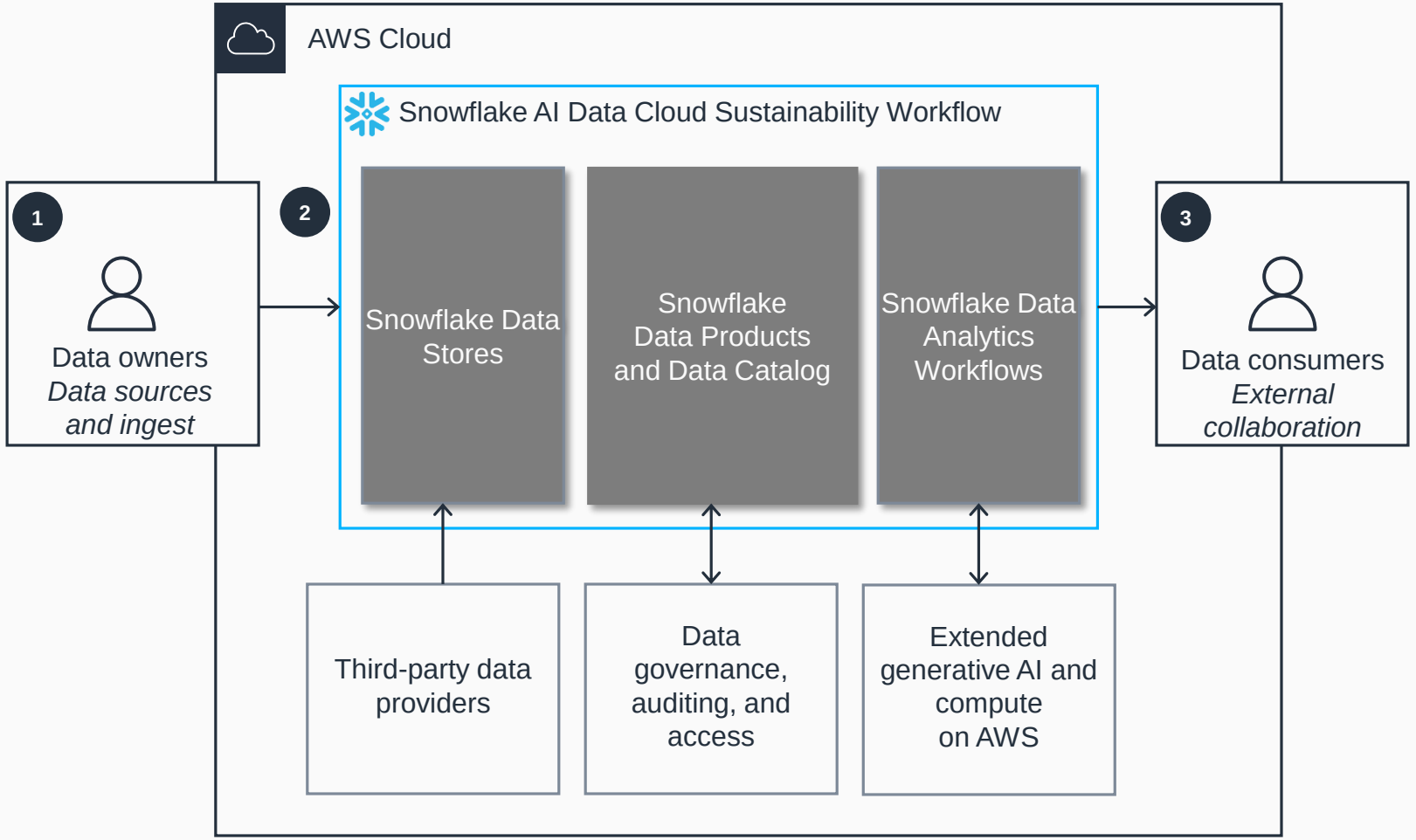
Guidance for Building a Sustainability Data Fabric with Snowflake on AWS

Overview

This architecture diagram illustrates how to use Snowflake for integrating, processing, and analyzing sustainability data, with the unified platform encompassing carbon accounting, energy optimization, and supply chain transparency. The subsequent slides depict data ingestion and data management processes.



