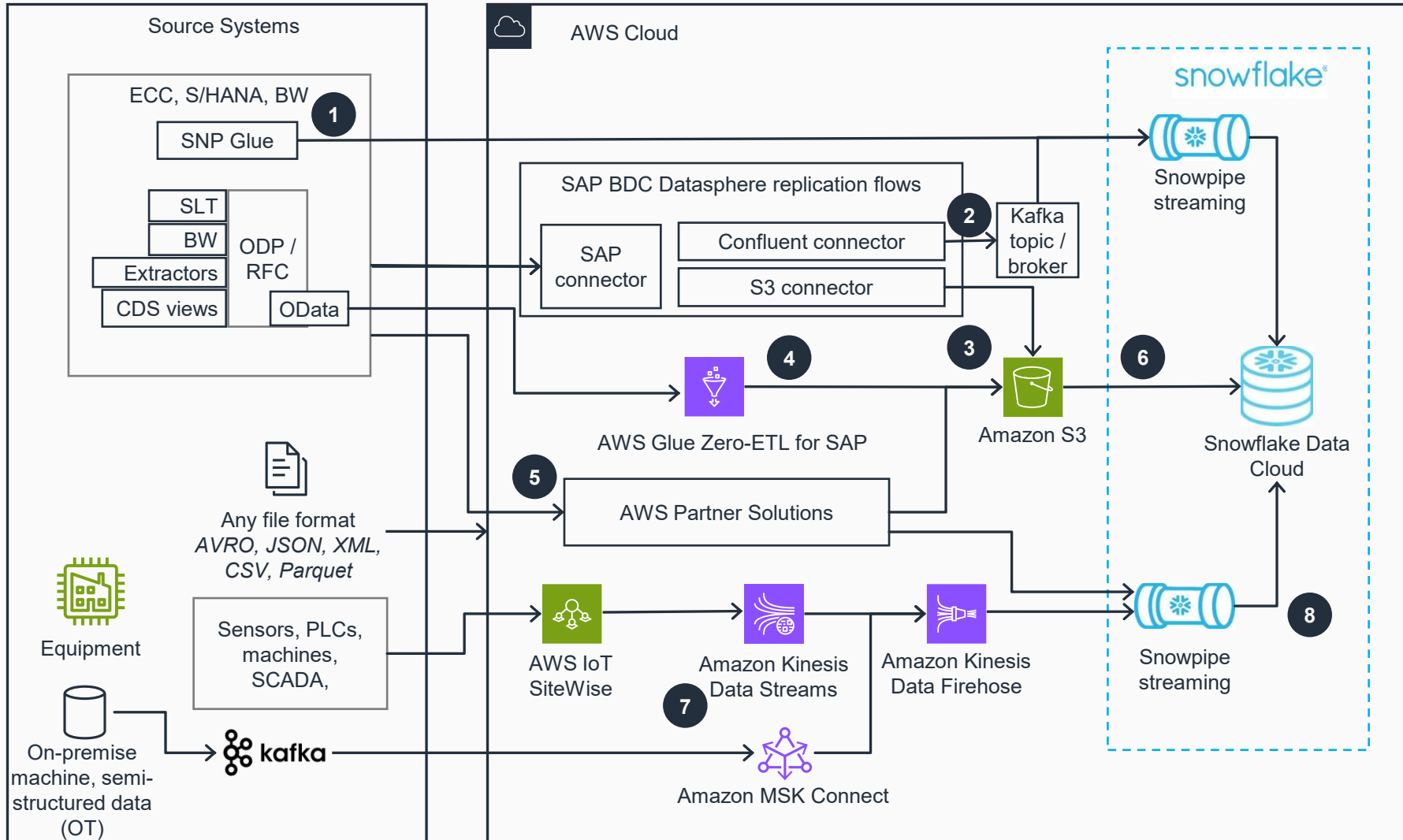


Guidance for Integrating SAP and Non-SAP Data using Snowflake on AWS

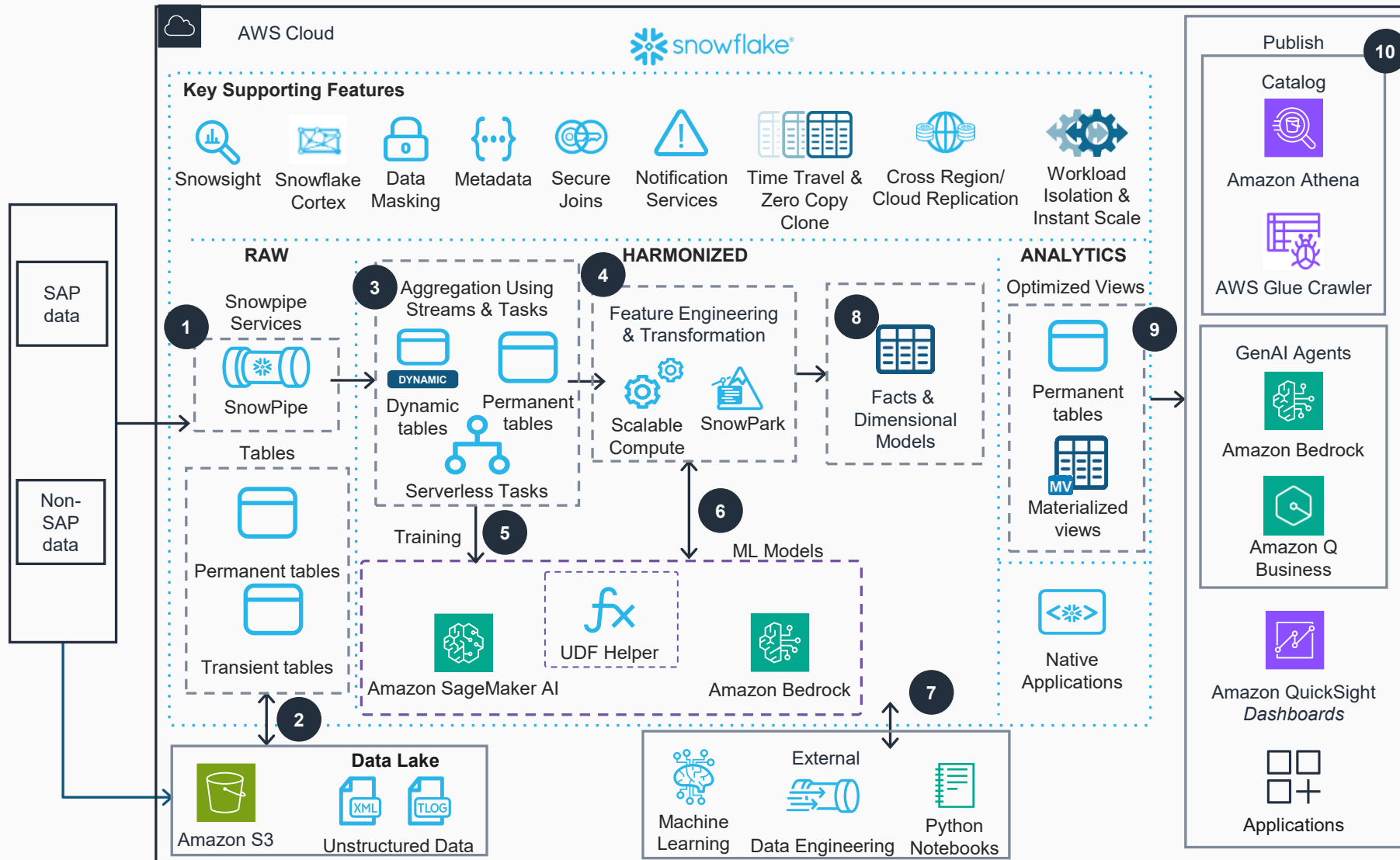
This architecture diagram illustrates how to effectively integrate SAP and non-SAP data using AWS, SAP, and Partner solutions with Snowflake.



