Sqoop: Used for efficiently transferring data between Hadoop and relational databases Sqoop helps to import data from RDBMS to Hadoop Distributed File System (HDFS), and export data from HDFS to RDBMS.

Flume: Apache Flume is a tool for collecting and moving very large amounts of stream data from various sources. The main purpose of Apache Flume is to update the collection and intake of large datasets into HDFS or other data stores and for data logging.

Flume agent - A Java virtual machine that hosts the process components through event flows. The agent is responsible for collecting and filtering data to a destination

Three main components of a flume agent –

* Sources
* Channels
* Sinks

Flume Source - A component that is responsible for taking in data from various sources and makes it ready for processing by the flume agent.

Flume Sink - A component of the flume agent it is used for storing data into HDFS

Oozie: A workflow scheduler system that manages Apache Hadoop Jobs. The purpose of Oozie is to simplify the process of Hadoop jobs by allowing users to build and execute workflows of Hadoop jobs a lot easier.

Fork Node - used to divide the path of execution into multiple paths of execution.

Workflow - A collection of actions organized in a Directed Acyclic Graph(DAG). A workflow can contain a variety of actions such as Hadoop MapReduce jobs and many others.