timestamp);
```

```
create table sales(  
  salesid integer not null,  
  listid integer not null distkey,  
  sellerid integer not null,  
  buyerid integer not null,  
  eventid integer not null,  
  dateid smallint not null sortkey,  
  qty sold smallint not null,  
  pricepaid decimal(8,2),  
  commission decimal(8,2),  
  saletime timestamp);
```

2. Chargez des exemples de données depuis Amazon S3 en utilisant l'instruction COPY.

Note

Nous recommandons d'utiliser l'instruction COPY pour charger de grands jeux de données dans Amazon Redshift à partir d'Amazon S3 ou d'Amazon DynamoDB. Pour plus d'informations sur la syntaxe COPY, consultez [COPY](#) dans le Guide du développeur de la base de données Amazon Redshift.

- a. Téléchargez le fichier [ticketdb.zip](#), qui contient des exemples de fichier individuel de données.
- b. Décompressez et chargez les fichiers individuels dans un dossier `ticket` de votre compartiment Amazon S3 dans votre Région AWS.
- c. Modifiez les commandes COPY de ce didacticiel de manière à pointer vers les fichiers de votre compartiment Amazon S3. Pour plus d'informations sur la gestion des fichiers avec Amazon S3, consultez [Création et configuration d'un compartiment S3](#) dans le Guide de l'utilisateur Amazon Simple Storage Service.
- d. Pour charger les exemples de données, vous devez fournir l'authentification de votre cluster afin qu'il accède à Amazon S3 en votre nom. Vous fournissez l'authentification en faisant référence au rôle IAM que vous avez créé et que vous avez défini comme rôle par défaut pour votre cluster au cours des étapes précédentes.

Les commandes COPY incluent un espace réservé pour le nom de ressource Amazon (ARN) pour le rôle IAM, le nom de votre compartiment et un Région AWS, comme indiqué dans l'exemple suivant.

```
copy users from 's3://<myBucket>/ticket/allusers_pipe.txt'  
iam_role default  
delimiter '|' region '<aws-region>;
```

Votre commande COPY doit être similaire à l'exemple suivant.

```
copy users from 's3://<myBucket>/ticket/allusers_pipe.txt'  
iam_role default  
delimiter '|' region '<aws-region>;
```

Pour charger les exemples de données, remplacez *<myBucket>* , et *<aws-region>* par vos valeurs dans les instructions COPY suivantes. Si vous utilisez l'éditeur de requêtes Amazon Redshift, exécutez individuellement les instructions suivantes.

```
copy users from 's3://<myBucket>/ticket/allusers_pipe.txt'  
iam_role default  
delimiter '|' region '<aws-region>;
```

```
copy venue from 's3://<myBucket>/ticket/venue_pipe.txt'  
iam_role default  
delimiter '|' region '<aws-region>;
```

```
copy category from 's3://<myBucket>/ticket/category_pipe.txt'  
iam_role default  
delimiter '|' region '<aws-region>;
```

```
copy date from 's3://<myBucket>/ticket/date2008_pipe.txt'  
iam_role default  
delimiter '|' region '<aws-region>;
```

```
copy event from 's3://<myBucket>/ticket/allevvents_pipe.txt'  
iam_role default  
delimiter '|' timeformat 'YYYY-MM-DD HH:MI:SS' region '<aws-region>;
```

```
copy listing from 's3://<myBucket>/ticket/listings_pipe.txt'  
iam_role default
```

```
delimiter '|' region '<aws-region>';
```

```
copy sales from 's3://<myBucket>/tickit/sales_tab.txt'  
iam_role default  
delimiter '\t' timeformat 'MM/DD/YYYY HH:MI:SS' region '<aws-region>';
```

Étape 5 : Essayer des exemples de requêtes à l'aide de l'éditeur de requêtes

Maintenant, essayez quelques exemples de requêtes, comme indiqué ci-dessous. Pour plus d'informations sur l'utilisation de l'instruction SELECT, consultez [SELECT](#) dans le Guide du développeur Amazon Redshift.

```
-- Get definition for the sales table.  
SELECT *  
FROM pg_table_def  
WHERE tablename = 'sales';  
  
-- Find total sales on a given calendar date.  
SELECT sum(qtysold)  
FROM sales, date  
WHERE sales.dateid = date.dateid  
AND caldate = '2008-01-05';  
  
-- Find top 10 buyers by quantity.  
SELECT firstname, lastname, total_quantity  
FROM (SELECT buyerid, sum(qtysold) total_quantity  
FROM sales  
GROUP BY buyerid  
ORDER BY total_quantity desc limit 10) Q, users  
WHERE Q.buyerid = userid  
ORDER BY Q.total_quantity desc;  
  
-- Find events in the 99.9 percentile in terms of all time gross sales.  
SELECT eventname, total_price  
FROM (SELECT eventid, total_price, ntile(1000) over(order by total_price desc) as  
percentile  
FROM (SELECT eventid, sum(pricepaid) total_price  
FROM sales  
GROUP BY eventid)) Q, event E  
WHERE Q.eventid = E.eventid  
AND percentile = 1
```



