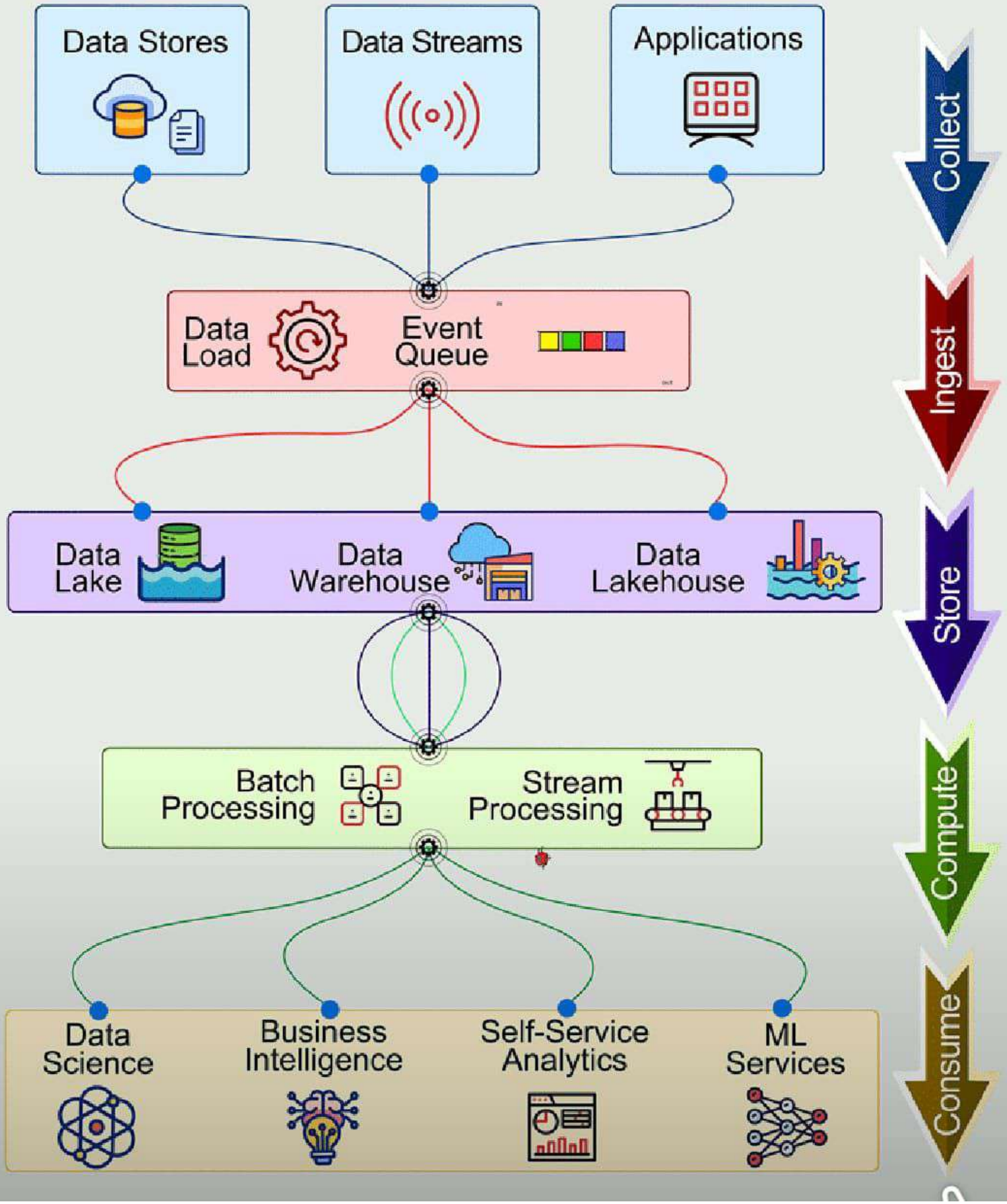
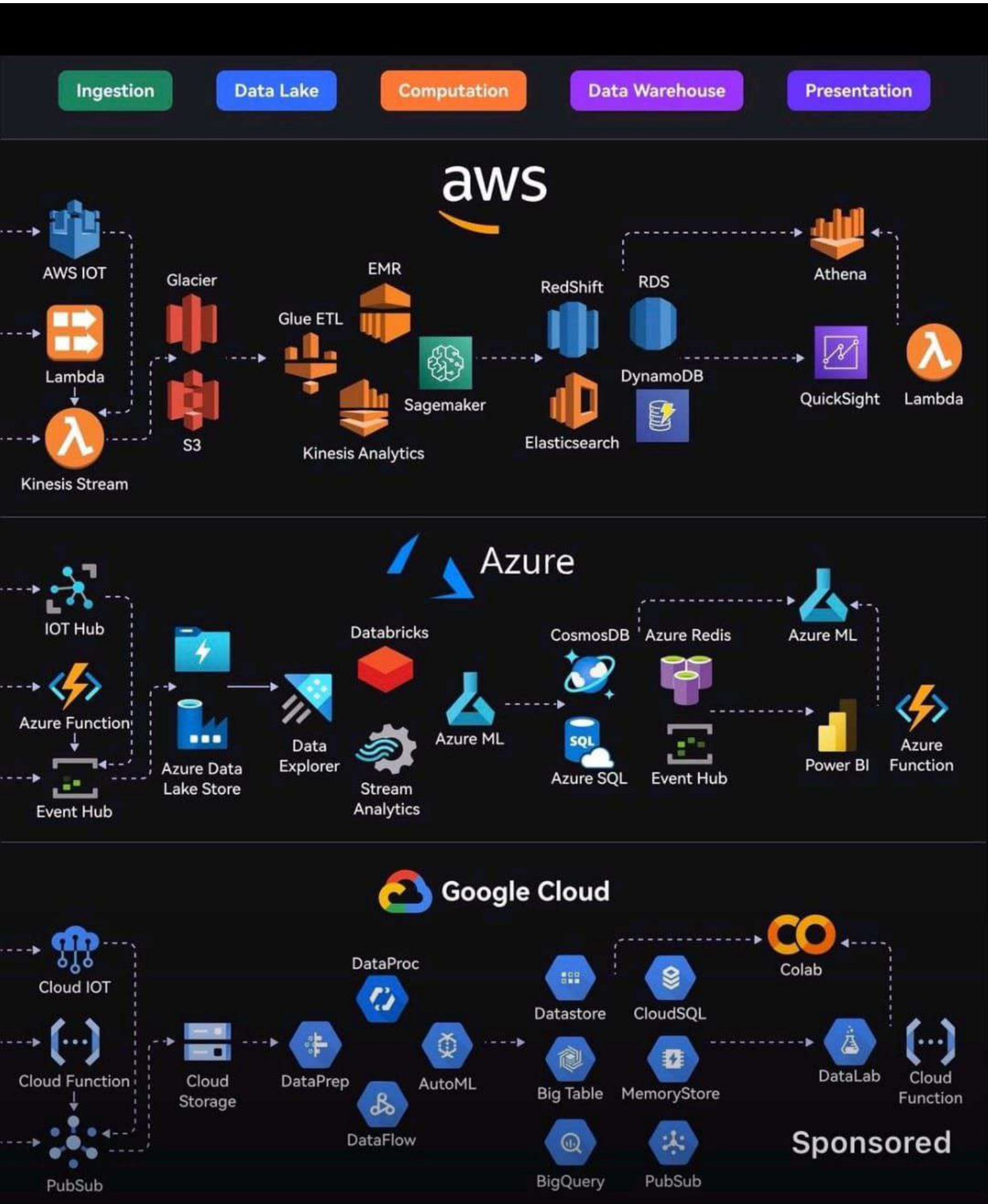
