

中软国际实训班预习内容安排

时间安排：

6月15号至6月28号（详见附件中的课表）

1、Java 基础

本学期已学习Java系列课程，基础知识掌握不牢固的同学自己抽时间再巩固，这里不再安排时间。供参考的资料：《JAVA程序设计教程第2版-清华大学-雍俊海》和《Java就业培训教程及源代码-张孝祥》。

2、版本控制系统Git

工具下载：<https://git-scm.com/downloads>

参考资料：

视频资料在学院实验中心主页的下载专区：<https://cslab.nankai.edu.cn/info/1049/1085.htm>

《廖雪峰Git教程》<https://www.liaoxuefeng.com/wiki/896043488029600>

Git Book: <https://git-scm.com/book/en/v2>

Git Doc: <https://git-scm.com/doc>

《Python编程：从入门到实践》这本书的附录D部分也有介绍Git

2.1、Git简介

Git是目前世界上最先进的分布式版本控制系统，在处理各种项目时都十分高效，而且非常的高大上。

SVN是集中式版本控制系统，版本库是集中放在中央服务器的，而干活的时候，用的都是自己的电脑，所以首先要从中央服务器哪里得到最新的版本，然后干活，干完后，需要把自己做完的活推送到中央服务器。集中式版本控制系统是必须联网才能工作。

Git是分布式版本控制系统，它就没有中央服务器的，每个人的电脑就是一个完整的版本库，这样，工作的时候就不需要联网了，因为版本都是在自己的电脑上。

2.2、Git 的多人协作 / 远程仓库

<https://www.cnblogs.com/mkl7/p/10854151.html>

2.3、Git操作命令手册

<https://gitee.com/liaoxuefeng/learn-java/raw/master/teach/git-cheatsheet.pdf>

3、Linux基础

https://blog.csdn.net/weixin_38407447/article/details/90581454

Linux菜鸟教程: <https://www.runoob.com/linux/linux-tutorial.html>

基础操作命令: https://blog.csdn.net/qq_41923771/article/details/81452529

- 安装虚拟机、在虚拟机中安装Centos7.0
- Linux基本命令: 查看目录和文件、find命令、查找特定命令、用户及版本信息查看、帮助命令
- linux目录结构, 建立文件目录, 文档的移动、复制、删除, 改变文档所有权, 改变文档权限, 建立链接
- linux的账号与群组, 账号管理, 群组管理, 用户间切换
- 查看进程, 实时监控进程, kill进程
- rpm/yum, 安装JDK, 安装mysql, 从eclipse生成jar包, 在linux执行
- Vi/Vim编辑器
- 配置网络ip, 查看网络ip地址等常用命令

4、Python语言

参考书籍《Python编程: 从入门到实践》, 预计时间2-3天, Python易上手, 容易学, 前边大家已有C、C++、Java语言的基础, 学习起来不难, 2-3天掌握Python的基本用法, 会写会用, 有些不熟的地方, 具体语法用的时候再查即可。参考的书籍上已经很详细了, 有针对不同系统的安装教程, 有代码演示和源码以及比较详细的代码解释。

推荐文本编辑器: VSCode[下载地址](#)

使用VSCode开发Python教程: [官方文档](#)

5、Hadoop

这里引入知乎上一个高赞的解释

链接：<https://www.zhihu.com/question/333417513/answer/74246581>

4

Hadoop是什么

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。

使用Hadoop

1、搭建Hadoop集群

无论是在windows上装几台虚拟机玩Hadoop，还是真实的服务器来玩，说简单点就是把Hadoop的安装包放在每一台服务器上，改改配置，启动就完成了Hadoop集群的搭建。

2、上传文件到Hadoop集群

Hadoop集群搭建好以后，可以通过web页面查看集群的情况，还可以通过Hadoop命令来上传文件到hdfs集群，通过Hadoop命令在hdfs集群上建立目录，通过Hadoop命令删除集群上的文件等等。

3、编写map/reduce程序

通过集成开发工具（例如eclipse）导入Hadoop相关的jar包，编写map/reduce程序，将程序打成jar包扔在集群上执行，运行后出计算结果。

关于Hadoop的学习路线图、编程实践以及案例分析

<https://www.zhihu.com/question/19795366>

6、关于测验和提交学习报告

测验：在问卷星中发布测验题目，旨在帮助大家巩固所学习的一些基础知识。

学习报告：每天晚上12点之前把一天学习内容做一个回顾，总结当天的学习内容，掌握了哪些知识，还有哪些内容不会，做一下记录，也可以写写自己的心得体会，以**Git**方式提交到**GitHub**，小组长负责监督、查看小组成员提交情况以及学习进度，并汇总后提交给朱静雯老师。