西京学院教师授课计划

20XX - 20XX学年第 X 学期

课程名称： 大数据技术原理及应用 课程性质： 专业核心课程 编写教师： 徐鲁辉 职称： 副教授

总学时： 48 其中理论学时： 32 实验学时： 16 其它学时： 0 考核方式： 考试

教材名称： Hadoop大数据原理与应用 主编： 徐鲁辉 出版社：西安电子科技大学出版社

适用专业、年级: 数据科学与大数据技术20XX级

要求：1.授课计划由教研室主任（团队负责人）召集相关主讲教师集体研究后编写；

2.每个授课单元（2学时）的内容占一格，表格不够可另加附页；

3.任课教师须严格按授课计划教学，进度提前或推迟超过4学时视为教学事故；

4.“讲课”、“实验”、“其它”的填写，应在相应栏画“√”，如果有其它教学内容必须在备注栏进一步说明；“课程性质”填写公共必修课程、公共选修课程、学科基础与专业核心课程、专业方向课程、专业任选课程、专业实践环节；

5.授课计划经审核后使用，报院系教务科备案，授课教师人手一份。

教研室主任签字: 院（系、中心）负责人签字:

教学单位盖章

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **序号** | **周** | **周课次** | **讲课** | **实验** | **其它** | **教 学 内 容** | | | **备 注**  **（如课堂讨论、习题等）** |
| **章** | **节** | **（章节名称、教学内容提要、实验的名称、课堂讨论的题目）** |
| 1 | 1 | 1 | √ |  |  | 1 | 1.1-1.7 | 开学第一课  1.自我介绍  2.课程地位、先修及后续课程  3.教学目标及知识图谱  4.课时分配（理论+实验）及进度、教材、参考资料、上机软件  5.考核方式：平时+实验+期末  6.雨课堂/云班课使用方法  一、大数据概述  1.大数据内涵、特征、关键技术、产业  2.大数据与物联网、云计算、人工智能、5G的关系  3.大数据岗位介绍  4.大数据学习路线 | 头脑风暴，问卷调查，在线复习并测试Java知识，导入必备Linux基础操作  在线测试1 |
| 2 | 1 | 2 | √ |  |  | 2 | 2.1  2.5 | 二、初识Hadoop  1.Hadoop简介、发展简史、特点、版本  2.Hadoop生态系统  3.Hadoop体系架构  4.Hadoop应用现状  5.部署和运行Hadoop（准备工作） | 在线测试2  预习实验1实验指导书 |
| **序号** | **周** | **周课次** | **讲课** | **实验** | **其它** | **教 学 内 容** | | | **备 注**  **（如课堂讨论、习题等）** |
| **章** | **节** | **（章节名称、教学内容提要、实验的名称、课堂讨论的题目）** |
| 3 | 2 | 1 |  | √ |  | 2 | 2.5 | 实验1 全分布模式下的Hadoop集群部署  1.规划部署。  2.准备机器。  3.准备软件环境：配置静态IP；修改主机名；编辑域名映射；安装和配置Java；安装和配置SSH免密登录。  4.获取和安装Hadoop。  5.配置全分布模式Hadoop集群。  6.关闭防火墙。  7.格式化文件系统。  8.启动和验证Hadoop。  9.关闭Hadoop。 | 实验报告1 |
| 4 | 2 | 2 | √ |  |  | 3 | 3.1-3.4 | 三、分布式文件系统HDFS  1.HDFS简介  2.HDFS体系结构  3.HDFS文件存储机制  4.HDFS数据读写过程 |  |
| 5 | 3 | 1 | √ |  |  | 3 | 3.5-3.6 | 三、Hadoop分布式文件系统HDFS  5.实战HDFS：HDFS Web UI、HDFS Shell、HDFS Java API编程  6.HDFS高可靠性机制 | 在线测试3  预习实验2实验指导书 |
| 6 | 3 | 2 |  | √ |  | 3 | 3.5 | 实验2 实战HDFS | 实验报告2 |
| 7 | 4 | 1 | √ |  |  | 4 | 4.1-4.5 | 四、分布式计算框架MapReduce  1.MapReduce简介  2.第一个MapReduce案例：WordCount  3.MapReduce作业执行流程  4.MapReduce数据类型与格式  5.Shuffle机制 |  |
| 8 | 4 | 2 | √ |  |  | 4 | 4.6-4.9 | 四、分布式计算框架MapReduce  6.在MapReduce中自定义组件  7.实战MapReduce：MapReduce Web UI、MapReduce Shell、MapReduce Java API编程  8.MapReduce调优  9.其他主流计算框架：Spark、Storm、Flink等 | 在线测试4  预习实验3实验指导书 |
