



亚马逊云科技白皮书

将 Microsoft Power BI 与 AWS 云结合使用



将 Microsoft Power BI 与 AWS 云结合使用: 亚马逊云科技白皮书

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

Amazon 的商标和商业外观不得用于任何非 Amazon 的商品或服务，也不得以任何可能引起客户混淆或者贬低或诋毁 Amazon 的方式使用。所有非 Amazon 拥有的其他商标均为各自所有者的财产，这些所有者可能附属于 Amazon、与 Amazon 有关联或由 Amazon 赞助，也可能不是如此。

Table of Contents

将 Microsoft Power BI 与 AWS 云结合使用	1
摘要	1
介绍	2
Microsoft Power BI 套件	3
Power BI Desktop	3
Power BI 服务	3
Power BI 报表服务器	4
本地部署数据网关	4
将 Microsoft Power BI Desktop 连接到 AWS 数据源	5
在本地部署中使用 Power BI Desktop	5
通过 Internet 连接到数据源	5
通过 AWS VPN 连接到数据源	8
通过 AWS Direct Connect 连接到数据源	11
在 AWS 云中使用 Microsoft Power BI Desktop	14
选项 1：在 Amazon EC2 实例上安装 Microsoft Power BI Desktop	15
选项 2：在 Amazon WorkSpaces 环境中安装 Microsoft Power BI	15
选项 3：在 Amazon AppStream 2.0 环境中安装 Microsoft Power BI	15
Microsoft Power BI Desktop 连接选项摘要	17
将 Microsoft Power BI 服务连接到 AWS 数据源	19
建议的配置	19
其他注意事项	20
使用 Amazon QuickSight	24
总结	25
贡献者	26
延伸阅读	27
文档修订	28
附录：Microsoft Power BI 支持的 AWS 数据源	29
Amazon Redshift	29
Amazon RDS	30
Amazon Athena	31
Amazon OpenSearch Service (Amazon Elasticsearch Service 的后继者)	32
Amazon Lake Formation	32
声明	34

将 Microsoft Power BI 与 AWS 云结合使用

将 Microsoft Power BI 与 AWS 云结合使用

发布日期：2021 年 11 月 3 日 ([文档修订](#))

摘要

这份白皮书讨论了如何将 Microsoft Power BI (桌面、报表服务器、服务和本地部署数据网关) 与 Amazon Web Services (AWS) 集成和结合使用。它为希望将 Microsoft Power BI 产品连接到 AWS 服务 (如 Amazon Redshift、Amazon Athena、Amazon RDS、Amazon OpenSearch 和 AWS Lake Formation) 的客户提供了选项，重点关注连接性、安全性、性能和成本优化。

本白皮书面向希望快速了解 Microsoft Power BI 概念以及在使用 AWS 服务作为数据源时利用这些技术的具体选项的 IT 决策者和架构师。

引言

拥有各种规模企业的客户都在使用 AWS 产品和服务来可靠、经济高效且安全地存储其数据。这在一定程度上要归功于可用的成熟数据存储和分析服务的广泛生态系统。其中一些服务包括以下服务：

- [Amazon Simple Storage Service](#) (Amazon S3) 提供简单、可扩展、安全且经济高效的数据存储库。它已成为存储应用程序数据的行业标准，也是客户数据湖的首选。
- [Amazon Athena](#) 是一种交互式查询服务，可使用此服务通过标准 SQL 在 Amazon S3 中轻松分析数据。
- [Amazon Relational Database Service](#) (Amazon RDS) 让您能够在云中轻松设置、操作和扩展关系数据库。它在自动执行耗时的管理任务（如硬件预置、数据库设置、修补和备份）的同时，提供经济实用的可调容量。SQL Server、Oracle 数据库、MySQL、MariaDB 和 PostgreSQL 引擎都可用。
- [Amazon Redshift](#) 是完全托管式且可大规模扩展的数据仓库，可轻松分析结构化和非结构化数据集。
- [Amazon QuickSight](#) 是一项采用云技术的快速商业智能服务，可以轻松向组织中的所有人提供见解。
- [Amazon OpenSearch](#)（Amazon Elasticsearch Service 的继任者）是一项完全托管的服务，方便您大规模经济高效地部署、保护和运行 Elasticsearch。
- [AWS Lake Formation](#) 是一项服务，可以在几天内轻松建立安全的数据湖。

为了更好地了解服务之间的关系，我们经常将数据服务标记为数据源或数据使用者。数据源允许客户和应用程序存储和检索服务中的数据。通常，数据源还具有内置计算功能，可以提供计算分析和筛选功能。但最终，数据会加载到这些数据源中，最终由数据使用者从这些数据源中检索数据。Amazon S3、Amazon Athena 和 Amazon Redshift 都是很好的数据源示例。

另一方面，数据使用者可以访问来自数据源的数据，并且通常会对其进行处理。它们也可以选择显示它。Amazon QuickSight 和 Microsoft Power BI 套件是很好的数据使用者示例。它们从数据源中读取，然后协助分析、可视化和发布信息。

AWS 为客户提供了充分的灵活性，可以混合使用他们偏爱的技术来满足其数据需求。尽管许多客户选择 Amazon QuickSight 来满足其商业智能 (BI) 需求，但其他客户则选择 Microsoft Power BI、Tableau 和 Qlik 等供应商。

本文档重点介绍 Microsoft Power BI 产品和服务套件，以及如何将它们与 AWS 服务结合使用。

Microsoft Power BI 套件

为了减少由于产品命名相似性而引起的混淆，本白皮书介绍了每个 Microsoft Power BI 产品和服务的含义。

Power BI Desktop

Power BI Desktop 是您安装在本地计算机上的免费应用程序。它使您可以连接、转换和可视化您的数据。借助 Power BI Desktop，您可以连接到多个不同的数据源，并将它们（通常称为建模）合并为一个数据模型。借助此数据模型，您可以构建视觉对象和视觉对象集合，这些视觉对象可以作为报表与组织内的其他人员共享。

Power BI Desktop 可以连接到本地或通过网络提供的任何受支持的数据源。有关支持的数据源，请参阅[附录：Microsoft Power BI 支持的 AWS 数据源](#)。

从事商业智能项目的大多数用户都使用 Power BI Desktop 来创建报表。然后，他们将内容推送到 Power BI 报表服务器或 Power BI 服务，以便与其他人共享其报表。将内容从 Power BI Desktop 推送到 Power BI 报表服务器或 Power BI 服务的行为称为发布。有关详细信息，请参阅[什么是 Power BI Desktop？](#)

Note

Power BI Desktop 是仅适用于 Windows 的应用程序。它不适用于 Linux、macOS 或其他操作系统。

Power BI 服务

Power BI 是软件服务、应用程序和连接器的集合，它们协同工作，以最有效地为您和您的企业服务的方式帮助您创建、共享和使用业务见解。Power BI 服务（有时也称为在线 Power BI）是 Power BI 的软件即服务 (SaaS) 部分。有关详细信息，请参阅[什么是 Power BI 服务？](#)

Power BI 服务是一项基于云的服务。它支持团队和组织的简单报表编辑和协作。您也可以连接到 Power BI 服务中的数据源，但建模会受到限制。

从事商业智能项目的大多数报表设计人员均使用 Power BI Desktop 创建报表，然后使用 Power BI 服务与其他人分发其报表。有关此关键组件的更多信息，请参阅[将 Microsoft Power BI 服务连接到 AWS 数据源](#)。

Power BI 报表服务器

Power BI 报表服务器是一个私有报表服务器，其中包含一个 Web 门户，您可以在其中显示和管理报表和 KPI。如果客户不希望将其报表发布到 Power BI 服务，则可以使用 Power BI 报表服务器。尽管 Power BI 报表服务器最初是为本地部署环境设计的，但它也可以在 AWS 上运行。有关更多信息，请参阅[在 AWS 中使用 Microsoft Power BI 报表服务器](#)。

