



如何从 Amazon Lake House 智能湖仓中获取洞见

发布于 2021 年 10 月 12 日

Derive Insights from Amazon Lake House

阅读须知

客户有责任利用自身信息独立评估本文档中的内容。本文档：（a）仅供参考；（b）仅代表当前亚马逊云科技产品与实践，如有更改，恕不另行通知；（c）不构成亚马逊云科技及其附属公司、供应商或许可人的任何承诺或保证。文中对一切亚马逊云科技产品或服务皆依“原文”理解，不得以任何明示或暗示形式做出延伸性保证、陈述或解读。亚马逊云科技对客户承担的责任与义务遵循亚马逊云科技协议条款，本文档不属于亚马逊云科技与客户间的协议内容，亦不影响协议内的既有约定。

© 2021 亚马逊云科技有限公司或其附属公司版权所有。

目录

摘要.....	4
内容简介.....	4
智能湖仓是什么?.....	5
为什么使用 Amazon 进行智能湖仓分析?.....	6
亚马逊云科技专用分析服务.....	7
可扩展数据湖.....	7
性能与成本效益.....	8
无缝化数据移动.....	9
集中治理.....	9
亚马逊云科技上的智能湖仓架构.....	10
亚马逊云科技上的智能湖仓分析模式.....	10
获取洞见-由内而外的数据移动.....	11
获取洞见-由外而内的数据移动.....	14
获取洞见-数据环湖运动.....	16
主要优势.....	20
总结.....	20
贡献者.....	22
备注.....	22
文档修订.....	22

摘要

智能湖仓架构允许您跨越数据仓库、数据湖与运营数据库进行数据查询与分析，以前所未有的方式更快、更深入获取洞见。本份白皮书针对不同用户角色或工作职能提供多种云上智能湖仓架构最佳实践与架构设计指导，旨在帮助云架构师、数据科学家及开发人员如何玩转亚马逊云科技云端智能湖仓架构，并在搭建起这样一套架构之后，如何进行高效数据分析，获取洞见。

本份白皮书在结尾处展示了以智能湖仓为中心的分析场景，同时使用亚马逊云科技上智能湖仓架构所需要的其他各类资源与方法。

内容简介

随着整个人类社会持续迈向数字化新时代，我们创建及收集到的数据总量也在不断增长。企业需要收集、分析更多数据，以确保在变化发生时可以快速做出良好决策。与之对应，以往用于数据存储、数据管理与分析的传统本地解决方案已经无法跟上数据的快速发展步伐。低效高成本的协同解决方案导致我们无法对不同孤岛内的数据加以整合，更遑论全面执行有效的分析梳理。这一难题不仅限制了组织的敏捷性也降低了从数据中获取更多洞察及价值的的能力，同时也拖慢了根据需求变化而采用高级分析工具（比如 AI）的进度。亚马逊云科技也在观察中注意到，希望通过云迁移实现数据与分析基础设施现代化的客户数量一直在持续增长。

企业的各组织部门之间通常会以相互隔离的方式构建数据仓库与数据湖解决方案，每种解决方案采用自己独立的数据摄取、存储、管理与治理层。这种数据仓库与数据湖生态层面的严重脱节，最终会导致数据孤岛、数据集成复杂度过高、数据一致性冲突导致的数据移动与合并困难。这一切又可能延缓数据驱动型决策的速度、拉高决策成本，导致我们无法对相关数据进行全面分析并获取深入洞见。

本份白皮书介绍了[智能湖仓](#)中以角色为中心的使用模式，可帮助您更简单、集成度更高的方式收集、管理、处理并分析一切结构化与非结构化数据。智能湖

仓架构还允许您将全部数据引入各类用例，包括交互式 SQL、商务智能 (BI)、机器学习 (ML)、流与大数据分析等。

白皮书结构：

本份白皮书首先探讨了智能湖仓解决方案在概念层面与数据仓库及数据湖的联系和区别。之后，本文还展示了三种智能湖仓模式，可供您从智能湖仓中获取洞见结论。

智能湖仓是什么？

众多企业正在将自有数据从各种孤岛处转移到数据湖中，并在这里应用机器学习与分析处理技术，亚马逊云科技的数据湖方案的核心是 [Amazon Simple Storage Service \(Amazon S3\)](#) 存储，作为亚马逊云科技第一个云服务，Amazon S3 树立了对对象存储的标准，且绝大多数数据湖都建立在 Amazon S3 之上。此外，也有不少客户选择使用针对特定使用场景优化的专用分析存储，希望能够在集中式数据湖与周边专用分析服务之间自由的实现数据无缝、安全、合规移动，进而快速、敏捷地获取洞见结论。我们将这种现代分析方法，总结为智能湖仓架构。



