Atitit 大数据系列技术概要hadoop spark hive

为什么使用 Hive

直接使用 MapReduce 所面临的问题：

　　1、人员学习成本太高

　　2、项目周期要求太短

　　3、MapReduce实现复杂查询逻辑开发难度太大

为什么要使用 Hive：

　　1、更友好的接口：操作接口采用类 SQL 的语法，提供快速开发的能力

　　2、更低的学习成本：避免了写 MapReduce，减少开发人员的学习成本

1. 更好的扩展性：可自由扩展集群规模而无需重启服务，还支持用户自定义函数

2. Hive：hive是基于Hadoop的一个数据仓库工具，可以将结构化的数据文件（或者非结构化的数据）映射为一张数据库表，并提供简单的sql查询功能，可以将sql语句转换为MapReduce任务进行运行。 其优点是学习成本低，可以通过类SQL语句快速实现简单的MapReduce统计，不必开发专门的MapReduce应用，十分适合数据仓库的统计分析。

使用Hive，就不用去写MapReduce，而是写sql语句就行了。

作者：ashi

链接：https://juejin.cn/post/6844903984268836871

来源：掘金

著作权归作者所有。商业转载请联系作者获得授权，非商业转载请注明出处。