本地部署数据网关

Microsoft 本地部署数据网关是一个常用部署的组件，可以提高 Power BI 部署的安全性和性能。它允许 Power BI 服务访问私有化数据源，这些数据源位于另一个设施中，可通过数据源和数据网关之间的内部网络连接进行访问。尽管它通常作为服务器组件安装，但您也可以在本地上作为应用程序安装个人模式。本白皮书仅重点介绍标准（服务器）模式。有关更多信息，请参阅[将 Microsoft Power BI 服务连接到 AWS 数据源](#)。

将 Microsoft Power BI Desktop 连接到 AWS 数据源

大多数情况下，使用 Microsoft Power BI Desktop 的客户都对如何从其本地部署计算机和网络连接到 AWS 数据源感兴趣。桌面应用程序通常在其本地 Windows 笔记本电脑上运行，而与 AWS 数据源的物理和逻辑连接被视为进入的最大障碍。

但是，还有另一种选择，那就是在 AWS 云中运行 Microsoft Power BI Desktop。此选项可显著减少与 AWS 数据源的连接障碍，但仍需要一些额外的注意事项。本节将讨论这两种模型。我们将研究每种方法对连接性、安全性、性能和成本的影响，以便您可以决定哪种方法最适合您。本节中介绍的选项演示了 Amazon RDS、Amazon Redshift 和 Amazon Athena。有关所有 AWS 数据源的完整讨论，请参阅[附录：Microsoft Power BI 支持的 AWS 数据源](#)。

在本地部署中使用 Power BI Desktop

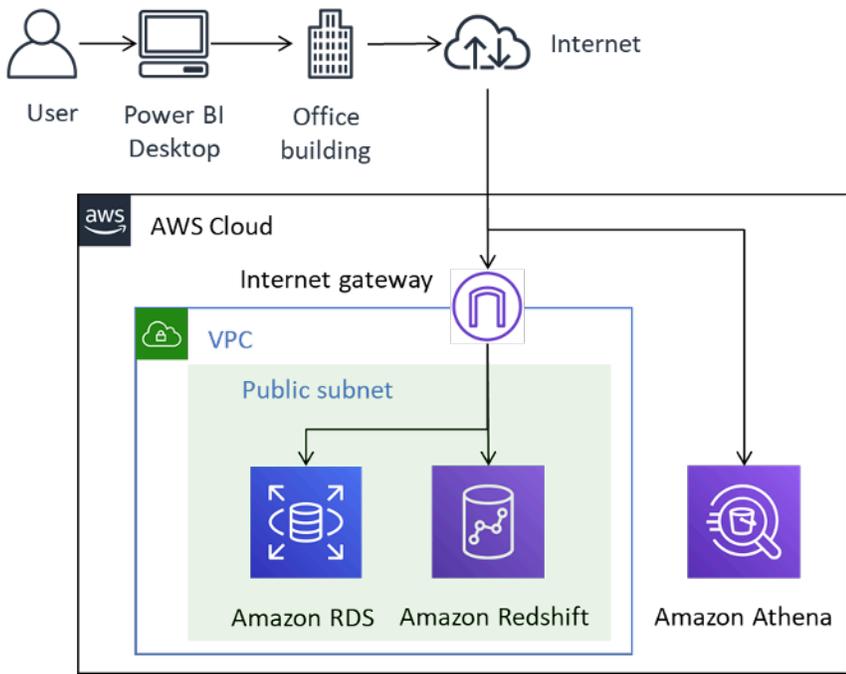
如果您计划在本地部署中使用 Power BI desktop，并且数据源存储在 AWS 云中，则 Power BI 可以通过以下三种方式之一访问这些数据源：

- 使用 Internet 连接到数据源。
- 使用 AWS 虚拟专用网络 (AWS AWS VPN) 连接到数据源。
- 使用 AWS Direct Connect 连接到数据源。

以下各节详细介绍了每种方法。

通过 Internet 连接到数据源

在此模型中，Power BI Desktop 应用程序放置一个出站连接，该连接通过 Internet 路由到可通过 Internet 访问的 AWS 数据源的 IP 地址。例如，Amazon RDS 和 Amazon Redshift 在客户的 Amazon Virtual Private Cloud (Amazon Amazon VPC) 中实例化，它们支持公开可访问性选项，以使实例可通过 Internet 访问。可以使用特定区域的服务终端节点直接从 Internet 查询 Amazon Athena。



Power BI 通过 Internet 与 AWS 数据源的连接

尽管这种连接方法在技术上是可行的，但除了少数用户之外，我们不建议其他任何人使用这种方法。下表列出了重要的注意事项。

表 1 — 通过 Internet 访问 AWS 数据源的注意事项

标准	通过 Internet 访问 AWS 数据源的注意事项
网络连接	可通过连接到 VPC 中的私有 IP 地址，或使用区域或 VPC 服务终端节点来获取数据源。Power BI Desktop 通过 VPN 连接并直接访问数据源（Amazon RDS、Amazon Redshift、基于 Amazon EC2 的数据源），或者使用私有 VPC 终端节点或区域终端节点访问具有区域终端节点 (Amazon Athena) 的服务，具体取决于 DNS 配置。
安全性	IP 访问控制 安全组充当实例的虚拟防火墙以控制入站和出站流量。为了限制对可信任实体的访问，请将安全

标准	通过 Internet 访问 AWS 数据源的注意事项
	<p>组配置为仅允许与已知的无类域间路由 (CIDR) 范围关联的入站 IP 范围。</p> <p>传输中加密</p> <p>AWS 建议您为使用公有 IP 地址的任何数据源 (例如 Amazon RDS、Amazon Redshift 或任何基于 Amazon EC2 的数据源) 配置加密。这可确保降低数据或凭证在传输过程中被泄露的风险。未能配置加密会带来很大的风险。请勿忽视这方面的情况。</p> <p>诸如 Amazon Athena 之类的区域服务终端节点经 TLS 加密。此外, 流式传输到 JDBC 或 ODBC 客户端的 Amazon Athena 查询结果使用传输层安全性 (TLS) 进行加密。</p> <p>身份验证和授权</p> <p>AWS 建议您使用提供对数据集的只读访问权限的凭证, 并根据贵公司的策略设置轮换凭证的流程。</p>

