**大数据学院**

**2018年10月24日**



**[第一单元 大数据概论和分布式介绍]**

Hadoop标准化课件设计

**《大数据离线计算》**

**第一单元 Hadoop伪分布式安装**

1. **本单元教学目标**
2. 知识目标：

**·**  了解现代社会的数据特征和常见解决手段；

**·**  掌握大数据及分布式的概念，了解大数据及分布式的应用场景

**·** 掌握Hadoop的特性和优点；

**·**  了解知名企业的Hadoop应用案例，了解Hadoop的三种部署方式；

**·**  掌握Hadoop生态系统及其核心模块构成；

**·** 掌握常见的大数据相关面试问题。

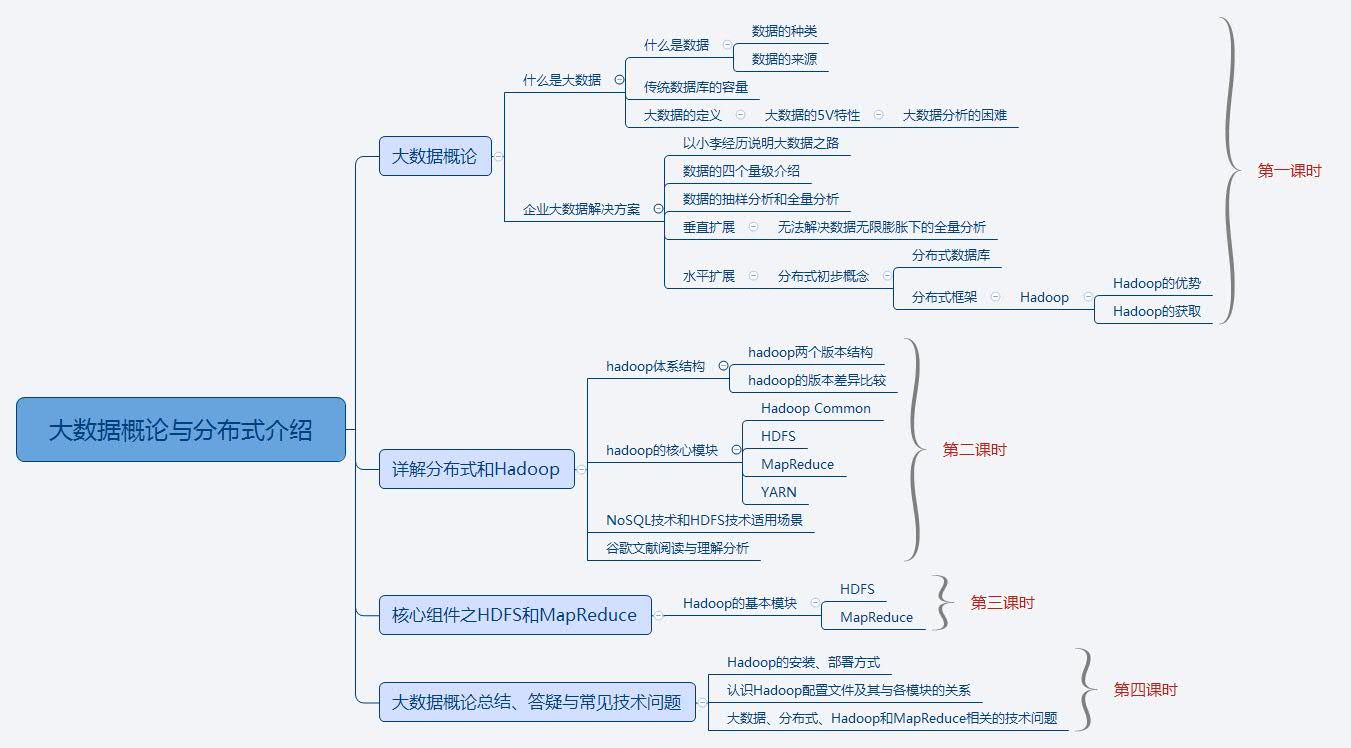
1. 能力目标：

**·** 能够画出Hadoop生态环境技术框架图；

**·** 能够针对企业中的大数据相关问题给出基于Hadoop的解决方案；

**·** 初步掌握搭建Hadoop伪分布式环境的步骤；

1. **课时分配**

****

1. **第一课时：大数据概论**

本课时通过对大数据和分布式的宏观介绍，使学生初步建立大数据知识体系，使得学生能为学习以后的课程打下基础。

1. 知识目标：

**·** 掌握大数据和数据量级的概念、大数据的5V特性、抽样分析和量化分析的概念；

**·** 掌握分布式处理问题的原理和分布式的基本要素；

**·** 了解大数据相关的知名应用、知名企业，大数据岗位职业素养和要求；

**·** 掌握Hadoop的特性和Hadoop在企业中的的基本应用和部署方式。

1. 能力目标：

**·** 能够针对企业的不同数据量级提出初步解决方案；

**·** 能够画出Hadoop生态圈的模块结构图；

**·** 初步培养学生的互联网资料查询和获取能力，消除学生对英文资料和英文网站的畏惧感，能够从Hadoop官网获取Hadoop相关资料。

1. **第二课时： 详解分布式和Hadoop**
