Spark = many times faster than hadoop 10-20 times faster

Spark sits on top of hadoop

Principle: why to load intermediate results on secondary storage especially hdfs (why not on primary memory) and during down-time, automatically the framework shud store data on secondary storage.

Real time processing: possible in spark , not hadoop (spark streaming)

Spark has features of Hive.

Spark is written in scala.

Scala runs in JVM

Only one sparkcontext per jvm

Scala is more efficient than java

Scala has its own build tool: sbt

Build.sbt file (like pom.xml for mvn)

Spark does lazy computing

Emeditor: good for big files