Spark 系统及其编程技术

Spark 系统

Scala 语言

Scala 是一种搭载在 JVM 上的语言，继承了 OOP 特性和函数式特性

变量分为可变引用和不变引用

编译为字节码运行在 JVM 上

为什么要有 Spark

MapReduce 的缺陷：高延时，难以共享数据，大量磁盘 IO，对复杂计算支持不足

后 Hadoop 时代：内存计算为核心，需要数据共享

RDDs (Resilient Distributed Datasets)：弹性分布式数据集，是 Spark 的核心，能跨集群所有节点进行并行计算的分区元素集合

RDD 可以从文件创建，或是通过一个已有的 RDD 转换得到

RDD 基本的算子是 transformation 和 action

RDD 是只读且可分区的，其全部或者部分可以缓存在内存，在多次计算中重用，弹性是指内存不足时可以交换到磁盘

RDD 可以构成计算流图

Spark 基本构架

Spark 的基本是 MapReduce，RDD 和 FP（函数式编程），底层的文件系统可以是 HDFS 等

节点分为 Master node 和 Worker node，分别负责控制和计算，每个 worker 上有一个 executor，负责完成任务执行

一个 Spark 应用包含了一个 Driver 和多个 executor，一个应用中有多个 job，由 Spark action 产生，而一个 job 中有众多 task，将 task 分组，每个组称为 stage 或是 taskset（MapReduce 中可以分为 map 的 stage 和 reduce 的 stage），task 是基本的执行单元，在 executor 上执行（executor 是多线程环境，可以执行多个 task）

driver 负责启动程序和初始化环境

executor 和 worker 一一对应，但是 worker node 上可以有多个 worker（对应不同的应用）

Spark 程序执行过程

spark context 由 driver 启动，是 Spark 运行的核心模块，是 Spark 程序最基本的初始化，可以用于连接相应集群的配置来分配资源，然后分发代码至各 executor

从 RDD 角度来看，一个 job 就是一张 RDD 的 DAG，即一组 transformation 和一个 action 的集合，每执行一次 action 就会提交一个 job

stage 分为 shuffle stage 和 final stage，每个 job 只有一个 final stage，如果有 shuffle 操作则会生成 shuffle stage，shuffle 只在宽依赖触发

job 和 stage 都是针对一个 RDD 的划分，而 task 是对 RDD 中某个分区的执行

Spark 技术特点

惰性计算：RDD 的 transformation 并不进行作业提交，直到 action 才触发作业提交

lineage：通过 RDD 的世系关系记录其转变过程，如果数据丢失了可以重新计算

线程调度：使用线程池，避免线程启动和切换的开销

API：使用 Scala 开发，十分简洁，同时也支持 Java 和 python

可以部署在多种底层平台

Spark 编程模型

分布式数据集的抽象 RDD

RDD 是只读记录的集合，只能通过两种方式创建

- 将一个已存在的集合并行化或是引用外部数据
- 通过其他 RDD 的确定性操作创建

RDD 的两种操作为

- transformation：惰性，只定义不计算
- action：立即计算 RDD 的值并返回给程序，或是写入外存

RDD 有两种容错方式

- lineage：记录 RDD 之间的变换关系
- checkpoint：对于很长的 lineage，记录检查点

Spark 中有两种依赖方式

- 窄依赖：父 RDD 中一个 partition 最多被子 RDD 中一个 partition 依赖
- 宽依赖：父 RDD 中一个 partition 被子 RDD 中多个 partition 依赖

可以使用多种方法存储 RDD：如直接将 Java 对象存在内存，或是序列化后存入内存，或是存入磁盘

每个 RDD 包含

- 一组 partition，是数据集的原子结构
- lineage 信息，即依赖的父 RDD
- 函数，指明在父 RDD 上如何计算
- 元数据

Spark 和集群管理工具的结合

集群管理工具可以提供资源的管理，共享，隔离，以及良好的可扩展性和容错

Spark 可以与多种集群管理工具共同工作，如 yarn，mesos 等

Spark 可以打包在 docker 中，进行快速部署

Spark 系统其他组件

Spark SQL：处理结构化数据的分布式 SQL 引擎

Spark Streaming：流式数据处理系统，将数据流表示为一系列连续的 RDD

GraphX：对图表示和处理的组件，使用点分割存储，图的一条边存储在一台机器上，但是点存储在多个机器，有一个点作为主点，其余作为虚点，主点更新后再发送给各个虚点

MLlib：分布式机器学习算法库