2. 知识目标：

**·** 了解hadoop的两个版本(1.0和2.x)；

**·** 掌握Hadoop1.0和Hadoop2.x的体系结构及其区别；

**·** 掌握Hadoop的核心模块及其基本功能；

**·** 掌握HDFS基本概念及基本原理，了解NoSQL技术和HDFS技术各自的适用场景，对谷歌大数据技术文献的快速阅读和理解分析。

1. 能力目标：

**·** 能够根据系统模块分辨是Hadoop1.0还是Hadoop2.x系统；

**·** 能够针对不同的客户存储需求，确定是否适合使用HDFS作为存储系统

**·** 深刻理解HDFS的使用局限性，初步锻炼高新技术资料的阅读和知识获取能力。

1. **第三课时： 核心组件之HDFS和MapReduce**
2. 知识目标：

**·** 掌握Hadoop的基本模块；

1. 能力目标：

**·** 熟悉Hadoop基本模块的基本功能及其相互关系；

**·** Hadoop的工作原理

**·** 掌握MapReduce和HDFS的工作原理；

1. **第四课时： 大数据概论总结、答疑与常见技术问题**
2. 知识目标：

**·** 了解Hadoop的安装、部署方式，认识Hadoop配置文件及其与各模块的关系；

**·** 掌握大数据中关于大数据、分布式、Hadoop和MapReduce相关的技术问题。

1. 能力目标：

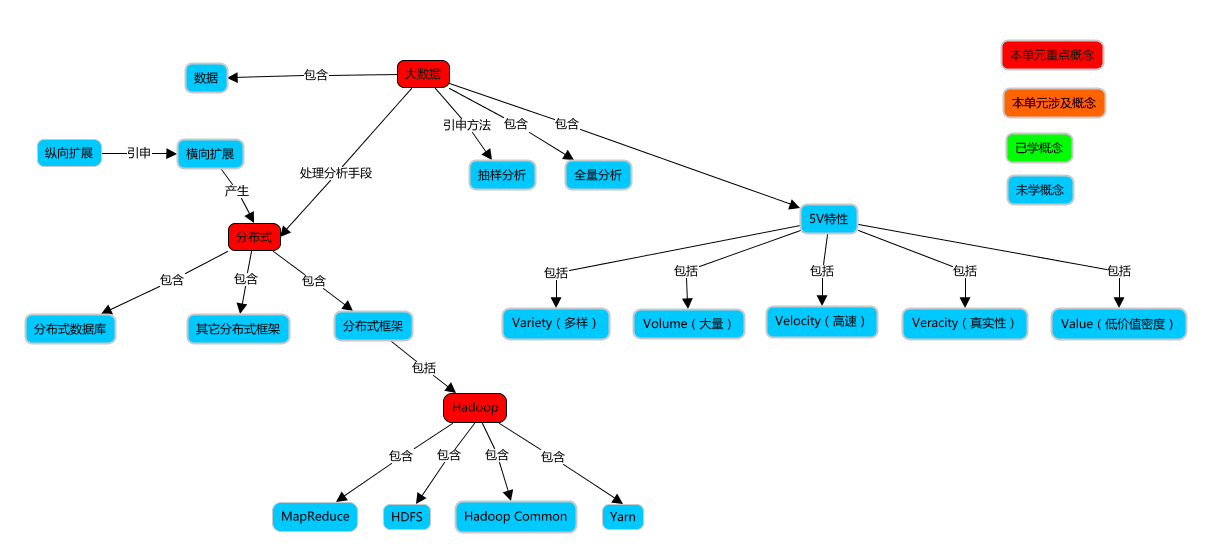
**·** 了解Hadoop的架构、Hadoop的搭建过程；

**·** 解答常见的Hadoop相关概念性问题；

**·** 能够大致了解Hadoop环境搭建和部署的流程，对常见大数据瓶颈问题提出解决方案；

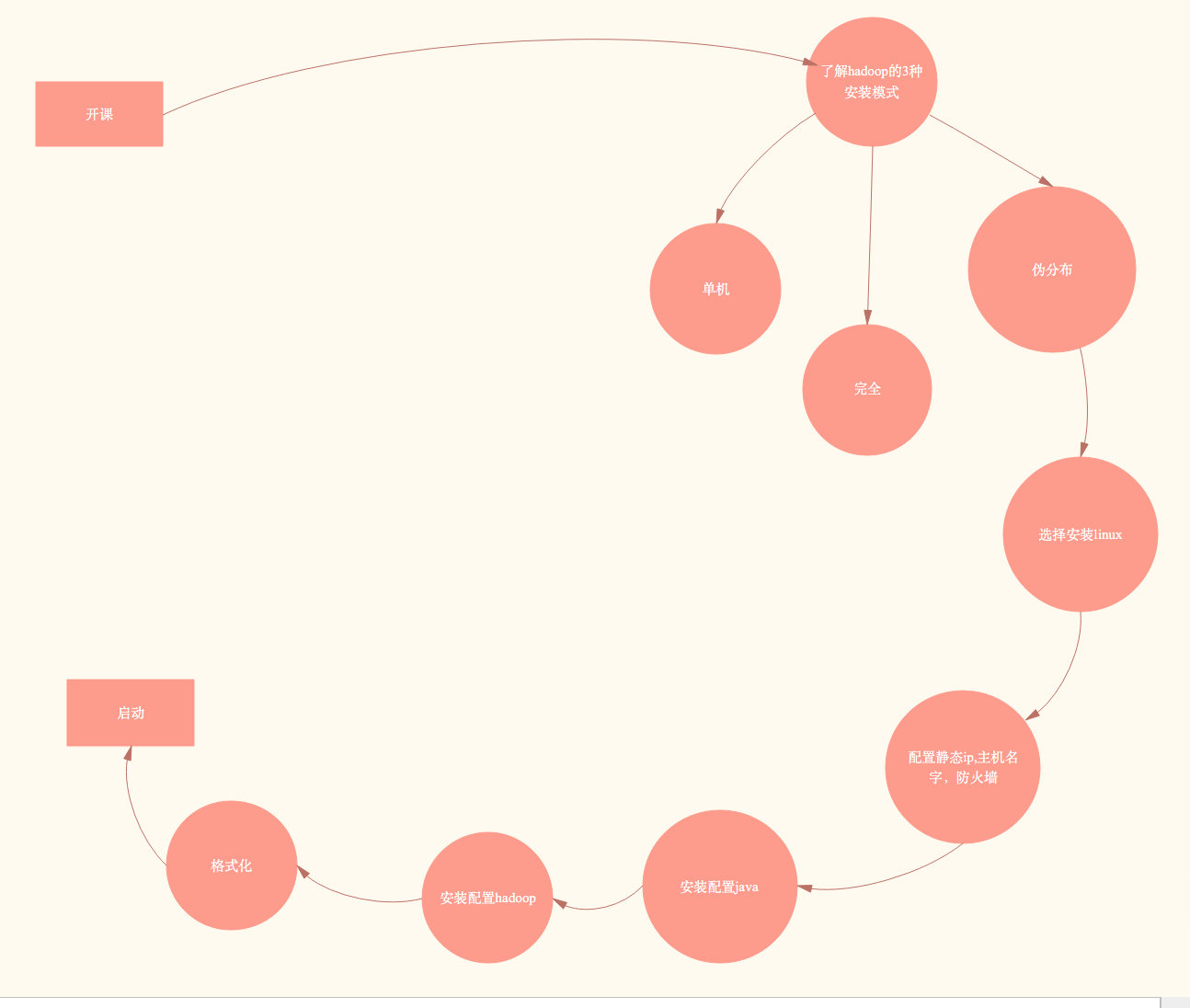
**·** 熟练画出Hadoop生态结构图。

**三、本单元概念图**



**四、教学过程脚本设计—第四课时**

1. **脚本设计思路**