亚马逊云科技智能湖仓架构

智能湖仓架构是由数据仓库与基于数据湖的解决方案所演变而来。下表从数据与性能特征的角度展示了这一演变趋势。

表 1：智能湖仓的数据与分析架构演变

	数据仓库	数据湖	智能湖仓
数据	来自事务系统、交易数据库、运营数据库以及业务应用行的关系数据	所有数据，涵盖结构化、半结构化与非结构化数据	智能湖仓是整个演变的下一阶段，能够跨越数据仓库、数据湖与数据库实现数据查询
性能	使用本地存储加快结果查询速度	使用低成本存储配合计算-存储解耦加快结果查询速度	无需移动数据，即可更快获取深层洞见

为什么使用 Amazon 进行智能湖仓分析？

客户构建的数据库、数据仓库与数据湖解决方案往往相互隔离，其各自拥有独立的数据摄取、存储、管理与治理层。这种数据仓库与数据湖生态层面的严重脱节，最终会导致数据与处理孤岛、过高的数据集成复杂度、严重的数据移动与数据一致性冲突等。这些问题导致客户难以获取深入洞见。为了克服难题、轻松实现数据移动，亚马逊云科技引入了[智能湖仓新方法](#)。“智能湖仓”架构不止打通湖与仓而是湖、仓、专用数据存储整合为一体。

亚马逊云科技提供广泛的托管服务平台，可帮助您使用智能湖仓方法快速构建、保护并无缝扩展端到端数据分析应用。无需采购硬件、无需维护或扩展基础设施——相反，您只需要指定用于数据收集、存储、处理及分析的必要设备。亚马逊云科技将提供足以容纳体量庞大且快速增长的大型数据集的分析解决方案，并借此对您的业务体系做出深度剖析。

亚马逊云科技专用分析服务

Amazon 为您提供广泛且深入的专用分析服务组合，包括 [Amazon Athena](#)、[Amazon EMR](#)、[Amazon OpenSearch Service](#)、[Amazon Kinesis](#) 以及 [Amazon Redshift](#) 等，全面支持您的独特分析用例。这些服务在设计方面充分吸纳最佳实践，努力帮助您在使用时避免做出性能、规模与成本方面的平衡。

例如，[Amazon Redshift 的性价比较其他云数据仓库高出 3 倍](#)；[EMR 上的 Apache Spark 运行速度相较于标准 Apache Spark 3.0 快 1.7 倍](#)，意味着您可以用不到一半的成本达成与传统 PB 级本地分析解决方案相同的结果。

可扩展数据湖

成千上万客户已经在亚马逊云科技上运行自己的数据湖。时至今日，数据湖的常见设置与管理方法仍然涉及大量只能手动操作、且极为耗时的任务。Amazon Lake Formation 能够自动执行这些任务，帮助您将数据湖的整个构建与保护周期从几个月缩短为几天。

在数据湖存储方面，Amazon S3 凭借着以下特性成为理想的存储方案：

- 无与伦比的 99.999999999%持久性与 99.99%可用性。
- 在对象层级的审计日志记录与访问控制方面提供卓越的安全性、合规性及审计功能保障。
- 五个存储层，灵活覆盖各类实际需求。
- 低使用成本，起步价不足\$1/TB/Month。

S3 还在数据访问、成本、复制及保护等管理环节中为您提供强大助力。



Amazon S3 – 构建数据湖的最佳选择

性能与成本效益

亚马逊云科技致力于在一切分析服务中以低成本提供最佳性能，同时通过持续不断的功能创新来提升服务性价比。除了行业领先的分析服务性价比之外，Amazon S3 智能分层的特性还能将数据存储成本节约达 70%。[Amazon EC2](#) 针对 200 多种实例类型提供行业领先的对应方案选项，提供高达每秒 1000 Gbps 的网络传输带宽，同时允许您在按需实例、预留实例与竞价实例之间灵活做出选择。

借助配备托管存储的 [Amazon Redshift RA3 实例](#)，您可以根据自身性能要求选择节点数量，且只需要为您实际使用的托管存储容量付费。[高级查询加速器](#)（Advanced Query Accelerator，简称 AQUA）则是 Amazon Redshift 的分析查询加速器，采用定制设计的硬件加快大型数据集的查询扫描速度。这种硬件加速型缓存机制使得 Amazon Redshift 的运行速度提升达 10 倍，得以横向扩展并跨多个节点并行处理数据。另外，各个节点还能对压缩、加密以及常见的多种数据处理任务（例如 Scan、聚合与过滤）进行加速。