1. Sustainability data can originate from sources external to Snowflake and AWS. This data can represent physical data objects, such as utility bills, or digital assets, such as emissions database tables. AWS services can be used to ingest both structured and unstructured data into AWS and Snowflake.

2. The Snowflake AI Data Cloud is used to create and run sustainability workloads to gain and share insights across your organization. Snowflake also provides connectors to adopt services outside of the Snowflake AI Data Cloud. Third-party resources, such as weather data and emission factors, can be supplemented through Snowflake Marketplace, **AWS Marketplace**, and **AWS Data Exchange**.

3. Data consumers can request access to data assets from the data owner. Data owners can grant the request and apply rules on how the data is used to facilitate external collaboration.

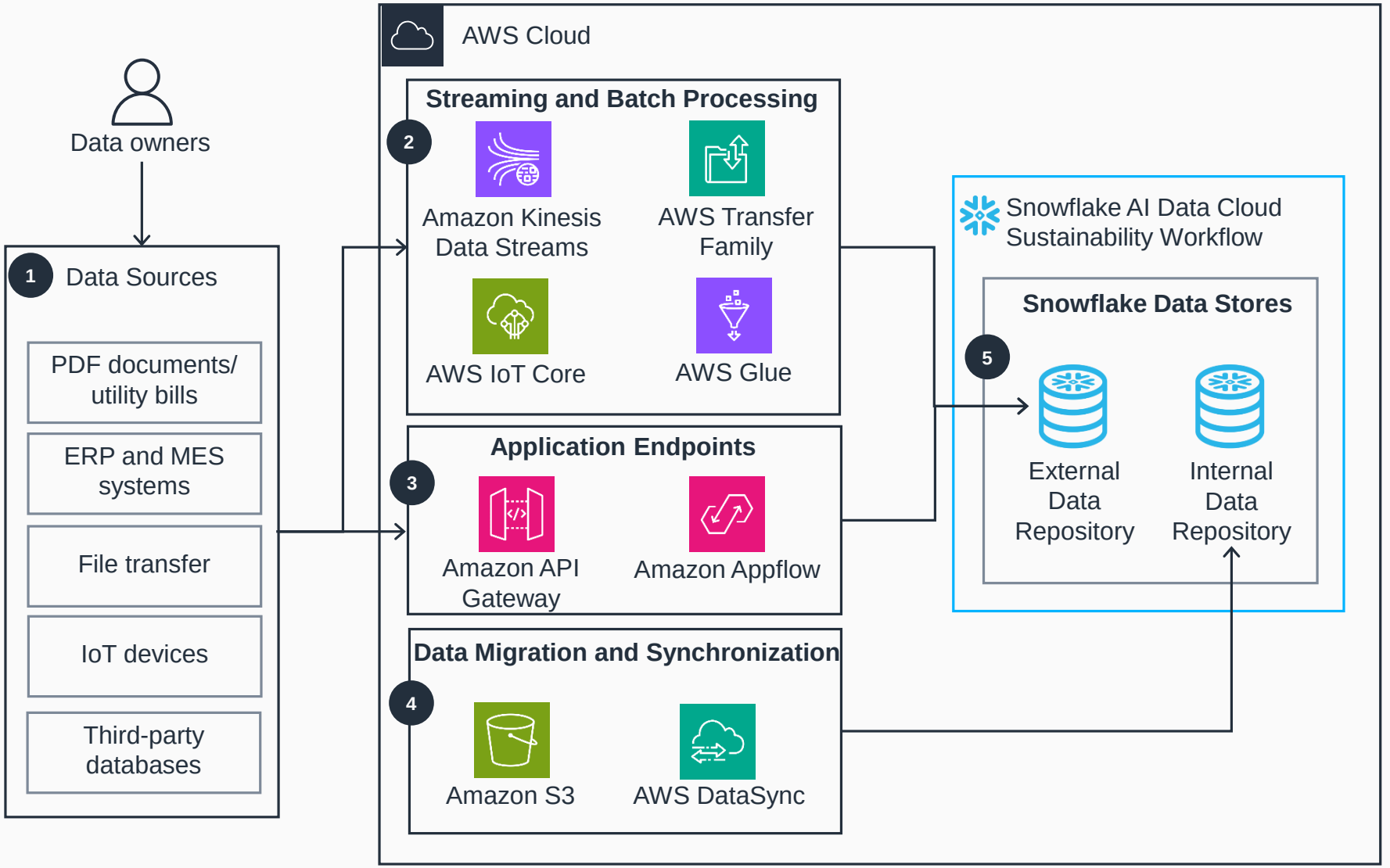
Reviewed for technical accuracy February 12, 2025