1. **脚本切片设计思路**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **序号** | **内容** | **小节目标** | **设计亮点** | **时间** |
| **1** | **课堂礼仪** | 集中学生的注意力，喊出士气，以饱满的精神状态投入本节课的学习 | 整齐、声音洪亮、铿锵有力，提升精气神！ | 30” |
| **2** | **Hadoop的三种安装模式介绍** | hadoop的三种安装模式的区别； | 通过应用场景了解三种模式； |  |
| **3** | **Hadoop伪分布式安装步骤** | 介绍安装步骤 | 安装步骤概览 |  |
| **4** | **宿主操作系统选择** | linux | Hadoop是为开源而设计的，几乎所有的操作系统都选择了linux的相关版本。 | 2' |
| **5** | **配置静态ip** | 为什么配置静态ip并掌握 | 配置的详细步骤 | 3' |
| **6** | **修改主机名** | 修改机器名字 | 修改机器名字的步骤 |  |
| **7** | **修改主机名与ip的映射关系** | 修改ip映射 | 修改ip映射的步骤 |  |
| **8** | **关闭防火墙** | 关闭防火墙 | 关闭防火墙的位置 |  |
| **9** | **安装配置jdk** | 为什么要配置jdk以及配置步骤并掌握 | Jdk配置 | 2'30" |
| **10** | **安装解压** | 掌握解压安装hadoop | 解压安装的详细过程 | 4'30" |
| **11** | **配置hadoop环境变量** | 掌握如何配置hadoop环境变量 | 配置详细方法 |  |
| **12** | **Hadoop伪分布配置** | 掌握伪分布式的配置文件配置的节点的意义 | 配置详细步骤 | 5' |