```
ORDER BY total_price desc;
```

Vous avez créé avec succès un cluster Amazon Redshift et interrogé les données de votre propre jeu de données à l'aide de l'éditeur de requêtes Amazon Redshift.

Étape 6 : Réinitialiser votre environnement

Lorsque vous avez terminé ce didacticiel, nous vous recommandons de réinitialiser votre environnement à l'état précédent en supprimant l'exemple de cluster. Vous continuez à payer des frais pour le service Amazon Redshift tant que vous n'avez pas supprimé le cluster.

Cependant, vous souhaitez peut-être conserver l'exemple de cluster en cours d'exécution si vous avez l'intention de tester des tâches d'autres manuels Amazon Redshift.

Pour supprimer un cluster

1. [Connectez-vous à la console Amazon Redshift AWS Management Console et ouvrez-la à l'adresse `https://console.aws.amazon.com/redshiftv2/`.](https://console.aws.amazon.com/redshiftv2/)
2. Dans le menu de navigation, choisissez Clusters pour afficher la liste des clusters.
3. Choisissez le cluster **examplecluster**. Pour Actions, choisissez Supprimer. La page Supprimer un cluster s'affiche.
4. Confirmez le cluster à supprimer, puis choisissez Supprimer le cluster.

Sur la page de la liste des clusters, l'état du cluster est mis à jour afin de refléter la suppression du cluster.

Après avoir terminé ce didacticiel, vous pourrez trouver des informations supplémentaires sur Amazon Redshift et les étapes suivantes dans [Ressources supplémentaires](#).

Tâches courantes liées aux bases de données

Vous trouverez ci-dessous une description et une marche à suivre pour les tâches courantes afin que vous puissiez commencer à utiliser les bases de données Amazon Redshift.

Une fois que vous vous êtes connecté au cluster initial dev, vous pouvez créer une nouvelle base de données. Que vous choisissiez d'utiliser l'exemple de jeu de données ou d'apporter vos propres données à Amazon Redshift lors de la création d'un cluster, Amazon Redshift crée la base de données dev.

Les exemples de cette section supposent ce qui suit :

- Vous avez créé un cluster Amazon Redshift. Pour plus d'informations, consultez [Clusters Amazon Redshift et chargement de données](#).
- Vous avez établi une connexion au cluster à partir de votre outil client SQL, tel que l'éditeur de requêtes de la console Amazon Redshift. Pour plus d'informations, consultez [Étape 3 : Accorder des privilèges d'accès aux éditeurs de requête et exécuter des requêtes](#).

Important

Le cluster que vous avez déployé pour cet exercice s'exécute dans un environnement en direct. Tant qu'il fonctionne, il est débité sur votre AWS compte. Pour obtenir des informations sur les prix, consultez la [page de tarification Amazon Redshift](#).

Pour éviter des frais inutiles, supprimez votre cluster lorsque vous avez terminé. L'étape finale de l'exercice explique comment faire.

Tâche 1 : Créer une base de données

Après avoir vérifié que votre cluster fonctionne, vous pouvez créer votre propre première base de données. Cette base de données est l'emplacement où vous créez des tables, chargez des données et exécutez des requêtes. Un seul cluster peut héberger plusieurs bases de données. Par exemple, vous pouvez avoir une base de données SALESDB et une base de données ORDERSDB sur le même cluster.

Par exemple, pour créer une base de données nommée **SALESDB**, exécutez l'instruction suivante dans votre outil de client SQL.

```
CREATE DATABASE SALESDB;
```

Pour cet exercice, acceptez les valeurs par défaut. Pour plus d'informations sur davantage d'options de commande, consultez [CREATE TABLE](#) dans le Guide du développeur de la base de données Amazon Redshift.

Une fois que vous avez créé la base de données SALESDB, vous pouvez vous connecter à la nouvelle base de données depuis votre client SQL. Utilisez les mêmes paramètres de connexion que ceux dont vous vous êtes servi pour votre connexion en cours, mais modifiez le nom de base de données en SALESDB.

Tâche 2 : Créer un utilisateur

Par défaut, seul l'utilisateur administrateur que vous avez créé lorsque vous avez lancé le cluster a accès à la base de données initiale du cluster. Pour accorder l'accès à d'autres utilisateurs, créez un ou plusieurs comptes. Les comptes d'utilisateur de base de données appartiennent à toutes les bases de données d'un cluster, et non à une base de données spécifique.

Utilisez l'instruction `CREATE USER` pour créer un nouvel utilisateur de base de données. Lorsque vous créez un nouvel utilisateur, vous spécifiez le nom du nouvel utilisateur et un mot de passe. Nous vous recommandons de spécifier un mot de passe pour l'utilisateur. Celui-ci doit comprendre entre 8 et 64 caractères et inclure au moins une lettre majuscule, une lettre minuscule et un chiffre.

Par exemple, pour créer un utilisateur nommé **GUEST** avec le mot de passe **ABCd4321**, exécutez la commande suivante.

```
CREATE USER GUEST PASSWORD 'ABCd4321';
```

Pour vous connecter à la base de données SALESDB en tant qu'utilisateur GUEST, utilisez le même mot de passe que lors de la création de l'utilisateur, tel que ABCd4321.

Pour plus d'informations sur les autres options de commande, consultez [CREATE USER](#) dans le Guide du développeur de la base de données Amazon Redshift.

Tâche 3 : Créer un schéma

Après avoir créé une base de données, vous pouvez créer un schéma dans la base de données actuelle. Un schéma est un espace de noms qui contient des objets de base de données nommés tels que des tables, des vues et des fonctions définies par l'utilisateur (UDF). Une base de données peut contenir un ou plusieurs schémas, et chaque schéma appartient à une seule base de données. Deux schémas peuvent avoir des objets différents qui partagent le même nom.

Vous pouvez créer plusieurs schémas dans la même base de données pour organiser les données comme vous le souhaitez ou pour regrouper vos données de manière fonctionnelle. Par exemple, vous pouvez créer un schéma pour stocker toutes vos données intermédiaires et un autre schéma pour stocker toutes les tables de reporting. Vous pouvez également créer différents schémas pour stocker des données pertinentes pour différents groupes professionnels qui se trouvent dans la même base de données. Chaque schéma peut stocker différents objets de base de données, tels que des tables, des vues et des fonctions définies par l'utilisateur (UDF). De plus, vous pouvez créer des schémas avec la clause `AUTHORIZATION`. Cette clause donne la propriété à un utilisateur

spécifié ou définit un quota sur la quantité maximale d'espace disque que le schéma spécifié peut utiliser.

Amazon Redshift crée automatiquement un schéma appelé `public` pour chaque nouvelle base de données. Lorsque vous ne spécifiez pas le nom du schéma lors de la création d'objets de base de données, les objets sont placés dans le schéma `public`.

Pour accéder à un objet dans un schéma, qualifiez l'objet en utilisant la notation `schema_name.table_name`. Le nom qualifié du schéma se compose du nom du schéma et du nom de la table séparés par un point. Par exemple, vous pouvez avoir un schéma `sales` qui possède une table `price` et un schéma `inventory` qui possède également une table `price`. Lorsque vous faites référence à la table `price`, vous devez la qualifier comme `sales.price` ou `inventory.price`.

L'exemple suivant crée un schéma nommé **SALES** pour l'utilisateur `GUEST`.

