Databases

Messaging Queue

api

Flow logs

Kafka Streaming Platform

Able to handle Different formats/schema, real-time as well as batch with high throughput. Can be deployed in any cloud vendor

Ingestion Layer

High throughput Spark jobs
Platform can be anything - AWS EMR, GCP Data Proc, HDP distribution

Lamdas
If some file based triggered light weight processing is required

Transforming the data (cleaning, formatting, deduplication)

Kafka APIs for streaming data transformations
These will be simple Java applications can be deployed on kubernetes Engine

Stand Alone applications
If any very custom data processing need is there. Can be deployed on cloud servers/kubernetes

Data Transformation/Processing Layer

Big Query   Hive   HBase   Elastic Search

S3, GCS, HDFS

Data/Metadata Storage Layer

Tablue   QlikView   Power BI

BI/Reporting Tools