



永州职业技术学院
Yongzhou Vocational Technical College

永州职业技术学院 学生专业技能考核标准

| | |
|-------|---------|
| 专业代码: | 510201 |
| | 计算机应用技术 |
| 专业名称: | (大数据应用) |
| 二级学院: | 信息工程学院 |

永州职业技术学院
2021年8月

目 录

| | |
|-------------------------------|----|
| 一、专业名称 | 1 |
| 1. 专业名称: | 1 |
| 2. 适用对象 | 1 |
| 二、考核目标 | 1 |
| 三、考核内容 | 1 |
| 模块一、大数据开发基础 | 1 |
| 项目 1: 大数据编程基础 | 1 |
| 项目 2: 大数据平台部署与基础开发 | 3 |
| 模块二、数据采集与存储 | 3 |
| 项目 1: 基于 Flume 的数据采集 | 3 |
| 项目 2: 基于 Kafka 的消息队列数据采集 | 4 |
| 项目 3: 基于 HBase 的非结构化数据存储 | 5 |
| 模块三、数据清洗与分析 | 5 |
| 项目 1: 基于 pandas 的数据清洗 | 5 |
| 项目 2: 基于 kettle 的数据清洗 | 6 |
| 项目 3: 基于 hive 的数据分析 | 7 |
| 项目 4: Spark 大数据处理与分析 | 7 |
| 模块四、数据可视化 | 8 |
| 项目 1: 基于 matplotlib 的数据分析和可视化 | 8 |
| 项目 2: 基于 pyecharts 的数据可视化 | 9 |
| 四、评价标准 | 10 |
| 五、考核方式 | 19 |
| 六、附录 | 19 |

一、专业名称

1. 专业名称

计算机应用技术（大数据应用）（专业代码：510201）

2. 适用对象

高职全日制在籍毕业年级学生 2021 级

二、考核目标

设置大数据开发基础模块、数据采集与存储模块、数据清洗与分析模块、数据可视化模块等 4 个技能考核模块。测试学生的编程能力、数据采集能力、大数据存储能力、数据清洗能力、数据分析能力、数据可视化能力、项目管理能力，以及从事大数据技术工作的团队协作、成本控制、质量效益、安全规范等职业素养。学校加强专业教学基本条件建设，深化课程教学改革，强化实践教学环节，增强学生创新创业能力，促进学生个性化发展，提高专业教学质量和专业办学水平，培养适应信息时代发展需要的大数据技术高素质技术技能人才。

三、考核内容

根据本专业的人才培养目标和实际工作内容，本专业技能考核分专业基本技能模块和岗位核心技能模块。测试范围包括：用 Python 或 Java 语言编程解决实际问题；MySQL 数据库开发；hadoop 完全分布式平台部署；使用 HDFS 对 hadoop 平台进行基础开发；以大数据技术 flume、kafka 实现数据采集，其中 flume 技术采集端口数据、文件数据、变化数据，kafka 采集实时的消息；以 HBase 数据库实现非结构化数据存储；用 pandas 对数据进行清洗与分析；使用 kettle 实现数据 ETL 操作；hive 数据分析、spark 数据分析，实现批量与实时数据的清洗处理与分析；以 Python、Pycharm 为工具，实现 matplotlib、pyecharts 数据可视化操作；从事大数据技术工作的团队协作，安全规范等职业素养。

模块一、大数据开发基础

项目 1：编程基础

1. 技能要求

(1) 掌握 MySQL 数据定义语言 (DDL) 的开发工作, 包括创建数据库、数据表;

(2) 掌握 MySQL 数据操作语言 (DML) 的开发工作, 包括插入数据到表;

(3) 掌握 MySQL 数据查询语言 (DQL) 的开发工作;

(5) 掌握基本的开发技能、排错技能。

2. 素养要求

(1) 具有较强的分析问题和解决问题的能力;

(2) 具备程序员的严谨细致、规范的工作态度和正确的价值观;

(3) 体现良好的工作习惯, 设计文件的准备和有效性确认、将作品存放到特定的位置以及为文件命名、随时保存设计文档等。在工作区域不可有任何与工作无关的物品。设计任务完成后整理工作台, 保持工作台面干净整洁; 凳子放回原位, 按顺序退出考场;