```
CREATE SCHEMA SALES AUTHORIZATION GUEST;
```

Pour plus d'informations sur davantage d'options de commande, consultez [CREATE SCHEMA](#) dans le Guide du développeur de la base de données Amazon Redshift.

Pour afficher la liste des schémas de votre base de données, exécutez l'instruction suivante.

```
select * from pg_namespace;
```

La sortie doit ressembler à ce qui suit :

nspname	nspowner	nspacl
sales	100	
pg_toast	1	
pg_internal	1	
catalog_history	1	
pg_temp_1	1	
pg_catalog	1	{rdsdb=UC/rdsdb,=U/rdsdb}
public	1	{rdsdb=UC/rdsdb,=U/rdsdb}
information_schema	1	{rdsdb=UC/rdsdb,=U/rdsdb}

Pour plus d'informations sur la façon d'interroger les tables de catalogue, consultez [Interroger les tables de catalogue](#) dans le Guide du développeur de la base de données Amazon Redshift.

Utilisez l'instruction `GRANT` pour octroyer des privilèges aux utilisateurs pour les schémas.

L'exemple suivant octroie à l'utilisateur GUEST le privilège de sélectionner des données dans toutes les tables ou vues de SALESHEMA à l'aide d'une instruction SELECT.

```
GRANT SELECT ON ALL TABLES IN SCHEMA SALES TO GUEST;
```

L'exemple suivant octroie tous les privilèges disponibles en une seule fois à l'utilisateur GUEST.

```
GRANT ALL ON SCHEMA SALES TO GUEST;
```

Tâche 4 : Créer une table

Une fois que vous avez créé votre base de données, créez les tables qui contiendront vos données. Spécifiez les informations sur les colonnes lorsque vous créez la table.

Par exemple, exécutez la commande suivante pour créer une table nommée **DEMO**.

```
CREATE TABLE Demo (  
  PersonID int,  
  City varchar (255)  
);
```

Vous pouvez également créer une table en utilisant la notation `schema_name.object_name` afin de créer la table dans le schéma SALES.

```
CREATE TABLE SALES.DEMO (  
  PersonID int,  
  City varchar (255)  
);
```

Pour afficher et inspecter les schémas et leurs tables, vous pouvez utiliser l'éditeur de requêtes Amazon Redshift. Vous pouvez également voir la liste des tables dans les schémas à l'aide des vues système. Pour plus d'informations, consultez [Tâche 6 : Interroger les tables système](#).

Par défaut, les nouveaux objets de base de données, tels que les tables, sont créés dans le schéma par défaut nommé `public`, généré lors de la création du cluster. Vous pouvez utiliser un autre schéma pour créer des objets de base de données. Pour plus d'informations sur les schémas, consultez [Gestion de la sécurité des bases de données](#) dans le Manuel du développeur de base de données Amazon Redshift.

Les colonnes `encoding`, `distkey` et `sortkey` sont utilisés par Amazon Redshift pour le traitement parallèle. Pour plus d'informations sur la conception de tables qui intègrent ces éléments, consultez [Bonnes pratiques Amazon Redshift pour la conception de tables](#).

Insérer des lignes de données dans une table

Une fois que vous avez créé une table, insérez des lignes de données dans cette table.

Note

La commande [INSERT](#) insère des lignes dans une table. Pour les charges en bloc standard, utilisez la commande [COPY](#). Pour plus d'informations, consultez [Utilisation d'une commande COPY pour charger les données](#).

Par exemple, pour insérer des valeurs dans la table `DEMO`, exécutez l'instruction suivante.

```
INSERT INTO DEMO VALUES (781, 'San Jose'), (990, 'Palo Alto');
```

Pour insérer des données dans une table qui se trouve dans un schéma, exécutez la commande suivante.

```
INSERT INTO SALES.DEMO VALUES (781, 'San Jose'), (990, 'Palo Alto');
```

Sélectionner les données dans une table

Une fois que vous avez créé une table et l'avez remplie avec des données, utilisez une instruction `SELECT` pour afficher les données de la table. L'instruction `SELECT *` renvoie tous les noms de colonne et toutes les valeurs de ligne de toutes les données d'une table. L'utilisation de `SELECT` est un bon moyen de vérifier que les données récemment ajoutées ont été correctement insérées dans la table.

Pour afficher les données que vous avez saisies dans la table **DEMO**, saisissez la commande suivante.

```
SELECT * from DEMO;
```

Le résultat doit avoir l'aspect suivant.

