

AWS Whitepaper

Using Microsoft Power BI with the AWS Cloud



Using Microsoft Power BI with the AWS Cloud: AWS Whitepaper

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

Amazon's trademarks and trade dress may not be used in connection with any product or service that is not Amazon's, in any manner that is likely to cause confusion among customers, or in any manner that disparages or discredits Amazon. All other trademarks not owned by Amazon are the property of their respective owners, who may or may not be affiliated with, connected to, or sponsored by Amazon.

Table of Contents

Abstract	i
Introduction	2
The Microsoft Power BI Suite	4
Power BI Desktop	4
Power BI Service	4
Power BI Report Server	5
On-premises data gateway	5
Connecting Microsoft Power BI Desktop to AWS data sources	6
Using Power BI Desktop on premises	6
Connecting to data sources through the internet	6
Connecting to data sources through AWS VPN	10
Connecting to data sources via AWS Direct Connect	14
Using Microsoft Power BI Desktop in the AWS Cloud	16
Option 1: Install Microsoft Power BI Desktop on an Amazon EC2 instance	17
Option 2: Install Microsoft Power BI in an Amazon WorkSpaces environment	18
Option 3: Install Microsoft Power BI in an Amazon AppStream 2.0 environment	18
Summary of Microsoft Power BI Desktop connectivity options	21
Connecting the Microsoft Power BI service to AWS data sources	22
Recommended configuration	22
Additional considerations	23
Using Amazon QuickSight	28
Conclusion	30
Contributors	31
Further reading	32
Document revisions	33
Appendix: Microsoft Power BI supported AWS data sources	34
Amazon Redshift	34
Amazon RDS	35
Amazon Athena	36
Amazon OpenSearch Service	37
AWS Lake Formation	38
Notices	39

Using Microsoft Power BI with the AWS Cloud

Publication date: **November 03, 2021** ([Document revisions](#))

This whitepaper discusses how to integrate and use Microsoft Power BI (Desktop, Service, and on-premises data gateway) with the Amazon Web Services (AWS) Cloud. It presents options for customers looking to connect Microsoft Power BI products to AWS services such as Amazon Redshift, Amazon Athena, Amazon RDS, Amazon OpenSearch Service, and AWS Lake Formation, with a focus on connectivity, security, performance, and cost optimization.

This whitepaper is for IT decision makers and architects looking to quickly understand Microsoft Power BI concepts and what options exist to make use of those technologies when using AWS services as data sources.

Introduction

Customers with businesses of all sizes are using AWS products and services to store their data reliably, cost effectively, and securely. This is due in part to the broad ecosystem of mature data storage and analytics offerings that are available. Some of these offerings include the following services:

- [Amazon Simple Storage Service](#) (Amazon S3) provides a simple, scalable, secure, and cost-effective data repository. It has become an industry standard for storing application data, as well as a first choice for customer data lakes.
- [Amazon Athena](#) is an interactive query service that makes it easy to analyze data in Amazon S3 using standard SQL.
- [Amazon Relational Database Service](#) (Amazon RDS) makes it easy to set up, operate, and scale a relational database in the cloud. It provides cost-efficient and resizable capacity while automating time-consuming administration tasks such as hardware provisioning, database setup, patching, and backups. SQL Server, Oracle Database, MySQL, MariaDB, and PostgreSQL engines are available.
- [Amazon Redshift](#) is fully managed, massively scalable data warehouse that makes it easy to analyze both structured and unstructured datasets.
- [Amazon QuickSight](#) is a fast, cloud-powered business intelligence service that makes it easy to deliver insights to everyone in your organization.
- [Amazon OpenSearch Service](#) makes it easy for you to perform interactive log analytics, near real-time application monitoring, website search, and more.
- [AWS Lake Formation](#) is a service that makes it easy to set up a secure data lake in days.

To better understand how services relate to one another, we often label data services as either being *data sources* or *data consumers*. A data source allows customers and applications to store and retrieve data from the service. Frequently, data sources also have built-in compute and can provide computational analysis and filtering. But ultimately, data is loaded into these data sources and eventually data is retrieved from them by data consumers. Amazon S3, Amazon Athena, and Amazon Redshift are good examples of data sources.

Data consumers, on the other hand, access the data from data sources and, typically, process it. They might optionally display it too. Amazon QuickSight and the Microsoft Power BI suite are

good examples of data consumers. They read from data sources, and then assist in the analysis, visualization, and publication of information.

AWS gives customers full flexibility in mixing the technologies they prefer for their data needs. While many customers choose Amazon QuickSight for their business intelligence (BI) needs, other customers choose vendors such as Microsoft Power BI, Tableau, and Qlik.

This document focuses on the Microsoft Power BI suite of products and services, and how to use them in combination with AWS services.

The Microsoft Power BI suite

To reduce confusion due to product naming similarities, this whitepaper presents what each Microsoft Power BI product and service is.

Power BI Desktop

Power BI Desktop is a free application you install on your local computer. It lets you connect to, transform, and visualize your data. With Power BI Desktop, you can connect to multiple different sources of data and combine them (often called *modeling*) into a data model. This data model lets you build visuals and collections of visuals you can share as reports with other people inside your organization.

Power BI Desktop can connect to any supported data source that is available locally or over the network. For supported data sources, see the [Appendix: Microsoft Power BI supported AWS data sources](#).

Most users who work on business intelligence projects use Power BI Desktop to create reports. Then they push content to either Power BI Report Server or the Power BI service in order to share their reports with others. The act of pushing content from Power BI Desktop to the Power BI Report Server or the Power BI service is known as *publishing*. For more information, see [What is Power BI Desktop?](#)

Note

Power BI Desktop is a Windows-only application. It is not available for Linux, macOS, or other operating systems.

Power BI Service

Power BI is a collection of software services, apps, and connectors that work together to help you create, share, and consume business insights in a way that serves you and your business most effectively. The Power BI service, sometimes referred to as Power BI online, is the *software as a service* (SaaS) part of Power BI. For more information, see [What is the Power BI service?](#)

The Power BI service is a cloud-based service. It supports light report editing and collaboration for teams and organizations. You can connect to data sources in the Power BI service too, but modeling is limited.

Most report designers who work on business intelligence projects use Power BI Desktop to create reports, and then use the Power BI service to distribute their reports with others. For additional information on this crucial component, refer to [Connecting the Microsoft Power BI service to AWS data sources](#).

Power BI Report Server

Power BI Report Server is a private report server with a web portal in which you display and manage reports and KPIs. Customers use Power BI Report Server in cases where they do not want their reports published to the Power BI service. Due to Microsoft licensing restrictions, Power BI Report Server cannot be deployed within AWS.

On-premises data gateway

The Microsoft on-premises data gateway is a commonly-deployed component that can increase the security and performance of Power BI deployments. It allows the Power BI service to access privatized data sources, which are located in another facility and accessible by internal network connectivity between the data source and the data gateway. Although it is typically installed as a server component, you can also install a personal mode on your local computer as an application. This whitepaper focuses only on the standard (server) mode. For additional information, refer to [Connecting the Microsoft Power BI service to AWS data sources](#).