AWS Reference Architecture

Guidance for Building a Sustainability Data Fabric with Snowflake on AWS

Data sources and ingestion

This architecture diagram illustrates the ingest, preparation, and storage of unstructured sustainability data using AWS services.



1. Sustainability data can take many forms, such as near real-time streaming data from industrial processes or CSV files. This information is organized and collected from data providers.

2. Organizations ingesting data from streaming sources, such as sensors and industrial machinery, can use services like **Amazon Kinesis Data Streams** and **AWS IoT Core** to bring data into AWS. **AWS Transfer Family** helps automate data transfer processing steps, such as copying, tagging, scanning, filtering, compression, and encryption. **AWS Glue** is a managed service to perform extract, transform, and load (ETL) operations on the data.

3. **Amazon API Gateway** helps you build secure HTTP, REST, and WebSocket endpoints to process thousands of API calls. **Amazon AppFlow** is an application integration service to securely transfer data between software as a service (SaaS) applications and AWS services.

4. For data existing on AWS, such as **Amazon Simple Storage Service (Amazon S3)** objects, **AWS DataSync** can be used to synchronize data automatically with the Snowflake AI Data Cloud Sustainability Workflow.

5. Both internal and external data is collected in Snowflake data stores.

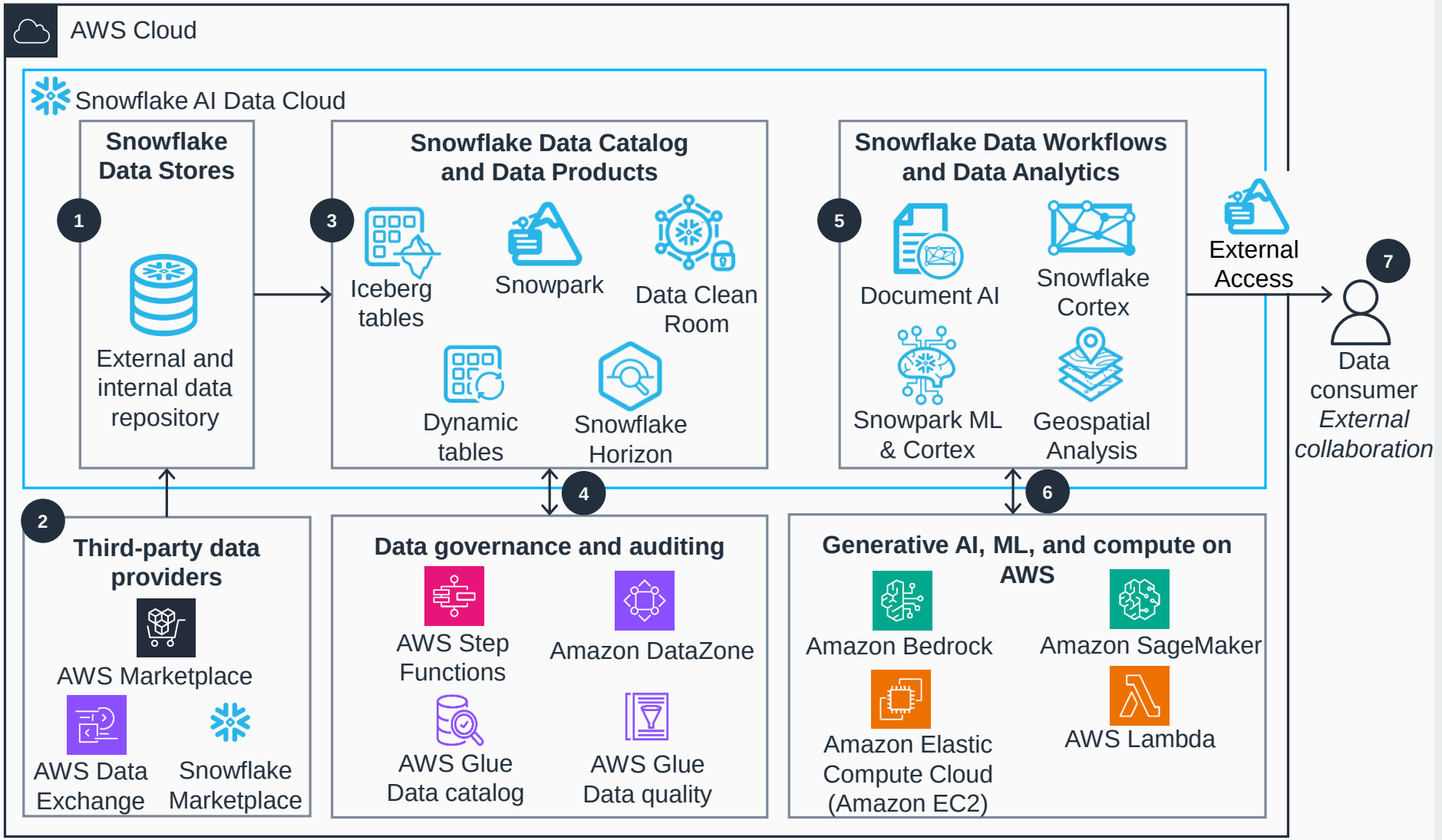
Reviewed for technical accuracy February 12, 2025

AWS Reference Architecture

Guidance for Building a Sustainability Data Fabric with Snowflake on AWS