```
personid | city
```

```
-----+-----  
      781 | San Jose  
      990 | Palo Alto  
(2 rows)
```

Pour plus d'informations sur l'utilisation de l'instruction SELECT pour interroger les tables, consultez [SELECT](#).

Tâche 5 : Charger des exemples de données

La plupart des exemples de ce guide utilisent l'exemple de jeu de données TICKIT. Vous pouvez télécharger le fichier [ticketdb.zip](#), qui contient des exemples de fichiers individuels de données.

Vous pouvez charger les exemples de données dans votre propre compartiment Amazon S3.

Pour charger les exemples de données de votre base de données, commencez par créer les tables. Utilisez ensuite la commande COPY pour charger les tables avec des exemples de données stockés dans un compartiment Amazon S3. Pour connaître les étapes pour créer les tables et charger les exemples de données, consultez [Étape 4 : Charger les données d'Amazon S3 vers Amazon Redshift](#).

Tâche 6 : Interroger les tables système

Outre les tables que vous créez, votre base de données contient un certain nombre de tables système. Ces tables système contiennent des informations sur votre installation et sur les différents processus et requêtes qui s'exécutent sur le système. Vous pouvez interroger les tables système pour recueillir les informations sur votre base de données.

Note

La description de chaque table dans cette documentation indique si une table est visible par tous les utilisateurs ou par les super-utilisateurs uniquement. Connectez-vous en tant que super-utilisateur pour interroger les tables visibles uniquement des super-utilisateurs.

Amazon Redshift permet d'accéder aux types de tables système suivants :

- [Tables STL](#)

Ces tables système sont générées à partir des fichiers journaux Amazon Redshift pour fournir un historique du système. Les tables de journalisation ont un préfixe STL.

- [Tables STV](#)

Ces tables sont des tables système virtuelles qui contiennent les instantanés des données système actuelles. Les tables d'instantanés ont un préfixe STV.

- [Vues système](#)

Les vues système contiennent un sous-ensemble de données disponibles dans plusieurs des tables système STL et STV. Les vues système ont un préfixe SVV ou SVL.

- [Tables catalogue système](#)

Les tables catalogue système stockent les métadonnées du schéma, telles que les informations sur les tables et les colonnes. Les tables catalogue système ont un préfixe PG.

Vous devrez peut-être spécifier l'ID de processus associé à une requête pour récupérer les informations de la table système sur cette requête. Pour plus d'informations, consultez [Déterminer l'ID de processus d'une requête en cours d'exécution](#).

Afficher la liste des noms de table

Pour afficher la liste de toutes les tables d'un schéma, vous pouvez interroger la table catalogue système PG_TABLE_DEF. Vous pouvez d'abord examiner le paramétrage de search_path.

```
SHOW search_path;
```

Le résultat devrait être similaire à ce qui suit :

```
search_path
-----
$user, public
(1 row)
```

L'exemple suivant ajoute le schéma SALES au chemin de recherche et affiche toutes les tables du schéma SALES.

```
set search_path to '$user', 'public', 'sales';

SHOW search_path;
search_path
```



```

-----
"$user", public, sales
(1 row)

select * from pg_table_def where schemaname = 'sales';
schemaname | tablename | column | type | encoding | distkey |
sortkey | notnull
-----+-----+-----+-----+-----+-----
+-----+-----
sales | demo | personid | integer | az64 | f |
0 | f
sales | demo | city | character varying(255) | lzo | f |
0 | f
(2 rows)

```

L'exemple suivant illustre une liste de toutes les tables appelées DEMO dans tous les schémas de la base de données actuelle.

```

select * from pg_table_def where tablename = 'demo';
schemaname | tablename | column | type | encoding | distkey |
sortkey | notnull
-----+-----+-----+-----+-----+-----
+-----+-----
public | demo | personid | integer | az64 | f |
0 | f
public | demo | city | character varying(255) | lzo | f |
0 | f
sales | demo | personid | integer | az64 | f |
0 | f
sales | demo | city | character varying(255) | lzo | f |
0 | f
(4 rows)

```

Pour plus d'informations, consultez [PG_TABLE_DEF](#).

Vous pouvez également utiliser l'éditeur de requêtes v2 pour afficher toutes les tables d'un schéma spécifié en choisissant d'abord une base de données à laquelle vous souhaitez vous connecter.

Afficher les utilisateurs

Vous pouvez interroger la table catalogue PG_USER pour afficher la liste de tous les utilisateurs, ainsi que l'ID utilisateur (USESYSID) et les privilèges utilisateur.

```
SELECT * FROM pg_user;
  username | usesysid | usecreatedb | usesuper | usecatupd | passwd | valuntil |
  useconfig
-----+-----+-----+-----+-----+-----+-----
+-----
 rdsdb    |        1 | true        | true     | true     | ***** | infinity |
 awsuser  |       100 | true        | true     | false    | ***** |          |
 guest    |       104 | true        | false    | false    | ***** |          |
(3 rows)
```

Le nom d'utilisateur `rdsdb` est utilisé en interne par Amazon Redshift pour effectuer les tâches habituelles d'administration et de maintenance. Vous pouvez filtrer votre requête pour afficher uniquement les noms d'utilisateur définis par l'utilisateur en ajoutant `where usesysid > 1` à votre instruction `SELECT`.

```
SELECT * FROM pg_user WHERE usesysid > 1;
  username | usesysid | usecreatedb | usesuper | usecatupd | passwd | valuntil |
  useconfig
-----+-----+-----+-----+-----+-----+-----
+-----
 awsuser   |       100 | true        | true     | false    | ***** |          |
 guest    |       104 | true        | false    | false    | ***** |          |
(2 rows)
```

Afficher les requêtes récentes

Dans l'exemple précédent, l'ID utilisateur (`USESYSID`) de `adminuser` est 100. Pour afficher les cinq requêtes les plus récentes exécutées par `adminuser`, vous pouvez interroger la vue `SVL_QLOG`.

La vue `SVL_QLOG` est un sous-ensemble d'informations plus convivial de la table `STL_QUERY`. Vous pouvez utiliser cette vue pour trouver l'ID de requête (`QUERY`) ou l'ID de processus (`PID`) d'une requête récemment exécutée. Vous pouvez également l'utiliser pour connaître le temps mis par une requête pour s'exécuter. `SVL_QLOG` inclut les 60 premiers caractères de la chaîne de requête (`SUBSTRING`) pour vous aider à rechercher une requête spécifique. Utilisez la clause `LIMIT` avec votre instruction `SELECT` pour limiter les résultats à cinq lignes.