Connecting Microsoft Power BI Desktop to AWS data sources

Most often customers who start with Microsoft Power BI Desktop are interested in how they can connect to AWS data sources from their on-premises computers and network. The desktop application is typically running on their local Windows laptop and physical and logical connectivity to AWS data sources are the biggest perceived barriers to entry.

However, another option exists, which is to run the Microsoft Power BI Desktop in the AWS Cloud. This option significantly reduces connectivity barriers to AWS data sources, but also requires some additional considerations. Both models are discussed in this section. We examine the implications of each in relation to connectivity, security, performance, and costs so that you can decide which option is best for you. The options presented in this section illustrate Amazon RDS, Amazon Redshift, and Amazon Athena. For a full discussion of all AWS data sources, refer to [Appendix: Microsoft Power BI supported AWS data sources](#).

Using Power BI Desktop on premises

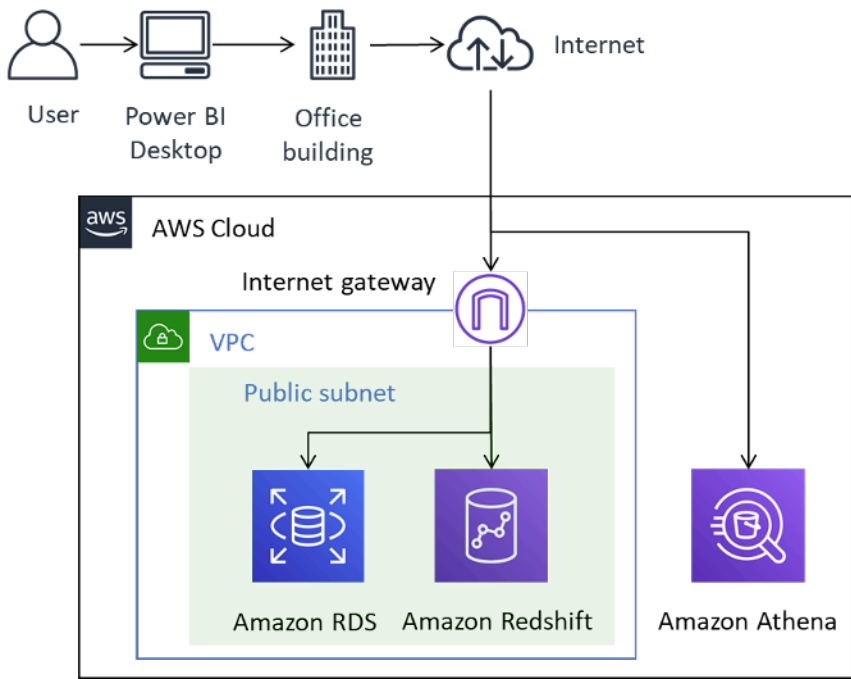
If you plan on using Power BI desktop on premises with data sources that are stored within the AWS Cloud, Power BI can access these sources in one of three ways:

- Connecting to data sources using the internet.
- Connecting to data sources using AWS Virtual Private Network (AWS VPN).
- Connecting to data sources using AWS Direct Connect.

Each method is detailed in the following sections.

Connecting to data sources through the internet

In this model, the Power BI Desktop application places an outbound connection that is routed over the internet to an IP address of an internet-accessible AWS data source. For example, Amazon RDS and Amazon Redshift, which are instantiated within a customers' Amazon Virtual Private Cloud (Amazon VPC), support the *public accessibility* option to make instances accessible over the internet. Amazon Athena can be queried directly from the internet by using the service endpoint for your specific Region.



Power BI connectivity to AWS data sources over the internet

Although this method of connectivity is technically possible, we don't recommend it for anything other than a small number of users. The following table lists important considerations.

Table 1 — Considerations for accessing AWS data sources over the internet

Criteria	Considerations for accessing AWS data sources over the internet
<p>Network connectivity</p>	<p>Data sources are available by connecting to private IP addresses in a VPC, or using a regional or VPC service endpoint. Power BI Desktop connects via VPN and either accesses data sources directly (Amazon RDS, Amazon Redshift, Amazon EC2-based data sources), or for services with a regional endpoint (Amazon Athena) by using a private VPC endpoint or the regional endpoint, depending on the DNS configuration.</p>
<p>Security</p>	<p>IP access control</p>

Criteria	Considerations for accessing AWS data sources over the internet
	<p>A <i>security group</i> acts as a virtual firewall for your instance to control inbound and outbound traffic. In order to limit access to trusted entities, configure security groups to only allow inbound IP ranges associated with known Classless Inter-Domain Routing (CIDR) ranges.</p> <p>Encryption in transit</p> <p>AWS recommends that you configure encryption for any data sources that use public IP addresses, such as Amazon RDS, Amazon Redshift, or any Amazon EC2-based data sources. This ensures that the risk of data or credentials being compromised while in transit, is reduced. Failure to configure encryption represents a significant risk. Do not overlook this aspect.</p> <p>Regional service endpoints, such as Amazon Athena, are TLS encrypted. In addition, Amazon Athena query results that stream to JDBC or ODBC clients are encrypted using Transport Layer Security (TLS).</p> <p>Authentication and authorization</p> <p>AWS recommends that you use credentials that provide read-only access to datasets, and set up processes to rotate credentials per your company policy.</p>

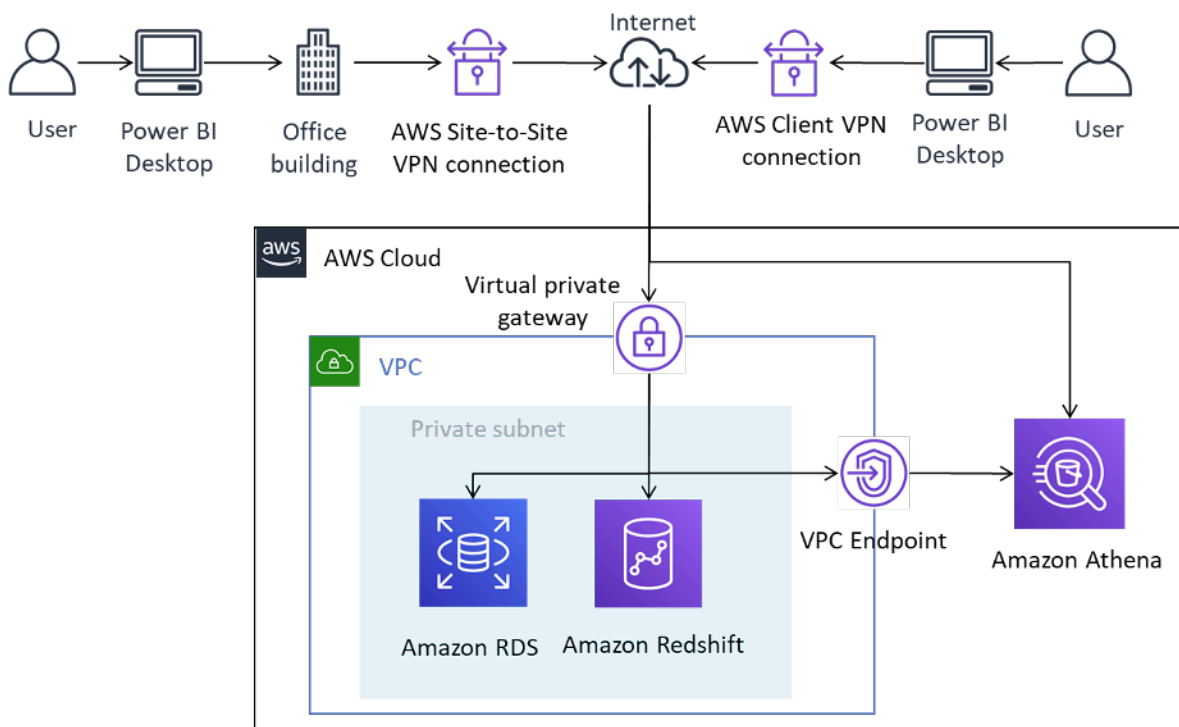
Criteria	Considerations for accessing AWS data sources over the internet
Performance	<p>Some factors that might impact the overall Power BI Desktop performance when accessing AWS data sources over the internet include:</p> <ul style="list-style-type: none">• The size of the dataset being accessed. Larger datasets take longer to retrieve. We recommend limiting queries and using filters to reduce the amount of data retrieved over the internet.• The quality of the internet connection, including bandwidth, latency, and packet loss. Where possible, access data in AWS Regions which you are geographically close to in order to reduce the effect of latency. If your internet is shared, consider loading data sources at off peak times and ensuring that enough bandwidth is available. <p>In general, AWS recommends testing the experience at different times of the day, with different datasets, and with progressively larger number of users.</p>
Cost	<p>Data sources that reside in a VPC and are queried using public IP address over the internet incur standard Amazon VPC data egress charges. In order to reduce costs, we recommend limiting queries and using filters to reduce the amount of data retrieved over the internet.</p>

