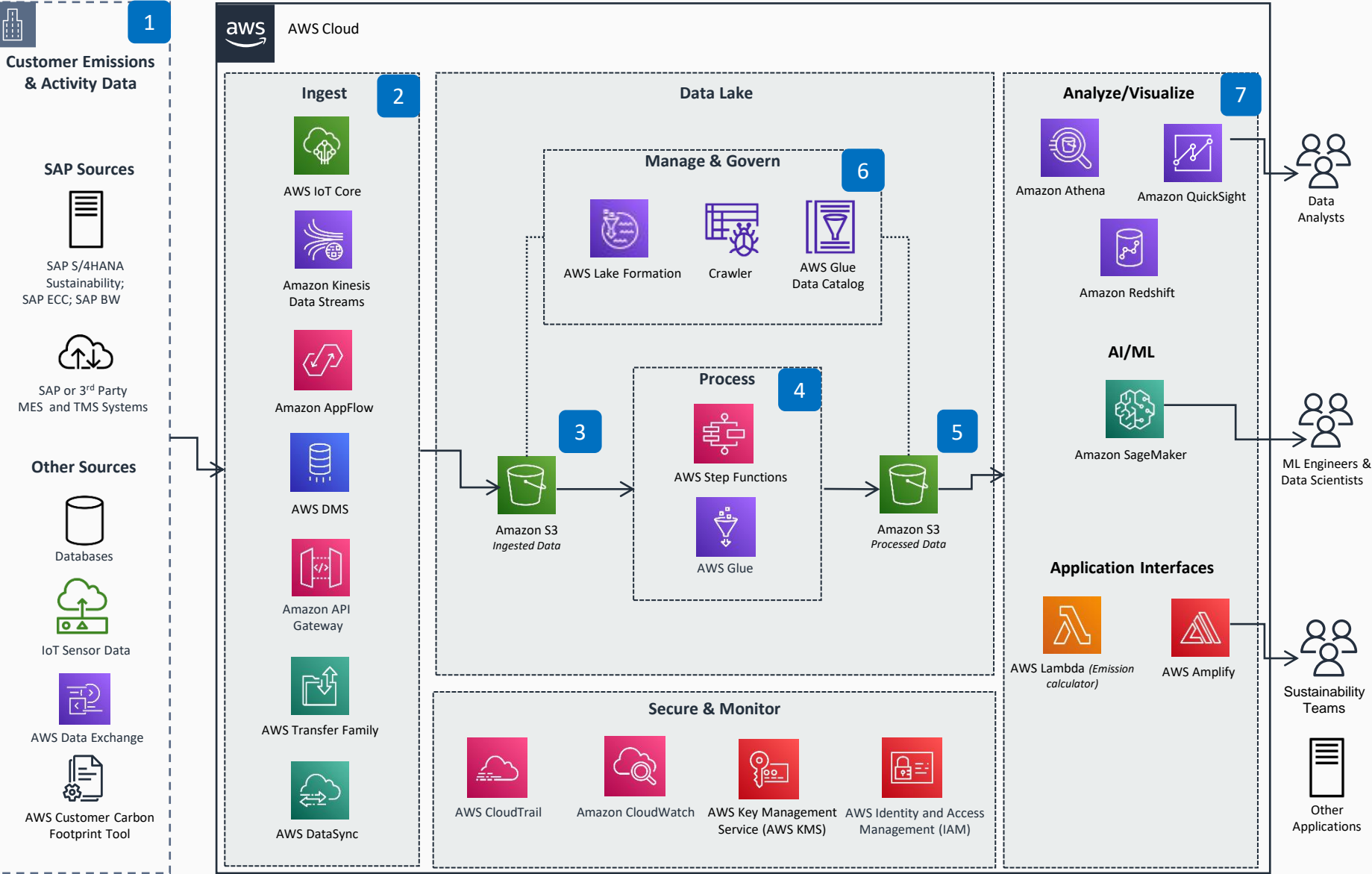# Guidance for SAP Sustainability Data Lake on AWS

This diagram displays various options to process emissions and business activity data from SAP using AWS services.

1 Customer Emissions & Activity Data can be sourced from various systems, including SAP S/4HANA Sustainability, SAP ERP Central Component (SAP ECC)/ SAP Business Warehouse (BW), SAP Manufacturing Execution System (MES), and SAP Transportation Management and Logistics System (TMS). Data can also be sourced from Software-as-a-Service (SaaS) apps, file shares, **AWS Data Exchange**, AWS Customer Carbon Footprint Tool, and Internet of Things (IoT) devices.

2 Data is ingested into the customer's account through various ingestion mechanisms, depending on the source. Data can be ingested using **AWS IoT Core**, **Amazon Kinesis Data Streams**, **Amazon AppFlow**, **AWS Database Migration Service** (AWS DMS), **Amazon API Gateway**, **AWS Transfer Family**, or **AWS DataSync**.

3 **Amazon Simple Storage Service** (Amazon S3) provides a single landing area for all ingested emissions and business activity data. Data ingress to the landing zone bucket triggers the data pipeline.

4 **AWS Step Functions** orchestrates the data pipeline that includes data quality checks, data compaction, transformation, standardization, and enrichment using **AWS Glue**.

5 The enriched emission data is then stored in **Amazon S3** in a format optimized for consumption and made available to various downstream consumers.

6 Manage the data lake using **AWS Glue** crawlers to infer new table schemas from objects in **Amazon S3** for storage in the AWS Glue Data Catalog. **AWS Lake Formation** enables permissions access controls to govern the access to the data lake objects.

7 Analyze and visualize your data using **Amazon Athena** and **Amazon QuickSight**, or load the data into **Amazon Redshift** for powerful data warehousing uses.

Empower your artificial intelligence and machine learning (AI/ML) workloads using **Amazon SageMaker.**

Application interface stacks can use **AWS Lambda** for calculating emissions and **AWS Amplify** for preconfigured web application management.

AWS Reference Architecture