(4) 在测试时间内完成任务, 体现良好的时间管理能力。

项目 2: 数据库基础

1. 技能要求

(1) 能熟练使用主流的软件开发平台以及进行相关参数的配置; 能使用软件开发平台进行项目的创建、开发、编译、运行及调试;

(2) 对软件开发平台故障分析判断准确、突发情况处理及时;

(3) 掌握 Python、Java 的基本语法知识, 主要包括语法规则、注释、输入和输出等内容;

(4) 能使用 Python、Java 的数据类型、变量、常量、运算符、表达式、函数、并结合顺序、分支、循环三种控制结构实现项目需求的具体功能;

(5) 掌握基本的编程技能, 设计算法、程序设计、排错以及 Python 软件包查找、使用技能。

2. 素养要求

(1) 具有较强的分析问题和解决问题的能力;

(2) 具有良好的编程习惯;

(3) 具备利用程序语言去建模、解决将来实际工作、日常生活等方面的能力;

(4) 具备程序员的严谨细致、规范的工作态度和正确的价值观;

(5) 体现良好的工作习惯,设计文件的准备和有效性确认、将作品存放到特定的位置以及为文件命名、随时保存设计文档等。在工作区域不可有任何与工作无关的物品。设计任务完成后整理工作台,保持工作台面干净整洁;凳子放回原位,按顺序退出考场;

(6) 在测试时间内完成任务,体现良好的时间管理能力。

项目 3: 大数据平台部署与基础开发

1. 技能要求

- (1) 能熟练安装 Hadoop 大数据平台及进行相关参数的配置;
- (2) 能熟练启动 Hadoop 平台,并进行 web 端访问;
- (3) 能熟练掌握使用 hdfs 命令进行 Hadoop 文件操作;
- (4) 能熟练掌握使用 MapReduce 对数据进行处理;
- (5) 能熟练使用 YARN 命令查看 Hadoop 集群状态;
- (6) 能迅速定位 Hadoop 大数据分析项目运行过程中的问题并解决;
- (7) 对数据 Hadoop 数据分析平台故障分析判断准确、突发情况处理及时。

2. 素养要求

- (1) 代码编写格式规范、变量命名规范,注释规范;
- (2) 具有较强的分析问题和解决问题的能力;
- (3) 具有较强的知识扩展能力;
- (4) 能够遵守电子数据的收集提取等相关法律法规;
- (5) 有高度的敬业精神及工作激情,工作态度积极乐观;
- (6) 在测试时间内完成任务,体现良好的时间管理能力。

模块二、数据采集与存储

项目 1: 基于 Flume 的数据采集

1. 技能要求

- (1) 能熟练使用常用的专业数据采集工具;
- (2) 能熟练掌握数据采集工具的场景适用性;
- (3) 能使用数据采集框架进行项目的创建、开发、配置、运行及调试;
- (4) 了解 Flume 的体系结构、能够熟练掌握 Event、Client 和 Agent 等 Flume