```
SELECT query, pid, elapsed, substring from svl_qlog
WHERE userid = 100
ORDER BY starttime desc
LIMIT 4;
```

Le résultat ressemble à ce qui suit.

```

query|  pid  |  elapsed  |      substring
-----+-----+-----
+-----+-----+-----
892  | 21046 | 55868    | SELECT query, pid, elapsed, substring from svl_qlog WHERE
us
620  | 17635 | 1296265  | SELECT query, pid, elapsed, substring from svl_qlog WHERE us
610  | 17607 | 82555    | SELECT * from DEMO;
596  | 16762 | 226372   | INSERT INTO DEMO VALUES (100);)

```

Déterminer l'ID de processus d'une requête en cours d'exécution

Il se peut que vous ayez besoin de trouver le PID d'une requête qui est toujours en cours d'exécution. Par exemple, vous avez besoin du PID si vous devez annuler une requête qui prend trop de temps pour s'exécuter. Vous pouvez interroger la table système STV_RECENTS pour obtenir la liste des ID de processus des requêtes en cours d'exécution, ainsi que la chaîne de requête correspondante. Si votre requête retourne plusieurs PID, vous pouvez examiner le texte de la requête pour déterminer le PID dont vous avez besoin.

Pour déterminer le PID d'une requête en cours d'exécution, exécutez l'instruction SELECT suivante.

```

SELECT pid, user_name, starttime, query
FROM stv_recents
WHERE status='Running';

```

Tâche 7 : Annuler une requête

Si vous exécutez une requête qui prend trop de temps ou consomme des ressources de cluster excessives, annulez-la. Par exemple, vous créez une liste de vendeurs de billets comprenant le nom du vendeur et la quantité de billets vendus. La requête suivante sélectionne les données des tables SALES et USERS, et joint les deux tables en associant SELLERID et USERID dans la clause WHERE.

```

SELECT sellerid, firstname, lastname, sum(qtysold)
FROM sales, users
WHERE sales.sellerid = users.userid
GROUP BY sellerid, firstname, lastname
ORDER BY 4 desc;

```

Le résultat ressemble à ce qui suit.

sellerid	firstname	lastname	sum
48950	Nayda	Hood	184
19123	Scott	Simmons	164
20029	Drew	Mcguire	164
36791	Emerson	Delacruz	160
13567	Imani	Adams	156
9697	Dorian	Ray	156
41579	Harrison	Durham	156
15591	Phyllis	Clay	152
3008	Lucas	Stanley	148
44956	Rachel	Villarreal	148

Note

Il s'agit d'une requête complexe. Pour ce didacticiel, vous n'avez pas à vous préoccuper de la façon dont la requête est construite.

La requête précédente s'exécute en quelques secondes et retourne 2 102 lignes.

Supposons vous oubliez de placer la clause WHERE.

```
SELECT sellerid, firstname, lastname, sum(qtysold)
FROM sales, users
GROUP BY sellerid, firstname, lastname
ORDER BY 4 desc;
```

L'ensemble de résultats contient le nombre de toutes les lignes de la table SALES multiplié par le nombre de toutes les lignes de la table USERS (49989 * 3766). Cette jointure est appelée jointure cartésienne et n'est pas recommandée. Le résultat est supérieur à 188 millions de lignes et son exécution prend beaucoup de temps.

Pour annuler une requête en cours d'exécution, utilisez la commande CANCEL avec le PID de la requête.

Pour obtenir l'ID du processus, démarrez une nouvelle séance et interrogez la table STV_RECENTS, comme illustré dans l'étape précédente. L'exemple suivant montre comment vous pouvez rendre les résultats plus lisibles. Pour ce faire, utilisez la fonction TRIM pour couper les espaces de fin de ligne et n'afficher que les 20 premiers caractères de la chaîne de requête.

```
SELECT pid, trim(user_name), starttime, substring(query,1,20)
FROM stv_recents
WHERE status='Running';
```

Le résultat ressemble à ce qui suit.

```
 pid | btrim | starttime | substring
-----+-----+-----+-----
 610 | adminuser | 2013-03-28 18:39:49.355918 | select sellerid, fir
(1 row)
```

Pour annuler la requête avec le PID 610, entrez la commande suivante.

```
CANCEL 610;
```

Note

La commande CANCEL n'arrête pas une transaction. Pour arrêter ou restaurer une transaction, utilisez la commande ABORT ou ROLLBACK. Pour annuler une requête associée à une transaction, annulez d'abord la requête, puis arrêtez la transaction.

Si la requête que vous avez annulée est associée à une transaction, utilisez la commande ABORT ou ROLLBACK. pour annuler la transaction et ignorer toutes les modifications apportées aux données :