| **13** | **SSH免密码** | 为什么配置ssh;  掌握ssh免密码登录的方法 | 免密码登录的方法 | 5' |
| **14** | **Hadoop格式化** | 为什么格式化；  掌握格式化的命令使用方法 | 详细的命令使用操作 | 3' |
| **15** | **hadoop启动** | 掌握hadoop的启动方法  掌握启动hdfs  掌握启动yarn | 掌握hadoop的启动步骤 | 2'30" |
| **16** | **启动日志查看** | 了解如何在命令行查看hadoop的启动日志。 | 查看hadoop的启动日志的方法 |  |

1. **脚本切片一：[课堂礼仪](file:///D:\\llq\\八维\\工作\\教学标准化计划\\切片(总共14片)\\切片1.mp4)**

|  |  |  |  |
| --- | --- | --- | --- |
| **课堂模块** | **教学内容/台词** | **教学标准/素材** | **时长** |
| **课堂礼仪** | 使学生注意力集中并进入上课状态 |  |  |
| **讲师：**上课！①  **班长：**起立！  **讲师：**同学们好！  **学生：**老师好！  **班长：**我们的信念是：  **全体：**知识^很简单，学习^很快乐，潜力^无极限！  **班长：**我们的口号是：  **全体：**辉煌八维大数据，引领时代高科技！人工智能显神奇，高薪就业创佳绩！  **讲师：**请坐！② | 1. 进入教室前整理着装，提前30秒面带自信的笑容的走进教室。，给学生打招呼。当上课铃声响起，喊出“上课”，声音高于授课声音。 2. 细节要求：学生发型整齐、脸部整洁、面带微笑或者礼仪笑（八颗牙齿），严禁大笑，举止合理，避免指手画脚。   IMG_256IMG_256   1. 学生礼仪内容：学习信念，学院口号，班级口号。   讲师授课过程中要面带微笑。   1. 插入合理手势，避免课堂枯燥。   ①动作表情：上课起立手势  ②动作表情：全体请坐手势 | **0:30** |