的核心组件；

(5)掌握多种 source、多种 sink 和 channel 的组合,形成不同类型的 agent,采集不同类别的数据；

(6)能熟练掌握 Flume agent 的启动命令,掌握数据采集的过程；

(7)对数据采集框架平台故障分析判断准确、突发情况处理及时。

2. 素养要求

(1)代码编写格式规范、变量命名规范,注释规范；

(2)具有较强的分析问题和解决问题的能力；

(3)具有较强的知识扩展能力；

(4)能够遵守电子数据的收集提取等相关法律法规；

(5)有高度的敬业精神及工作激情,工作态度积极乐观；

(6)在测试时间内完成任务,体现良好的时间管理能力。

项目 2: 基于 Kafka 的消息队列数据采集

1. 技能要求

(1)能熟练根据项目描述特性选择合适的 Flume source 组件；

(2)能熟练掌握 Flume 各组件的配置文件编写；

(3)能熟练掌握数据采集框架中,Kafka 组件的配置和设计；

(4)能熟练启动 kafka 的服务端和命令端；

(5)能熟练掌握 Flume agent 的启动命令和使用 Kafka 相关的脚本命令；

(6)能熟练掌握 Kafka 的常用操作,如扩容,删除和增加 topic；

(7)能熟练掌握数据采集的过程；

(8)能迅速定位数据采集项目运行过程中的问题并解决。

2. 素养要求

(1)具备数据采集技术解决实际应用中不同类型数据源的收集能力；

(2)具有良好的编程习惯,较强的分析问题的能力和良好的发散思维及创新能力；

(3)具有较强的逻辑思维能力和综合运用知识的能力；

(4)能够遵守电子数据的收集提取等相关法律法规；

(5)具备程序员的严谨细致、规范的工作态度和正确的价值观；

(6) 在测试时间内完成任务，体现良好的时间管理能力。

项目 3：基于 HBase 的非结构化数据存储

1. 技能要求

(1) 能理解非结构化数据的概念，熟练掌握 HBase 的启动，利用 HBase 存储非结构化数据。

(2) 能熟练掌握利用 HBase 数据定义语言，创建命名空间、表；

(3) 能熟练掌握 HBase 数据操作语言，存储非结构化数据；

(4) 能熟练掌握 HBase 数据查询语言，对非结构化数据进行查询、分析操作；

(5) 能熟练掌握 HBase 过滤器操作，对非结构化数据进行过滤操作；

(6) 能迅速定位非结构化数据存储项目运行过程中的问题并解决。

2. 素养要求

(1) 具备掌握非结构化数据存储技术，解决实际应用中不同类型非结构化数据源的存储能力；

(2) 具有良好的编程习惯，较强的分析问题的能力和良好的发散思维及创新能力；

(3) 具有较强的逻辑思维能力和综合运用知识的能力；

(4) 能够遵守电子数据的收集提取等相关法律法规；

(5) 具备程序员的严谨细致、规范的工作态度和正确的价值观；

(6) 在测试时间内完成任务，体现良好的时间管理能力。

模块三、数据清洗与分析

项目 1：基于 pandas 的数据清洗

1. 技能要求

(1) 能熟练使用 pandas 读写数据文件，保存为 DataFrame 数据；

(2) 能熟练使用 pandas 统计数据文件的属性、数据分布情况；

(3) 能熟练使用 pandas 分析数据的分布规律；

- (4) 能熟练利用 pandas 检测数据中的脏数据；
- (5) 能熟练利用 pandas 清洗数据中的脏数据；
- (6) 能定位和解决数据清洗项目运行过程中的异常。

2. 素养要求

- (1) 具备利用数据清洗技术解决实际问题中不同类型数据的清洗能力；
- (2) 具有良好的编程习惯，较强的分析问题的能力和良好的发散思维及创新能力；
- (3) 遵守职业道德和法律法规等相关知识产权，保护数据，不盗取数据；
- (4) 有较强的需求理解、整体设计、功能实现能力；
- (5) 具有较强的逻辑思维能力和综合运用知识的能力；
- (6) 在测试时间内完成任务，体现良好的时间管理能力。

项目 2：基于 kettle 的数据清洗

1. 技能要求

- (1) 能熟练安装 kettle 工具及进行相关软件参数的配置；
- (2) 能分析项目描述选择合适的 kettle 步骤组件，搭建完整的转换；
- (3) 能熟练掌握配置 kettle 步骤组件的参数，使其完成相应的清洗或校验功能；
- (4) 能熟练掌握数据在 kettle 中的导入与导出；
- (5) 能熟练掌握使用 kettle 清洗和分析 csv 数据；
- (6) 能定位和解决数据清洗项目运行过程中的异常。

2. 素养要求

- (1) 具备利用数据清洗技术解决实际问题中不同类型数据的清洗能力；
- (2) 具有良好的转换搭建习惯，较强的分析问题的能力和良好的发散思维及创新能力；
- (3) 遵守职业道德和法律法规等相关知识产权，保护数据，不盗取数据；
- (4) 有较强的需求理解、整体设计、功能实现能力；
- (5) 具有较强的逻辑思维能力和综合运用知识的能力；
- (6) 在测试时间内完成任务，体现良好的时间管理能力。

项目 3：基于 hive 的数据处理与分析

1. 技能要求

- (1) 能熟练掌握 hive 的启动；
- (2) 能理解内、外部表的概率，熟练使用 Hive 建立外部表；
- (3) 能熟练使用 hive HQL 对表进行数据关联，管理数据；
- (4) 能熟练使用 hive HQL 对数据进行分析，统计数据的分布情况；
- (5) 能定位和解决项目运行过程中的异常。

