Architecture Diagrams

Modern Data Analytics Reference Architecture on AWS

Modern Data Analytics Reference Architecture on AWS: Architecture Diagrams

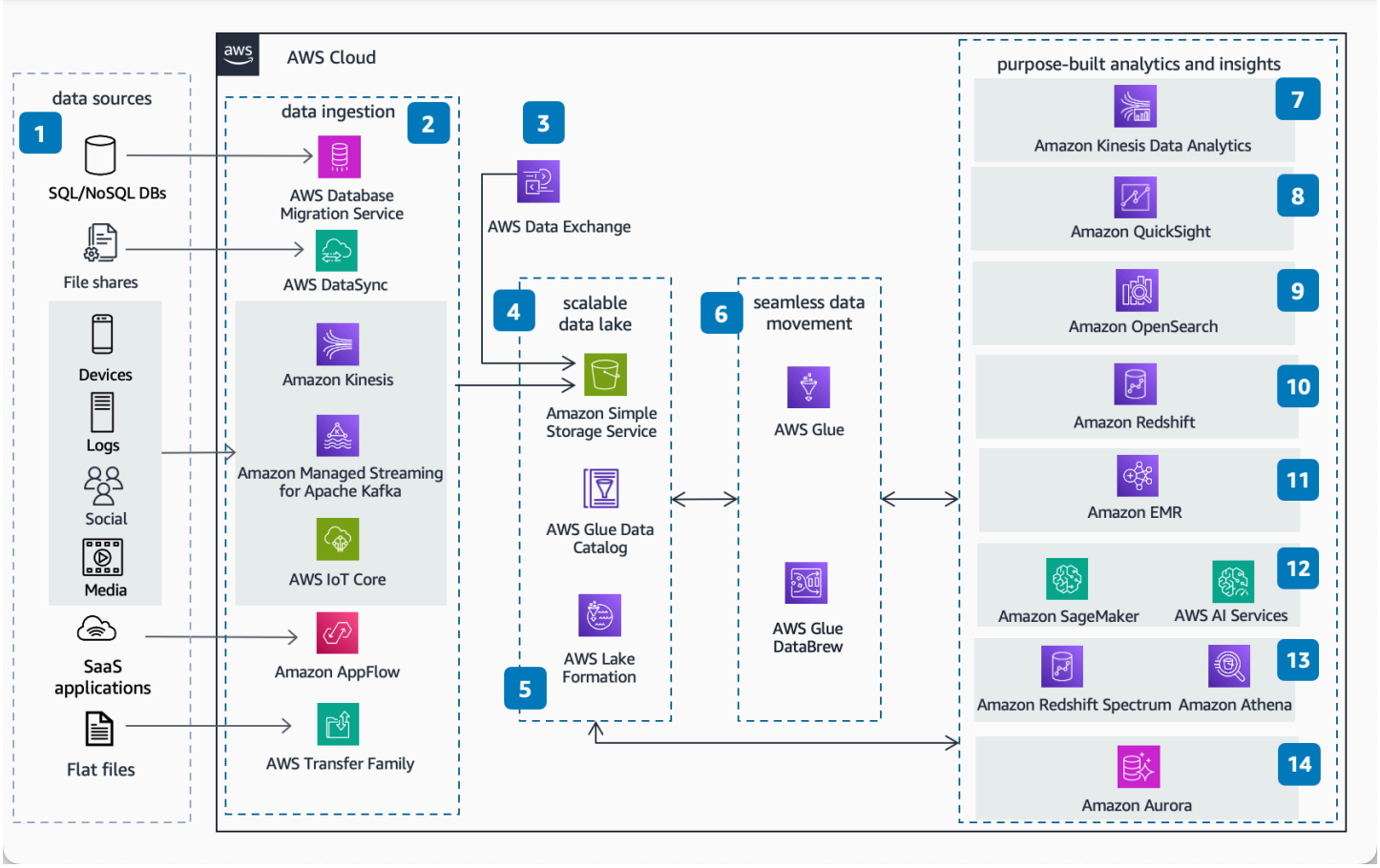
Table of Contents

Modern Data Analytics Reference Architecture on AWS Diagram

Publication date: **May 31, 2022 (Diagram history)**

This architecture enables customers to build data analytics pipelines using a Modern Data Analytics approach to derive insights from the data.

Modern Data Analytics Reference Architecture on AWS



1. Data is collected from multiple data sources across the enterprise, SaaS applications, edge devices, logs, streaming media, flat files, and social networks.
2. Based on the type of the data source, **AWS Database Migration Service** (AWS DMS), **AWS DataSync**, **Amazon Kinesis**, **Amazon Managed Streaming for Apache Kafka**, **AWS IoT Core**, **Amazon AppFlow**, and **AWS Transfer Family** ingest the data into a data lake in AWS.
3. **AWS Data Exchange** integrates third-party data into the data lake.

4. **AWS Lake Formation** builds the scalable data lake, and **Amazon S3** is used as the data lake storage. **AWS Glue Data Catalog** is a centralized metadata repository.
5. **AWS Lake Formation** also enables unified governance to centrally manage the security, access control, and audit trails.
6. **AWS Glue** and **AWS Glue DataBrew** catalog, transform, enrich, move, and replicate data across multiple data stores and the data lake.
7. **Amazon Managed Service for Apache Flink** is used to transform and analyze streaming data in real time.
8. **Amazon QuickSight** provides machine learning (ML)-powered business intelligence.
9. **Amazon OpenSearch Service** offers operational analytics.
10. **Amazon Redshift** is a cloud data warehouse. With federated queries, you can query and analyze data across operational databases, data warehouses, and data lakes.
11. **Amazon EMR** provides the cloud big data platform for processing vast amounts of data using open-source tools.
12. **Amazon SageMaker** and **AWS AI services** can build, train and deploy ML models and add intelligence to your applications.
13. **Amazon Redshift Spectrum** and **Amazon Athena** enable interactive querying, analyzing, and processing capabilities. **Athena** supports Apache Iceberg for data and **AWS Glue** data catalog.
14. **Amazon Aurora** offers high performance and availability at global scale. **Aurora** supports zero-ETL integration with **Amazon Redshift**.

Download editable diagram

To customize this reference architecture diagram based on your business needs, download the ZIP file which contains an editable PowerPoint.

Create a free AWS account

Further reading

For additional information, refer to

- AWS Architecture Icons
- AWS Architecture Center
- AWS Well-Architected

Diagram history

To be notified about updates to this reference architecture diagram, subscribe to the RSS feed.

Change	Description	Date
Initial publication	Reference architecture diagram first published.	May 31, 2022

Note

To subscribe to RSS updates, you must have an RSS plugin enabled for the browser you are using.