1. **脚本切片一**：**hadoop的三种安装模式**

|  |  |  |  |
| --- | --- | --- | --- |
| **课堂模块** | **教学内容/台词** | **教学标准/素材** | **时长** |
| **hadoop的三种安装模式** | **单机模式**  **伪分布模式**  **完全分布模式** |  |  |
| 老师：  什么是单机模式？  单机模式在一台单机上运行，没有分布式文件系统，而是直接读写本地[操作系统](https://www.baidu.com/s?wd=%E6%93%8D%E4%BD%9C%E7%B3%BB%E7%BB%9F&tn=SE_PcZhidaonwhc_ngpagmjz&rsv_dl=gh_pc_zhidao" \t "_blank)的文件系统。  什么是伪分布模式？  伪分布模式也是在一台单机上运行，但用不同的Java进程模仿分布式运行中的各类结点(NameNode,DataNode,JobTracker,TaskTracker,SecondaryNameNode)  完全分布式？  Hadoop的各个节点(NameNode,DataNode,JobTracker,TaskTracker,SecondaryNameNode)在多台独立的服务器上运行协作工作 | 单机模式：一般在本地开发用，运行在hadoop上的简单的mapreduce程序用这种方式  伪分布：一般在本地开发运行各种hadoop的程序以及开发操作运行于hadoop上的各种框架程序用这种方式  完全分布式：一般在hadoop测试环境或生产环境采用这种方式  Hadoop测试环境，模拟测试开发工程师开发的各种程序在测试环境下的输入输出是否符合预期，是否符合要求。  Hadoop生产环境下是测试通过后部署到线上的环境。 | 3' |

1. **脚本切片二：[hadoop](file:///D:\\llq\\八维\\工作\\教学标准化计划\\切片(总共14片)\\切片1.mp4)伪分布式安装步骤**

|  |  |  |  |
| --- | --- | --- | --- |
| **课堂模块** | **教学内容/台词** | **教学标准/素材** | **时长** |
| **课堂礼仪** | **了解安装步骤** |  |  |
| 安装步骤如下：  选择安装Linux  配置linux  安装配置java  安装配置hadoop伪分布式  格式化  启动  查看启动节点与启动日志 |  | **0:30** |

1. **脚本切片三：宿主操作系统选择**

|  |  |  |  |
| --- | --- | --- | --- |
| **模块** | **教学内容/台词** | **教学标准/素材** | **时长** |
| **宿主操作系统选择** | 选择安装的操作系统 |  |  |
| 讲师：hadoop是一个框架（软件），要用hadoop，首先得进行安装，安装要安装到操作系统上。那么应该选择什么操作系统那？（老师问）  讲师：操作系统选择一般选择有如下几大类：  ubuntu,centos,redhat  当然其它的linux也可以。咱们选择centos  。 |  | 2' |

1. **脚本切片四：**配置静态ip

|  |  |  |  |
| --- | --- | --- | --- |
| **课堂模块** | **教学内容/台词** | **教学标准/素材** | **时长** |
| **配置静态ip** | 为安装hadoop坐在的linux配置静态ip |  |  |
| 讲师：为什么配置静态ip  因为配置文件要用到ip，如果ip经常变化，那么需要经常去改配置文件，这样很麻烦，修改成静态ip后，ip固定不变了，所以不管什么时候只要配置好了配置文件一次，以后就不用再修改了。 |  | 3' |

1. **脚本切片五：修改主机名**

|  |  |  |  |
| --- | --- | --- | --- |
| **课堂模块** | **教学内容/台词** | **教学标准/素材** | **时长** |
| **修改主机名** | **修改linux主机名** |  |  |
| 教师:如何修改主机名？（问学生）  vi /etc/sysconfig/network | NETWORKING=yes HOSTNAME=hadoop  HOSTNAME后面的内容修改成自己希望要的名字。没有限制，最好是起一个有意义的名字。 | 4'30" |

1. **脚本切片六：修改主机名与ip映射**

|  |  |  |  |
| --- | --- | --- | --- |
| **课堂模块** | **教学内容/台词** | **教学标准/素材** | **时长** |
| **修改主机名与ip映射** | **修改linux主机名** |  |  |
| 教师：为什么配置主机名与ip映射？  因为做完映射后，用主机名就可以指代ip.  如：用ping命令ping主机名,如下：  ping hadoop  教师:如何修改主机名与ip映射？（问学生）  vi /etc/hosts |  | 4'30" |

