**学习日志**

姓名：李奥 学号： 1812998 日期： 6.20

**学习内容安排：（以6.15为例）**

**了解Hadoop 、 安装Apache Hadoop学习反馈：**

已掌握知识：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，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来访问应用程序的数据，适合那些有着超大数据集的应用程序。HDFS放宽了POSIX的要求，可以以流的形式访问 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。

安装Hadoop

1.将hadoop-2.7.2.tar.gz导入到linux

2.解压hadoop-2.7.2.tar.gz

tar -zxvf hadoop-2.7.2.tar.gz -C /opt/module/

3.配置hadoop的环境变量

(1)获取hadoop的安装路径

pwd

1

/opt/module/hadoop-2.7.2

(2)打开/etc/profile.d/env.sh文件

sudo vim /etc/profile.d/env.sh

1

(3)在文件末尾加上Hadoop路径

#HADOOP\_HOME

export HADOOP\_HOME=/opt/module/hadoop-2.7.2

export PATH=PATH:PATH:PATH:HADOOP\_HOME/bin

export PATH=PATH:PATH:PATH:HADOOP\_HOME/sbin

(4)刷新配置文件让文件生效

source /etc/profile

1

(5)测试Hadoop是否安装成功

hadoop version

未掌握知识：

学习心得：Hadoop分布式文件系统中很好用