2. 素养要求

- (1) 具备分析数据，搭建数据仓库的能力
- (2) 具有良好的编程习惯，较强的分析问题的能力和良好的发散思维及创新能力；
- (3) 遵守职业道德和法律法规等相关知识产权，保护数据，不盗取数据；
- (4) 有较强的需求理解、整体设计、功能实现能力；
- (5) 具有较强的逻辑思维能力和综合运用知识的能力；
- (6) 在测试时间内完成任务，体现良好的时间管理能力。

项目 4：Spark 大数据处理与分析

1. 技能要求

- (1) 能熟练安装 Spark 大数据平台及进行相关参数的配置。
- (2) 能熟练掌握数据文件读取并转为 RDD 的操作。
- (3) 能熟练掌握在 IDEA 中开发 Spark 的编写能力。
- (4) 能熟练掌握 SparkRDD 数据处理的基础函数操作和进阶函数操作。
- (5) 能熟练掌握 Spark Shell、SparkSQL 的命令行常用操作。
- (6) 能熟练掌握从 MySQL 数据库中读取写入数据的命令行操作。
- (7) 能熟练掌握数据文件转 DataFrame，RDD 转 DataFrame 的操作。
- (8) 能熟练掌握 SparkSQL 中 SQL 语言和 DSL 语言开发。
- (9) 能迅速定位 Spark 大数据分析项目运行过程中的问题并解决。

2. 素养要求

- (1) 具备 Spark 大数据处理与分析解决实际问题的能力；

- (2) 具有良好的编程习惯，较强的分析问题的能力和良好的发散思维及创新能力；
- (3) 具有较强的逻辑思维能力和综合运用知识的能力；
- (4) 能够遵守大数据分析处理等相关法律法规；
- (5) 具备程序员的严谨细致、规范的工作态度和正确的价值观。

模块四、数据可视化

项目 1: matplotlib 数据可视化

1. 技能要求

- (1) 能熟练完成 matplotlib 绘图工具包和 pandas、numpy 模块的安装与导入；
- (2) 能熟练掌握 pyplot 常用绘图参数的调节方法；
- (3) 能熟练掌握子图的绘制方法；
- (4) 能熟练掌握绘制图形的保存与展示方法；
- (5) 能熟练掌握散点图的作用与绘制方法；
- (6) 能熟练掌握折线图的作用与绘制方法；
- (7) 能熟练掌握纵向柱状图、横向柱状图的作用与绘制方法；
- (8) 能熟练掌握饼图的作用与绘制方法；
- (9) 能熟练掌握 pandas 中 Series 对象的创建，Series 常用属性的访问，Series 数据的基本操作；
- (10) 能熟练掌握 pandas 中 Data Frame 对象的创建，Data Frame 常用属性的访问，Data Frame 数据的基本操作；
- (11) 能熟练使用 pandas 内置函数读写 CSV 等文件；
- (12) 能定位和排除图表生成过程中的异常。

2. 素养要求

- (1) 具备利用可视化技术解决实际问题中不同类型数据展示的能力；
- (2) 具备良好的数据操作能力、数据统计分析和理解能力；
- (3) 具有良好的视觉审美能力、创新设计理念；

(4) 具有良好的编程习惯，较强的分析问题的能力和良好的发散思维及创新能力；

(5) 遵守职业道德和法律法规等相关知识产权，保护数据，不盗取数据；

(6) 在测试时间内完成任务，体现良好的时间管理能力。

项目 2：pyecharts 数据可视化

1. 技能要求

(1) 能熟练完成 pyecharts 绘图工具包和 pandas、numpy 的安装与导入；

(2) 能熟练掌握图形参数的配置；

(3) 能熟练掌握全局配置项和系列配置项的设置；

(4) 能熟练掌握绘制图形的保存方法；

(5) 能熟练掌握散点图的作用与绘制方法；

(6) 能熟练掌握折线图的作用与绘制方法；

(7) 能熟练掌握柱状图、横向柱状图、双柱状图的作用与绘制方法；

(8) 能熟练掌握 pandas 中 Series 对象的创建，Series 常用属性的访问，Series 数据的基本操作；

(9) 能熟练掌握 pandas 中 Data Frame 对象的创建，Data Frame 常用属性的访问，Data Frame 数据的基本操作；

(10) 能熟练使用 pandas 内置函数读写 CSV、Excel 等文件；

(11) 能定位和排除图表生成过程中的异常。

2. 素养要求

(1) 具备利用可视化技术解决实际问题中不同类型数据展示的能力；

(2) 具备良好的数据分析和理解能力；

(3) 具有良好的视觉审美能力、创新设计理念；

(4) 具有良好的编程习惯，较强的分析问题的能力和良好的发散思维及创新能力；

(5) 遵守职业道德和法律法规等相关知识产权，保护数据，不盗取数据；

(6) 在测试时间内完成任务，体现良好的时间管理能力。

四、评价标准

本专业技能考核采取过程考核与结果考核相结合，技能考核与职业素养考核相结合。根据考生操作的规范性、熟练程度和用时量等因素评价过程成绩；根据设计作品、运行测试结果和提交文档质量等因素评价结果成绩。