1. **脚本切片七：关闭防火墙**

|  |  |  |  |
| --- | --- | --- | --- |
| **课堂模块** | **教学内容/台词** | **教学标准/素材** | **时长** |
| **关闭防火墙** | **关闭防火墙的步骤** |  |  |
| 教师：如何关闭防火墙？（问学生）  教师：为什么关闭防火墙？  要是不关闭防火墙，就不能用ftp上传文件了等等。因为防火墙默认将很多端口给禁了。当然可以打开，但是这样太麻烦。所以便于讲课，就建议将防火前关闭。但是在公司部署的话一本不这样做，一般都是通过配置防火墙开放访问端口来做到。 | #查看防火墙的状态  service iptables status  #停止防火墙  service iptables stop  #查看防火墙开机启动状态  chkconfig iptables --list  #关闭防火墙开机启动  chkconfig iptables off | 4'30" |

1. **脚本切片八：安装配置jdk**

|  |  |  |  |
| --- | --- | --- | --- |
| **课堂模块** | **教学内容/台词** | **教学标准/素材** | **时长** |
|  | 安装配置jdk |  |  |
| **安装配置jdk** | 教师：安装hadoop为什么要配置jdk?  因为hadoop是java开发的软件框架，运行要依赖于java环境。  教师：安装jdk的版本应该是多少那？  最好是1.8以上  教师：环境变量在哪里配置？(找个学生回答)  教师：配置完环境文件，怎么是配置生效？（问学生）  教师：source /etc/profile,必须执行完这条命令才能使刚修改的配置文件里的环境变量生效 | jdk-8u77-linux-x64.tar  vi /etc/profile 修改指定的配置文件  cat /etc/profile 查看配置文件的内容  配置内容如下：  JAVA\_HOME=/home/gcs/user/java/jdk1.8.0\_111  PATH=$JAVA\_HOME/bin:$PATH  CLASSPATH=.:$JAVA\_HOME/lib/dt.jar:$JAVA\_HOME/lib/tools.jar  export JAVA\_HOME  export PATH  export CLASSPATH |  |

1. **脚本切片九：安装解压hadoop**

|  |  |  |  |
| --- | --- | --- | --- |
| **课堂模块** | **教学内容/台词** | **教学标准/素材** | **时长** |
| **安装解压hadoop** | **安装解压hadoop** |  |  |
| 教师：解压的方法？（问学生）  教师：解压如下：  首先用ftp软件将hadoop-2.8.1.tar.gz包上传到linux里，然后解压  tar -zxvf hadoop-2.8.1.tar.gz |  | 4'30" |

1. **脚本切片十：配置hadoop环境变量**

|  |  |  |  |
| --- | --- | --- | --- |
| **课堂模块** | **教学内容/台词** | **教学标准/素材** | **时长** |
| **配置hadoop环境变量** | **Hadoop环境变量配置**  **PATH** |  |  |
| 教师：环境变量的配置方式模仿jdk.  配置文件是/etc/profile | export HADOOP\_HOME=/opt/hadoop-2.8.1  export PATH=$PATH:$HADOOP\_HOME/sbin  export PATH=$PATH:$HADOOP\_HOME/bin | 4'30" |