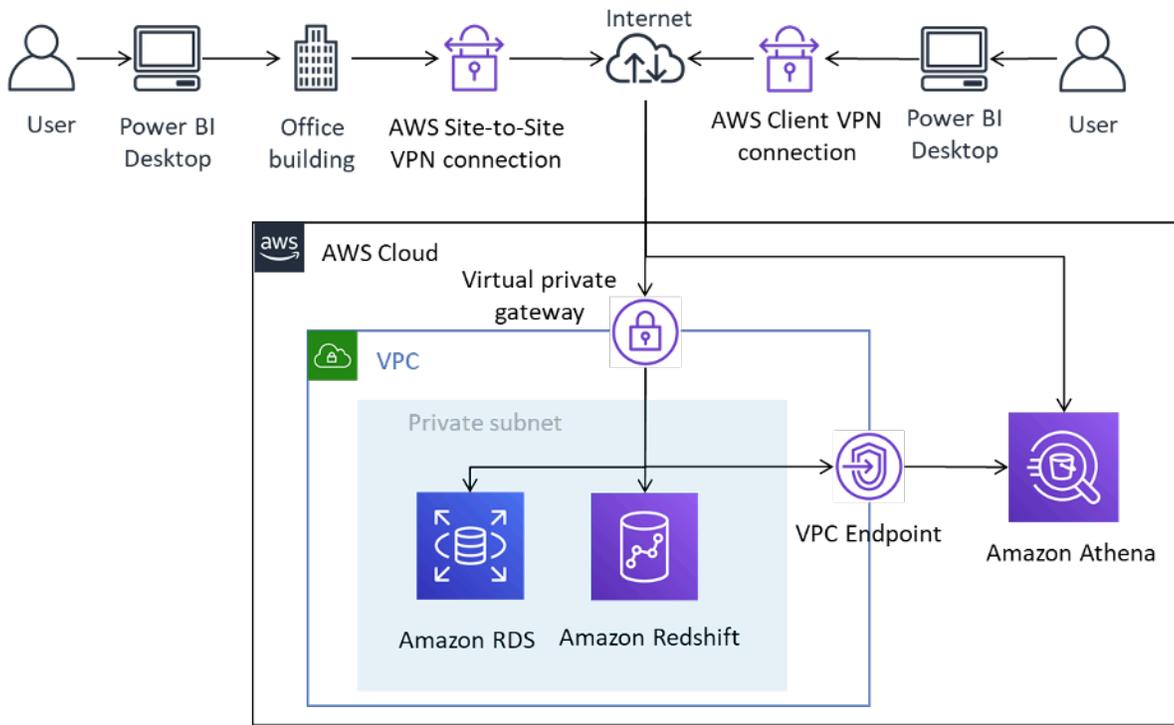
标准	通过 Internet 访问 AWS 数据源的注意事项
性能	<p>通过 Internet 访问 AWS 数据源时，可能会影响 Power BI Desktop 整体性能的一些因素包括：</p> <ul style="list-style-type: none"> 正在访问的数据集的大小。较大的数据集需要更长的时间来检索。我们建议限制查询并使用筛选条件来减少通过 Internet 检索的数据量。 Internet 连接的质量，包括带宽、延迟和丢包。在可能的情况下，访问您在地理位置上较近的 AWS 区域中的数据，以减少延迟的影响。如果您的 Internet 是共享的，请考虑在非高峰时间加载数据源，并确保有足够的带宽可用。 <p>通常，AWS 建议在一天中的不同时间、使用不同的数据集以及在用户数量逐渐增加的情况下测试体验。</p>
成本	<p>驻留在 VPC 中并通过 Internet 使用公有 IP 地址查询的数据源会产生标准 Amazon VPC 数据出口费用。为了降低成本，我们建议限制查询并使用筛选条件来减少通过 Internet 检索的数据量。</p>

通过 AWS VPN 连接到数据源

在此模型中，Power BI Desktop 安装使用以下两种 AWS VPN 方法之一连接到 AWS 网络中的数据源：AWS Site-to-Site VPN 或 AWS Client VPN。每种连接类型均能提供高度可用、有弹性的托管云 VPN 解决方案，以保护您的网络流量安全。

Site-to-Site VPN 在您的网络和 AWS VPN 或 AWS Transit Gateway 之间创建加密隧道。Client VPN 可使用免费的 VPN 软件客户端将您的用户连接到 AWS 或本地资源。

来自 Site-to-Site VPN 和 Client VPN 连接的 VPN 流量都会在您的 VPC 中停止。因此，它可以路由到私有 IP 地址，因此您的实例不再需要公开 IP 地址。对于具有可从公开服务终端节点（例如 Athena）访问的数据路径的服务，这些服务请求可以通过 Internet 进行路由，也可以通过 VPN 连接和 VPC 终端节点进行路由。



通过 Site-to-Site VPN 和 Client VPN 将 Power BI Desktop 连接到 AWS 数据源

Site-to-Site VPN 还可以连接到 AWS Transit Gateway，从而方便访问分布在多个 VPC 中的数据源。

使用 AWS VPN 的优势是，在访问存储在 AWS 中的数据源时采用加密，而无需显式配置每个数据源。配置完成后，VPN 技术基本上对终端用户来说是无缝的。

表 2 — 使用 AWS VPN 访问 AWS 数据源的注意事项

标准	使用 AWS VPN 访问 AWS 数据源的注意事项
网络连接	可通过连接到 VPC 中的私有 IP 地址，或使用区域或 VPC 服务终端节点来获取数据源。Power BI Desktop 通过 VPN 连接并直接访问数据源（Amazon RDS、Amazon Redshift、基于 Amazon EC2 的数据源），或者使用私有 VPC 终端节点或区域终端节点访问具有区域终端节点 (Amazon Athena) 的服务，具体取决于 DNS 配置。
安全	IP 访问控制

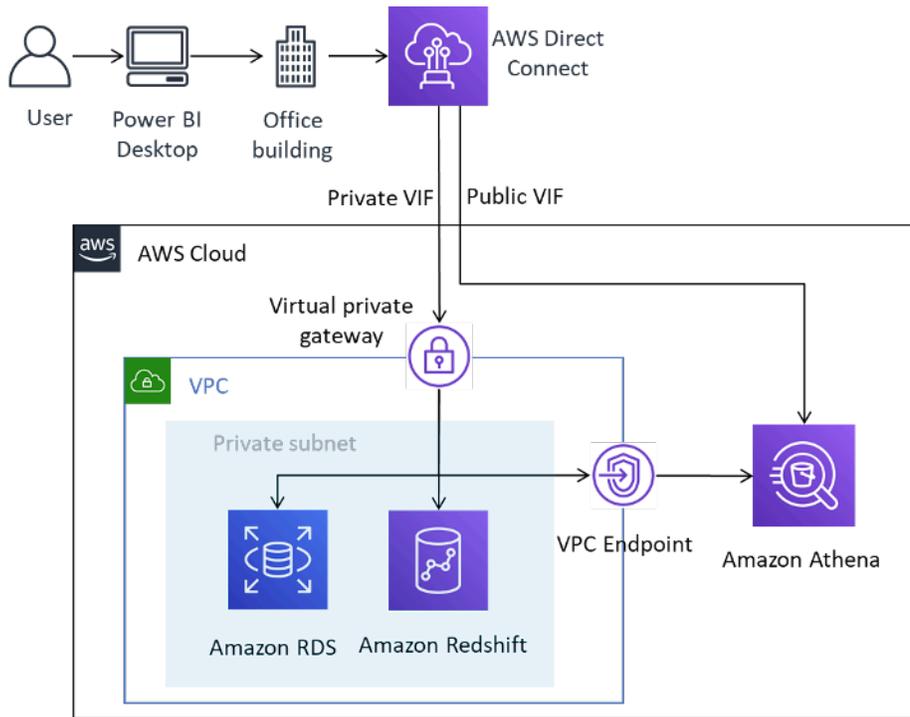
标准	使用 AWS VPN 访问 AWS 数据源的注意事项
	<p>您可以结合使用路由组和安全组来控制对存储在 AWS 云中的数据源的访问。</p> <p>传输中加密</p> <p>这两种类型的 AWS VPN 均使用 IPsec 加密，这意味着传输的数据在 AWS 与本地部署之间传输时会被加密。这样可以确保即使数据源未配置为使用加密通信，也可以在遍历 Internet 时保护数据。</p> <p>身份验证</p> <p>Site-to-Site VPN 需要一次性配置，一旦建立后，对用户来说就是无缝的。终端用户无需进行身份验证即可使用 Site-to-Site VPN，但他们需要对数据源进行身份验证。</p> <p>另一方面，Client VPN 确实需要终端用户进行身份验证才能建立连接。Client VPN 身份验证可通过 Active Directory (基于用户)、相互身份验证 (基于证书) 或 SAML SSO (基于用户) 进行。通过身份验证后，连接对终端用户来说是无缝的。添加到 Power BI Desktop 的 AWS 数据源需要进行身份验证。</p> <p>AWS 建议您使用仅对所需数据集具有只读访问权限的身份对 AWS 数据源进行身份验证。</p>