```
ABORT;
```

Sauf si vous êtes connecté comme super-utilisateur, vous pouvez annuler uniquement vos propres requêtes. Un super-utilisateur peut annuler toutes les requêtes.

Annuler une requête à partir d'une autre séance

Si votre outil de requête ne prend pas en charge l'exécution simultanée de requêtes, démarrez une nouvelle séance pour annuler la requête. Par exemple, l'éditeur de requêtes que nous utilisons dans le Guide de démarrage Amazon Redshift ne prend pas en charge les requêtes multiples simultanées. Pour démarrer une nouvelle séance à l'aide de l'éditeur de requêtes, sélectionnez File (Fichier), New Window (Nouvelle fenêtre), puis connectez-vous avec les mêmes paramètres de connexion. Ensuite, vous pouvez trouver le PID et annuler la requête.

Annuler une requête à l'aide de la file d'attente super-utilisateur

Si votre séance en cours a un trop grand nombre de requêtes s'exécutant en même temps, il se peut que vous ne soyez pas en mesure d'exécuter la commande CANCEL tant qu'une autre requête n'est pas terminée. Dans ce cas, exécutez la commande CANCEL en utilisant une autre file d'attente de requêtes de gestion de la charge de travail.

La gestion de la charge de travail vous permet d'exécuter les requêtes dans différentes files d'attente de afin que vous n'ayez pas besoin d'attendre qu'une autre requête se termine. Le gestionnaire de la charge de travail crée une file d'attente distincte, appelée file d'attente super-utilisateur, que vous pouvez utiliser à des fins de dépannage. Pour utiliser la file d'attente super-utilisateur, connectez-vous en tant que super-utilisateur et définissez le groupe de requêtes avec la valeur « superuser » à l'aide de la commande SET. Après l'exécution de vos commandes, réinitialisez le groupe de requêtes à l'aide de la commande RESET.

Pour annuler une requête à l'aide de la file d'attente super-utilisateur, exécutez ces commandes.

```
SET query_group TO 'superuser';
CANCEL 610;
RESET query_group;
```

Tâche 8 : Nettoyer vos ressources

Si vous avez déployé un cluster pour réaliser cet exercice, lorsque vous avez fini ce dernier, supprimez le cluster. En supprimant le cluster, vous évitez qu'il n'entraîne des frais sur votre compte AWS .

Pour supprimer le cluster, suivez les étapes de la rubrique [Suppression d'un cluster](#) dans le Guide de gestion Amazon Redshift.

Si vous voulez garder le cluster, conservez les exemples de données à titre de référence. La plupart des exemples de ce guide utilisent les tables que vous avez créées dans cet exercice. La taille des données n'aura pas d'effet significatif sur votre stockage disponible.

Si vous voulez conserver le cluster, mais souhaitez nettoyer les exemples de données, exécutez la commande suivante pour supprimer la base de données SALESDB.

```
DROP DATABASE SALESDB;
```

Si vous n'avez pas créé de base de données SALESDB, ou si vous ne voulez pas supprimer la base de données, exécutez les commandes suivantes pour supprimer uniquement les tables.

