# 学习心得

1813003 李昱龙

Hadoop：

数据的存储与分析思路：

只有一个硬盘的百分之一似乎很浪费。我们可以存储100个数据集，让他们共享磁盘的访问

Hadoop提供了一个稳定的共享存储和分析系统。存储由HDFS实现，分析由MapReduce实现。

MapReduce：

采用蛮力，针对每一个查询，每一个数据集都会被处理

安装Hadoop

按照网上的教程，先安装jdk，再安装hadoop

安装SSH

安装 Java 环境

Java 环境可选择 Oracle 的 JDK，或是 OpenJDK（可看作 JDK 的开源版本），现在一般 Linux 系统默认安装的基本是 OpenJDK，这里安装的是 OpenJDK1.8.0版本的。

有的 CentOS 6.4 默认安装了 OpenJDK 1.7，这里我们可以使用命令检查一下，和 Windows 下的命令一样，还可以查看 JAVA\_HOME 这个环境变量的值。

安装 Hadoop

tar -zxf ~/下载/hadoop-2.6.5.tar.gz -C /usr/local # 解压到/usr/local目录中

cd /usr/local/ # 切换当前目录为 /usr/local 目录

mv ./hadoop-2.6.5/ ./hadoop # 将文件夹名改为hadoop

chown -R root:root ./hadoop # 修改文件权限，root 是当前用户名

Hadoop 安装方式有三种，分别是单机模式，伪分布式模式，分布式模式。

单机模式：Hadoop 默认模式为非分布式模式（本地模式），无需进行其他配置即可运行。非分布式即单 Java 进程，方便进行调试。

伪分布式模式：Hadoop 可以在单节点上以伪分布式的方式运行，Hadoop 进程以分离的 Java 进程来运行，节点既作为 NameNode 也作为 DataNode，同时，读取的是 HDFS 中的文件。

分布式模式：使用多个节点构成集群环境来运行Hadoop，需要多台主机，也可以是虚拟主机。