标准	使用 AWS VPN 访问 AWS 数据源的注意事项
性能	<p>AWS VPN 的使用发生在 Internet 上。因此，其性能范围类似于所介绍的第一个场景。通过 Internet 访问 AWS 数据源时，一些因素可能会影响 Power BI Desktop 的整体性能。其中包括：</p> <ul style="list-style-type: none"> 正在访问的数据集的大小。较大的数据集需要更长的时间来检索。我们建议限制查询并使用筛选条件来减少通过 Internet 检索的数据量。 Internet 连接的质量，包括带宽、延迟和丢包。在可能的情况下，访问您在地理位置上较近的 AWS 区域中的数据。这样可减少延迟的影响。如果您的 Internet 是共享的，请考虑在非高峰时间加载数据源和/或确保有足够的带宽可用。 <p>通常，AWS 建议在一天中的不同时间、使用不同的数据集以及在用户数量逐渐增加的情况下测试体验。</p>
成本	<p>驻留在 VPC 中并使用 AWS VPN 查询的数据源会产生标准 AWS VPN 数据传输费用。为了降低成本，我们建议限制查询并使用筛选条件来减少通过 Internet 检索的数据量。</p>

通过 AWS Direct Connect 连接到数据源

AWS Direct Connect 通过标准的以太网光纤电缆将您的内部网络链接到 AWS Direct Connect 位置。电缆的一端接到您的路由器，另一端接到 AWS Direct Connect 路由器。借助此连接，您就可以创建直接连接到公有 AWS 服务（如连接到 Amazon S3）或连接到 Amazon VPC 的虚拟接口，从而绕过您的网络路径中的 Internet 服务提供商。AWS Direct Connect 位置在其关联的区域提供对 AWS 的访问权限。您可以使用公有区域或 AWS GovCloud（美国）中的单个连接访问所有其他公有区域中的公有 AWS 服务。

在此模型中，客户的本地部署网络通过 AWS Direct Connect 直接连接到 AWS 网络。虽然有多种方法可以配置 AWS Direct Connect，但在最简单的模式下，您可以使用称为私有虚拟接口（私有 VIF）的结构访问 VPC 中的 IP 范围。您可以使用公有虚拟接口（公有 VIF）访问 Internet/公有 IP 范围。



通过 AWS Direct Connect 将 Power BI Desktop 连接到 AWS 数据源

在 Power BI 中添加数据源时，如果私有 IP 地址位于 VPC 中，则选择私有 IP 地址，或者选择服务的私有 VPC 终端节点，具体取决于您的 DNS 配置方式。

表 3 — 使用 AWS Direct Connect 访问 AWS 数据源的注意事项

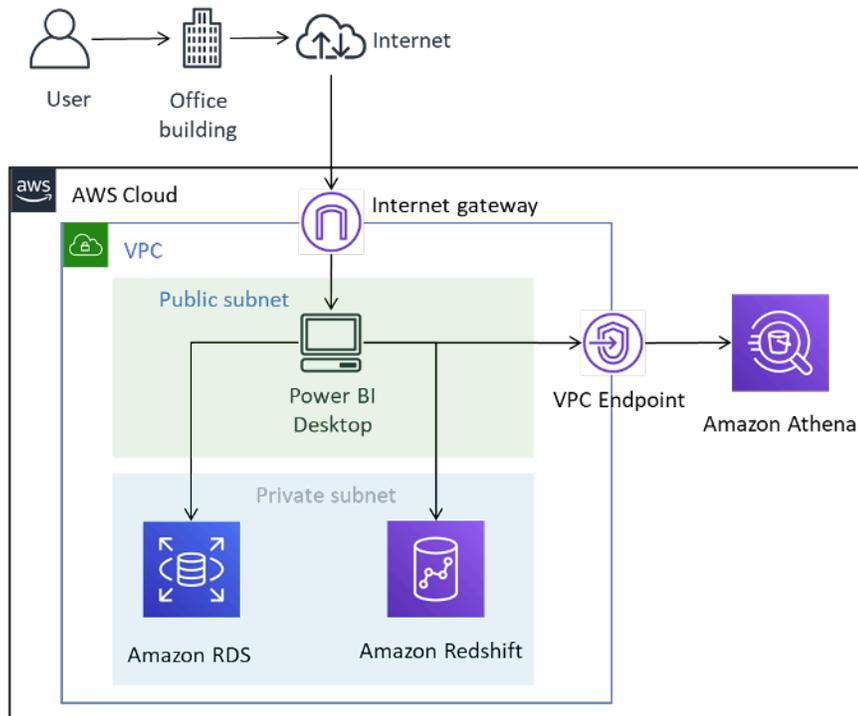
标准	使用 AWS Direct Connect 访问 AWS 数据源的注意事项
网络连接	配置 Direct Connect 后，它可以通过连接到 VPC 中的私有 IP 或使用区域服务终端节点来访问数据源。
安全性	<p>IP 访问控制</p> <p>您可以结合使用路由组和安全组来控制对存储在 AWS 云中的数据源的访问。</p> <p>传输中加密</p>

<p>标准</p>	<p>使用 AWS Direct Connect 访问 AWS 数据源的注意事项</p>
	<p>Direct Connect 不提供数据的行级别加密。我们建议您在数据源级别启用 TLS 加密，或者将 Site-to-Site VPN 与 Direct Connect 结合使用，以确保数据和凭证在传输过程中不会泄露。</p> <p>身份验证</p> <p>配置后，Direct Connect 不需要任何其他身份验证。</p> <p>AWS 建议您使用仅对所需数据集具有只读访问权限的身份对 AWS 数据源进行身份验证。</p>
<p>性能</p>	<p>使用 Direct Connect 时，可以使用 1 Gbps 和 10 Gbps 的端口。您可以从任何支持 Direct Connect 的 APN 合作伙伴处订购 50 Mbps、100 Mbps、200 Mbps、300 Mbps、400 Mbps 和 500 Mbps 的速度。</p> <p>与基于 Internet 的连接相比，Direct Connect 连接通常可提供更高的带宽并减少延迟。根据连接的容量，加载数据源时 Power BI Desktop 的性能可能不再受网络的限制。因此，在访问与 Direct Connect 关联的同一区域中的数据集时，您的用户可以期望查询具有良好的性能。</p> <p>请注意加载大型数据集，并注意 Power BI Desktop 的数据集限制为 10 GB。</p>

<p>标准</p>	<p>使用 AWS Direct Connect 访问 AWS 数据源的注意事项</p>
<p>成本</p>	<p>如 AWS Direct Connect 定价 中所述，除了标准 Direct Connect 端口费用外，使用 Direct Connect 连接访问的数据源还会产生标准 Direct Connect 数据传输费用。通过 Direct Connect 发送到 AWS 云的数据不会产生任何费用。</p> <p>为了降低成本，AWS 建议限制查询并使用筛选条件来减少检索的数据量。</p>

在 AWS 云中使用 Microsoft Power BI Desktop

在 AWS 云中使用 Microsoft Power BI Desktop 是应对上一节中描述的许多挑战的常用解决方案。在此模型中，客户在 AWS 云中托管 Microsoft Power BI Desktop，然后在本地部署中远程访问它。下图显示了一个示例。



在 AWS 云中部署的 Microsoft Power BI Desktop

尽管此图描述了用户通过 Internet 连接到该桌面的情况，但 AWS VPN 和 Direct Connect 也均是有效的连接类型。由于仅传输了图形管理流量，因此带宽要求非常适合典型的 Internet 连接。

在此模型中，Microsoft Power BI Desktop 托管在公有子网中的 Amazon VPC 内，并与具有私有 IP 地址的数据源（如 Amazon RDS 和 Amazon Redshift）直接建立网络连接。您可以使用 VPC 终端节点连接作为目标（如图所示）或使用区域公有服务终端节点来连接到 Amazon Athena 和其他区域服务。

在 AWS 云中托管 Microsoft Power BI Desktop 有多种选项：

选项 1：在 Amazon EC2 实例上安装 Microsoft Power BI Desktop

在此选项中，创建一个或多个安装了 Microsoft Windows Server 的 Amazon EC2 实例。您的用户使用远程桌面协议 (RDP) 应用程序连接到该实例，然后像在笔记本电脑上安装一样安装、配置和使用 Microsoft Power BI。默认情况下，Windows Server 安装限制为两个并发 RDP 客户端会话。如果需要更多，可以从 Microsoft 经销商处购买客户端访问许可证。您可能会发现，增加单个服务器上的并发用户数会导致资源争用。您可以通过增加实例的数量来获得更好的用户体验，其中每个实例都有更少的用户。