Connecting to data sources via AWS VPN

In this model, Power BI Desktop installations connect to data sources in the AWS network using one of two AWS VPN methods: AWS Site-to-Site VPN or AWS Client VPN. Each connection type delivers a highly available, managed, and elastic cloud VPN solution to protect your network traffic.

Site-to-Site VPN creates encrypted tunnels between your network and your AWS VPN or AWS Transit Gateway. Client VPN connects your users to AWS or on-premises resources using a free VPN software client.

VPN traffic from both Site-to-Site VPN and Client VPN connections stops in your VPC. As such, it can route to private IP addresses so your instances no longer need public-facing IP addresses. For services with a data path accessible from a publicly facing service endpoint, such as Athena, these service requests can either be routed over the internet, or over the VPN connection and through a VPC endpoint.



Connecting Power BI Desktop to AWS data sources over Site-to-Site VPN and Client VPN

Site-to-Site VPN can also connect to AWS Transit Gateway, facilitating access to data sources spread across multiple VPCs.

Using AWS VPN provides the benefit of employing encryption when accessing data sources stored in AWS, without requiring that each data source to be explicitly configured. Once configured, VPN technology is largely seamless to end users.

Table 2 — Considerations for accessing AWS data sources using AWS VPN

Criteria	Considerations for accessing AWS data sources using AWS VPN
Network connectivity	<p>Data sources are available by connecting to private IP addresses in a VPC, or using a regional or VPC service endpoint. Power BI Desktop connects via VPN and either access data sources directly (Amazon RDS, Amazon Redshift, Amazon EC2-based data sources), or for services with a regional endpoint (Amazon Athena) by using a private VPC endpoint or the regional endpoint, depending on the DNS configuration.</p>
Security	<p>IP access control</p> <p>You can use a combination of routing and security groups to control access to data sources stored in the AWS Cloud.</p> <p>Encryption in transit</p> <p>Both types of AWS VPN use IPsec encryption, meaning that data transferred is encrypted as it travels between AWS and on premises. This ensures that even if data sources are not configured to use encrypted communications, that data is protected while traversing the internet.</p> <p>Authentication</p>

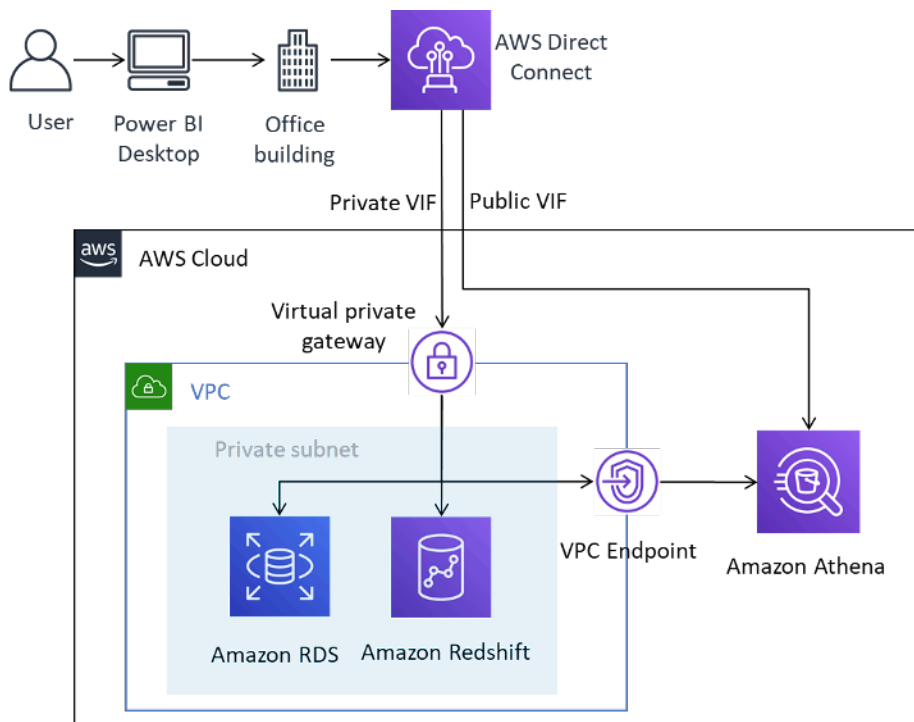
Criteria	Considerations for accessing AWS data sources using AWS VPN
	<p>Site-to-Site VPN requires a one-time configuration and, once established, is seamless to users. End users are not required to authenticate to use the Site-to-Site VPN, but they require authentication to data sources.</p> <p>On the other hand, Client VPN does require authentication by the end users in order to establish the connection. Client VPN authentication can take place via Active Directory (user-based), mutual authentication (certificate-based), or SAML SSO (user-based). Once authenticated, the connection is seamless to the end user. AWS data sources added to Power BI Desktop require authentication.</p> <p>AWS recommends that you authenticate with AWS data sources using an identity that has read-only access only to the datasets required.</p>

Criteria	Considerations for accessing AWS data sources using AWS VPN
Performance	<p>The use of AWS VPN occurs over the internet. As such, its performance envelope is similar to the first scenario presented. Some factors can impact the overall Power BI Desktop performance when accessing AWS data sources over the internet. They include:</p> <ul style="list-style-type: none">• The size of the dataset being accessed. Larger datasets take longer to retrieve. We recommend limiting queries and using filters to reduce the amount of data retrieved over the internet.• The quality of the internet connection, including bandwidth, latency, and packet loss. Where possible, access data in AWS Regions that you are geographically close to. This reduces the effect of latency. If your internet is shared, consider loading data sources at off-peak times and/or ensuring that enough bandwidth is available. <p>In general, AWS recommends testing the experience at different times of the day, with different datasets, and with progressively larger number of users.</p>
Cost	<p>Data sources that reside in a VPC and are queried using AWS VPN incur standard AWS VPN data transfer charges. To reduce costs, we recommend limiting queries and using filters to reduce the amount of data retrieved over the internet.</p>

Connecting to data sources via AWS Direct Connect

AWS Direct Connect links your internal network to an AWS Direct Connect location over a standard Ethernet fiber-optic cable. One end of the cable is connected to your router, the other to an AWS Direct Connect router. With this connection, you can create virtual interfaces directly to public AWS services (for example, to Amazon S3) or to Amazon VPC, bypassing internet service providers in your network path. An AWS Direct Connect location provides access to AWS in the Region with which it is associated. You can use a single connection in a public Region or AWS GovCloud (US) to access public AWS services in all other public Regions.