本专业技能考核满分为 100 分，其中专业技能占 80 分，职业素养占 20%（其中专业素养和道德规范各占 10 分）。

根据模块中考核项目的不同，重点考核学生对该项目所必须掌握的技能和要求。虽然不同考试题目的技能侧重点有所不同，但完成任务的工作量和难易程度基本相同。各模块和项目的技能评价要点内容如表 1 所示。

表1 大数据应用专业技能考核评价要点

| 类别 | 模块 | 项目 | 评价要点 |
|--------|-------------|------------|--|
| 专业基本技能 | 模块一、大数据开发基础 | 项目 1: 编程基础 | 专业技能： 1. 正确配置软件开发环境； 2. 开发环境配置过程符合平台搭建与运维的职业规范； 3. 项目的设计步骤清晰，方法科学合理，设计过程符合职业规范； 4. 正确使用 Python、Java 的基本语法知识； 5. 正确定义变量、常量、名称符合命名规范； 6. 正确使用运算符、表达式和函数进行编程； 7. 正确使用顺序、分支和循环三种控制结构实现项目需求的具体功能； 8. 正确使用数组和列表等基本数据结构进行编程； 9. 正确将数组和列表等基本数据结构的查找、排序等算法应用到代码的编写中，实现项目性能的提升。 10. 掌握基本的编程技能，如设计算法、程序设计、排错以及 Python 软件包查找、使用技能； |

| | | | |
|--|--|--------------------|--|
| | | | <p>11. 正确使用文件实现数据的输入和输出、存储和读取；</p> <p>12. 正确将程序设计思维运用到项目设计中，有效降低代码的冗余度，提高代码的复用性。</p> <p>专业素养和道德规范：</p> <ol style="list-style-type: none"> 1. 严格遵守开发平台搭建与运维的工作规范，硬件设备操作符合电子设备安全操作规范； 2. 具有较强的分析问题和解决问题的能力； 3. 具有良好的编程习惯； 4. 具备利用程序语言去建模、解决将来实际工作、日常生活等方面的能力； 5. 具备程序员的严谨细致、规范的工作态度和正确的价值观。 |
| | | 项目 2: 大数据平台部署与基础开发 | <p>专业技能：</p> <ol style="list-style-type: none"> 1. 正确安装 Hadoop 大数据平台及进行相关参数的配置； 2. 正确启动 Hadoop 平台，并进行 web 端访问； 3. 正确使用 hdfs 命令进行 Hadoop 文件操作； 4. 正确使用 MapReduce 对数据进行处理； 5. 正确使用 YARN 命令查看 Hadoop 集群状态； 6. 正确定位 Hadoop 大数据分析项目运行过程中的问题并解决； 7. 正确解决 Hadoop 环境的运维问题； 8. 正确完成答案与结果的保存。 <p>专业素养和道德规范：</p> <ol style="list-style-type: none"> 1. 代码编写格式规范、变量命名规范，注释规范； 2. 具有较强的分析问题和解决问题的能力； 3. 具有较强的知识扩展能力； 4. 能够遵守电子数据的收集提取等相关法律法规； |

