**Spark Introduction**

* **Spark** is an execution engine and not storage service, it is similar to but better than MapReduce from HDFS.
* Spark includes **core spark, SparkSQL, MLA (Machine Learning), GraphX (Graph Representation)**
* Spark can also read from **HIVE**. **Hive** is used only for storage, process is done using SparkSQL
* **Spark** can also run on **Kubernetes**
* **ZooKeeper** is used to communicate between machines, it comes with Spark.

**Hadoop** includes **Active node** and **StandBy node.** How do you know which is active? ZooKeeper does it. If Active node goes down then ZooKeeper brings up another node. Services like YARN, Hbase, Spark are always on Active node and Standby node.