In this model, the customer's on-premises network is connected through AWS Direct Connect directly to the AWS network. While there are multiple methods for configuring AWS Direct Connect, in its simplest mode, you are given access to IP ranges within a VPC using a construct known as a private virtual interfaces (private VIF). You access to the internet/public IP ranges using a public virtual interface (public VIF).



Connecting Power BI Desktop to AWS data sources over AWS Direct Connect

When adding data sources in Power BI, you select the private IP address if it's located in a VPC, or a private VPC endpoint for the service, depending on how your DNS is configured.

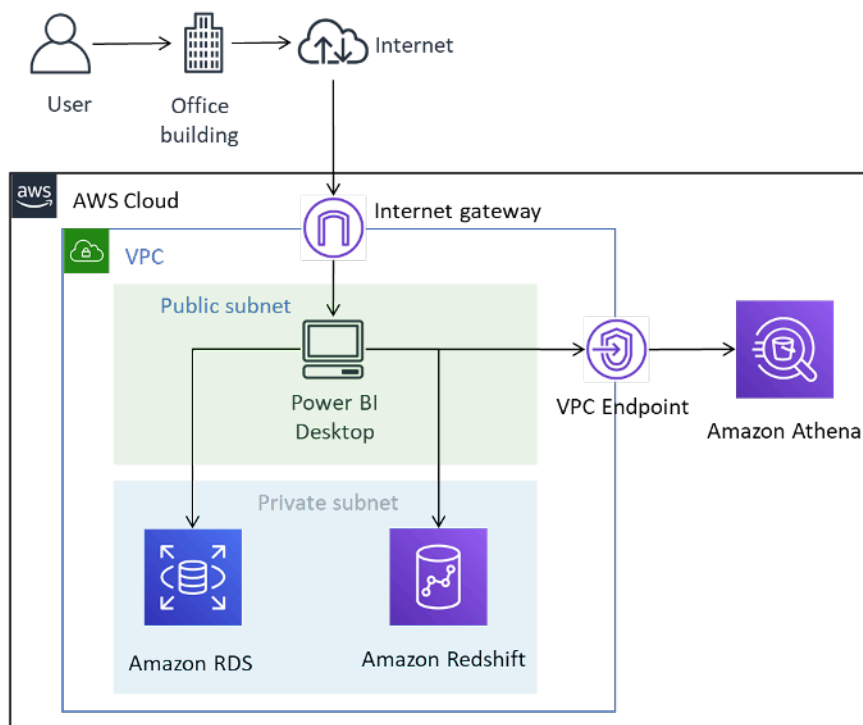
Table 3 — Considerations for accessing AWS data sources using AWS Direct Connect

Criteria	Considerations for accessing AWS data sources using AWS Direct Connect
<p>Network connectivity</p>	<p>After you configure Direct Connect, it can access data sources by connecting to private IPs in a VPC, or by using a regional service endpoint.</p>
<p>Security</p>	<p>IP access control</p> <p>You can use a combination of routing and security groups to control access to data sources stored in the AWS Cloud.</p> <p>Encryption in transit</p> <p>Direct Connect does not provide line-level encryption of data. We recommend that you either enable TLS encryption at the data source level, or combine Site-to-Site VPN with Direct Connect to ensure that data and credentials are not compromised during transmission.</p> <p>Authentication</p> <p>Direct Connect does not require any additional authentication once configured.</p> <p>AWS recommends that you authenticate AWS data sources using an identity that has read-only access only to the datasets required.</p>
<p>Performance</p>	<p>With Direct Connect, 1 Gbps and 10 Gbps ports are available. You can order speeds of 50 Mbps, 100 Mbps, 200 Mbps, 300 Mbps, 400 Mbps, and 500 Mbps from any APN Partners that support Direct Connect.</p>

Criteria	Considerations for accessing AWS data sources using AWS Direct Connect
	<p>When compared to internet-based connections, Direct Connect connections typically provide increased bandwidth and reduced latency. Depending on the capacity of the connection, Power BI Desktop performance might no longer be constrained by the network when loading data sources. As such, when accessing datasets in the same Region that the Direct Connect is associated to, your users can expect good performance for queries.</p> <p>Be mindful of loading large datasets, and note that Power BI Desktop has a 10 GB dataset limit.</p>
Cost	<p>In addition to standard Direct Connect port charges, data sources that are accessed using a Direct Connect connection incur standard Direct Connect data transfer charges, as explained in AWS Direct Connect pricing. Data sent to the AWS Cloud over Direct Connect does not incur any charges.</p> <p>To reduce costs, AWS recommends limiting queries and using filters to reduce the amount of data retrieved.</p>

Using Microsoft Power BI Desktop in the AWS Cloud

Using the Microsoft Power BI Desktop in the AWS Cloud is a popular solution for many of the challenges described in the previous section. In this model, customers host the Microsoft Power BI Desktop in the AWS Cloud, and then access it remotely on premises. The following diagram shows an example.



Microsoft Power BI Desktop deployed in the AWS Cloud

Although the diagram depicts user connectivity to the desktop occurring through the internet, AWS VPN and Direct Connect are both valid connection types too. Because only graphical management traffic is transmitted, the bandwidth requirements are well suited for typical internet connections.

In this model the Microsoft Power BI Desktop is hosted within the Amazon VPC in a public subnet and has direct network connectivity to data sources with private IP addresses, such as Amazon RDS and Amazon Redshift. You can connect to Amazon Athena and other regional services by using a VPC endpoint connection as the destination (pictured in the diagram), or by using the regional public service endpoint.

There are a number of options for hosting Microsoft Power BI Desktop in the AWS Cloud:

Option 1: Install Microsoft Power BI Desktop on an Amazon EC2 instance

In this option, you create one or more Amazon EC2 instances with Microsoft Windows Server installed. Your users connect to the instance using a Remote Desktop Protocol (RDP) application and install, configure, and use Microsoft Power BI as they would as if it was installed on their

laptop. By default, Windows Server installations are limited to two concurrent RDP client sessions. If you require more, you can buy Client Access Licenses from a Microsoft reseller. You might find that increasing the number of simultaneous users on a single server causes resource contention. You may achieve a better user experience by increasing the number of instances, each with fewer users.

