

《企业级 Spark 1.x 应用开发课程》大纲

一、环境搭建

Spark 生态系统介绍

Spark 对于 MapReduce

Spark 运行模式

源码下载

源码编译

Spark 环境搭建：scala、Spark

spark-shell、spark-submit 使用

Spark IDEA 开发环境使用

二、编程模型

RDD、transformation、action、cache

Lineage、Checkpoint、Spark 容错

Dependency

Broadcast Variables、Accumulators

案例

三、Spark Streaming

概述

和 Storm 对比

DStream

Operations

Persistence

Checkpoint

Fault-tolerance Semantics

Performance Tuning

案例

四、Spark SQL

概述

架构

External Data Source

案例

五、Spark 内核

DAGScheduler、TaskSeduler、Task

容错性、推测机制

数据本地性

六、Tachyon

Tachyon 原理

Tachyon 案例

七、综合

Spark 优化案例分析

Kafka/HDFS、Spark Streaming、Spark SQL 集合 External Data Source
案例分析