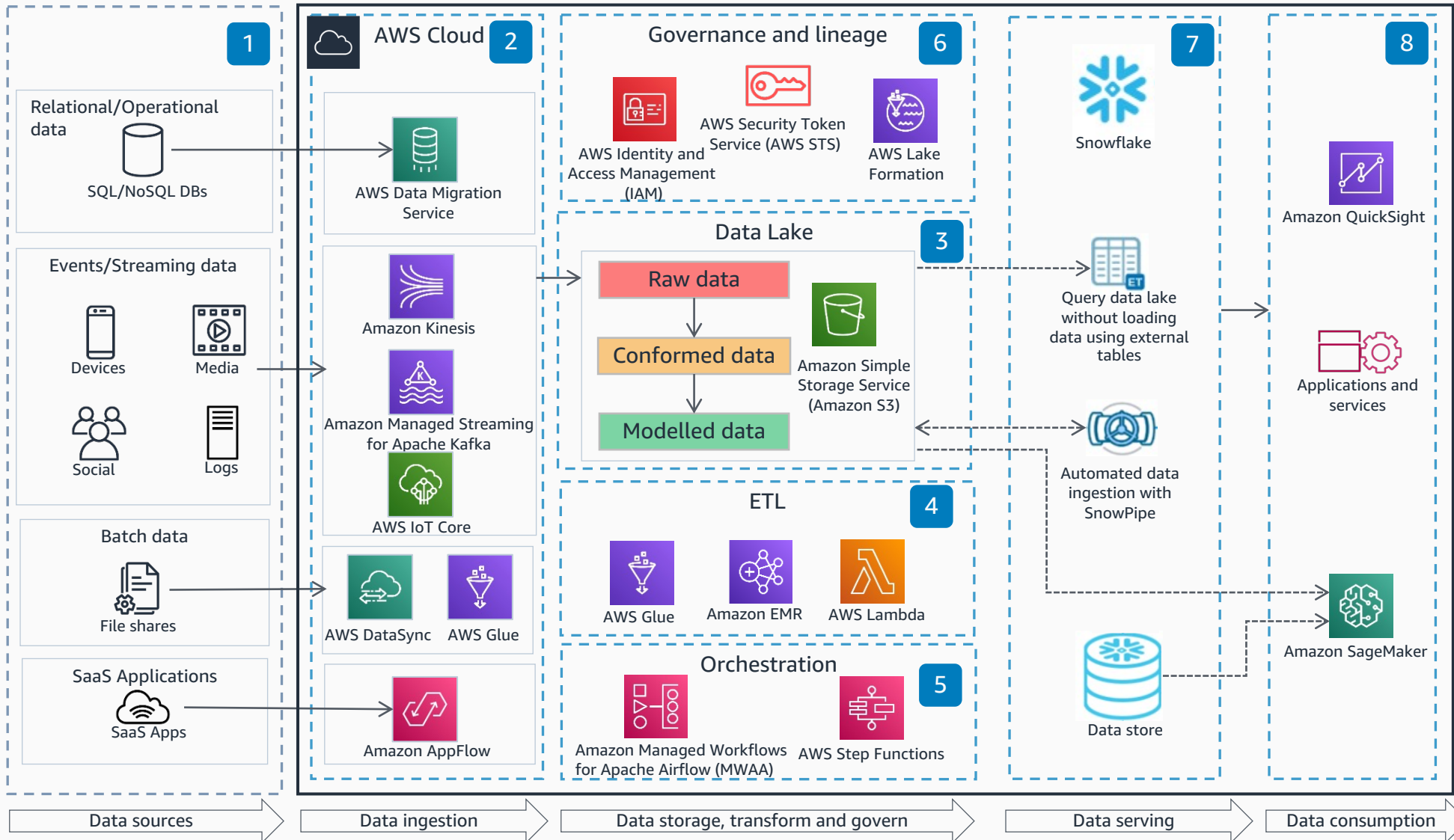# Modern Data Platform using AWS and Snowflake

This architecture enables customers to build end-to-end modern data analytics platforms using AWS and Snowflake.

1

Data sources

Relational/Operational data
- SQL/NoSQL DBs

Events/Streaming data
- Devices
- Media
- Social
- Logs

Batch data
- File shares

SaaS Applications
- SaaS Apps

2 AWS Cloud

Data ingestion

- AWS Data Migration Service
- Amazon Kinesis
- Amazon Managed Streaming for Apache Kafka
- AWS IoT Core
- AWS DataSync
- AWS Glue
- Amazon AppFlow

6 Governance and lineage

- AWS Identity and Access Management (IAM)
- AWS Security Token Service (AWS STS)
- AWS Lake Formation

3 Data Lake

- Raw data
- Conformed data
- Modelled data
- Amazon Simple Storage Service (Amazon S3)

4 ETL

- AWS Glue
- Amazon EMR
- AWS Lambda

5 Orchestration

- Amazon Managed Workflows for Apache Airflow (MWAA)
- AWS Step Functions

Data storage, transform and govern

7 Data serving

- Snowflake
- Query data lake without loading data using external tables
- Automated data ingestion with SnowPipe
- Data store

8 Data consumption

- Amazon QuickSight
- Applications and services
- Amazon SageMaker

---

1 Data is collected from multiple data sources across the enterprise, software as a service (SaaS) applications, edge devices, logs, streaming data, and social media networks.

2 Based on the type of data source, **AWS Database Migration Service**, **AWS DataSync**, **Amazon Kinesis**, **Amazon Managed Streaming for Apache Kafka**, **AWS IoT Core**, **AWS Glue** and **Amazon AppFlow** are used to ingest the data into the data lake in AWS.

3 **Amazon S3** is used for fully managed, highly available and scalable data lake storage.

4 **AWS Glue** is used to extract, transform and ingest data across multiple data stores. **Amazon EMR** provides the cloud big data platform for processing vast amounts of data using open source analytics framework. **AWS Lambda** and **Amazon EC2** provide compute capability for data enrichment needs.

5 Am**azon Managed Workflows for Apache Airflow (MWAA)** or **AWS Step Functions** is used for orchestrating end-to-end data pipelines.

6 **AWS Lake Formation** makes it easy to build, secure and manage your data lake, providing single place to enforce data classification and manage fine-grained access. **AWS IAM** and **AWS STS** provides ability to manage access permissions and temporary credentials.

7 Snowflake is used as virtual data warehouse with ability to query **Amazon S3** using external tables, and automated and continuous data ingestion using SnowPipe.

8 **Amazon SageMaker** can be used to build, train, and deploy machine learning (ML) models, and add intelligence to your Applications. **Amazon QuickSight** provides ML-powered business intelligence (BI).

AWS Reference Architecture