Snowflake AI Data Cloud sustainability workflow

This architecture diagram shows how you can manage sustainability workloads across your organization by applying data governance best practices, integrating workflows with third-party data, and using compute instances to extract additional insights.



1. Snowflake offers data storage capabilities that allow for the collection of raw and unstructured data into a unified and scalable data management platform, providing a single access point to initiate a sustainability workflow.
2. Snowflake Marketplace, **AWS Marketplace** and **AWS Data Exchange** provide you with access to a wide range of sustainability-related datasets, such as emission factors and weather data.
3. Data products on Snowflake perform ETL operations with data quality checks.
4. Assets created by Snowflake data products are registered in the data catalog using Snowflake Horizons. You can orchestrate asset registration with **AWS Step Functions**.
5. With machine learning (ML) technology, data analytics workflow products on Snowflake help you automate sustainability workloads, such as carbon accounting and environmental, social, and governance (ESG) information.
6. You can use **Amazon Bedrock** to map environmental impact factors at scale, as outlined in Guidance for Environmental Impact Factor Mapping on AWS. Additionally, you can use **Amazon SageMaker AI** to optimize energy consumption as described in Guidance for Monitoring and Optimizing Energy on AWS.
7. The Snowflake External Access service provides secure access between the Snowflake AI Data Cloud and external data consumers.

Reviewed for technical accuracy February 12, 2025

AWS Reference Architecture