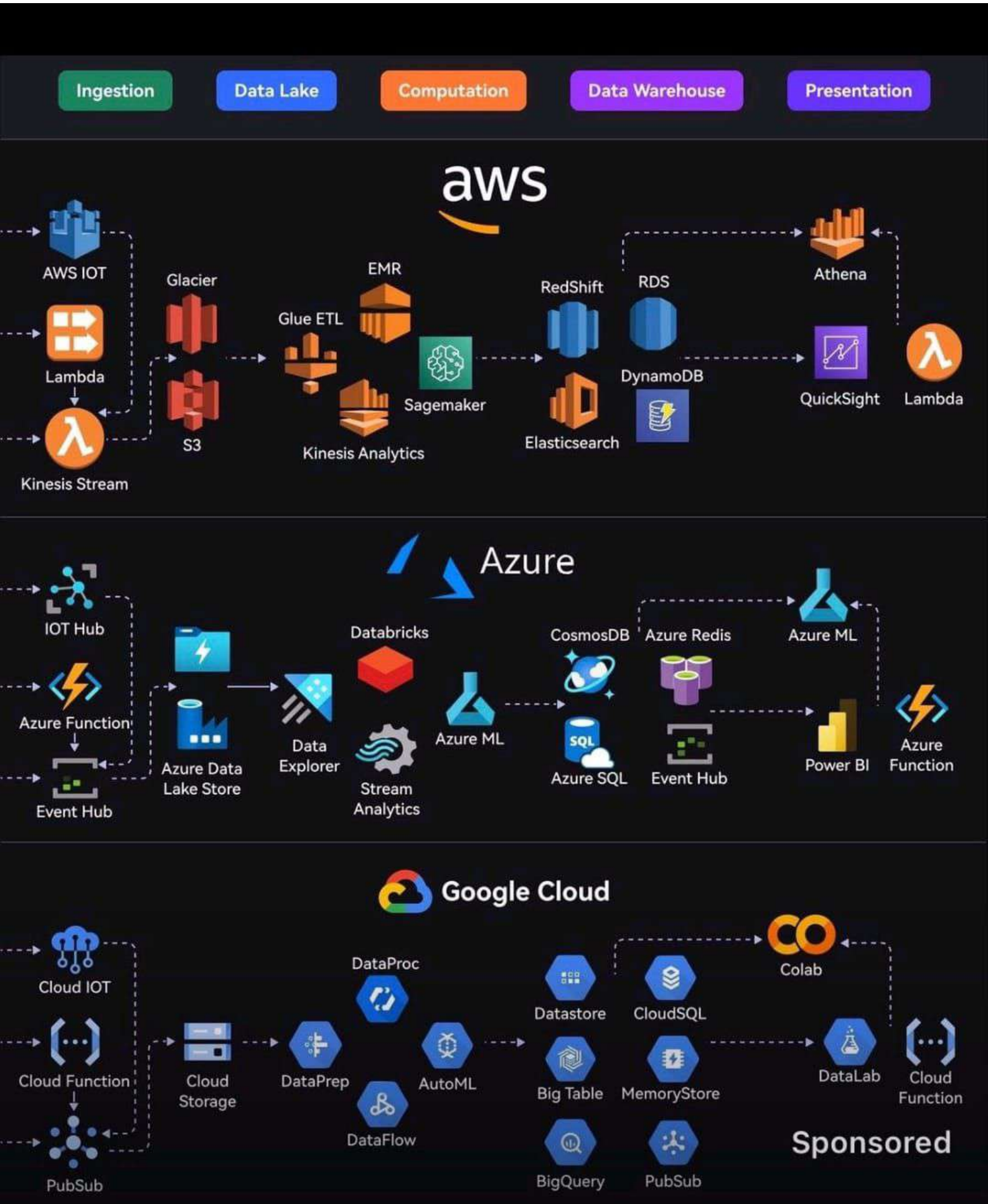
Data Pipeline Overview