```
DROP TABLE DEMO;  
DROP TABLE users;  
DROP TABLE venue;  
DROP TABLE category;  
DROP TABLE date;  
DROP TABLE event;  
DROP TABLE listing;  
DROP TABLE sales;
```

Présentation conceptuelle d'Amazon Redshift

Amazon Redshift sans serveur vous permet d'accéder aux données et de les analyser sans toutes les configurations d'un entrepôt des données provisionné. Les ressources sont automatiquement provisionnées et la capacité de l'entrepôt des données est intelligemment mise à l'échelle afin d'offrir des performances rapides, même pour les charges de travail les plus exigeantes et les plus imprévisibles. Vous ne payez pas de frais lorsque l'entrepôt des données est inactif, vous ne payez donc que ce que vous utilisez. Vous pouvez charger des données et commencer à effectuer des requêtes immédiatement dans l'éditeur de requête Amazon Redshift v2 ou dans votre outil d'informatique décisionnelle (BI) préféré. Profitez des meilleures performances en termes de prix et des fonctionnalités SQL habituelles dans un environnement facile à utiliser et sans administration.

Si vous utilisez Amazon Redshift pour la première fois, nous vous recommandons de commencer par lire les sections suivantes :

- [Présentation des fonctions d'Amazon Redshift sans serveur](#) : dans cette rubrique, vous trouverez une présentation d'Amazon Redshift sans serveur et de ses principales fonctionnalités.
- [Éléments principaux du service et prix](#) : sur cette page détaillée du produit, vous trouverez des détails sur les services et les prix d'Amazon Redshift sans serveur.
- [Premiers pas avec Amazon Redshift sans serveur](#) : dans cette rubrique, vous pouvez en savoir plus sur la façon de créer un entrepôt de données Amazon Redshift sans serveur, et commencer à interroger des données à l'aide de l'éditeur de requête v2.

Si vous préférez gérer manuellement vos ressources Amazon Redshift, vous pouvez créer des clusters provisionnés pour vos besoins en matière d'interrogation de données. Pour plus d'informations, consultez [Clusters Amazon Redshift](#).

Si votre organisation est éligible et que votre cluster est créé dans une Région AWS où Amazon Redshift sans serveur n'est pas disponible, vous pourrez peut-être créer un cluster dans le cadre du programme d'essai gratuit d'Amazon Redshift. Choisissez Production ou Essai gratuit pour répondre à la question Quelle est l'utilisation prévue de ce cluster ? Lorsque vous choisissez Essai gratuit, vous créez une configuration avec le type de nœud dc2.large. Pour plus d'informations sur le choix d'un essai gratuit, accédez à [Essai gratuit d'Amazon Redshift](#). Pour obtenir la liste des Régions AWS où Amazon Redshift sans serveur est disponible, consultez les points de terminaison répertoriés pour [l'API Redshift sans serveur](#) dans le Référence générale d'Amazon Web Services.

Vous trouverez ci-après quelques concepts clés d'Amazon Redshift sans serveur.

- Espace de noms : collection d'objets et d'utilisateurs de la base de données. Les espaces noms regroupent toutes les ressources que vous utilisez dans Amazon Redshift sans serveur, telles que les schémas, les tables, les utilisateurs, les unités de partage des données et les instantanés.
- Groupe de travail : une collection de ressources informatiques. Les groupes de travail hébergent des ressources de calcul qu'Amazon Redshift sans serveur utilise pour exécuter des tâches de calcul. Les unités de traitement Redshift (RPU), les groupes de sécurité et les limites d'utilisation sont quelques exemples de ces ressources. Les groupes de travail disposent de mises en réseau et de paramètres de sécurité que vous pouvez configurer à l'aide de la console Amazon Redshift sans serveur, de la AWS Command Line Interface, ou des API Amazon Redshift sans serveur.

Pour plus d'informations sur la configuration des ressources des espaces noms et des groupes de travail, consultez [Utilisation des espaces noms](#) et [Utilisation des groupes de travail](#).

Vous trouverez ci-dessous quelques concepts clés des clusters Amazon Redshift provisionnés :

- Cluster : le composant principal de l'infrastructure d'un entrepôt des données Amazon Redshift est un cluster.

Un cluster est composé d'un ou plusieurs nœuds de calcul. Les nœuds de calcul exécutent le code compilé.

Si un cluster est alloué avec deux nœuds de calcul ou plus, un nœud principal supplémentaire coordonne les nœuds de calcul. Le nœud principal gère la communication externe avec des applications, telles que les outils de business intelligence et les éditeurs de requêtes. Votre application cliente n'interagit directement qu'avec le nœud principal. Les nœuds de calcul sont transparents pour les applications externes.

- Base de données : un cluster contient une ou plusieurs bases de données.

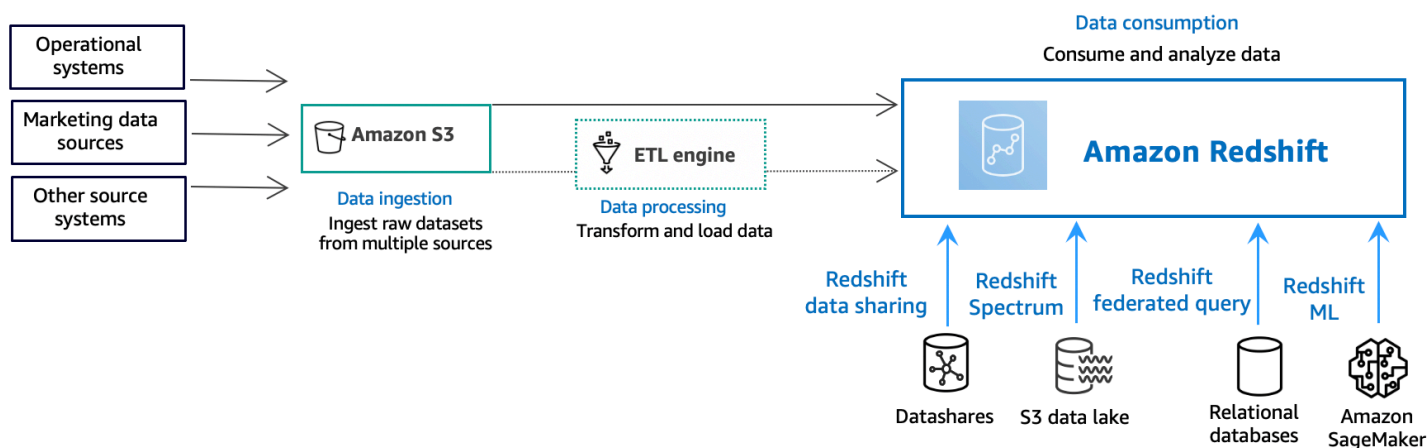
Les données utilisateur sont stockées dans une ou plusieurs bases de données sur les nœuds de calcul. Votre client SQL communique avec le nœud principal, qui à son tour coordonne l'exécution des requêtes avec les nœuds de calcul. Pour plus de détails sur les nœuds principaux et les nœuds de calcul, veuillez consulter [Architecture système de l'entrepôt de données](#). Dans une base de données, les données utilisateur sont organisées en un ou plusieurs schémas.

Amazon Redshift est un système de gestion de base de données relationnelle (SGBDR), compatible de ce fait avec d'autres applications SGBDR. Il fournit les mêmes fonctionnalités qu'un SGBDR classique, y compris les fonctions de traitement transactionnel en ligne (OLTP) comme

l'insertion et la suppression de données. Amazon Redshift est également optimisé pour l'analyse par lots hautes performances et la création de rapports de jeux de données très volumineux.

Vous trouverez ci-dessous une description du flux de traitement des données typique dans Amazon Redshift, ainsi que des descriptions des différentes parties du flux. Pour plus d'informations sur l'architecture système Amazon Redshift, veuillez consulter [Architecture système de l'entrepôt de données](#).

Le diagramme suivant illustre un flux de traitement des données standard dans Amazon Redshift.



Un entrepôt des données Amazon Redshift est un système de gestion et de requête de base de données relationnelle de niveau entreprise. Amazon Redshift prend en charge les connexions client avec de nombreux types d'applications, notamment les outils de business intelligence (BI), de reporting, de données et d'analytique. Lorsque vous exécutez des requêtes analytiques, vous extrayez, comparez et évaluez de grandes quantités de données dans le cadre d'opérations à plusieurs étapes, afin d'obtenir un résultat final.