Option 2: Install Microsoft Power BI in an Amazon WorkSpaces environment

Amazon WorkSpaces is a managed, secure Desktop-as-a-Service (DaaS) solution. You can use Amazon WorkSpaces to provision either Windows or Linux desktops in just a few minutes and quickly scale to provide thousands of desktops to workers across the globe.

In this option, you provision Microsoft Windows Desktops with Microsoft Power BI Desktop installed for your users. Users can access the Windows desktop using a lightweight client application that is available for Android, iOS, Fire, Mac, PC, Chromebook, and Linux devices.

Using Amazon WorkSpaces, Microsoft Power BI Desktop users have dedicated resources, and desktops can be automatically stopped to save costs. This option also has the advantage of allowing non-Windows devices to access Amazon WorkSpaces, as well as providing users a desktop environment for performing other tasks, such as accessing Microsoft Office or other Windows applications.

Option 3: Install Microsoft Power BI in an Amazon AppStream 2.0 environment

Amazon AppStream 2.0 is a fully-managed application streaming service. You centrally manage your desktop applications on AppStream 2.0 and securely deliver them to any computer. You can easily scale to any number of users across the globe without acquiring, provisioning, and operating hardware or infrastructure.

In this option, you provide your on-premises users just the Microsoft Power BI Desktop application, streamed to an HTML5-compliant web browser, with no plugins required. Like the Amazon WorkSpaces option, non-Windows devices can connect and users have access to dedicated resources.

One downside to this option is that by default, AppStream 2.0 applications are limited to 1 GB of storage capacity, which may not be suitable for saving larger models and reports. The capacity

can be increased by combining file storage from Amazon FSx for Windows File Server, but this introduces additional complexity that customers should be aware of.

Table 4 — Considerations for Power BI Desktop running in the AWS Cloud

Criteria	Considerations for Power BI Desktop running in the AWS Cloud
<p>Network connectivity</p>	<p>Network connectivity to data sources is straightforward because both the data consumer and the data sources reside in the AWS Cloud. Data sources that reside in an Amazon VPC, such as Amazon RDS and Amazon Redshift, can be accessed directly. Data sources that use regional endpoints can be accessed through the Amazon VPC internet gateway, or an Amazon VPC endpoint.</p> <p>Connectivity to the Microsoft Power BI Desktop occurs over the internet, AWS VPN, or AWS Direct Connect. All three options have modest requirements that most internet connections can meet.</p>
<p>Security</p>	<p>IP access control</p> <p>Customers can use a combination of routing and security groups to control access to data sources stored in the AWS Cloud.</p> <p>For the first option (using Amazon EC2), you can additionally use a combination of routing and security groups to allow only specific on-premises CIDR ranges.</p> <p>Encryption in transit</p> <p>We recommend that data sources within an Amazon VPC are configured to use encryptio</p>

Criteria	Considerations for Power BI Desktop running in the AWS Cloud
	<p>n for transmission of data. Regional services already make use of TLS encryption.</p> <p>Management access is encrypted for all three options.</p> <p>Authentication</p> <p>AWS recommends that you authenticate with AWS data sources using an identity that has read-only access only to the datasets required.</p> <p>All three options require the use of an authentication mechanism in order for the remote user to prove their identity before accessing the environment that hosts Microsoft Power BI Desktop. This mechanism is typically sign-in credentials, but MFA options are available as well.</p>
Performance	<p>When running in the AWS Cloud, Microsoft Power BI Desktop performance is typically better than when run on premises. Both networking and compute instances are likely to be modern and have higher specifications, and there is likely to be low latency between the application and the data sources.</p>

Criteria	Considerations for Power BI Desktop running in the AWS Cloud
Cost	<p>By placing Microsoft Power BI Desktop in the AWS Cloud, data transfer charges can be significantly reduced or removed entirely. Use care when accessing data sources across Availability Zones, in different VPC's, or in different Regions, because data transfer charges can apply.</p> <p>For each option, there are additional cost considerations that need to be factored in. Amazon EC2, Amazon WorkSpaces, and Amazon AppStream 2.0 are priced based on usage. For more information, see the pricing page for each service.</p>

Summary of Microsoft Power BI Desktop connectivity options

For a small number of users with light dataset requirements, running Microsoft Power BI Desktop on premises and connecting securely over the internet, or using AWS VPN, might be an adequate solution. Make sure that security is configured and maintained in this model. We also recommend testing this configuration to determine if it meets users' performance expectations

As the number of users increase, we recommend that you consider connectivity through AWS Direct Connect. Direct Connect provides a better user experience when loading larger datasets. Make sure that users are aware of the cost implications of transferring large datasets.

We recommend that you evaluate running Microsoft Power BI Desktop in the AWS Cloud. This is likely to provide both the best performance experience for the end user and the best management experience for cloud administrators. Amazon WorkSpaces in particular can scale from a small number of users to thousands of users. These Services also provide significant security and management benefits.

Connecting the Microsoft Power BI service to AWS data sources

Microsoft Power BI service (SaaS) can be connected directly to internet-accessible data sources, or to private data sources in an Amazon VPC. Connection to private data sources requires an application component called Microsoft on-premises data gateway. The Microsoft on-premises data gateway is downloaded and installed on an Amazon EC2 instance in the VPC and configured with Microsoft Power BI credentials. The gateway establishes an outbound connection to the Microsoft Azure Service Bus over the internet, and is configured in Microsoft Power BI to connect to data sources that it can access. Larger deployments can use multiple on-premises data gateways to balance load or increase fault tolerance.

Use of the Microsoft on-premises data gateway provides a number of substantial benefits which have been reported by AWS customers:

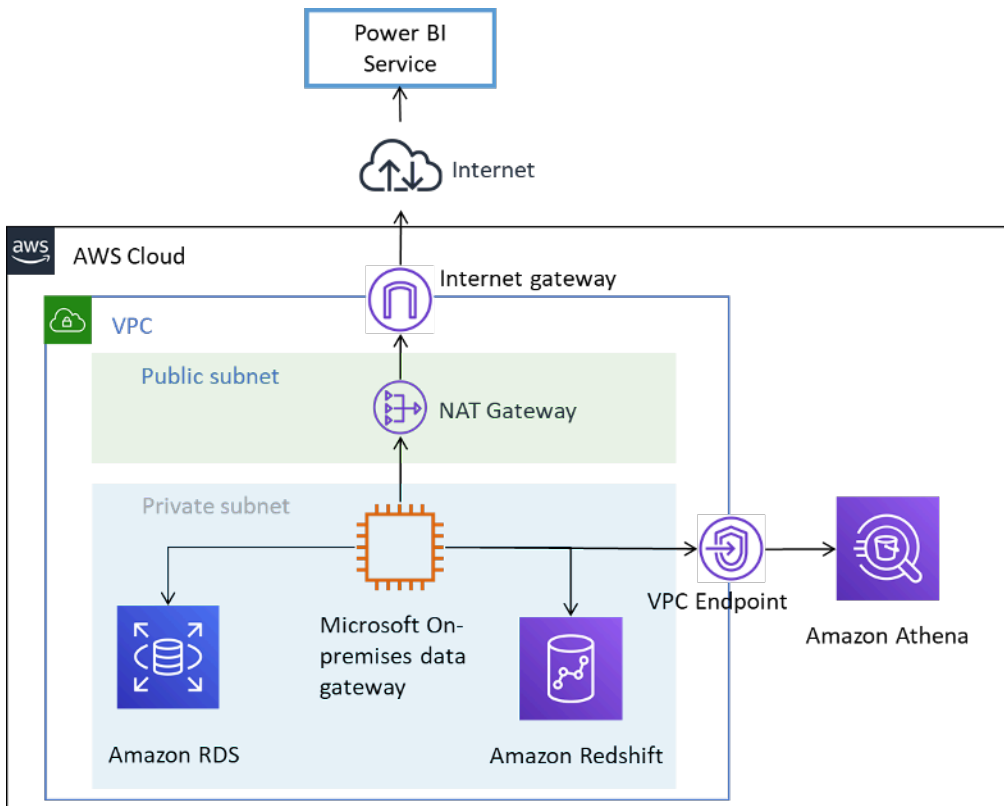
- **Improved security posture:** The [Microsoft on-premises data gateway](#) does not accept inbound connections from the Microsoft Azure Cloud, and only initiates outbound connections to the Azure Service Bus. This one-way traffic model allows you to keep your data sources private, and not expose them on the internet.
- **Reduced data transfer out:** When connecting to a data source, the Microsoft on-premises data gateway retrieves the entire result set and stores it locally in a process called spooling. Before the results are transmitted to the Power BI service the data is compressed. Users commonly report 10:1 compression ratios, which reduce not only the time to transmit the data across the internet, but reduce egress charges.
- **Reduced solution costs:** When Microsoft on-premises data gateway is used, some of the data processing required by the service is performed by the gateway instead. Using Amazon EC2, in combination with cost-reduction plans like Savings Plans, or Reserved Instances, may help reduce your overall BI solution costs.

Recommended configuration

AWS recommends that you install the Microsoft on-premises data gateway on an Amazon EC2 instance in the private subnet that contains your data sources. This subnet is configured to route requests to the internet via an [Amazon VPC NAT gateway](#) installed in a public subnet. You can use a network address translation (NAT) gateway to enable instances in a private subnet to connect to

the internet or to other AWS services, but prevent the internet from connecting to those instances. If you require a highly available data gateway implementation, we recommend using a cluster of on-premises data gateways installed across multiple EC2 instances that span different AWS Availability Zones. For information, see [Add another gateway to create a cluster](#).

The options presented in this section illustrate Amazon RDS, Amazon Redshift, and Amazon Athena. For a full discussion of all AWS data sources, refer to [Appendix: Microsoft Power BI supported AWS data sources](#).



Connecting AWS data sources to the Microsoft Power BI service

Additional considerations

Table 5 — Considerations for Microsoft Power BI service with data sources in the AWS Cloud

Criteria	Considerations for Microsoft Power BI service with data sources in the AWS Cloud
Network connectivity	Microsoft on-premises data gateway connectivity to data sources is straight forward because

Criteria	Considerations for Microsoft Power BI service with data sources in the AWS Cloud
	<p>both the data consumer and the data sources reside within the AWS Cloud. Data sources that live in an Amazon VPC, such as Amazon RDS and Amazon Redshift, can be accessed directly. Data sources that use regional endpoints can be accessed through the Amazon VPC internet gateway, or by an Amazon VPC endpoint.</p> <p>Microsoft on-premises data gateway connectivity to the Microsoft Power BI service occurs over the internet and is an outbound connection only.</p>

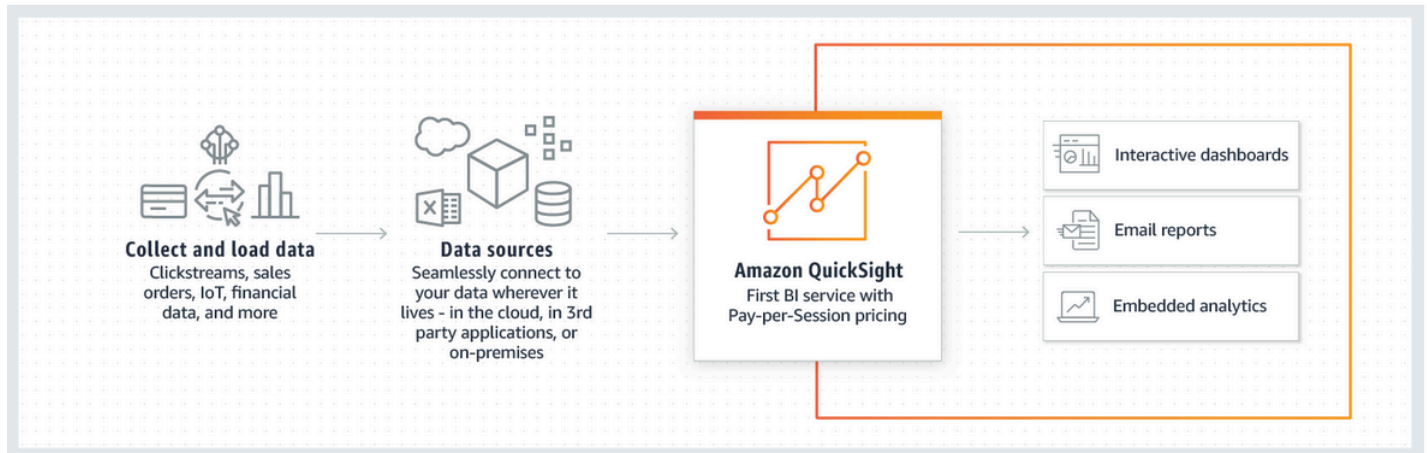
Criteria	Considerations for Microsoft Power BI service with data sources in the AWS Cloud
Security	<p>IP access control</p> <p>You can use a combination of routing and security groups to control access to data sources stored within the AWS Cloud.</p> <p>Because Microsoft on-premises data gateway is installed on an Amazon EC2 instance, it will have an associated security group that can be used to limit inbound access to the operating system. The gateway does not accept inbound requests. The instance does not need a public IP address, and should not be configured with one.</p> <p>Encryption in transit</p> <p>We recommend that data sources within an Amazon VPC are configured to use encryption for transmission of data. Regional services already make use of TLS encryption.</p> <p>Microsoft on-premises data gateway connectivity can be configured to connect to the Microsoft Azure Service Bus using HTTPS instead of TCP. We recommend using the HTTPS mode for communication. This is also the default for new gateway installations since the June 2019 gateway software version release.</p> <p>Authentication</p> <p>AWS recommends that you authenticate with AWS data sources using an identity that has read-only access only to the datasets</p>

Criteria	Considerations for Microsoft Power BI service with data sources in the AWS Cloud
	<p>required. The credentials that you enter for a data source are encrypted and stored in the gateway cloud service. The credentials are decrypted at the gateway on premises. (The credentials that you enter for a data source are encrypted and stored in the gateway cloud service.)</p> <p>Make sure that Microsoft Power BI credentials are securely controlled. Access to the service permits access to AWS data sources and potentially sensitive information they might contain.</p>
Performance	Microsoft on-premises data gateway in the AWS Cloud typically performs well due to the ability to size and scale up the Amazon EC2 instance. It also performs fast in Region networking and connectivity to the internet.