选项 2：在 Amazon WorkSpaces 环境中安装 Microsoft Power BI

Amazon WorkSpaces 是一种安全的托管式桌面即服务 (DaaS) 解决方案。您可以使用 Amazon WorkSpaces 在短短几分钟内预置 Windows 或 Linux 桌面，并快速扩展，从而为全球各地的工作人员提供数千个桌面。

在该选项中，为您的用户预置安装了 Microsoft Power BI Desktop 的 Microsoft Windows 桌面。用户可以使用适用于 Android、iOS、Fire、Mac、PC、Chromebook 和 Linux 设备的轻量级客户端应用程序访问 Windows 桌面。

使用 Amazon WorkSpaces，Microsoft Power BI Desktop 用户拥有专用资源，并且可以自动停止桌面以节省成本。此选项的优势还在于允许非 Windows 设备访问 Amazon WorkSpaces，并为用户提供桌面环境来执行其他任务，例如访问 Microsoft Office 或其他 Windows 应用程序。

选项 3：在 Amazon AppStream 2.0 环境中安装 Microsoft Power BI

Amazon AppStream 2.0 是一种完全托管式应用程序流式传输服务。您可以在 AppStream 2.0 上集中管理桌面应用程序，并将它们安全地交付到任何计算机上。您可以轻松扩展到全球任意数量的用户，而无需获取、预置和操作硬件或基础设施。

在此选项中，您只需向本地用户提供流式传输到符合 HTML5 标准的 Web 浏览器的 Microsoft Power BI Desktop 应用程序，而无需任何插件。与 Amazon WorkSpaces 选项一样，非 Windows 设备可以连接，用户也可以访问专用资源。

此选项的一个缺点是，默认情况下，AppStream 2.0 应用程序的存储容量限制为 1 GB，这一点可能不适合保存较大的模型和报告。可以通过合并来自 Amazon FSx for Windows File Server 的文件存储来增加容量，但这样会增加客户应注意的额外复杂性。

表 4 — 在 AWS 云中运行的 Power BI Desktop 的注意事项

标准	在 AWS 云中运行的 Power BI Desktop 的注意事项
<p>网络连接</p>	<p>与数据源的网络连接非常简单，因为数据使用者和数据源均驻留在 AWS 云中。可以直接访问驻留在 Amazon VPC 中的数据源，例如 Amazon RDS 和 Amazon Redshift。使用区域终端节点的数据源可通过 Amazon VPC 互联网网关或 Amazon VPC 终端节点进行访问。</p> <p>与 Microsoft Power BI Desktop 的连接是通过 Internet、AWS VPN 或 AWS Direct Connect 进行的。这三种选择都有适度的要求，大多数 Internet 连接都能满足。</p>
<p>安全</p>	<p>IP 访问控制</p> <p>客户可以使用路由和安全组的组合来控制对存储在 AWS 云中的数据源的访问。</p> <p>对于第一个选项（使用 Amazon EC2），您还可以使用路由和安全组的组合，以便仅允许特定的本地部署 CIDR 范围。</p> <p>传输中加密</p> <p>我们建议将 Amazon VPC 中的数据源配置为使用加密来传输数据。区域服务已经在使用 TLS 加密。</p> <p>所有三个选项的管理访问权限都是加密的。</p> <p>身份验证</p>

标准	在 AWS 云中运行的 Power BI Desktop 的注意事项
	<p>AWS 建议您使用仅对所需数据集具有只读访问权限的身份对 AWS 数据源进行身份验证。</p> <p>所有这三个选项都要求使用身份验证机制，以便远程用户在访问托管 Microsoft Power BI Desktop 的环境之前证明其身份。此机制通常是用户名和密码，但也可以使用 MFA 选项。</p>
性能	在 AWS 云中运行时，Microsoft Power BI Desktop 的性能通常比在本地运行时要好。联网和计算实例都可能是现代化的，并且具有更高的规格，并且应用程序和数据源之间的延迟可能很低。
成本	<p>通过在 AWS 云中放置 Microsoft Power BI Desktop，可以显著降低或完全消除数据传输费用。跨可用区、在不同 VPC 或不同区域访问数据源时请务必小心，因为可能会收取数据传输费用。</p> <p>对于每种选项，都需要考虑额外的成本考虑因素。Amazon EC2、Amazon WorkSpaces 和 Amazon AppStream 2.0 是根据使用量定价的。有关更多信息，请参阅每项服务的定价页面。</p>

Microsoft Power BI Desktop 连接选项摘要

对于少数具有轻量数据集要求的用户，在本地部署中运行 Microsoft Power BI Desktop 并通过互联网安全连接或使用 AWS VPN，可能是一个合适的解决方案。确保在此模型中配置和维护了安全性。我们还建议测试此配置，以确定它是否符合用户的性能预期

随着用户数量的增加，我们建议您考虑通过 AWS Direct Connect 进行连接。加载较大的数据集时，Direct Connect 可提供更好的用户体验。确保用户了解传输大型数据集所带来的成本影响。

我们建议您评估在 AWS 云中运行 Microsoft Power BI Desktop 的情况。这样既可以为终端用户提供最佳的性能体验，也可能为云管理员提供最佳的管理体验。特别是 Amazon WorkSpaces 可以从少数用户扩展到数千个用户。这些服务还提供了显著的安全和管理优势。

将 Microsoft Power BI 服务连接到 AWS 数据源

Microsoft Power BI 服务 (SaaS) 可以直接连接到 Internet 可访问的数据源或 Amazon VPC 中的私有数据源。连接到私有数据源需要一个名为 Microsoft 本地部署数据网关的应用程序组件。微软本地数据网关将下载并安装在 VPC 中的 Amazon EC2 实例上，并使用 Microsoft Power BI 凭证进行配置。网关通过 Internet 建立与 Microsoft Azure 服务总线的出站连接，并在 Microsoft Power BI 中配置为连接到它可以访问的数据源。较大型部署可以使用多个本地部署数据网关来平衡负载或提高容错能力。

使用 Microsoft 本地部署数据网关可提供许多实质性的优势，AWS 客户已报告了这些优势：

1.改善了安全态势

Microsoft 本地部署数据网关不接受来自 Microsoft Azure 云的入站连接，只启动与 Azure 服务总线的出站连接。这种单向流量模型允许您保持数据源的私有性，而不会将其暴露在公共 Internet 上。

2.降低了数据传出

连接到数据源时，Microsoft 本地部署数据网关会检索整个结果集，并将其存储在本地称为假脱机的流程中。在将结果传输到 Power BI 服务之前，会对数据进行压缩。用户通常会报告 10:1 的压缩率，这样不仅可缩短通过 Internet 传输数据的时间，还可减少出站费用。

