大数据概念

1.2.1 单位换算

1.3.3 hadoop是什么

1.3.4 谁先提出来的

1.3.5 发布版本

1.4.1 组成，重点

1.5.2、3

1.5.4 bin相当于接口，用户怎么使用 sbin怎么启动namenode这些节点的，配置文件存在etc/hadoop

生成密钥！ssh-keygen -t rsa

（选填判断）\*10 大题\*2

大数据概念

4v 单位换算 应用发展组织架构场景不考

Hadoop解决什么 人名 基于什么 三大发行版本 优势（4高） \*架构

解决存储（HDFS由什么组成）和计算（Yarn，计算形式MR，app master，容器）的问题

虚拟机不考 jdk和hadoop的安装 配置什么环境变量 hadoop目录结构要知道（bin放接口，把文件传进系统，和sbin是hadoop系统的组件的工具）

hadoop配置文件存在哪 三种方式的配置文件core-site集群，不背数值，要知道这一项是在这个配置文件里的吗，hdfs，yarn资源调用框架，maprdsite与mr计算有关，历史服务器，一共四个site，三个env，hadoop，yarn，mr，还有一个slaves写从机的配置）

maven不考 官方案例不考命令是什么 伪分布 yarn的命令要清楚 有可能问怎么启动namenode 怎么格式化 所有的默认端口 hdfs dfs的命令，看文件，ls，put，cat这个考大题，创建文件，写信息，传到hdfs，查传没传，再下载，或者看看

启动和停止stop yarn把start换成stop jps看启没启动成功 历史服务器的配置文件和启动 日志聚合知道在文件就行 期末考关闭nmrmhm和重启

和Linux结合：是否有个简便的方法写个伪分布集群启动停止的脚本 权限问题