ByteByteGo




Ingestion
Data Lake
Computation
Data Warehouse
Presentation
aws
AWS IOT
Lambda
Kinesis Stream
Glacier
S3
EMR
Glue ETL
Kinesis Analytics
Sagemaker
RedShift
Elasticsearch
RDS
DynamoDB
Athena
QuickSight
Lambda
Azure
IOT Hub
Azure Function
Event Hub
Azure Data Lake Store
Data Explorer
Databricks
Stream Analytics
Azure ML
CosmosDB
Azure SQL
Azure Redis
Event Hub
Azure ML
Power BI
Azure Function
Google Cloud
Cloud IOT
Cloud Function
PubSub
Cloud Storage
DataPrep
DataProc
DataFlow
AutoML
Datastore
Big Table
BigQuery
CloudSQL
MemoryStore
PubSub
Colab
DataLab
Cloud Function
Sponsored


Ingestion
Data Lake
Computation
Data Warehouse
Presentation
aws
AWS IOT
Lambda
Kinesis Stream
Glacier
S3
EMR
Glue ETL
Kinesis Analytics
Sagemaker
RedShift
Elasticsearch
RDS
DynamoDB
Athena
QuickSight
Lambda
Azure
IOT Hub
Azure Function
Event Hub
Azure Data Lake Store
Data Explorer
Databricks
Stream Analytics
Azure ML
CosmosDB
Azure SQL
Azure Redis
Event Hub
Azure ML
Power BI
Azure Function
Google Cloud
Cloud IOT
Cloud Function
PubSub
Cloud Storage
DataPrep
DataProc
DataFlow
AutoML
Datastore
Big Table
BigQuery
CloudSQL
MemoryStore
PubSub
Colab
DataLab
Cloud Function
Sponsored

