|  |  |  |  |
| --- | --- | --- | --- |
| **模块** | **任务** | **主要知识与技能点** | **分值** |
| 模块A:大数据平台搭建（容器环境） | 任务一：Hadoop伪分布式安装配置 | Hadoop伪分布式下的JDK的解压安装、JDK环境变量配置、Hadoop配置文件修改、运行测试等 | 7 |
| 任务二：Flume安装配置 | Flume的解压安装、环境变量配置、配置文件修改、运行测试等 | 4 |
| 任务三：Hive安装配置 | Hive的解压安装、环境变量配置、运行测试等 | 4 |
| 小计 | | 15 |
| 模块B:离线数据处理 | 任务一：数据抽取 | 利用Sqoop,从MySQL中进行离线数据抽取，包括全量数据抽取和增量数据抽取操作 | 10 |
| 任务二：数据清洗 | 利用Spark及MR,从ods到dwd的数据清洗，包括数据排序、去重、数据类型转换等操作 | 8 |
| 任务三：指标计算 | 利用Hive SQL及Spark,在dwd及dws层对数据进行相关数据指标的统计、计算等操作 | 7 |
| 小计 | | 25 |
| 模块C:数据挖掘 | 任务一：特征工程 | 对工业数据集进行特征提取及数据预处理等操作 | 5 |
| 任务二：报警预测 | 基于随机森林的设备报警预测的设计开发操作 | 5 |
| 小计 | | 10 |
| 模块D:数据采集与实时计算 | 任务一：实时数据采集 | 基于Flume和Kafka的实时数据采集，包括Flume采集文件及配置、数据注入Kafka等操作 | 8 |
| 任务二：使用Flink处理Kafka中的数据 | 使用Flink消费Kafka中的数据进行实时计算，包括实时数据统计计算、Redis基本操作、Kafka基本操作等 | 12 |
| 小计 | | 20 |
| 模块E：数据可视化 | 任务一：用饼状图展示某设备当天各状态时间 | 正确使用Vue.js框架，结合Echarts绘制饼状图 | 2 |
| 任务二：用柱状图展示各车间设备每月平均运行时长 | 正确使用Vue.js框架，结合Echarts绘制柱状图 | 2 |
| 任务三：用折线图展示机器运行时长 | 正确使用Vue.js框架，结合Echarts绘制折线图 | 2 |
| 任务四：用散点图展示PM2.5浓度变化 | 正确使用Vue.js框架，结合Echarts绘制散点图 | 3 |
| 任务五：用条形图展示各机器每日产量 | 正确使用Vue.js框架，结合Echarts绘制条形图 | 3 |
| 任务六：用折柱混合图展示机器日均产量和车间日均产量 | 正确使用Vue.js框架，结合Echarts绘制折柱混合图 | 3 |
| 小计 | | 15 |
| 模块F:综合分析 | 任务一：HDFS存储小文件问题 | 正确分析HDFS存储小文件问题 | 4 |
| 任务二：如何解决数据倾斜问题 | 正确分析如何解决数据倾斜问题 | 3 |
| 任务三：简要描述任务过程中的问题并进行总结 | 合理描述任务过程中的问题并进行总结 | 3 |
| 小计 | | 10 |
| 模块G:职业素养 | 考察职业素养 | 竞赛团队分工明确合理、操作规范、文明竞赛 | 5 |
| 小计 | | 5 |
| 总分 | | | 100 |