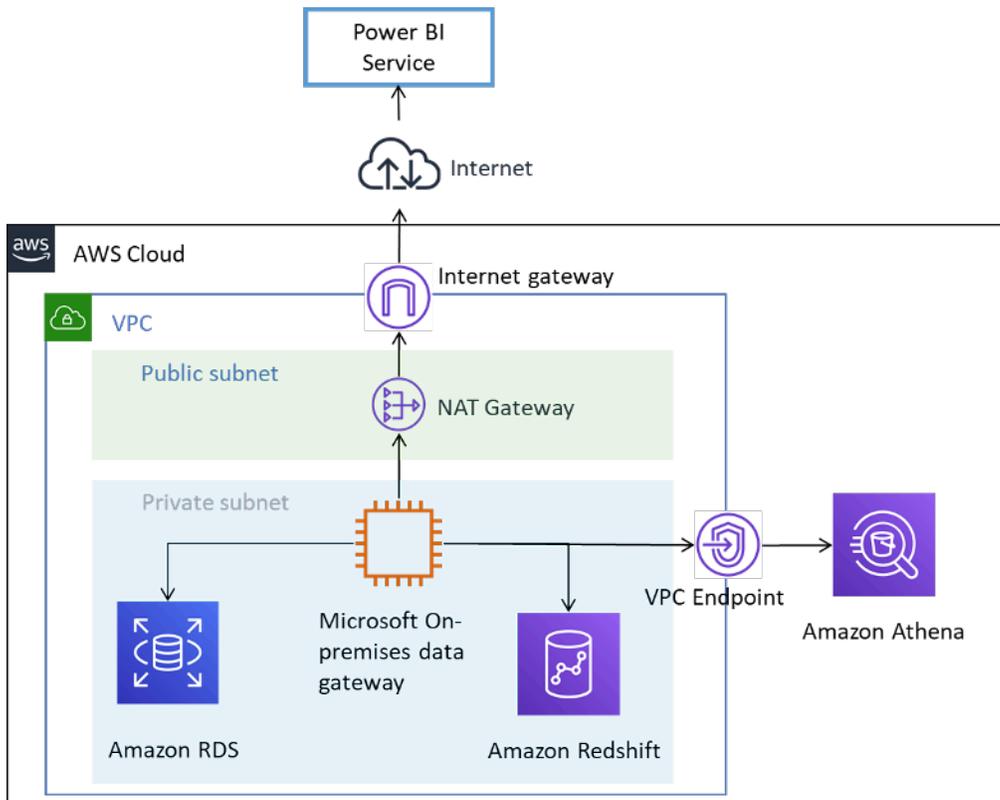
3.降低解决方案成本

使用 Microsoft 本地部署数据网关时，服务所需的某些数据处理将由网关执行。将 Amazon EC2 与成本削减计划（如节省计划或预留实例）结合使用，可能有助于降低 BI 解决方案的总体成本。

建议的配置

AWS 建议您在包含数据源的私有子网中的 Amazon EC2 实例上安装 Microsoft 本地部署数据网关。此子网配置为通过安装在公有子网中的 [Amazon VPC NAT 网关](#) 将请求路由到 Internet。您可以使用网络地址转换 (NAT) 网关允许私有子网中的实例连接到 Internet 或其他 AWS 服务，但阻止 Internet 连接到这些实例。如果您需要高度可用的数据网关实施，我们建议使用安装在跨越不同 AWS 可用区的多个 EC2 实例上的本地部署数据网关集群。有关信息，请参阅[添加其他网关以创建集群](#)。

本节中介绍的选项演示了 Amazon RDS、Amazon Redshift 和 Amazon Athena。有关所有 AWS 数据源的完整讨论，请参阅[附录：Microsoft Power BI 支持的 AWS 数据源](#)。



将 AWS 数据源连接到 Microsoft Power BI 服务

其他注意事项

表 5 — 在 AWS 云中使用数据源的 Microsoft Power BI 服务的注意事项

标准	在 AWS 云中使用数据源的 Microsoft Power BI 服务的注意事项
网络连接	Microsoft 本地部署数据网关与数据源的连接非常简单，因为数据使用者和数据源都驻留在 AWS 云中。可以直接访问位于 Amazon VPC 中的数据源，例如 Amazon RDS 和 Amazon Redshift。使用区域终端节点的数据源可通过 Amazon VPC Internet 网关或 Amazon VPC 终端节点进行访问。

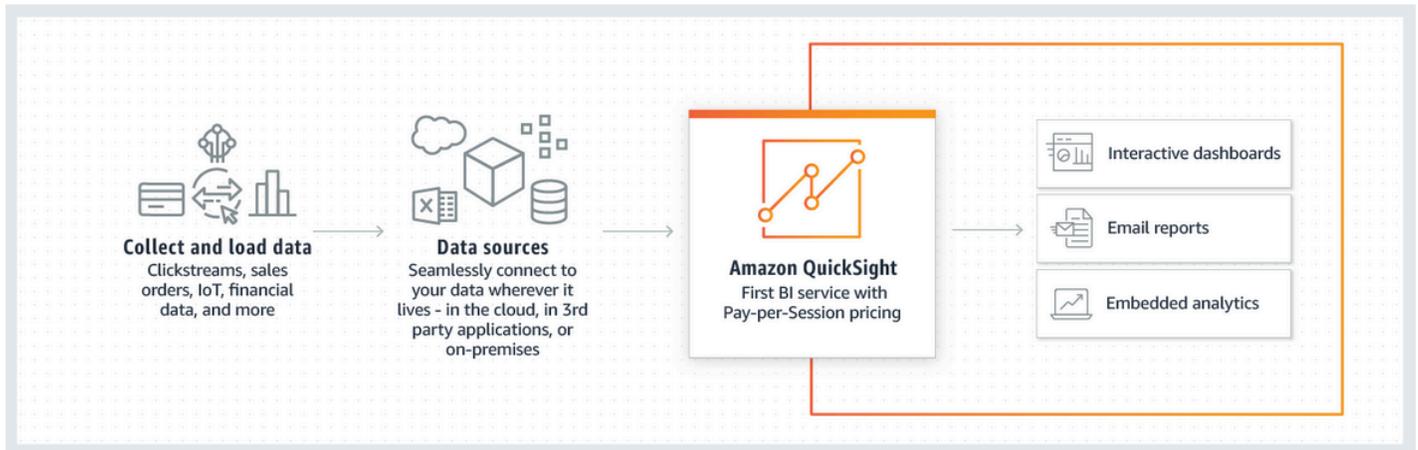
标准	在 AWS 云中使用数据源的 Microsoft Power BI 服务的注意事项
	Microsoft 本地部署数据网关与 Microsoft Power BI 服务的连接是通过 Internet 进行的，并且仅是出站连接。

标准	在 AWS 云中使用的 Microsoft Power BI 服务的注意事项
安全	<p>IP 访问控制</p> <p>您可以结合使用路由和安全组来控制对存储在 AWS 云中的数据源的访问。</p> <p>由于 Microsoft 本地部署数据网关安装在 Amazon EC2 实例上，因此它将具有可用于限制对操作系统入站访问的关联安全组。网关不接受入站请求。实例不需要公有 IP 地址，也不应配置公有 IP 地址。</p> <p>传输中加密</p> <p>我们建议将 Amazon VPC 中的数据源配置为使用加密来传输数据。区域服务已经在使用 TLS 加密。</p> <p>可以将 Microsoft 本地部署数据网关连接配置为使用 HTTPS (而不是 TCP) 连接到 Microsoft Azure 服务总线。我们建议使用 HTTPS 模式进行通信。自 2019 年 6 月网关软件版本发布以来，这也是新网关安装的默认设置。</p> <p>身份验证</p> <p>AWS 建议您使用仅对所需数据集具有只读访问权限的身份对 AWS 数据源进行身份验证。您为数据源输入的凭证经过加密并存储在网关云服务中。凭证在本地部署网关处解密。(您为数据源输入的凭证经过加密并存储在网关云服务中。凭证在本地部署网关处解密。)</p> <p>确保 Microsoft Power BI 凭据受到安全控制。对该服务的访问权限允许访问 AWS 数据源及其可能包含的潜在敏感信息。</p>