| | | | |
|--------|-------------|--------------------------|--|
| | | | <ol style="list-style-type: none"> 5. 有高度的敬业精神及工作激情，工作态度积极乐观； 6. 在测试时间内完成任务，体现良好的时间管理能力。 |
| 岗位核心技能 | 模块二、数据采集与存储 | 项目 1: 基于 Flume 的数据采集 | <p>专业技能：</p> <ol style="list-style-type: none"> 1. 正确使用数据采集框架进行项目的创建、开发、配置、运行及调试； 2. 正确使用多种 source、多种 sink 和 channel 的组合，形成不同类型的 agent，采集不同类别的数据； 3. 正确掌握 Flume agent 的启动命令，掌握数据采集的过程； 4. 对数据采集框架平台故障分析判断准确、突发情况处理及时。 5. 正确完成答案与结果的保存。 <p>专业素养和道德规范：</p> <ol style="list-style-type: none"> 1. 代码编写格式规范、变量命名规范，注释规范； 2. 具有较强的分析问题和解决问题的能力； 3. 具有较强的知识扩展能力； 4. 能够遵守电子数据的收集提取等相关法律法规； 5. 有高度的敬业精神及工作激情，工作态度积极乐观； 6. 在测试时间内完成任务，体现良好的时间管理能力。 |
| | | 项目 2: 基于 kafka 的消息队列数据采集 | <p>专业技能：</p> <ol style="list-style-type: none"> 1. 正确根据项目描述特性选择合适的 Flume source 组件； 2. 正确配置 Flume 各组件的文件； 3. 正确掌握数据采集框架中，Kafka 组件的配置和 |
| | | | |

| | | | |
|--|---------------------------------|---|--|
| | | <p>设计；</p> <ol style="list-style-type: none"> 4. 正确启动 kafka 的服务端和命令端； 5. 正确掌握 Flume agent 的启动命令和使用 Kafka 相关的脚本命令； 6. 正确掌握 Kafka 的常用操作，如扩容，删除和增加 topic； 7. 正确掌握数据采集的过程； 8. 正确定位数据采集项目运行过程中的问题并解决。 | <p>专业素养和道德规范：</p> <ol style="list-style-type: none"> 1. 具备数据采集技术解决实际问题中不同类型数据源的收集能力； 2. 具有良好的编程习惯，较强的分析问题的能力和良好的发散思维及创新能力； 3. 具有较强的逻辑思维能力和综合运用知识的能力； 4. 能够遵守电子数据的收集提取等相关法律法规； 5. 具备程序员的严谨细致、规范的工作态度和正确的价值观； 6. 在测试时间内完成任务，体现良好的时间管理能力。 |
| | <p>项目 3: 基于 HBase 的非结构化数据存储</p> | | <p>专业技能：</p> <ol style="list-style-type: none"> 1. 正确启动 HBase。 2. 正确使用 HBase 数据定义语言，创建命名空间、表； 3. 正确使用 HBase 数据操作语言，存储非结构化数据； 4. 正确使用 HBase 数据查询语言，对非结构化数据进行查询、分析操作； |

| | | | |
|--|--|-----------------------|--|
| | | | <ol style="list-style-type: none"> 5. 正确使用 HBase 过滤器操作，对非结构化数据进行过滤操作； 6. 能迅速定位非结构化数据存储项目运行过程中的问题并解决。 <p>专业素养和道德规范：</p> <ol style="list-style-type: none"> 1. 具备掌握非结构化数据存储技术，解决实际应用中不同类型非结构化数据源的存储能力； 2. 具有良好的编程习惯，较强的分析问题的能力和良好的发散思维及创新能力； 3. 具有较强的逻辑思维能力和综合运用知识的能力； 4. 能够遵守电子数据的收集提取等相关法律法规； 5. 具备程序员的严谨细致、规范的工作态度和正确的价值观； 6. 在测试时间内完成任务，体现良好的时间管理能力。 |
| | | 项目 1: 基于 pandas 的数据清洗 | <p>专业技能：</p> <ol style="list-style-type: none"> 1. 正确使用 pandas 读写数据文件，保存为 DataFrame 数据； 2. 正确使用 pandas 统计数据文件的属性、数据分布情况； 3. 正确使用 pandas 分析数据的分布规律； 4. 正确利用 pandas 检测数据中的脏数据； 5. 正确利用 pandas 清洗数据中的脏数据； 6. 正确定位和解决数据清洗项目运行过程中的异常。 <p>专业素养和道德规范：</p> <ol style="list-style-type: none"> 1. 具备利用数据清洗技术解决实际应用中不同类型数据的清洗能力； |