Au niveau de la couche d'ingestion de données, différents types de sources de données chargent en continu des données structurées, semi-structurées ou non structurées vers la couche de stockage de données. Cette zone de stockage de données sert de zone de transit qui stocke les données dans différents états de préparation à la consommation. Un compartiment Amazon Simple Storage Service (Amazon S3) peut être un exemple de stockage.

Au niveau de la couche de traitement des données facultative, les données source passent par le prétraitement, la validation et la transformation à l'aide de pipelines d'extraction, de transformation, de chargement (ETL) ou d'extraction, de chargement, de transformation (ELT). Ces jeux de données bruts sont ensuite affinés à l'aide d'opérations ETL. Un exemple de moteur ETL est AWS Glue.

Au niveau de la couche de consommation des données, les données sont chargées dans votre cluster Amazon Redshift, où vous pouvez exécuter des charges de travail analytiques.

Pour des exemples de charges de travail analytiques, consultez [Interroger des sources de données externes](#).

Ressources supplémentaires

Pour plus d'informations sur Amazon Redshift sans serveur, nous vous recommandons de continuer à approfondir les concepts présentés dans ce guide en utilisant les ressources Amazon Redshift suivantes :

- Vidéos sur les fonctions : ces vidéos vous en apprennent davantage sur les fonctions d'Amazon Redshift.
 - Pour comprendre Amazon Redshift sans serveur à un niveau avancé, regardez la vidéo suivante. [Amazon Redshift sans serveur expliqué en 90 secondes](#).
 - Pour découvrir comment configurer un entrepôt des données sans serveur et commencer à interroger des données, regardez la vidéo suivante. [Démarrer avec Amazon Redshift sans serveur](#).
- [Guide de gestion Amazon Redshift](#) : ce guide repose sur ce Guide de démarrage d'Amazon Redshift. Il fournit des informations approfondies sur les concepts et les tâches de création, de gestion et de surveillance des clusters Amazon Redshift sans serveur et Amazon Redshift provisionnés.
- [Guide du développeur de bases de données Amazon Redshift](#) : ce guide repose également sur le Guide de démarrage d'Amazon Redshift. Il fournit aux développeurs de bases de données des informations sur la conception, le développement, l'interrogation et la maintenance des bases de données qui constituent votre entrepôt de données.
 - [Référence SQL](#) : cette rubrique décrit les commandes SQL et les références de fonctions pour Amazon Redshift.
 - [Tables et vues système](#) : cette rubrique décrit les tables système et les vues d'Amazon Redshift.
- Didacticiels pour Amazon Redshift : cette rubrique présente des tutoriels sur les fonctionnalités d'Amazon Redshift.
 - [Chargement des données depuis Amazon S3](#) : ce didacticiel explique comment charger des données dans vos tables de base de données Amazon Redshift à partir de fichiers de données dans un compartiment Amazon S3.
 - [Partage de données dans Amazon Redshift Serverless](#) : cette section décrit comment partager et accéder aux données dans d'autres clusters Amazon Redshift.
 - [Utilisation de fonctions SQL spatiales avec Amazon Redshift](#) : ce didacticiel montre comment utiliser certaines fonctions SQL spatiales avec Amazon Redshift.

- [Interrogation des données imbriquées avec Amazon Redshift Spectrum](#) : ce didacticiel explique comment utiliser Redshift Spectrum pour interroger des données imbriquées dans les formats de fichiers Parquet, ORC, JSON et Ion à l'aide de tables externes.
- [Configuration des files d'attente de gestion manuelle de la charge de travail \(WLM\)](#) : ce didacticiel explique comment configurer la gestion manuelle de la charge de travail dans Amazon Redshift.
- [Premiers pas avec Amazon Redshift ML](#) : cette section décrit comment les utilisateurs peuvent créer, entraîner et déployer des modèles de machine learning à l'aide de commandes SQL familières.
- [Nouveautés](#) : cette page web répertorie les nouvelles fonctions et les mises à jour des produits Amazon Redshift.

Historique du document

Note

Pour obtenir une description des nouvelles fonctionnalités d'Amazon Redshift, consultez la section [Nouveautés](#).

Le tableau suivant décrit les modifications importantes apportées à la documentation du guide de démarrage Amazon Redshift.

Modification	Description	Date de publication
Mise à jour de la documentation	Mise à jour du guide pour inclure de nouvelles sections sur la prise en main des tâches de base de données courantes, l'interrogation de votre lac de données, l'interrogation de données sur des sources distantes, le partage de données et l'entraînement de modèles de machine learning avec les données Amazon Redshift.	30 Juin 2021
Nouvelle fonctionnalité	Mise à jour du guide pour décrire la nouvelle procédure de chargement des exemples.	4 juin 2021
Mise à jour de la documentation	Mise à jour du guide pour supprimer la console Amazon Redshift d'origine et améliorer le flux des étapes.	14 août 2020
Nouvelle console	Mise à jour du manuel pour décrire la nouvelle console Amazon Redshift.	11 novembre 2019
Nouvelle fonctionnalité	Mise à jour du guide pour décrire la procédure de lancement rapide du cluster.	10 août 2018
Nouvelle fonctionnalité	Mise à jour du guide pour lancer les clusters depuis le tableau de bord Amazon Redshift.	28 juillet 2015

Modification	Description	Date de publication
Nouvelle fonctionnalité	Mise à jour du guide pour utiliser de nouveaux noms de types de nœud.	9 juin 2015
Mise à jour de la documentation	Mise à jour des captures d'écran et de la procédure de configuration des groupes de sécurité VPC.	30 avril 2015
Mise à jour de la documentation	Mise à jour des captures d'écran et des procédures pour correspondre à la console actuelle.	12 novembre 2014
Mise à jour de la documentation	Déplacement des données de chargement depuis les informations Amazon S3 dans leur propre section et déplacement de la section des étapes suivantes dans l'étape finale pour plus de clarté.	13 mai 2014
Mise à jour de la documentation	Suppression de la page d'accueil et intégration du contenu dans la page de mise en route principale.	14 mars 2014
Mise à jour de la documentation	Ce document est une nouvelle version du Guide de démarrage Amazon Redshift qui traite des commentaires des clients et des mises à jour du service.	14 mars 2014
Nouveau guide	Il s'agit de la première version du Guide de démarrage Amazon Redshift.	14 février 2013

Les traductions sont fournies par des outils de traduction automatique. En cas de conflit entre le contenu d'une traduction et celui de la version originale en anglais, la version anglaise prévaudra.