标准	在 AWS 云中使用数据源的 Microsoft Power BI 服务的注意事项
性能	由于能够调整和扩展 Amazon EC2 实例的大小，因此 AWS 云中的 Microsoft 本地部署数据网关通常性能良好。它还可以在区域联网和 Internet 连接中快速运行。
成本	<p>需要考虑三个因素：Amazon EC2 实例费用、数据传输费用和 Amazon NAT 网关费用。</p> <p>根据 Microsoft 的要求调整您的 Amazon EC2 实例的大小。要降低成本，您可以购买 Amazon EC2 预留实例或AWS 节省计划。</p> <p>从 Microsoft 本地部署数据网关传输到 Microsoft BI 服务的数据将产生 VPC 出口费用。客户使用数据网关报告压缩率为 10:1，这样将减少流量，但我们建议您限制查询并使用筛选条件来确保仅传输相关数据。</p> <p>如果 Microsoft 本地部署数据网关连接到不同可用区或不同 AWS 区域中的数据源，则还将收取数据传输费用。</p> <p>如果 Microsoft 本地部署数据网关位于私有子网中并使用 AWS NAT 网关，则将适用按小时收费和数据处理费用。有关更多信息，请参阅Amazon VPC 定价。</p>

使用 Amazon QuickSight

我们鼓励考虑将 Microsoft Power BI 套件与 AWS 结合使用的客户将 [Amazon QuickSight](#) 作为替代方案进行评估。这种完全托管式云服务原生连接到 AWS 中的数据源，与其他 BI 解决方案相比，降低了复杂性和成本。



Amazon QuickSight 的工作原理

与其他 BI 解决方案相比，Amazon QuickSight 具有以下优势：

- 使用 Amazon QuickSight，您无需下载和安装客户端应用程序。所有功能（包括编写和报告）都可以通过 Web 浏览器从任何平台（Windows、Mac、Linux 等）访问。
- Amazon QuickSight 以完全托管式云原生 SaaS 应用程序的形式提供，易于构建控制面板并将其部署到生产环境。该服务是无服务器的，这意味着您不需要计算需要多少个节点和服务器来支持您的用户。QuickSight 还充分利用 AWS 提供的高可用性功能来实现恢复能力。
- 无论是小型设置还是大型设置，您都能通过 QuickSight 中的点击式界面添加用户，可以轻松上手。无需外部管理员干预。
- Amazon QuickSight 由超快、并行、内存中计算引擎 (SPICE) 提供支持，可实现快速响应时间（以毫秒为单位）和交互式可视化。目前，数据集最多可扩展到 200 GB。
- Amazon QuickSight 定价简单、低廉，包含两个组成部分：报表作者和报表读者。创建和发布交互式控制面板的报表作者按用户定价。如果用户在给定月份内没有登录，则不会对这些用户收取任何费用。报表读者按每 30 分钟的会话收费，每位读者每月最高为 5.00 美元。免费试用版允许您免费评估 Amazon QuickSight。有关更多信息，请参阅 [Amazon QuickSight 定价](#)。

总结

如果您希望使用 Microsoft Power BI Desktop，我们通常会发现客户开始在本机试用该软件，并通过 Internet 连接到数据源。尽管存在用于使用 AWS VPN 和 Direct Connect 的私有连接选项，但许多客户得出的结论是，在 Amazon WorkSpaces 中运行 Microsoft Power BI Desktop 可以提供更好的性能体验。

如果您希望将 AWS 中的数据源连接到 Microsoft Power BI 服务，则您应该可以肯定这是一种既定架构模式。您可以在 Amazon VPC 中安装 Microsoft 本地部署数据网关，然后将 Amazon RDS、Amazon Redshift、Amazon Athena、Amazon OpenSearch 和 AWS Lake Formation 等数据源无缝连接到该服务。

如果您想要一款能够提供相同业务成果的解决方案，同时又不增加安装、配置、修补和扩展自行管理 BI 解决方案的复杂性，我们建议使用 Amazon QuickSight。这种完全托管式服务将简单 Web 浏览器体验中的所有必需功能与按用户付费的定价相结合。无需安装任何内容，也不需要其他组件。

希望这只是您使用 AWS 的商业智能之旅的开始。有关帮助您入门的其他资源，请参阅[附录：Microsoft Power BI 支持的 AWS 数据源](#)。

贡献者

本文档的贡献者包括：

- Ralph Holm, Sr.Amazon Web Services 高级解决方案架构师
- Barret Newman, Sr.Amazon Web Services 高级解决方案架构师
- Fabrizio Napolitano , Amazon Web Services 数据库和分析专业解决方案架构师

延伸阅读

- [将 Power BI 与 Amazon Redshift 集成以获得洞察和分析](#)
- [在 Power BI Desktop 中连接到 Amazon Redshift 数据库](#)

文档修订

要获得有关此白皮书的更新通知，请订阅 RSS 源。

更新-历史记录-更改	更新-历史记录-描述	更新-历史记录-日期
已更新	已添加数据源：Amazon OpenSearch 和 AWS Lake Formation。Athena、RDS 和 Redshift 数据源更新。有关 Microsoft 本地部署数据网关的其他指南。	2021 年 11 月 3 日
已审核	已审核技术准确性	2021 年 3 月 1 日
已更新	已更新报表服务器许可	2021 年 1 月 15 日
初次发布	白皮书首次发布	2020 年 11 月 1 日

附录：Microsoft Power BI 支持的 AWS 数据源

支持的数据源的完整列表由 Microsoft 提供（请参阅 [Power BI 数据源](#)）；但是，每个 AWS 数据源的以下部分提供了可能对某些读者有所帮助的用法和配置指南。

Amazon Redshift

Amazon Redshift 是 AWS 云中一项完全托管的 PB 级数据仓库服务。Amazon Redshift 数据仓库是一个由称作节点的各种计算资源构成的集合，这些节点已整理到名为集群的组中。每个集群都运行 Amazon Redshift 引擎，并包含一个或多个数据库。

在以下情况下，您应该考虑使用 Amazon Redshift：

- 您正在构建或迁移到原生云数据仓库。
- 您可能需要将容量从数 TB 扩展到数百 TB。
- 您希望允许 Power BI 用户透明地访问存储在 Amazon S3 中的数据湖中的数据，并将其与数据仓库中的表连接起来。
- 您的查询工作负载包括：
 - 在大型（数 GB 和数 TB）表上计算聚合的查询。
 - 具有多个联接和子查询的极其复杂的 SQL。
 - 控制面板中使用的复杂分析查询和简单、高度过滤的查询的组合。

将 Amazon Redshift 与 Microsoft Power BI 配合使用时，请记住以下几点：

- 在 Microsoft Power BI Desktop 和 Power BI 服务中，Amazon Redshift 作为 Power BI 数据源均受本机支持，并且每种服务都支持导入和直接查询模式。
- 虽然 Redshift 集群可以在公有子网中启动并配置为允许从 Internet 进行访问，但大多数客户更喜欢在私有子网中启动它以提高安全性。使用私有子网时，请使用本地部署数据网关从 Power BI 服务连接到 Amazon Redshift。
- Redshift 连接器支持 Power BI 桌面和服务中的 Azure AD 身份验证。
- 通过 Spectrum 访问的外部表与原生 Redshift 表的处理方式没有任何区别，Power BI 无法区分它们。访问外部表中的数据时，请确保：
 - 包含字符串的列在 AWS Glue 数据目录中被编录为“VARCHAR”，而不是编录为“STRING”，否则 Power BI 将引发以下错误：Exception: OLE DB or ODBC error:

```
[Expression.Error] We couldn't fold the expression to the data source.  
Please try a simpler expression..
```

- 不支持包含复杂数据类型 (如 ARRAY) 的列。当使用包含复杂数据类型的列时, Power BI 将引发以下错误: Exception: ODBC: ERROR [42703] [Microsoft]Amazon Redshift Error occurred while trying to execute a query

如果您需要在模型中包含它们, 则可以 (在 Amazon Redshift 中) 在用户级别启用 JSON 序列化功能, 或者将复杂数据类型存储在原生表的 SUPER 列中。

Amazon RDS

利用 Amazon RDS, 能够轻松在云中设置、操作和扩展关系数据库。Amazon RDS 在多种类型的数据库实例 (针对内存、性能或 I/O 进行了优化的实例) 上均可用, 并提供六种常用的数据库引擎供您选择, 包括 Amazon Aurora、PostgreSQL、MySQL、MariaDB、Oracle Database 和 SQL Server。