无缝化数据移动

随着数据湖以及专用数据存储内数据总量的不断增长，您还需要轻松将一部分数据从一种数据存储移动至另一数据存储。Amazon 使您能够跨多种数据存储及数据湖，轻松实现数据的合并、移动与复制。

例如，[Amazon Glue](#) 提供全面的数据集成功能，可以轻松发现、准备并组合数据以供分析、机器学习与应用程序开发；而 Amazon Redshift 则可轻松查询 S3 数据湖内的数据。



Amazon Glue 是一套数据集成生态系统，用于加快智能湖仓的构建速度

Amazon Redshift 与 Amazon Athena 都支持联邦查询 ([Federated Query](#))，能够对存储在运营数据库、数据仓库以及数据湖内的数据执行查询，在无需移动数据也不必设置/维护复杂的摄取、转换与加载 (ETL) 管道的前提下，跨多个数据源获取洞见结论。

集中治理

现代分析架构中的一大重要组成部分，在于客户授权、管理及审计数据的能力。由于组织内整体数据存储的安全保障、访问控制与审计跟踪往往高度复杂、耗时且容易出错，因此相关管理工作往往困难重重。凭借集中访问控制与策略，外加列级数据过滤等功能，无需其他分析服务商的额外管理解决方案，

您就可以立足单一平台对跨数据湖及各类专用数据存储的所有数据进行访问管理。

借助集中式访问控制与策略，加之列级与行级过滤等功能，为您带来细粒度访问控制与治理能力，可通过单一控制点管理各类指向跨数据湖及专用数据存储的数据访问活动。

[Amazon Lake Formation 行级安全预览\(Row Level Access Control\)功能](#)，大大降低了在数据共享场景下对用户及应用程序访问活动的管理难度。行级(Row Level)安全机制，支持行级别的数据访问策略设置及过滤。

亚马逊云科技上的智能湖仓架构

随着数据湖、数据仓库以及专用存储中数据总量的不断增长，数据的移动难度也在持续提升。我们将这种情况称为数据引力。

为了快速敏捷地做出决策，我们需要使用集中数据湖并灵活应用以数据湖为中心的一系列专用数据服务。另外，您还需要以安全且符合监管要求的方式在各数据存储间轻松移动数据，借此深挖数据价值。这种现代分析方法也是智能湖仓架构的核心观点。关于更多详细信息，请参阅博文 [《如何在 Amazon 上构建智能湖仓》](#)。

亚马逊云科技上的智能湖仓分析模式

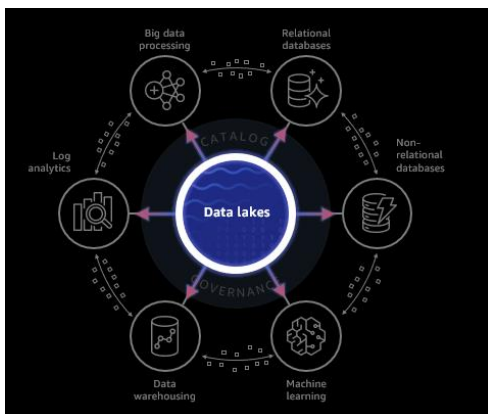
众多组织正在将全体数据由各个孤岛转移至统一位置，也就是数据湖，并在这里执行分析及机器学习。他们还会把数据存储于专门构建的数据存储之内，以便为特定用例提供更好的性能、规模与成本优势。此类数据存储包括但不限于数据仓库（用于快速对结构化数据执行复杂的查询）以及 OpenSearch（用于快速搜索并分析日志数据，进而监控生产系统的健康状况）等技术。在这种情况下，一刀切式的数据分析方法必然需要在某些环节上做出妥协，这样的方式早已不是最优解。

智能湖仓架构建立起一整套战略愿景，指明了如何将多种云上数据与分析服务组合起来，共同建立起统一的多用途数据处理与分析环境。通过使用亚马逊云科技智能湖仓解决方案，您可以通过以下三种分析模式获取洞见结论：

- [由内而外的数据移动](#)
- [由外而内的数据移动](#)
- [数据环湖移动](#)

获取洞见-由内而外的数据移动

为了充分利用您的数据湖与专用存储，大家必须能够轻松在不同系统之间移动数据。例如，您可以直接从数据湖中收集来自 Web 应用程序的点击流数据，并将部分数据移出数据仓库以生成日常报告。对于这类用例，我们将其称为由内而外移动数据。

