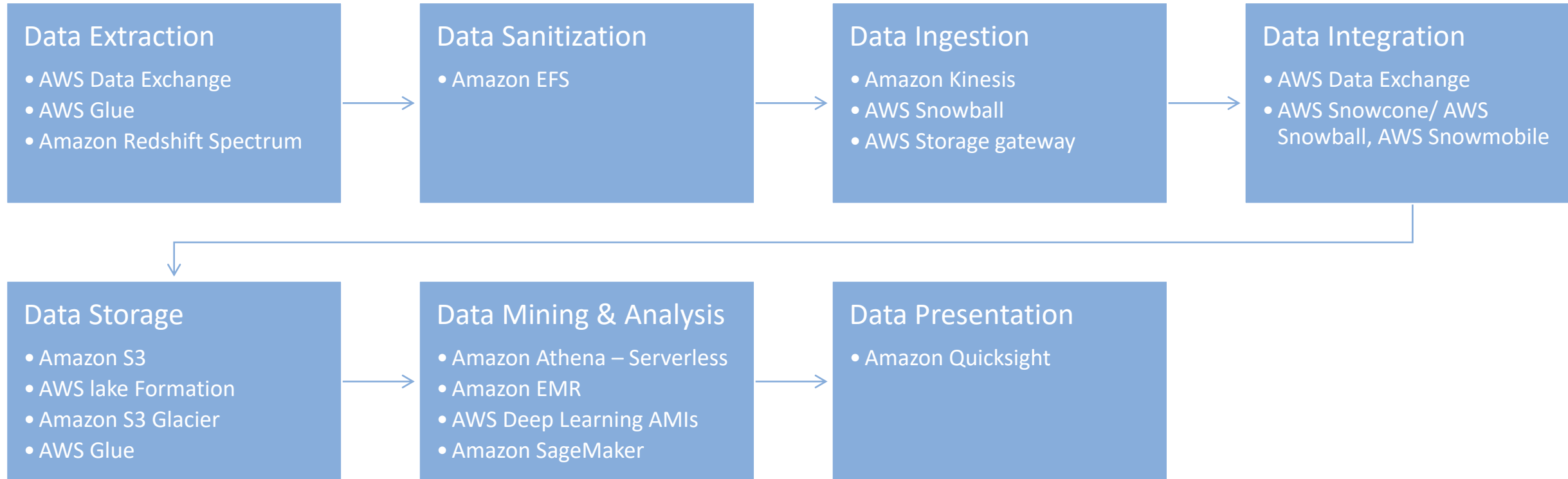


# AWS Ecosystem For Platform Development



Your Trusted Partner for Industry 4.0 Technologies

# Data Pipeline over AWS Ecosystem



**Data Security** – AWS role based IAM/ Security groups/ Configuration based security across all tools and services at each stage

**Data Governance** – Data Sanitization, Profiling, Auditing, Durability, Reliability across all tools and services at each stage

# Data Ingestion

---

## Amazon Kinesis Data Firehose

- Fully managed service for delivering real-time streaming data directly to Amazon S3.
- Automatically scales to match the volume and throughput of streaming data, and requires no ongoing administration

## AWS Snowball

- Securely and efficiently migrate bulk data from on-premises storage platforms and Hadoop clusters to S3 buckets

## AWS Storage Gateway

- Integrate legacy on-premises data processing platforms with an Amazon S3-based data lake.
- The File Gateway configuration of Storage Gateway offers on-premises devices and applications a network file share via an NFS connection

# Data Integration

---

## AWS Data Exchange

- Makes it easy to find, subscribe to, and use third-party data in the cloud

## AWS Snowcone/ AWS Snowball/ AWS Snowmobile

- Help physically transport up to exabytes of data into and out of AWS

## Talend

- Talend's end-to-end cloud data integration platform, Talend Cloud, offers over 70 connectors and components for AWS services, including self-service data preparation, data integration, data quality management, and native Apache Spark code generation capabilities for Amazon EMR.

# Data Storage

---

## Amazon S3

- Object storage built to store and retrieve any amount of data from anywhere – **Object storage**

## AWS lake Formation

- **Object Storage & Data Catalog**

## Amazon S3 Glacier & S3 Glacier Deep Archive

- Long-term, secure, durable Amazon S3 object storage classes for data archiving – **Backup and archive**

## AWS Backup

- Centrally manage and automate backups across AWS services – **Backup and archive**

## AWS Glue

- Fully managed extract, transform, and load (ETL) service – **Data Catalog**

# Data Mining & Analysis

---

## Amazon Athena – Serverless

- Interactive query service that makes it easy to analyze data in Amazon S3 using standard SQL – **Interactive Analysis**

## Amazon EMR

- Industry-leading cloud big data platform for processing vast amounts of data using open source tools such as Apache Spark, Apache Hive, Apache HBase, Apache Flink, Apache Hudi, and Presto. – **Big Data Processing**

## AWS Outposts

- Bring AWS services, infrastructure, and operating models to virtually any data center, co-location space, or on-premises facility. (Optional – depends on the architecture) – **Big Data Processing**

## Amazon Redshift

- Data Warehousing**

## Amazon Kinesis

- Makes it easy to collect, process, and analyze real-time, streaming data so you can get timely insights and react quickly to new information. - **Real-time analytics**

## Amazon Elasticsearch Service

- Operational Analytics**

## Amazon Kinesis Data Analytics

- Easiest way to analyze streaming data, gain actionable insights, and respond to your business and customer needs in real time – **Real Time Data movement and analytics**

## Amazon Kinesis Data Firehose

- Prepare and load real-time data streams into data stores and analytics tools - **Real Time Data movement and analytics**

## Amazon Kinesis Data Streams

- Collect streaming data, at scale, for real-time analysis - **Real Time Data movement and analytics**

## Amazon Kinesis Video Streams

- Capture, process, and store media streams for playback, analytics, and machine learning – **Real time Data movement and analytics**

## AWS Deep Learning AMIs

- Pre-configured environments to quickly build deep learning applications – **Predictive analytics and machine learning (Frameworks and interfaces)**

## Amazon SageMaker

- Fully managed service that provides every developer and data scientist with the ability to build, train, and deploy ML models quickly - **Predictive analytics and machine learning (Platform services)**

# Data Presentation/ Visualization

---

## Amazon Quicksight

- Fast, cloud-powered business intelligence service that makes it easy to deliver insights to everyone in your organization – **Dashboards & Visualization**

# Data Governance & Data Security

---

## Data Governance

- At each level, tools can be configured and managed for data sanitization, data profiling, data flow auditing etc. which ensured data governance

## Data Security

- AWS IAM role-based user authentication and authorization along with implementing security features at each level will secure the entire process



**Thank You**