Criteria	Considerations for Microsoft Power BI service with data sources in the AWS Cloud
Cost	<p>Three factors need to be considered: Amazon EC2 instance charges, data transfer charges, and Amazon NAT gateway charges.</p> <p>Size your Amazon EC2 instances according to Microsoft's requirements. To reduce costs, you can purchase Amazon EC2 Reserved Instances or AWS Savings Plans.</p> <p>Data transferred from the Microsoft on-premises data gateway to the Microsoft BI service incurs VPC egress charges. Customers report a 10:1 compression by using the data gateway which will reduce the amount of traffic, but we recommend that you limit queries and use filters to ensure that only relevant data is transferred.</p> <p>If the Microsoft on-premises data gateway connects to data sources in different Availability Zones or different AWS Regions, data transfer charges also apply.</p> <p>If the Microsoft on-premises data gateways are located in private subnets and make use of an AWS NAT gateway, hourly and data processing charges apply. For more information, see Amazon VPC pricing.</p>

Using Amazon QuickSight

Customers considering using the Microsoft Power BI Suite with AWS are encouraged to evaluate [Amazon QuickSight](#) as an alternative. This fully managed cloud service natively connects to data sources in AWS, reducing the complexity and cost when compared to other BI solutions.



How Amazon QuickSight works

When compared to other BI solutions, Amazon QuickSight has the following benefits:

- With Amazon QuickSight, there's no need to download and install a client application. All functionality, including authoring and reporting, can be accessed from any platform (Windows, Mac, Linux, and so forth) by a web browser.
- Amazon QuickSight is delivered as a fully managed, cloud-native SaaS application and is simple to build and deploy dashboards to production. The service is serverless, which means that you do not need to calculate how many nodes and servers you need to support your users. QuickSight also takes full advantage of high availability features provided by AWS for resiliency.
- It's easy to get started in small or large settings, with the ability to add users from a point-and-click interface within QuickSight. No external administrator intervention needed.
- Amazon QuickSight is powered by Super-fast, Parallel, In-memory Calculation Engine (SPICE) for a fast response time (in the milliseconds) and interactive visualizations. Datasets can currently scale up to 200 GB.
- Amazon QuickSight pricing is simple, inexpensive, and has two components: report authors and report readers. Report authors, who create and publish interactive dashboards, are priced per user. If users do not log in during a given month, there are no charges for those users. Report readers are charged per 30-minute session, with a maximum of \$5.00 per reader per month. A

free trial allows you to evaluate Amazon QuickSight without any charges. For more information, see [Amazon QuickSight Pricing](#).

Conclusion

If you're looking to use Microsoft Power BI Desktop, we generally find that customers start experimenting with the software on premises, connecting to data sources over the internet. While private connectivity options exist for using AWS VPN and Direct Connect, many customers have concluded that running Microsoft Power BI Desktop in Amazon WorkSpaces provides a better performing experience.

If you want to connect data sources in AWS to Microsoft Power BI Service, you should feel comfortable knowing that this is an established architectural pattern. You can install the Microsoft on-premises data gateway within an Amazon VPC and connect data sources such as Amazon RDS, Amazon Redshift, Amazon Athena, Amazon OpenSearch, and AWS Lake Formation seamlessly to the service.

If you plan on using Microsoft Power BI Report Server in AWS, there is also an established path forward. You can install the Microsoft Power BI Report Server within an Amazon VPC, close to AWS data sources, and there are connectivity options for both authors and report users.

If you want a solution that provides the same business outcomes, without the added complexity of installing, configuring, patching, and scaling self-managed BI solutions, we recommend Amazon QuickSight. This fully managed service combines all the required functionality in a simple web browser experience with pay-per-user pricing. There is nothing to install and no additional components are required.

Hopefully, this is just the start of your business intelligence journey with AWS. For additional resources to help you get started, see the [Appendix: Microsoft Power BI supported AWS data sources](#).

Contributors

Contributors to this document include:

- Ralph Holm, Sr. Solutions Architect, Amazon Web Services
- Barret Newman, Sr. Solutions Architect, Amazon Web Services
- Fabrizio Napolitano, Specialist Solutions Architect for Database and Analytics, Amazon Web Services

Further reading

- [Integrate Power BI with Amazon Redshift for insights and analytics](#)
- [Connect to an Amazon Redshift database in Power BI Desktop](#)

Document revisions

To be notified about updates to this whitepaper, subscribe to the RSS feed.

Change	Description	Date
Whitepaper updated	Report Server content removed due to Microsoft licensing restrictions.	August 29, 2023
Updated	Added data sources: Amazon OpenSearch and AWS Lake Formation. Updates to Athena, Amazon RDS, and Redshift data sources. Additional guidance on Microsoft on-premises data gateway.	November 3, 2021
Reviewed	Reviewed for technical accuracy	March 1, 2021
Updated	Updated Report Server Licensing	January 15, 2021
Initial publication	Whitepaper first published	November 1, 2020

Appendix: Microsoft Power BI supported AWS data sources

The full list of supported data sources is provided by Microsoft (refer to [Power BI data sources](#)); however, the following sections for each AWS data source provide usage and configuration guidance that may be helpful for some readers.

Amazon Redshift

Amazon Redshift is a fully-managed, petabyte-scale data warehouse service in the AWS Cloud. An Amazon Redshift data warehouse is a collection of computing resources called nodes, which are organized into a group called a cluster. Each cluster runs an Amazon Redshift engine and contains one or more databases.

You should consider using Amazon Redshift when:

- You are building or migrating to a native cloud data warehouse.
- You might need to scale from a few to hundreds of Terabytes.
- You want to allow the Power BI users to transparently access data from the data lake stored in Amazon S3 and join it with tables in the data warehouse.
- Your query workload includes:
 - Queries which compute aggregation on large (multi-gigabyte and multi-terabyte) tables.
 - Extremely complex SQL with multiple joins and sub-queries.
 - A mix of complex analytical queries and simple, highly-filtered queries used in Dashboards.

When using Amazon Redshift with Microsoft Power BI, keep the following points in mind:

- Amazon Redshift is natively-supported as a Power BI data source in both Microsoft Power BI Desktop and Power BI services, and each supports import and direct query modes.
- While a Redshift cluster can be launched in a public subnet and configured to allow access from the internet, the majority of customers prefer to launch it in a private subnet to increase security. When using a private subnet, make use of the on-premises data gateway to connect from the Power BI service to Amazon Redshift.
- The Redshift connector supports Azure AD Authentication in Power BI Desktop and service.

- External tables accessed via Spectrum are not treated any differently than native Redshift tables, and Power BI has no means to differentiate them. When accessing data in external tables, make sure that:
 - Columns that contain strings of characters are catalogued as 'VARCHAR' in the AWS Glue Data Catalog and not as 'STRING', otherwise Power BI will throw the following error: Exception: OLE DB or ODBC error: [Expression.Error] We couldn't fold the expression to the data source. Please try a simpler expression..
 - Columns containing complex data types such as ARRAY, are not supported. When columns containing complex data types are used, Power BI will throw the following error: Exception: ODBC: ERROR [42703] [Microsoft]Amazon Redshift Error occurred while trying to execute a query

If you need to include them in your model, you can either enable (in Amazon Redshift) the JSON serialization at the user level or store the complex data types in a SUPER column in a native table.

