**学习心得**

18183023 谭智元 6.22

**今天学习的内容：**

1. yum仓库配置：

YUM（全称为 Yellow dog Updater, Modified）是一个在Fedora和RedHat以及CentOS中的Shell前端软件包管理器。基于RPM包管理，能够从指定的服务器自动下载RPM包并且安装，可以自动处理依赖性关系，并且一次安装所有依赖的软件包，无须繁琐地一次次下载、安装。

在Linux上使用源码的方式安装软件非常满分，使用yum可以简化安装的过程。

常用命令：

yum install -y httpd：安装httpd并确认安装

yum list：列出所有可用的package和package组

yum clean all：清除所有缓冲数据

yum deplist httpd：列出一个包所有依赖的包

yum remove httpd：删除httpd

关联网络yum源和制作本地yum源的过程比较繁琐，耗时较长，不再赘述

2.Python类：

之前的python编程都是基于函数，并未接触过面向对象的python编程，因此借此机会粗略了解一下python的面向对象的特性。

通过学习掌握了一些关键点如下：

a. 在类定义声明的第一行代码之后可以立即书写一段儿字符串表示类的帮助文档；

b. 类的方法与普通的函数只有一个特别的区别——它们必须有一个额外的第一个参数名称, 按照惯例它的名称是 self。self代表类的示例，而不是类本身。

c. 类的构造方法用\_\_init\_\_ ( self [,args...] )表示，类的析构方法用\_\_del\_\_( self )表示。

d. python不使用new关键字来构造对象，而是直接调用构造方法构造新对象。

e. \_\_foo\_\_: 定义的是特殊方法，一般是系统定义名字 ，类似 \_\_init\_\_() 之类的；

\_foo: 以单下划线开头的表示的是 protected 类型的变量，即保护类型只能允许其本身与子类进行访问，不能用于 from module import \*；

\_\_foo: 双下划线的表示的是私有类型(private)的变量, 只能是允许这个类本身进行访问了。

3.了解Hadoop和安装Hadoop：

Hadoop就是存储海量数据和分析海量数据的⼯具。

Hadoop是由java语⾔编写的，在分布式服务器集群上存储海量数据并运⾏分布式分析应用的开源框架，其核⼼部件是HDFS与MapReduce。

HDFS是⼀个分布式⽂件系统：引⼊存放⽂件元数据信息的服务器Namenode和实际存放数据的服务器Datanode，对数据进⾏分布式储存和读取。

MapReduce是⼀个计算框架：MapReduce的核⼼思想是把计算任务分配给集群内的服务器里执行 。 通过对计算任务的拆分（Map计算/Reduce计 算 ） 再根据任务调度器（JobTracker）对任务进⾏分布式计算。

HDFS为海量的数据提供了存储，则MapReduce为海量的数据提供了计算。把HDFS理解为⼀个分布式的，有冗余备份的，可以动态扩展的⽤来存储⼤规模数据的⼤硬盘。把MapReduce理解成为⼀个计算引擎，按照MapReduce的规则编写Map计算/Reduce计算的程序，可以完成计算任务。

对Hadoop的安装只进行到下载安装包（外网源不稳定，下载速度慢）、创建Hadoop用户、配置SSH的无密码登录、以及jdk的安装部分。对于jdk与安装系统时自带的是openjdk还是仅仅是jre这一部分十分糊涂与困扰，因此卡在了这一步。

**反思：**

1.今天学习的时间主要花费在yum的配置上和查找Hadoop安装教程上，难度较大，需要花时间仔细研究

2.由于其余课程的期末结课作业，今天分配给预习的时间较少，可能需要等到结课作业提交之后状况才能有所缓解。