Cloud Services Comparison Cheatsheet

SERVICES	Azure	aws	Google Compute
Analytics	Azure stream Analytics	Amazon Kinesis	Cloud Dataflow
App Hosting	Azure cloud services	Amazon Elastic Beanstalk	Google App Engine
Automation	Azure Automation	AWS Opsworks	Compute Engine Management
Block Storage	Azure Managed Storage	Amazon Elastic Block Storage	Persistent Disk
Compliance	Azure Trust Center	AWS Cloud HSM	Google Cloud Platform security
Computing	Virtual Machines	Elastic Compute Cloud (EC2)	Compute Engine
Cloud Specific Container	Azure Container service	EC2 Container Service	Container Engine
Cloud Agnostic Container	Azure AKS	Amazon EKS	GKE
Content Delivery Network (CDN)	Azure CDN	Amazon CloudFront	Cloud CDN
DNS Services	Azure Traffic Manager	AWS Route 53	Cloud DNS
Identity & Access Management	Azure Active Directory	AWS Identity and Access Management	Cloud Identity Access Management
Key Management Services	Azure Key Vault	AWS KMS	Google Cloud KMS
Load Balancing	Load Balancing for Azure	Elastic Load Balancing	Cloud Load Balancing
Log Monitoring	Azure Operational Insights	Amazon CloudTrail	Cloud Logging
NoSQL Database Options	Azure DocumentDB	AWS DynamoDB	Cloud Datastore
Notifications	Azure Notification Hub	Amazon Simple Notification service	None
Object Storage	Azure Blob Storage	Amazon Simple Storage (S3)	Cloud Storage
Performance Monitoring	Azure Application Insights	Amazon CloudWatch	Stackdriver Monitioring
Private Connectivity	Azure Express Route	AWS Direct Connect	Cloud Interconnect
Relational Database	Azure Relational Database	Amazon RDS	Cloud SQL
Scaling Options	Azure Autoscale	Auto Scaling	Auto Scaler
Serverless Computing	Azure Functions	AWS Lambda	Google Cloud Functions
Virtual Network	Azure Virtual Network	Amazon VPC	Cloud Virtual Network

Database (DB) vs Data Warehouse (DW) vs Data Lake (DLH) vs Data Lakehouse (DLH

Database (DB)

Store current, structured, operational data

Purpose

- Highly structured
- OLTP
- Fast reads/writes
- Examples: MySQL PostgreSQL

Data Warehouse (DW)

Store historical, cleaned, aggregated data

Characteristics

- Schema-on-write
- OLAP (Online-Analytical Processing)
- Optimized for analytics
- Examples MySQL. PostgresqL

Data Lake (DL)

Store all raw data (structured, semi-structured, unstructured)

Characteristics

- Schema-on-read
- Stores everything raw
- Built on cheap storage
- Weak governance
- Commonly a „data swamp"

Data Lakehouse (DLH)

Combine flexibility of a data lake & ACID reliability of a data warehouse

Characteristics

- Open storage formats
- ACID transactions
- Streaming + batch
- Unified architecture
- Cheaper than DW

Delta Lake

Add ACID transactions + versioning to a data lake

Characteristics

- ACID transactions
- Time travel
- Schemaenforcement
- MERGE/UPDATE/DELETE
- Works on open formats

IT Help Desk (L1 Support)

Skills to learn:

- Windows troubleshooting
- Basic networking
- Windows troubleshooting
- Active Directory
- Group Policies
- File Server / Print Server

IT Support Engineer (L2 Support)

Skills to upgrade:

- OS installation & repair
- DNS/DHCP/IP troubleshooting

Core skills:

- Windows/Linux Server
- Active Directory
- Group Policies
- File Server / Print Server

System Administrator (L3 Support)

Core skills:

- Windows /Linux Server
- Active Directory

Goal:

- Manage servers + maintain infrastructcture

IT Infrastructure Engineer

Gulshan-kumar101

Advanced skills:

- Virtualization (VMware / Hyper-V)
- Storage (SAN/NAS)
- Firewalls & Security

Cloud skills:

- VMs, Storage, VNETs
- IAM / Security
- Backup & DR in cloud
- Migration

Cloud Engineer (Azure / AWS/ GCP)

Cloud skills:

- VMs, Storuge, VNETs
- IAM / Security
- Backup & DR in oloud

Goal:

- Autonation + scalable deployments

🎯 Conclusion:

Start from L1, build hands-on skills,
move to L2 → L3 → Infrastructure → Cloud → DevOps
This is the most stable & high-growth IT roadmap.