| | | | |
|--|--|--|---|
| | <p>模块 三、数 据清 洗与 分析</p> | | <ol style="list-style-type: none"> 2. 具有良好的编程习惯，较强的分析问题的能力和良好的发散思维及创新能力； 3. 遵守职业道德和法律法规等相关知识产权，保护数据，不盗取数据； 4. 有较强的需求理解、整体设计、功能实现能力； 5. 具有较强的逻辑思维能力和综合运用知识的能力； 6. 在测试时间内完成任务，体现良好的时间管理能力。 |
| | | <p>项目 2: 基 于 kettle 的数据清 洗</p> | <p>专业技能：</p> <ol style="list-style-type: none"> 1. 正确分析项目，选择合适的 kettle 步骤组件； 2. 正确搭建完整的转换； 3. 正确配置 kettle 步骤组件的参数，使其完成相应的清洗或校验功能； 4. 正确完成数据在 kettle 中的导入与导出； 5. 正确使用 kettle 清洗和分析 csv 数据； 6. 正确定位和解决数据清洗项目运行过程中的异常。 |
| | | | <p>专业素养和道德规范：</p> <ol style="list-style-type: none"> 1. 具备利用数据清洗技术解决实际应用中不同类型数据的清洗能力； 2. 具有良好的转换搭建习惯，较强的分析问题的能力和良好的发散思维及创新能力； 3. 遵守职业道德和法律法规等相关知识产权，保护数据，不盗取数据； 4. 有较强的需求理解、整体设计、功能实现能力； 5. 具有较强的逻辑思维能力和综合运用知识的能力； 6. 在测试时间内完成任务，体现良好的时间管理能 |

| | | | |
|--|--|----------------------|---|
| | | | 力。 |
| | | 项目 3: 基于hive 的数据分析 | <p>专业技能:</p> <ol style="list-style-type: none"> 1. 正确启动 hive; 2. 正确使用 Hive 建立外部表; 3. 正确使用 hive HQL 对表进行数据关联, 管理数据; 4. 正确使用 hive HQL 对数据进行分析, 统计数据的分布情况; 5. 正确定位和解决项目运行过程中的异常。 |
| | | | <p>专业素养和道德规范:</p> <ol style="list-style-type: none"> 1. 具备分析数据, 搭建数据仓库的能力 2. 具有良好的编程习惯, 较强的分析问题的能力和良好的发散思维及创新能力; 3. 遵守职业道德和法律法规等相关知识产权, 保护数据, 不盗取数据; 4. 有较强的需求理解、整体设计、功能实现能力; 5. 具有较强的逻辑思维能力和综合运用知识的能力; 6. 在测试时间内完成任务, 体现良好的时间管理能力。 |
| | | 项目 4: Spark 大数据处理与分析 | <p>专业技能:</p> <ol style="list-style-type: none"> 1. 正确安装 Spark; 2. 掌握读取数据文件、MySQL 数据库数据并转为 RDD 的操作; 3. 掌握 SparkRDD 数据处理的基础函数操作和进阶函数操作; 4. 掌握 Spark Shell、SparkSQL 的命令行操作; 5. 掌握数据文件转换成 DataFrame, RDD 转换成 |

| | | |
|--|--|--|
| | | <p>DataFrame 的操作;</p> <p>6. 掌握 SparkSQL 中 SQL 语言和 DSL 语言开发。</p> |
| | | <p>专业素养和道德规范:</p> <ol style="list-style-type: none"> 1. 具备利用 Spark 大数据处理与分析解决实际问题的能力; 2. 具有较强的逻辑思维能力和综合运用知识的能力, 能定位 Spark 大数据分析项目运行过程中的问题并解决的能力; 3. 具有良好的编程习惯, 较强的分析问题的能力和良好的发散思维及创新能力; 4. 具有较强的逻辑思维能力和综合运用知识的能力; 5. 能够遵守大数据分析处理等相关法律法规; 6. 具备程序员的严谨细致、规范的工作态度和正确的价值观; 7. 在测试时间内完成任务, 体现良好的时间管理能力。 |
| | <p>模块四、数据可视化</p> <p>项目 1: 基于 matplotlib 的数据可视化</p> | <p>专业技能:</p> <ol style="list-style-type: none"> 1. 正确完成 matplotlib 绘图工具包和 pandas、numpy 模块的安装与导入; 2. 正确掌握 pyplot 常用绘图参数的调节方法; 3. 正确掌握子图的绘制方法; 4. 正确掌握绘制图形的保存与展示方法; 5. 正确掌握散点图的作用与绘制方法; 6. 正确掌握折线图的作用与绘制方法; 7. 正确掌握纵向柱状图、横向柱状图的作用与绘制方法; 8. 正确掌握饼图的作用与绘制方法; 9. 正确创建 pandas 中 Series 对象, Series 常用属 |