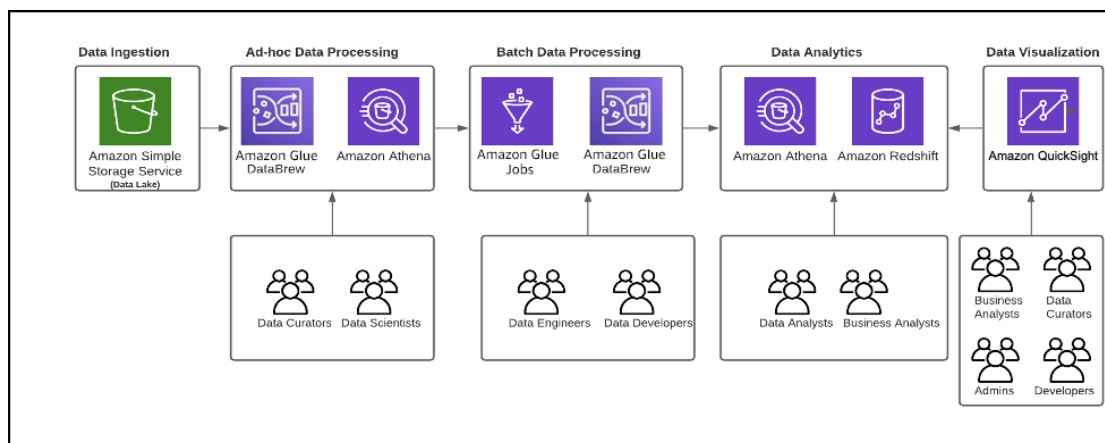


由内而外移动数据

使用 Amazon Redshift 与 Amazon QuickSight 从智能湖仓中获取基于事件的实时可视化洞见

客户往往希望在数据被摄取至数据湖之后，立即以可视化方式加以分析，以便快速敏捷地为下游业务价值提供决策依据。

下图所示，为智能湖仓如何使用 Amazon Redshift 与 [Amazon QuickSight](#) 的由内而外数据移动生成数据可视化洞察。



使用 Amazon Redshift 与 Amazon QuickSight
从智能湖仓中获取基于事件的实时可视化洞见

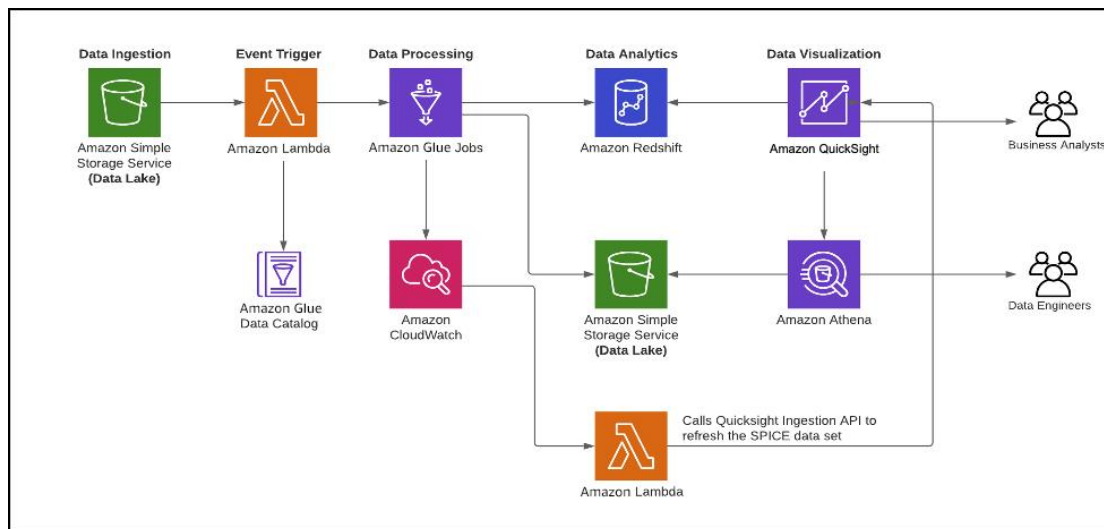
数据在智能湖仓架构当中将遵循以下操作步骤：

1. **数据摄取**——在 Amazon S3 中上传一个新数据文件，由 S3 事件触发 [Amazon Lambda](#) 函数。
2. **事件触发器**——Lambda 触发 Amazon Glue 工作流程以开始处理文件。Lambda 使用元数据变更对 [Glue 数据目录](#) 执行更新。
3. **数据处理**——将转换后的数据加载至 S3 与 Amazon Redshift 等目标数据存储当中。Amazon Glue 作业将日志和通知推送至 Amazon CloudWatch，CloudWatch 在 Amazon Glue 作业完成后触发 Lambda 函数。
4. **数据分析**——分析 Amazon Redshift 与数据湖（S3）中的数据。Lambda 调用 QuickSight 摄取 API 以刷新 [SPICE](#) 数据集。
5. **数据可视化**——新数据反映在 QuickSight 视觉当中。QuickSight 可以将 Amazon Redshift 与 Athena 中的数据组合起来以创建数据集，输出结果将存储在 SPICE 内以供快速分析。

