

零基础学习 Spark 1.x 应用开发系列课程

随堂笔记之 Lesson 01

1、 内容提纲

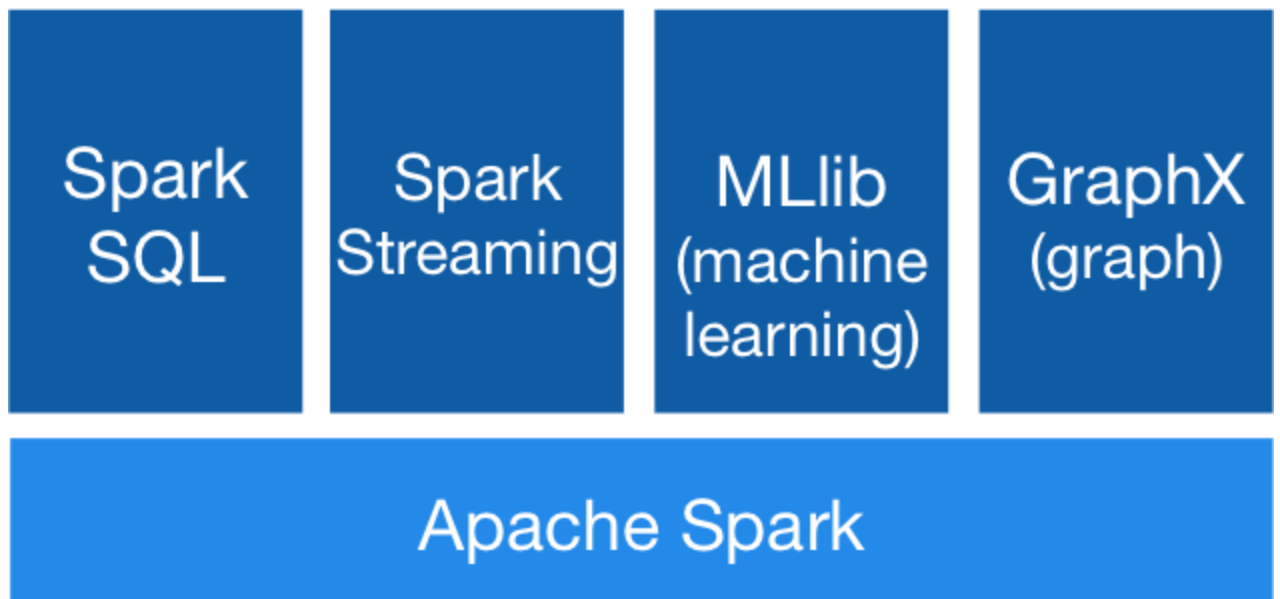


2、 具体内容

2.1、 知识点

1) 引擎

- a) 快速：DAG、Memory
- b) 通用：一栈式的大数据处理解决方案



2) 数据处理/数据分析

a) 大数据集

MapReduce 与 Spark 相比，有哪些异同点

1、 基本原理上：

- a) MapReduce: 基于磁盘的大数据批量处理系统
- b) Spark: 基于 RDD（弹性分布式数据集）数据处理，显式的将 RDD 数据存储到磁盘和内存中。

2、 模型上：

- a) MapReduce: 可以处理超大规模的数据，适合日志分析挖掘等较少的迭代的长任务需求；结合了数据的分布式的计算。
- b) Spark: 适合数据的挖掘，机器学习等多轮迭代式计算任务

容错性上:

- a) 数据容错性
- b) 节点容错性

Spark Linage