**Apache Spark:**

Spark was introduced by Apache Software Foundation for speeding up the Hadoop computational computing software process.

Spark is one of Hadoop’s sub project developed in 2009 in UC Berkeley’s AMPLab by Matei Zaharia.

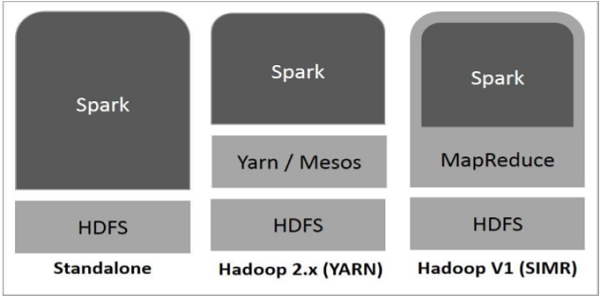
It was Open Sourced in 2010 under a BSD license.

It was donated to Apache software foundation in 2013, and now Apache Spark has become a top level Apache project from Feb-2014.

**Features of Apache Spark:**

1. Speed
2. Supports Multiple Languages
3. Advance analytics

**Spark Build on Hadoop:**

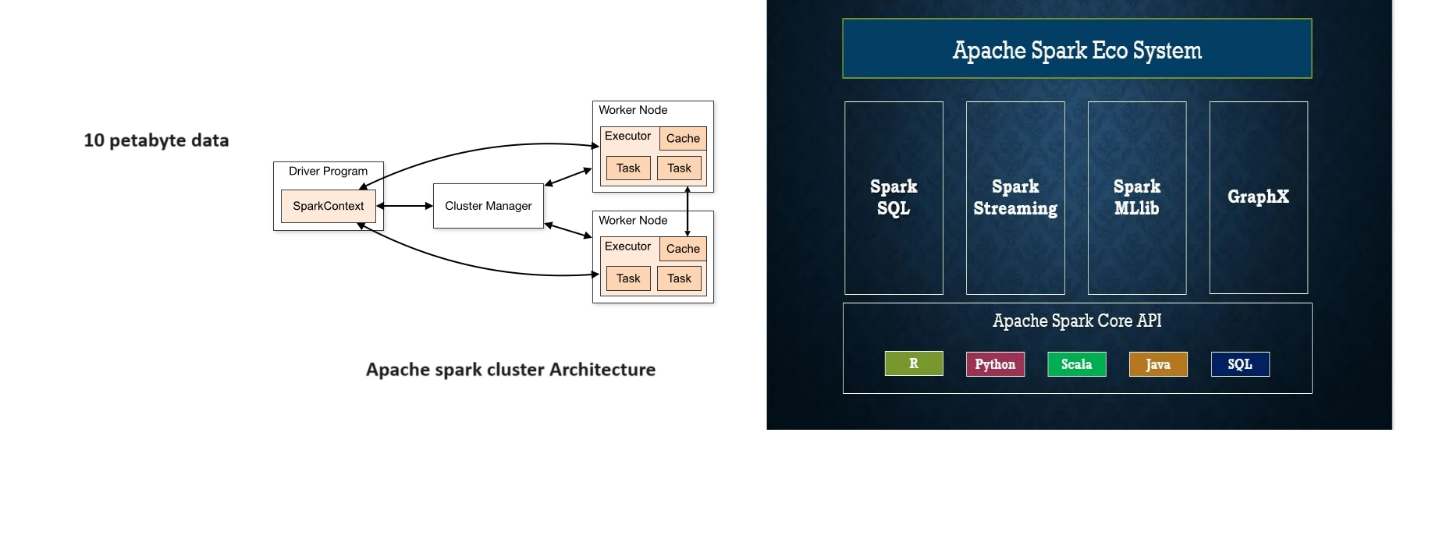


1. Standalone
2. Hadoop Yarn
3. Spark in MapReduce

**Components of Spark:**

1. Apache spark core
2. Spark SQL
3. Spark Streaming
4. MLib (Machine Learning)
5. GraphX (Graph)
6. SparkR

**Architecture:**



**Sample code that are run on pyspark:** 