暑期实训学习报告

1813012 柳鑫

一. 学习内容

- ① 了解 Hadoop
- ② 安装 Apache Hadoop
- ③ 尝试运行 Hadoop 自带示例 WorldCount

二. 学习结果

Hadoop 是一个由 Apache 基金会所开发的分布式系统基础架构。用户可以在不了解分布式底层细节的情况下,开发分布式程序。充分利用集群的威力进行高速运算和存储。Hadoop实现了一个分布式文件系统(Hadoop Distributed File System),简称 HDFS。HDFS 有高容错性的特点,并且设计用来部署在低廉的(low-cost)硬件上;而且它提供高吞吐量(highthroughput)来访问应用程序的数据,适合那些有着超大数据集(large data set)的应用程序。HDFS 放宽了(relax)POSIX 的要求,可以以流的形式访问(streaming access)文件系统中的数据。Hadoop 的框架最核心的设计就是:HDFS 和 MapReduce。HDFS 为海量的数据提供了存储,而 MapReduce 则为海量的数据提供了计算。

Hadoop

Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

Overview 'BigData02:9000' (active)