1. **脚本切片十一：配置hadoop配置文件**

|  |  |  |  |
| --- | --- | --- | --- |
| **课堂模块** | **教学内容/台词** | **教学标准/素材** | **时长** |
|  | hadoop-env.sh  core-site.xml  hdfs-site.xml  mapred-site.xml  yarn-site.xml |  |  |
| **配置hadoop配置文件** | **教师：配置hadoop配置文件如下，大家按照如下配置配置自己的hadoop,注意大小写，标记中要无空格，开始结束标记字母要一致，结束标记的斜线不能缺：**  **hadoop-env.sh**  **配置JAVA\_HOME环境变量的值为你的java安装路径。如：**  **export JAVA\_HOME=/opt/jdk1.8.0\_131**  **core-site.xml**  **配置hdfs的访问接口与hadoop**临时目录设定  <property>  <!—hdfs的访问接口-->  <name>fs.defaultFS</name>  <value>**hdfs://hadoop:9000**</value> </property>  <!--hadoop临时目录--> <property>  <name>hadoop.tmp.dir</name>  <value>**/install/hadoop-2.8.0/tmp**</value> </property>  hdfs-site.xml  <!—hdfs中文件块副本的数量-->  <property>  <name>dfs.replication</name>  <value>**1**</value>  </property>  mapred-site.xml  <!--  指定mr运行在yarn上，这个属性有三个取值：  Local:表示本地运行  Classic:表示使用经典的mapreduce框架（hadoop1.0）  Yarn:使用yarn框架（hadoop2.0）  -->  <property>  <name>mapreduce.framework.name</name>  <value>yarn</value>  </property>  yarn-site.xml  <!—指定yarn的老大（ResourceManager）的地址--> <property>  <name>yarn.resourcemanager.hostname</name>  <value>**hadoop**</value> </property>  <!—reducer获取数据的方式--> <property>  <name>yarn.nodemanager.aux-services</name>  <value>mapreduce\_shuffle</value> </property> |  | 5' |

1. **脚本切片十二：格式化hdfs系统**

|  |  |  |  |
| --- | --- | --- | --- |
| **课堂模块** | **教学内容/台词** | **教学标准/素材** | **时长** |
| **格式化hdfs系统** | **格式化命令使用** |  |  |
| 讲师：只有被格式化后，才能使用hdfs文件系统，才能存取文件。 | hadoop namenode -format | 5' |

1. **脚本切片十三：启动停止hadoop**

|  |  |  |  |
| --- | --- | --- | --- |
| **课堂模块** | **教学内容/台词** | **教学标准/素材** | **时长** |
| **启动hadoop** | 启动hdfs  启动yarn  启动所有 |  |  |
| **讲师：格式化完成就可以启动了，启动步骤如下：**  **启动：**  1：启动hadoop方式一  启动hdfs  start-dfs.sh  启动yarn  start-yarn.sh  2：启动hadoop方式二  start-all.sh  3：查看是否启动所有节点进程  Jps  讲师：启动后，要停止，按照如下步骤操作  **停止**  1：停止  Stop-all.sh  2：查看节点进程是否还在  Jps | 启动hadoop的两种方式 | 3' |

1. **脚本切片十四：**查看日志

|  |  |  |  |
| --- | --- | --- | --- |
| **课堂模块** | **教学内容/台词** | **教学标准/素材** | **时长** |
| **查看日志** | 用jps命令查看hadoop的伪分布式节点是否全部启动  查看hadoop每个节点的启动日志，确认启动没问题 |  |  |
| 讲师：查看hadoop启动后的所有进程的方式是用命令jps查看。  讲师：可以通过查看每个节点的启动日志，查看节点是否启动报错。查看方式可以用vi命令打开日志查看  讲师：如果发现某个节点的启动日志报错，请查看具体的配置文件或百度搜索查找相关解决方案。 |  | 3' |

1. **脚本切片十五：知识点总结**

|  |  |  |  |
| --- | --- | --- | --- |
| **课堂模块** | **教学内容/台词** | **教学标准/素材** | **时长** |
| **面试问题** | 使通过面试式提问把课堂知识贯彻。提高学生的面试能力。 |  |  |
| hadoop有几种部署模式？  简要叙述hadoop的几种部署模式的区别。  如何启动hadoop  Hadoop的格式化命令是什么？  Hadoop启动后都有哪些节点？  Hadoop启动后负责整个hadoop的资源管理的节点是什么节点？  如何查看hadoop启动后的所有节点进程?  如何停止hadoop?  免密码的命令是什么？  配置hadoop的环境变量在哪个文件里配置？  要使hadoop配置的环境变量生效，操作的命令是什么？  解压hadoop的tar包的命令是什么？  编辑hadoop配置文件的命令是什么？  配置修改hadoop伪分布式所涉及到的xml文件有哪几个文件？  Hdfs的访问接口在哪个文件里配置？  Hdfs的副本数在哪个文件里配置？ |  | **5:30** |