在以下情况下, 您应该考虑使用 RDS :

- 您正在构建运营数据存储。
- 您正在将 SQL Server 或 Oracle Database 数据仓库迁移到云中, 但对重构不感兴趣。
- 您的查询工作负载包括 :
 - 访问易于编制索引的表上经过高度筛选的数据的查询。
 - 对中小型表 (GB 级) 的分析查询。
 - 控制面板中使用的中等复杂度的分析查询和简单、高度筛选的查询的混合。

将 Amazon RDS 与 Microsoft Power BI 结合使用时, 请记住以下几点 :

- Amazon RDS 提供多个数据库引擎, 包括 SQL Server、MariaDB、MySQL、Oracle 数据库和 PostgreSQL。请注意, 数据库引擎列在 Power BI Desktop 和 Power BI 服务中, 而不是在 Amazon RDS 服务中。
- 对于 Amazon Aurora, 请使用 My SQL 或 PostgreSQL 连接类型, 具体取决于您选择的数据库引擎。
- 虽然 Amazon RDS 实例可以在公有子网中启动并配置为允许从 Internet 进行访问, 但大多数客户更喜欢在私有子网中启动它以提高安全性。使用私有子网时, 请使用本地部署数据网关从 Power BI 服务连接到 RDS。

- 利用 Amazon RDS，您可以部署多个版本的 SQL Server（2012、2014、2016、2017 和 2019），包括 Express、Web、Standard 和 Enterprise。

Amazon Athena

Amazon Athena 是一种交互式查询服务，可使用此服务通过标准 SQL 在 Amazon S3 中轻松分析数据。Athena 采用开箱即用设计，与 AWS Glue 数据目录相集成，从而允许您跨各种服务创建统一的元数据存储库，爬取数据源以发现架构，使用新的以及修改过的表和分区定义填充您的数据目录，并维护架构版本控制。

在以下情况下，您应该将 Athena 视为数据源：

- 您想直接查询数据湖。
- 您的查询工作负载包括：
 - 在大型（数 GB 和数 TB）表上计算聚合的查询
 - 交互式临时 SQL，用于探索目的。

将 Amazon Athena 与 Microsoft Power BI 结合使用时，请记住以下几点：

- 随着 2021 年 7 月 Microsoft Power BI 的发布，已经为 Amazon Athena 引入了经过 Microsoft 认证的连接器。您可以使用适用于 Amazon Athena 的 Microsoft Power BI 连接器在 Microsoft Power BI Desktop 中分析来自 Amazon Athena 的数据。将内容发布到 Power BI 服务后，您可以使用 Microsoft 本地部署数据网关通过按需刷新或计划刷新来使内容保持最新。
- 适用于 Amazon Athena 的 Microsoft Power BI 连接器同时支持导入和直接查询数据连接模式。在导入模式下，选定的表和列会导入到 Power BI Desktop 中进行查询。在直接查询模式下，不会将任何数据导入或复制到 Power BI Desktop 中，而是 Power BI Desktop 直接查询基础数据源。
- 有关适用于 Amazon Athena 的 Microsoft Power BI 连接器的更多信息，请参阅[使用 Amazon Athena Power BI 连接器](#)。
- 请注意，适用于 Amazon Athena 的 Microsoft Power BI 连接器需要在系统上使用 Amazon Athena ODBC 驱动程序和有效的 ODBC DSN 配置才能查询 Amazon Athena。要下载最新的 ODBC 驱动程序和配置信息，请参阅[使用 ODBC 连接到 Amazon Athena](#)。
- 有关使用适用于 Amazon Athena 的 Microsoft Power BI 连接器时的配置步骤和最佳实践的教程，请参阅[使用 Amazon Athena 在 Microsoft Power BI 上快速创建控制面板](#)。

Amazon OpenSearch Service (Amazon Elasticsearch Service 的后继者)

您可以使用 SQL 查询您的 Amazon OpenSearch Service，而不是使用基于 JSON 的搜索查询 DSL。如果您已经熟悉 SQL 语言或想要将您的域与使用该语言的应用程序（比如 Microsoft Power BI）集成，则使用 SQL 进行查询将非常有用。

在以下情况下，您应该将 Amazon OpenSearch Service 视为数据源：

- 您拥有半结构化数据（例如日志文件或 JSON 输出），需要快速搜索、分析或可视化这些信息。

将 Amazon OpenSearch Service 与 Microsoft Power BI 结合使用时，请记住以下几点：

- 连接到 Amazon OpenSearch Service 需要开放式数据库连接 (ODBC) 驱动程序，该驱动程序是适用于 Windows 和 macOS 的只读 ODBC 驱动程序，可让您将业务情报 (BI) 和数据可视化应用程序（如 [Tableau](#)、[Microsoft Excel](#) 和 [Power BI](#)）连接到集群上的 SQL 插件。该驱动程序可在 Open Distro for Elasticsearch [入门和下载](#) 网站上找到。有关配置说明，请参阅 [在 Windows 上配置 DSN](#)。
- 目前仅支持导入模式。
- Power BI Desktop 与 Amazon OpenSearch Service 的连接目前需要使用自定义连接器。要开始使用，请参阅 [将 Open Distro For Elasticsearch 连接到 Microsoft Power BI Desktop](#)
- 有关 Power BI 服务通过 Microsoft 本地部署数据网关连接到 Amazon OpenSearch Service 的信息，请参阅 [将 Open Distro For Elasticsearch 连接到 Microsoft Power BI Service](#) 以开始使用。

Amazon Lake Formation

Lake Formation 会帮助您从数据库和对象存储中收集并按目录分类数据，将数据移动到新的 [Amazon S3](#) 数据湖，使用机器学习算法清理和分类数据，并保护对敏感数据的访问权限。您的用户可以访问那些描述了可用数据集及其适当用法的集中[数据目录](#)。然后，用户可以通过所选的分析和机器学习服务利用这些数据集，例如 [Amazon Redshift](#)、[Amazon Athena](#) 和 [Amazon EMR for Apache Spark](#)（测试版）。Lake Formation 建立在 [AWS Glue](#) 中可用的功能之上。

如果您需要对数据湖进行细粒度（行和列）级别的访问，而不是基于 IAM 的传统控制，则应考虑使用 Lake Formation。

将 Lake Formation 与 Microsoft Power BI 结合使用时，请记住以下几点：

- 要使用 Power BI Desktop 或 Power BI 服务从 Lake Formation 数据目录中查询数据，请使用与在 Athena 中查询数据相同的流程和配置。如果您使用的是 Lake Formation 权限模型，请确保适用于 Amazon Athena 的 ODBC DSN 配置将“LakeformationEnabled”属性键设置为值“true”。这将指示 Amazon Athena ODBC 驱动程序使用 Lake Formation 服务进行授权，而不是直接使用 AWS Security Token Service。有关更多信息，请参阅[使用 ODBC 连接到 Amazon Athena](#) 文档。
- 为了与现有数据目录行为兼容而启用的“仅使用 IAM 访问控制”设置将提供完全兼容性。
- 将 AWS Glue 数据权限升级到 Lake Formation 模式可能会带来不兼容性，应在使用前进行测试。初步测试表明正在维持列级授予或拒绝，但作者尚未测试行和单元格级筛选，因为这仍为预览版且可能会发生变化。

声明

客户有责任对本文档中的信息，进行独立评估。本文档：(a) 仅供参考；(b) 代表当前提供的 AWS 产品和实践，如有更改，恕不另行通知；并且 (c) AWS 及其附属机构、供应商或许可方不做任何承诺或保证。AWS 产品或服务“按原样”提供，不提供任何形式的保证、陈述或条件，无论是明示还是暗示。AWS 对其客户的责任和义务由 AWS 协议决定，本文档与 AWS 和客户之间签订的任何协议无关，亦不影响任何此类协议。

© 2021 Amazon Web Services, Inc. 或其附属公司。保留所有权利。