使用 Amazon Glue DataBrew, Amazon Athena, Amazon Redshift 以及 Amazon QuickSight 从智能湖仓中获取以角色为中心的洞见

当您希望从呈指数增长的庞大数据当中获取洞见，帮助自身快速灵活地做出决策。组织需要使用集中数据湖配合一系列基于角色或工作职能的专用数据服务和数据仓库应对数据引力问题。

下图所示，为使用 Amazon [Glue DataBrew](#)、Amazon Athena、Amazon Redshift 以及 Amazon QuickSight 执行智能湖仓由内而外数据移动，借此实现以角色为中心的数据爆炸时代，万物快速演变

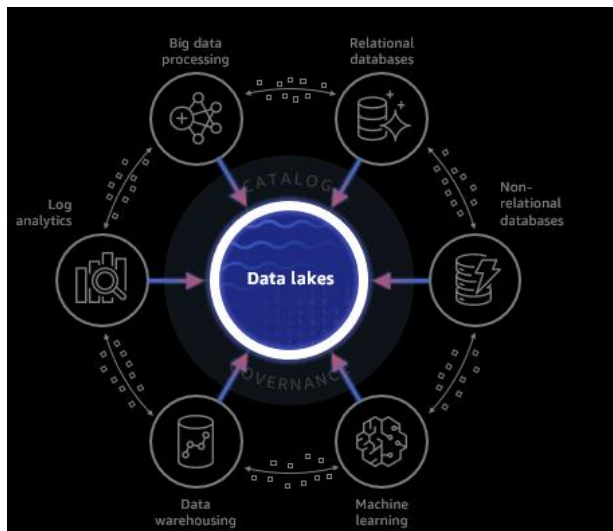


数据在架构中遵循的步骤如下：

1. **数据摄取**——从不同来源将数据摄取至 S3。
2. **即席数据处理**——数据管理员与数据科学家使用 Data Brew 验证、清理和丰富数据。Amazon Athena 还可运行即席查询以分析湖内数据。转换结果将共享给数据工程师以设置批处理作业。
3. **批量数据处理**——数据工程师或开发人员可以在 Amazon Glue 及 Amazon Glue DataBrew 当中设置批处理作业。这些作业可以由事件触发，也可以调度为定期运行。
4. **数据分析**——数据与业务分析现在可以对 Amazon Redshift 中、或者使用 Athena 对 S3 中的就绪数据集展开分析。
5. **数据可视化**——业务分析师可以在 QuickSight 中创建报表。数据管理员可以丰富来自多个来源的数据，并实施安全与数据治理方法。开发人员可以在应用程序中嵌入 QuickSight 仪表盘。

获取洞见-由外而内的数据移动

您也可以选择另外一种数据移动方向，即由外而内。例如，您可以将特定区域内产品销售的查询结果由数据仓库复制至数据湖内，再使用机器学习技术针对大规模数据集运行产品推荐算法。这种方式，即为由外而内的数据移动。

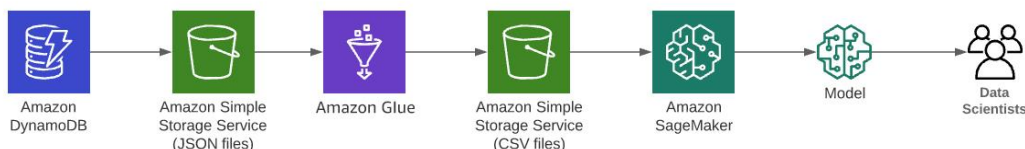


由外而内的数据移动

使用 Amazon SageMaker 对 Amazon DynamoDB 数据进行实时预测，借此获取洞见

[Amazon DynamoDB](#) 是一套快速 NoSQL 数据库，能够支持需要将延迟稳定控制在个位数毫秒的应用程序。客户希望将 DynamoDB 中的有价值数据移动至 S3 中以获取洞见。S3 中的这些数据可以成为理解客户过去行为、预测未来行为以及产生下游业务价值的主要来源。