| 9 | 5 | 1 |  | √ |  | 4 |  | 实验3 MapReduce编程之一 | 实验报告3 |
| **序号** | **周** | **周课次** | **讲课** | **实验** | **其它** | **教 学 内 容** | | | **备 注**  **（如课堂讨论、习题等）** |
| **章** | **节** | **（章节名称、教学内容提要、实验的名称、课堂讨论的题目）** |
| 10 | 5 | 2 | √ |  |  | 5 | 5.1-5.4 | 五、统一资源管理和调度框架YARN  1.初识YARN  2.YARN体系架构  3.YARN工作流程  4.实战YARN：YARN Web UI、YARN Shell、YARN Java API编程 |  |
| 11 | 6 | 1 | √ |  |  | 5 | 5.5-5.6 | 五、统一资源管理和调度框架YARN  5.YARN新特性  6.其它统一资源管理调度框架：Apache Mesos、Hadoop Corona、Google Borg/Omega/Kubernetes、Docker Swarm | 在线测试5  预习实验3实验指导书 |
| 12 | 6 | 2 |  | √ |  | 4 |  | 实验3 MapReduce编程之二 | 实验报告3 |
| 13 | 7 | 1 | √ |  |  | 6 | 6.1-6.4 | 六、分布式协调框架ZooKeeper  1.分布式协调技术  2.初识ZooKeeper  3.ZooKeeper系统模型  4.ZooKeeper工作原理 |  |
| 14 | 7 | 2 | √ |  |  | 6 | 6.5-6.7 | 六、分布式协调框架ZooKeeper  5.ZooKeeper典型应用场景  6.部署ZooKeeper集群（准备工作）  7.实战ZooKeeper：ZooKeeper四字命令、ZooKeeper Shell、ZooKeeper Java API编程 | 在线测试6  预习实验4实验指导书 |
| 15 | 8 | 1 |  | √ |  | 6 |  | 实验4 部署ZooKeeper集群和实战ZooKeeper | 实验报告4 |
| 16 | 8 | 2 | √ |  |  | 7 | 7.1-7.5 | 七、分布式数据库HBase  1.NoSQL简介  2.初识HBase  3.HBase数据模型  4.HBase体系架构  5.HBase运行机制 |  |
| 17 | 9 | 1 | √ |  |  | 7 | 7.6-7.8 | 七、分布式数据库HBase  6.部署HBase集群（准备工作）  7.实战HBase：HBase Web UI、HBase Shell、HBase Java API编程、在HBase中使用MapReduce  8.HBase性能优化 | 在线测试7  预习实验5实验指导书 |
| **序号** | **周** | **周课次** | **讲课** | **实验** | **其它** | **教 学 内 容** | | | **备 注**  **（如课堂讨论、习题等）** |
| **章** | **节** | **（章节名称、教学内容提要、实验的名称、课堂讨论的题目）** |
| 18 | 9 | 2 |  | √ |  | 7 |  | 实验5 部署全分布模式HBase集群和实战HBase | 实验报告5 |
| 19 | 10 | 1 | √ |  |  | 8 | 8.1-8.5 | 八、数据仓库Hive  1.初识Hive  2.Hive体系架构  3.Hive数据类型  4.Hive文件格式  5.Hive数据模型 |  |
| 20 | 10 | 2 | √ |  |  | 8 | 8.6-8.9 | 八、数据仓库Hive  6.Hive函数  7.部署Hive（准备工作）  8.实战Hive：HWI、Hive Shell、Hive Java API编程  9.Hive优化策略 | 在线测试8  预习实验6实验指导书 |
| 21 | 11 | 1 |  | √ |  | 8 |  | 实验6 部署Hive和实战Hive | 实验报告6 |
| 22 | 11 | 2 | √ |  |  | 9 | 9.1-9.4 | 九、大数据迁移和采集工具  1.数据迁移工具Sqoop  2.日志采集工具Flume  3.分布式流平台Kafka  4.ETL工具Kettle | 在线测试9 |
| 23 | 12 | 1 | √ |  |  | 10 | 10.1-10.4 | 十、数据可视化  1.可视化概念  2.可视化作用与意义  3.可视化图表类型  4.可视化工具：ECharts、Python、Tableau、阿里云DataX、D3.js | 在线测试10 |
| 24 | 12 | 2 |  |  | √ |  |  | 期末复习 |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |