

SPARK

SPARK Introduction

What is spark?

1. In memory cluster operations facilitates lightning fast processing.
2. Supports Built-in API in - Java
Scala
Python

Features of spark:

1. Faster disk operations
2. Faster in-memory operations
3. Built-in support for languages like scala, python, Java.
4. Compatible with SQL, streaming data, Machine learning, Graph algorithms - Helps for Analytics.

Evolution of an Apache Spark

2009

started as a Research Project at UC-Berkley AMP Lab, 2009.

The project was made open sourced

2010

kept evolving and expanding with various research papers.

2013

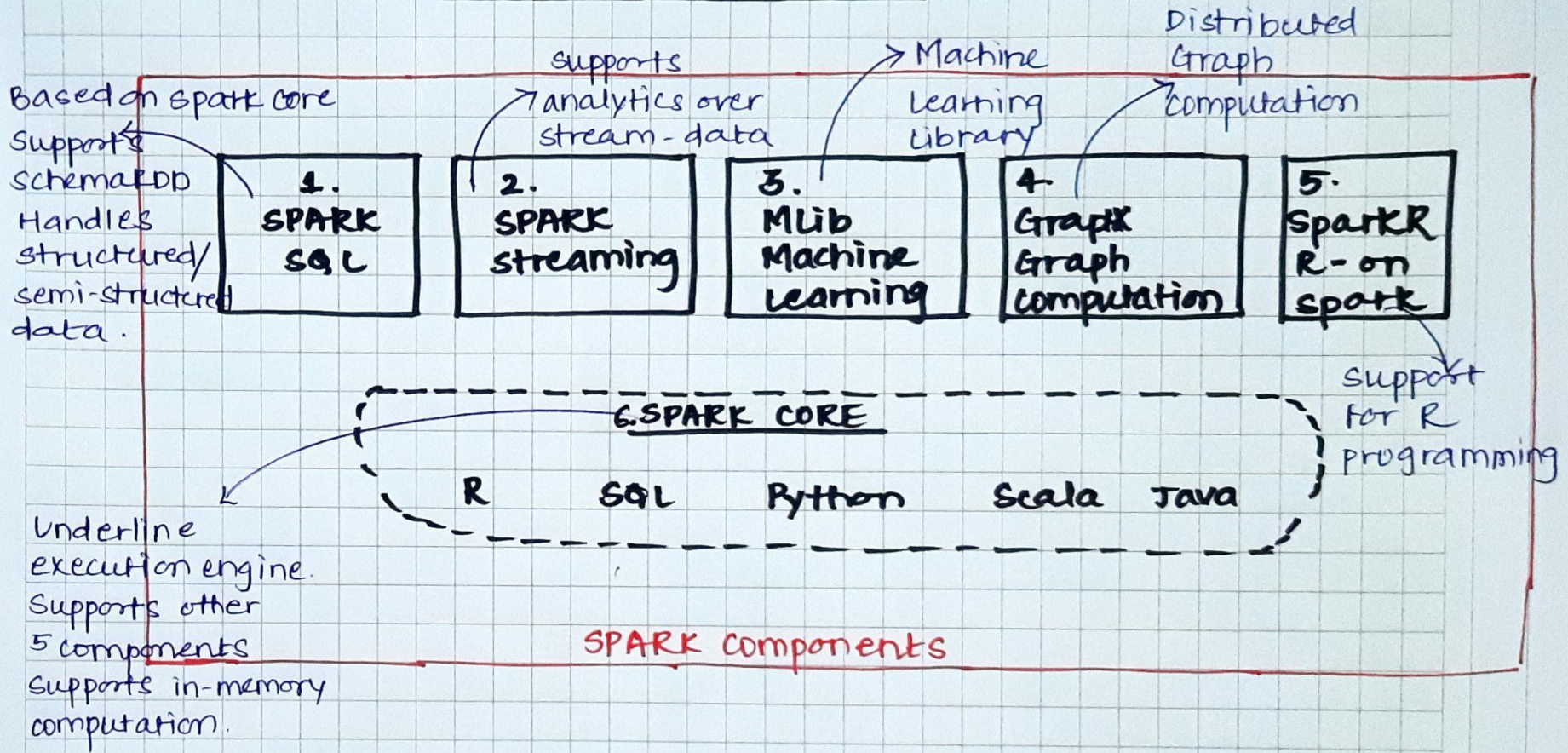
Transferred to an Apache Software Foundation.

Apache foundation's vital project

2014

Other research contributors (like databricks) used spark for key developments in this field.

SPARK ECOSYSTEM



Note: SparkR is compatible/supported after Apache spark 1.4 version.