Hadoop的前世今生

Google大数据技术

MapReduce，BigTable，GFS

Hadoop是什么？

开源的 分布式存储 分布式计算平台

两个核心：

HDFS：分布式文件系统，存储海量数据

MapReduce：并行处理框架，实现任务分解和调度

Hadoop可以用来做什么？

构建大型数据仓库，PB级数据的存储、处理、分析、统计等业务

Hadoop优势：

1. 高扩展
2. 低成本
3. 成熟的生态圈

Hadoop生态系统及版本

HIVE：

SQL语句转化为Hadoop任务，降低使用Hadoop的门槛

HBASE：

不是传统的关系型数据库，HBASE放弃事务的特性 追求更高的扩展

其提供对数据的随机读写和实时访问，实现对表数据的读写功能

Zookeeper：

监控Hadoop结点状态

选择学习稳定的1.2的版本