下图所示，为使用 DynamoDB 数据通过由外而内的智能湖仓数据移动获取个性化推荐。



使用 Amazon SageMaker 对 Amazon DynamoDB 数据进行实时预测，借此获取洞见

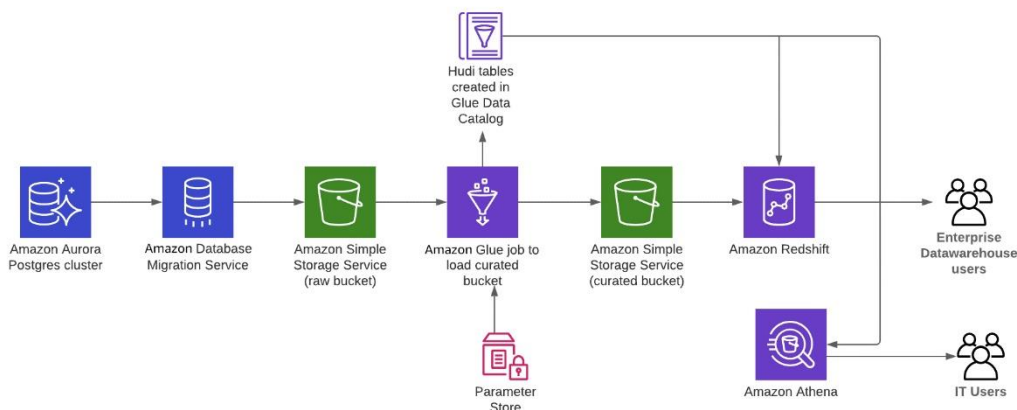
数据在架构中遵循的操作步骤如下：

1. [将 DynamoDB 各表以 JSON 的形式导出至 S3。](#)
2. 将导出的 JSON 文件转换为以逗号分隔的值 (.csv) 格式，进而作为 [Amazon SageMaker](#) 的数据源供 Amazon Glue 使用。
3. Amazon SageMaker 重用模型工件并更新端点。
4. 转换完成的.csv 文件即可供 Athena 用于即席查询。

使用 Apache Hudi, Amazon Glue, Amazon DMS 以及 Amazon Redshift 从 Amazon Aurora 数据中获取洞见

[Amazon 数据库迁移服务](#) (Amazon DMS) 能够将数据从源系统处复制至 S3。当数据处于 S3 中时，客户即根据自己的分析要求进行数据处理。一种典型的要求，是将 S3 中的数据与源系统数据进行更新同步。在关系数据库管理系统 (RDBMS) 上实现数据更新不难，但这样的 CDC 流程却很难在数据湖上实现。[Apache Hudi](#) 能够很好地解决这个问题。目前，您已经可以在 [Amazon EMR 上使用 Hudi](#) 创建 Hudi 表。

下图所示，为使用 Amazon Aurora Postgres 变更数据通过智能湖仓的由外而内数据移动以获取分析结果。



使用 Apache Hudi, Amazon Glue, Amazon DMS 以及 Amazon Redshift 从 Amazon Aurora 数据中获取洞见

数据在这套架构当中遵循以下操作步骤：

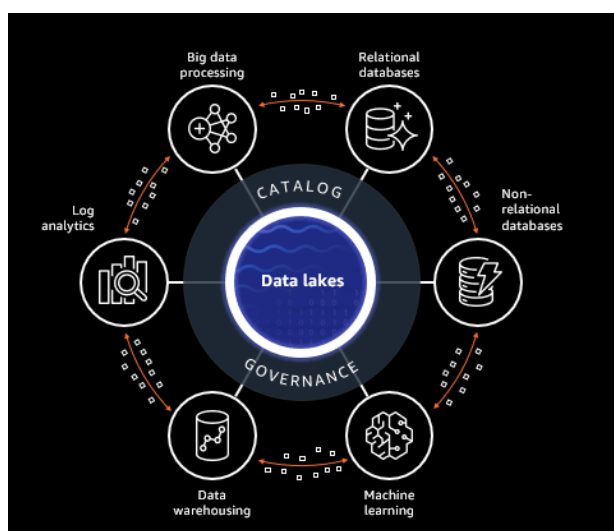
1. Amazon DMS 将 Aurora 集群中的数据复制至原始 S3 存储桶。

2. 使用 [Apache Hudi](#) 通过 [Amazon Glue](#) 作业在 Amazon Glue 数据目录中创建表。Amazon Glue 作业 (HudiJob) 被调度为定期运行，具体运行频率在 ScheduleToRunGlueJob 参数中设定。
3. 此作业从原始 S3 存储桶内读取数据，将结果写入至选定的 S3 存储桶，并在数据目录中创建 Hudi 表。
4. 此作业还在 [Amazon Redshift](#) 集群当中创建一个 Amazon Redshift 外部 schema。
5. 现在，您可以在 [Amazon Athena](#) 或者 [Amazon Redshift](#) 当中查询 Hudi 表了。