Amazon RDS

Amazon RDS makes it easy to set up, operate, and scale a relational database in the cloud. Amazon RDS is available on several database instance types (optimized for memory, performance, or I/O) and provides you with six familiar database engines to choose from, including Amazon Aurora, PostgreSQL, MySQL, MariaDB, Oracle Database, and SQL Server.

You should consider using RDS when:

- You are building an operational data store.
- You are migrating SQL Server or Oracle Database data warehouse to the cloud but not interested in refactoring.
- Your query workload includes:
 - Queries which access highly-filtered data on tables that can be easily indexed.
 - Analytics queries on small-to-medium sized tables (gigabytes).
 - A mix of medium-complexity analytical queries and simple, highly-filtered queries used in Dashboards.

When using Amazon RDS with Microsoft Power BI, keep the following points in mind:

- Amazon RDS provides multiple database engines including SQL Server, MariaDB, MySQL, Oracle Database, and PostgreSQL. Note that the database engines are listed in Power BI Desktop and Power BI service, not the Amazon RDS service.
- For Amazon Aurora, use the My SQL or PostgreSQL connection type, depending on your selected database engine.
- While an Amazon RDS instance can be launched in a public subnet and configured to allow access from the internet, the majority of customers prefer to launch it in a private subnet to increase security. When using a private subnet make use of the on-premises data gateway to connect from the Power BI service to RDS.
- With Amazon RDS, you can deploy multiple editions of SQL Server (2012, 2014, 2016, 2017, and 2019) including Express, Web, Standard, and Enterprise.

Amazon Athena

Amazon Athena is an interactive query service that makes it easy to analyze data in Amazon S3 using standard SQL. Athena is out-of-the-box integrated with AWS Glue Data Catalog, allowing you to create a unified metadata repository across various services, crawl data sources to discover schemas, populate your Data Catalog with new and modified table and partition definitions, and maintain schema versioning.

You should consider Athena as a data source when:

- You want to query your data lake directly.
- Your query workload includes:
 - Queries which compute aggregation on large (multi-gigabyte and multi-terabyte) tables
 - Interactive ad hoc SQL, for exploratory purposes.

When using Amazon Athena with Microsoft Power BI, keep the following points in mind:

- With the July 2021 release of Microsoft Power BI, a Microsoft-certified connector has been introduced for Amazon Athena. You can use the Microsoft Power BI connector for Amazon Athena to analyze data from Amazon Athena in Microsoft Power BI Desktop. After you publish content to the Power BI service, you can use the Microsoft on-premises data gateway to keep the content up to date through on-demand or scheduled refreshes.
- The Microsoft Power BI connector for Amazon Athena supports both Import and Direct Query data connectivity modes. With the Import mode, selected tables and columns are imported into

Power BI Desktop for querying. With the Direct Query mode, no data is imported or copied into Power BI Desktop, and instead Power BI Desktop queries the underlying data source directly.

- For more information on the Microsoft Power BI connector for Amazon Athena, refer to [Using the Amazon Athena Power BI Connector](#).
- Note that the Microsoft Power BI connector for Amazon Athena requires the use of the Amazon Athena ODBC driver and a valid ODBC DSN configuration on your system to query Amazon Athena. To download the latest ODBC driver and for configuration information, refer to [Connecting to Amazon Athena with ODBC](#).
- For a tutorial on the configuration steps and best practices when using the Microsoft Power BI connector for Amazon Athena, refer to [Creating dashboards quickly on Microsoft Power BI using Amazon Athena](#).

Amazon OpenSearch Service

You can use SQL to query your Amazon OpenSearch Service, rather than using the JSON-based search query DSL. Querying with SQL is useful if you're already familiar with the language or want to integrate your domain with an application that uses it, such as Microsoft Power BI.

You should consider Amazon OpenSearch Service as a data source when:

- You have semi-structured data such as log files or JSON output, and need to search, analyze, or visualize the information quickly.

When using Amazon OpenSearch Service with Microsoft Power BI, keep the following points in mind:

- Connectivity to Amazon OpenSearch Service requires the Open Database Connectivity (ODBC) driver, which is a read-only ODBC driver for Windows and macOS that lets you connect business intelligence (BI) and data visualization applications like [Tableau](#), [Microsoft Excel](#), and [Power BI](#) to the SQL plugin on your cluster. The driver is available on the OpenSearch [Download & Get Started](#) website. For configuration instructions, refer to the “Customizing the ODBC driver” section in [OpenSearch ODBC driver website](#).
- Only import mode is currently supported.
- Power BI connectivity to Amazon OpenSearch Service currently requires the use of a beta connector. Refer to [Microsoft Power Query Documentation - Connector reference: Amazon Opensearch Service \(Beta\)](#) to get started.

AWS Lake Formation

Lake Formation helps you collect and catalog data from databases and object storage, move the data into your new [Amazon S3](#) data lake, clean and classify your data using machine learning algorithms, and secure access to your sensitive data. Your users can access a centralized [data catalog](#) which describes available data sets and their appropriate usage. Your users then utilize these data sets with their choice of analytics and machine learning services, like [Amazon Redshift](#), [Amazon Athena](#), and (in beta) [Amazon EMR](#) for Apache Spark. Lake Formation builds on the capabilities available in [AWS Glue](#).

You should consider Lake Formation if you need fine-grained (row and column) level access to your data lake instead of the traditional IAM based controls.

When using Lake Formation with Microsoft Power BI, keep the following points in mind:

- To query data from the Lake Formation Data Catalog with Power BI Desktop or Power BI service, use the same process and configuration as querying data in Athena. If you are making use of the Lake Formation permission model, ensure that the ODBC DSN configuration for Amazon Athena has the "LakeformationEnabled" property key set to a value of "true". This tells the Amazon Athena ODBC driver to use the Lake Formation service for authorization, instead of AWS Security Token Service directly. For more information, refer to the documentation for [Connecting to Amazon Athena with ODBC](#).
- The "Use only IAM access control" setting enabled for compatibility with existing Data Catalog behavior will provide full compatibility.
- Upgrading AWS Glue Data Permissions to the Lake Formation Model may introduce incompatibilities and should be tested before use. Preliminary testing indicates that column level grant or deny are being honored, but row and cell level filtering have not been tested by the authors, as this is still in preview and subject to change.

Notices

Customers are responsible for making their own independent assessment of the information in this document. This document: (a) is for informational purposes only, (b) represents current AWS product offerings and practices, which are subject to change without notice, and (c) does not create any commitments or assurances from AWS and its affiliates, suppliers or licensors. AWS products or services are provided “as is” without warranties, representations, or conditions of any kind, whether express or implied. The responsibilities and liabilities of AWS to its customers are controlled by AWS agreements, and this document is not part of, nor does it modify, any agreement between AWS and its customers.

© 2021 Amazon Web Services, Inc. or its affiliates. All rights reserved.