Different Types of Databases

- RDBMS - Relational Database Management System
- Data Lake
- Data Warehouse
- NoSQL
- Graph
- Search

	Use Cases	Technologies	Volumes
RDBMS	OMS, POS	Oracle, SQL Server (SQL), IBM DB2, Postgres, MySQL	Small to Medium
Data Lake	Data from all relevant sources	AWS s3, Azure Storage Accounts (ADLS), GCP GCS	High
Data Warehouse	Reports and Dashboards	Snowflake, Databricks, Teradata, Oracle Exadata, AWS Redshift, Azure Synapse, GCP BigQuery	Medium (typical) to High
NoSQL	Operational Stores, chats, endorsements, recommendations, product catalogs	Cassandra, MongoDB, Kudu, AWS DynamoDB, Azure (mulitple), GCP BigTable	Medium (typical) to High
Graph	Networks	Neo4j	Small to Medium
Search	Search	Lucene, Elastic Search, SolR	Medium

Big Data

- Volume
- Variety
- Velocity

- GFS (Google File System)
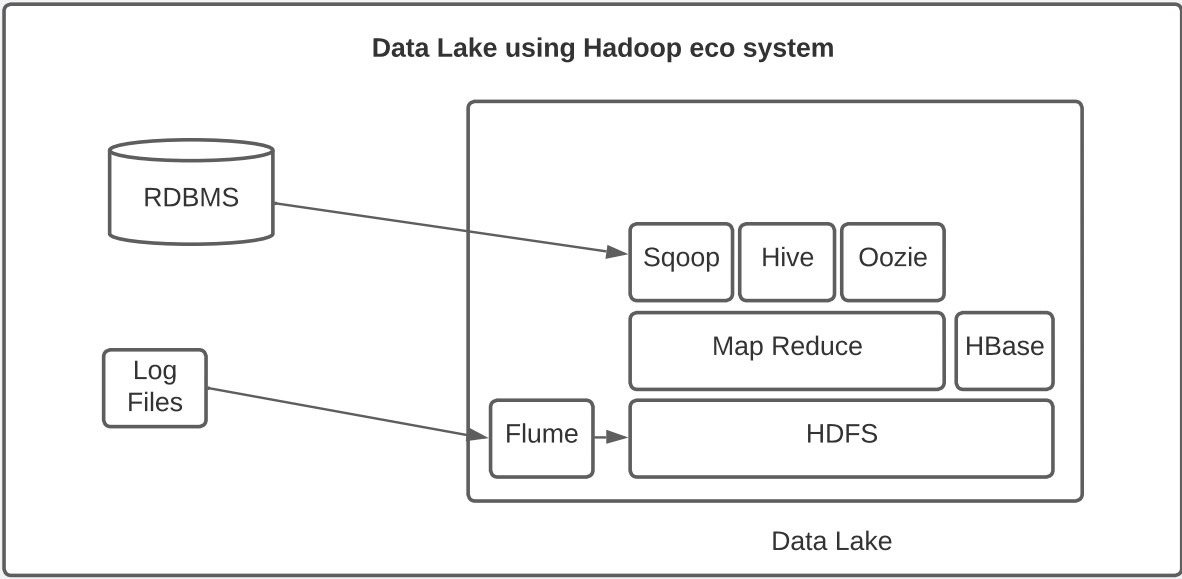- Google Map Reduce
- Google Big Table

Evolution of Big Data Technologies


- HDFS based on GFS
- Hadoop Map Reduce based on Google Map Reduce
- HBase based on Google Big Table
- Sqoop
- Hive
- Flume
- Oozie

Data Lake using Hadoop eco system

Data Lake

Challenges - Data Lake using Hadoop eco system

- Steep Learning Curve
- Setup Time and Costs
- Maintenance of the cluster
- Application Development Life Cycle
- Inefficient Usage of the cluster
- Slow Performance

Modern Data Lakes on Cloud
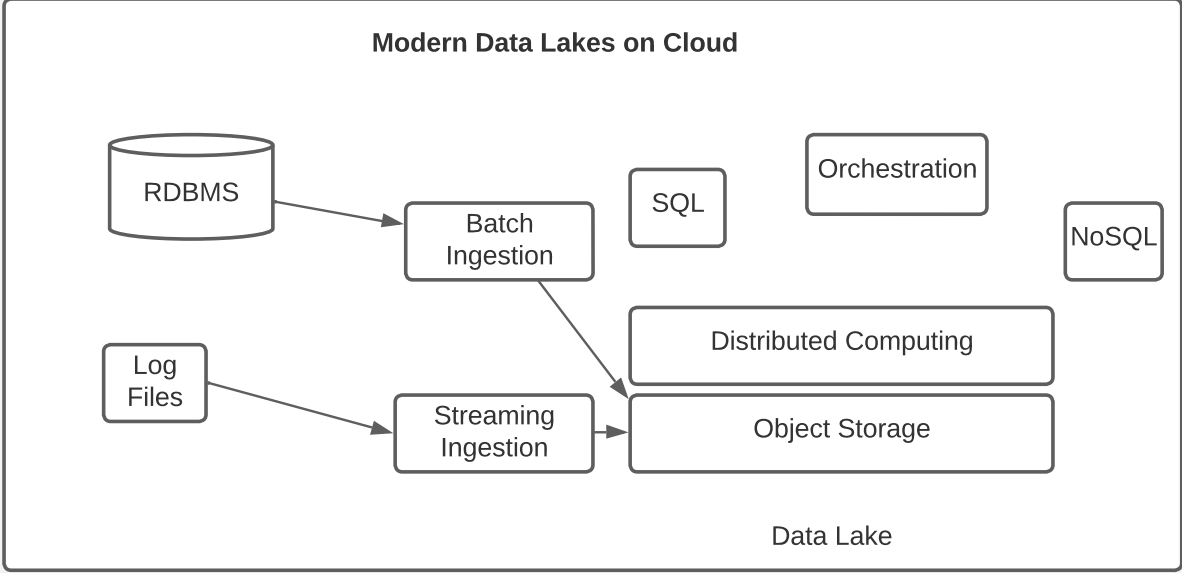
Data Lake

Implementation - Modern Data Lakes on Cloud

- **Cloud Native**
- Cloud Agnostic
- **Hybrid**

Major Cloud Platforms
- AWS
- Azure
- GCP

Advantages of Modern Data Lakes on Cloud

- Logical Separation of all the key components
- Low Cost and Decoupled Object Storage
- Serverless and Stateless Distributed Compute
- Low Setup Costs
- Low Maintenance Costs
- Faster Application Development Life Cycle
- Optimized Performance