关于更多详细信息，请参阅博文 [《使用 Apache Hudi、Amazon Glue、Amazon DMS 以及 Amazon Redshift 为智能湖仓数据复制管道创建源》](#)。

获取洞见-数据环湖运动

在其他情况下，您可能希望将数据从专门构建的数据存储移动至另一数据存储，即数据的周边移动。例如，您可以将存储在数据库内的产品目录数据复制到搜索服务当中，以便更轻松地浏览产品目录并为数据库分担搜索查询。我们将这种概念，理解为数据环湖运动。



数据环湖运动

从您的数据湖、数据仓库及运营数据库中获取洞见

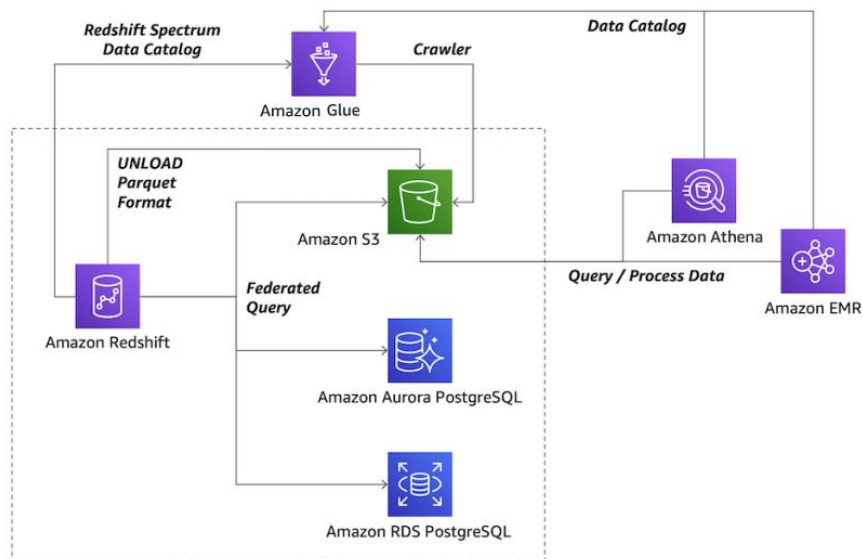
数据仓库是一种经过优化的数据库，有分布式架构也有集中式架构，用于分析来自事务系统和业务线应用程序的关系数据。[Amazon Redshift](#) 是一套快速、可扩展、完全托管的云上数据仓库，使用标准 SQL 配合现有商务智能 (BI) 工具以更简单、更加经济高效的方式执行数据分析。

要从不适合由数据仓库处理的非结构化数据中获取信息，构建[数据湖](#)是个理想的选择。数据湖是一套集中式存储库，允许您以任意规模存储一切结构化与非结构化数据。借助基于 Amazon S3 构建的数据湖，您可以轻松运行大数据分析，并使用机器学习从半结构化（例如 JSON、XML 格式）月光族非结构化数据集中获取洞见。

Amazon 还计划推出两项新功能，意在帮助您改进数据仓库与数据湖之间的集成管理方式：

- **数据湖导出 (Data Lake Export)**，用于将数据以 [Apache Parquet](#) 格式（一种针对分析经过优化的高效开放列式存储格式）从 Amazon Redshift 集群迁出至 S3。
- **联邦查询 (Federated Query)**，能够对 Amazon Redshift 集群中的数据执行以下查询：
 - 跨数据仓库集群存储的数据查询
 - 在您的 S3 数据湖内的数据查询
 - 在面向 PostgreSQL 及 Amazon Aurora PostgreSQL 数据库的一项或多项 Amazon Relational Database Service (Amazon RDS) 中的数据查询

下图所示，为使用 S3、Amazon Redshift、Amazon Aurora PostgreSQL 以及 Amazon EMR 通过智能湖仓的“数据环湖运动”方法获取分析结论。



从您的数据湖、数据仓库及运营数据库中获取洞见

数据在这套架构中遵循的操作步骤如下：

1. **使用 Redshift 数据湖导出**——您可以将 Redshift 查询的结果以 Apache Parquet 格式迁出至 S3 数据湖。与文本格式相比，Parquet 格式的选出速度可提高多达 2 倍，且将 S3 内的存储空间占用量缩减至六分之一。[Redshift Spectrum](#) 则允许您直接从 S3 中的文件查询数据，且无需任何数据移动。或者，您也可以使用 [Amazon Athena](#)、[Amazon EMR](#) 或者 [Amazon Sagemaker](#) 来分析数据。
2. **使用 Redshift 联邦查询**——您也可以直接通过 Amazon Redshift 数据仓库访问 Amazon RDS 与 Aurora PostgreSQL 存储中的数据。通过这种方式，您可以在数据可用时立即执行访问。在 Amazon Redshift 中执行联邦查询，您可以跨运营数据库、数据仓库及数据湖实现数据的查询和分析。

