

# Guidance for Data Lakes with SAP and Non-SAP Data on AWS

This architecture diagram provides a high-level overview of an enterprise data lake. Your organization can improve decision-making and operational processes with a holistic view of transformed and catalogued SAP and non-SAP data.

- 1 Data Integration & Management for SAP solutions on AWS let you extract SAP data by using Odata Operational Data Provisioning (ODP) and using SAP Business Warehouse (BW) extractors, ABAP Core Data Services (CDS) views, and SAP Landscape Transformation Replication Server (SLT) table replication.
- 2 You can also use **AWS Glue** to extract data from non-SAP systems and combine it with SAP data.
- 3 The raw layer in **Amazon Simple Storage Service (Amazon S3)** holds the extracted SAP data. The enriched layer contains a true representation of the data available in the source system, using the Apache Iceberg open table format for advanced data functionality, including time travel and UPSERT operations. The curated layer holds data that is ready for consumption (including calculated fields), such as for creating data warehouses or data marts.
- 4 **AWS Glue** is used for data processing. Data is propagated through the layers and inserted, updated, and merged based on applied business logic. **AWS Glue** jobs identify new, changed, and deleted records based on the indicators provided through the ODP framework. Before and after images, as well as duplicated records generated by CDS views, are also handled within these jobs so that all data is reconciled and data processing is optimized for performance and cost.
- 5 Curated data can be consumed by a variety of data and analytics services. For example, **Amazon Athena**, **Amazon Redshift** for data warehousing, **Amazon QuickSight** for data visualization and reporting (using **Amazon Q**), **Amazon SageMaker** for artificial intelligence and machine learning (AI/ML), and **Amazon Bedrock** for generative AI.
- 6 Data is cataloged in the AWS Glue Data Catalog for technical usage and **Amazon DataZone** in a business catalog by domain. Optionally, you can apply fine-grained access control through **AWS Lake Formation**.
- 7 Data pipelines are centrally orchestrated and monitored with **AWS Step Functions**.

