**SUMMARIES:**

**PIG:**

Pig tool consists of a SQL-like data flow language Pig Latin which supports distributed processing. It offers many query frameworks and user defined programs to achieve for fast and simple data manipulation. It has eased the process of writing complex java based map-reduce programs. It can be used for SQL querying and scripting without relying on the metadata.

**HIVE:**

Hive is a data-warehouse build on the Hadoop architecture. It consists of HiveQL query language to accomplish data summarization and data analysis. It provides a familiar interface to the SQL language. It processes queries based on the metadata.

**SQOOP:**

​​Sqoop is a data integration component, transfers data between Hadoop and relational databases. Sqoop parallelizes data transfer, mitigates excessive loads, allows data imports and transfers data quickly. This tool allows Map-Reduce, Hive, HBase, Cassandra, Pig to interact with traditional databases and performs data manipulation tasks.

**Flume:**

Flume is a data integration component which is used for the streaming ingestion into the Hadoop Distributed File System (HDFS). It integrates data like application logs, sensor data, machine data and geolocation data with the HDFS, allows tools like Hive and Pig to run interactive queries on the data.

**HBASE:**

Hbase is a table based NoSQL database which can scale to billions rows and columns to handle huge data set. It provides random and real time access to the data and allows to store multi structured and sparse data.

**FALCON:**

Falcon is a framework for managing data life cycle in Hadoop clusters. It addresses enterprise challenges related to Hadoop data replication, business continuity, and lineage tracing by deploying a framework for data management and processing. Falcon centrally manages the data lifecycle, facilitate quick data replication for business continuity and disaster recovery and provides a foundation for audit and compliance by tracking entity lineage and collection of audit logs

**ATLAS:**

Atlas allows exchange of metadata with other tools and processes within and outside of the Hadoop stack. It is designed to easily model new business processes and data assets with agility.

**RANGER:**

Ranger manages policies for access to files, folders, databases, tables, or column. These policies can be set for individual users or groups and then enforced consistently across HDP stack.