关于更多详细信息，请参阅博文 [《Amazon Redshift 新功能——数据湖导出与联邦查询》](#)。

使用 Glue Elastic Views 从您的数据湖、数据仓库和专门构建的分析存储中获取洞见

[Amazon Glue Elastic Views](#) 能够自动将数据由一个 Amazon 位置引流至另一位置，借此帮助数据工程师在无需编写复杂的摄取、转换与加载 (ETL) 或者摄取、

加载与转换（ELT）脚本的前提下，高效使用 Amazon 云端的数据资源。通过使用 CDC 技术，您可以保证随时从原始数据源处获取最新更改。

只需使用 SQL 创建一个视图并从数据库（例如 DynamoDB 或 Aurora）中提取数据，您就可以指定一个目标（例如 Amazon Redshift、S3 或者 ElasticSearch Service），之后所有变更都将自动传播。您还可以自动实现规模的横向和纵向伸缩。Amazon 会监控数据流是否发生了变更，因此大家不再需要为错误处理与监控工作而分神。这样一套体系，明显简化了跨服务间的数据移动方式。

Amazon Glue Elastic Views 建立在 Athena 的联邦查询功能之上，允许用户更轻松地了解最新数据，同时还让用户使用 SQL 在数据可能驻留的任意位置上查询数据。

Amazon Glue Elastic Views 预览版支持将 DynamoDB 与 Aurora 作为数据源，并支持将 Amazon Redshift 与 Elasticsearch 指定为目标。Amazon 的目标是随着时间推移而添加更多受支持的源与目标。此外，Amazon 还欢迎客户及合作伙伴使用 Elastic Views API 增强对其数据库和数据存储的支持能力。

下图所示，为使用 Amazon Glue Elastic Views 通过智能湖仓的“周边数据移动”方法获取洞见。



使用 Amazon Glue Elastic Views，
您可以从自己的数据湖、数据仓库以及专门构建的分析存储中获取洞见

主要优势

亚马逊云科技智能湖仓架构具有以下主要优势：

- 跨越数据库、数据仓库与数据湖实现统一分析
- 无需 ETL，使用 SQL 实现即可实现机器学习
- 全角色赋能——通过使用最合适的分析服务
- 跨数据湖实现全面的安全性、合规性与审计功能
- 经济高效且高持久性的数据，具备全局复制能力
- 提供一套全面的集成工具，为每位用户提供平等支持
- 以集中方式提供细粒度权限管理，为安全人员赋能
- 更精简的数据摄取与清洗流程，帮助数据工程师加快构建速度

总结

智能湖仓架构以一系列专门构建的服务为基础，可帮助您快速了解全体用户的所有数据。智能湖仓可帮助您面向未来实现构建，并随时将新兴分析方法与技术纳入架构当中。

本份白皮书介绍了多种专门构建的亚马逊云科技服务，您可以使用这些服务结合用户角色从智能湖仓中获取洞见。本文还引入多种选项，帮助您了解如何运用正确的亚马逊云科技服务为正确的工作提供灵活且丰富的功能。

即刻注册账号，开启您的智能湖仓之旅~

企业出海或个人体验

超过 200 项服务，
包含 80 余种产品免费试用



发展中国业务

近百项服务，
包含宁夏区域 20 余种产品免费试用



贡献者

本文档的贡献者包括：

- Raghavaram Sodabathina, 亚马逊云科技企业解决方案架构师
- Changbin Gong, 亚马逊云科技高级解决方案架构师

备注

关于构建智能湖仓架构中各层的具体架构模式、演练与示例代码，请参考以下资源：

- [使用 Amazon Analytics 释放数据的力量](#)
- 使用 Amazon Redshift 为智能湖仓架构构建 ETL 与 ELT 设计模式：[第一部分](#)与[第二部分](#)
- [使用 Apache Hudi、Amazon DMS 与 Amazon Redshift 为智能湖仓数据复制管道创建数据源](#)
- [使用 Amazon Redshift Concurrency Scaling 与 Spectrum 管理并控制成本](#)
- [使用 Apache Spark 与 Amazon Machine Learning 为 Amazon Redshift Analytics 提供支持](#)
- [使用 Amazon Redshift Data API 与 Amazon Redshift 集群进行交互](#)
- [使用 Amazon Redshift 物化视图加速 ELT 与商务智能查询](#)
- [使用 Redshift 联邦查询构建简化 ETL 与即度数据查询解决方案](#)