- 1 Install and configure SNP Glue ABAP add-on on the SAP ABAP-based source system (such as S/4HANA, Enterprise Central Component (ECC), Customer Relationship Management (CRM), or Business Warehouse (BW) to stream real-time data to Snowflake snowpipes. This enables point-to-point replication without additional hardware/software.
- 2 If currently using SAP data analytics products and looking for optimized SAP integration, configure SAP Business Data Cloud Datasphere data replication flow from SAP objects like Core Data Services (CDS) views using Confluent premium outbound connector to Kafka broker or Snowflake snowpipe.
- 3 SAP Datasphere replication flows support **Amazon Simple Storage Service (Amazon S3)** premium outbound connector.
- 4 For AWS fully managed replication, use **AWS Glue** SAP OData connector and Zero-ETL to replicate SAP CDS views and BW extractors using managed incremental data transfer to Apache Iceberg tables in **Amazon S3**.
- 5 Various partner solutions such Theobald, Qlik, Boomi can be used to extract data from SAP to **Amazon S3** using RFC and OData Protocols.
- 6 Load data from **Amazon S3** to Snowflake data cloud using external stage or volume functionality of Snowflake.
- 7 **AWS IoT SiteWise** collects and processes IoT and OT data from industrial sources. **Amazon Kinesis Data Streams** processes this real-time data using Kafka and services for analytics.
- 8 **Amazon Managed Streaming for Apache Kafka (Amazon MSK)** Connect facilitates the integration between Apache Kafka and Snowflake, while **Amazon Kinesis Data Firehose** transfers Kafka-streamed data into Snowflake for analytics and storage.

Guidance for Integrating SAP and Non-SAP Data using Snowflake on AWS

This diagram illustrates how to model and consume SAP and non-SAP data using Snowflake-curated data using an ELT framework.



1 Snowpipe ingestion loads both SAP and non-SAP data into Snowflake using Snowpipe in real-time or batches. This process ensures timely availability of enterprise data for decision-making and analytics.

2 Data lake storage stores raw SAP and non-SAP data in **Amazon S3** for processing and analytics. It serves as a central repository for structured and unstructured business data, enabling scalable and cost-efficient operations.

3 Aggregation Using Streams and Tasks processes raw data using Snowflake Streams and Serverless Tasks for transformation. It automates data pipelines using dynamic tables.

4 Feature Engineering and Transformation enhances data using Snowpark, SQL, and transformation logic. It refines raw data into meaningful business metrics for AI-driven insights and reporting.

5 Model training with **Amazon SageMaker AI** trains machine learning models on Snowflake data, empowering businesses to build predictive models for customer insights, fraud detection, and demand forecasting.

6 Machine learning inference deploys ML models using Snowflake user-defined functions (UDF), **Amazon SageMaker AI**, and **Amazon Bedrock**, which enables AI-driven automation, personalization, and real-time decision-making for enterprises.

7 External data access connects Snowflake with external environments such as Jupyter notebooks and facilitates collaboration between data scientists and business analysts for advanced analytics.

8 Facts & Dimensional Models structures harmonized data into dimensional models for analytics and reporting. This optimizes data for business intelligence, enabling better trend analysis and strategic decision-making.



Guidance for Integrating SAP and Non-SAP Data using Snowflake on AWS

This diagram illustrates how to model and consume SAP and non-SAP data using Snowflake-curated data using an ELT framework.

- 9 Permanent tables and materialized views provide optimized views, data monetization, and native applications for insights, accelerating time-to-insight by enabling self-service analytics for business users.
- 10 Publishing and consumption enable data access through catalogs, GenAI agents, dashboards, and applications and democratizes data access across the organization, driving innovation and informed decision-making.