| | | | |
|--|----------------------------------|---|--|
| | | <p>性的访问， Series 数据的基本操作；</p> <p>10. 正确创建 pandas 中 Data Frame 对象、访问吗 Data Frame 常用属性，正确掌握 Data Frame 数据的基本操作；</p> <p>11. 正确使用 pandas 内置函数读写 CSV 等文件；</p> <p>12. 正确定位和排除图表生成过程中的异常；</p> | <p>专业素养和道德规范：</p> <p>1. 具备利用可视化技术解决实际应用中不同类型数据展示的能力。</p> <p>2. 具备良好的数据操作能力、数据统计分析和理解能力。</p> <p>3. 具有良好的视觉审美能力、创新设计理念。</p> <p>4. 具有良好的编程习惯，较强的分析问题的能力和良好的发散思维及创新能力。</p> <p>5. 遵守职业道德和法律法规等相关知识产权，保护数据，不盗取数据；</p> <p>6. 在测试时间内完成任务，体现良好的时间管理能力。</p> |
| | <p>项目 2: 基于 pyecharts 的数据可视化</p> | | <p>专业技能：</p> <p>1. 正确完成 pyecharts 绘图工具包和 pandas、numpy 的安装与导入；</p> <p>2. 正确掌握图形参数的配置；</p> <p>3. 正确掌握全局配置项和系列配置项的设置；</p> <p>4. 正确掌握绘制图形的保存方法；</p> <p>5. 正确掌握散点图的作用与绘制方法；</p> <p>6. 正确掌握折线图的作用与绘制方法；</p> <p>7. 正确掌握柱状图、横向柱状图、双柱状图的作用与绘制方法；</p> <p>8. 正确创建 pandas 中 Series 对象，Series 常用属</p> |

| | | | |
|--|--|--|--|
| | | | <p>性的访问， Series 数据的基本操作；</p> <p>9. 正确创建 pandas 中 Data Frame 对象、访问吗 Data Frame 常用属性，正确掌握 Data Frame 数据的基本操作；</p> <p>10. 正确使用 pandas 内置函数读写 CSV、Excel 等文件；</p> <p>11. 正确定位和排除图表生成过程中的异常；</p> |
| | | | <p>专业素养和道德规范：</p> <p>1. 具备利用可视化技术解决实际问题中不同类型数据展示的能力；</p> <p>2. 具备良好的数据分析和理解能力；</p> <p>3. 具有良好的视觉审美能力、创新设计理念；</p> <p>4. 具有良好的编程习惯，较强的分析问题的能力和良好的发散思维及创新能力；</p> <p>5. 遵守职业道德和法律法规等相关知识产权，保护数据，不盗取数据；</p> <p>6. 在测试时间内完成任务，体现良好的时间管理能力。</p> |

五、考核方式

1. 模块抽取：本专业技能考核标准的四个模块均为必考模块。参考学生按 25%比例随机抽取考试模块。各模块考生人数按四舍五入计算，剩余的尾数考生随机在四个模块中抽取应试模块。

2. 项目抽取：学生在相应模块题库中随机抽取 1 套试题考核。

六、附录

1. 相关法律法规（摘录）

(1)《中华人民共和国计算机信息系统安全保护条例》第四章第二十三条规定：故意输入计算机病毒以及其他有害数据危害计算机信息系统安全的，或者未

经许可出售计算机信息系统安全专用产品的,由公安机关处以警告或者对个人处以 5000 元以下的罚款、对单位处以 15000 元以下的罚款;有违法所得的,除予以没收外,可以处以违法所得 1 至 3 倍的罚款。

(2)《中华人民共和国计算机信息系统安全保护条例》第二章第十三条规定:计算机信息系统的使用单位应当建立健全安全管理制度,负责本单位计算机信息系统的安全保护工作。

(3)《中华人民共和国计算机信息系统安全保护条例》第二章第十四条规定:对计算机信息系统中发生的案件,有关使用单位应当在 24 小时内向当地县级以上人民政府公安机关报告。

专业标准主要依据的计算机行业国家技术标准如表 2 所示。

表2 引用技术标准和规范

| 序号 | 标准号 | 中文标准名称 |
|----|----------------------|-----------------|
| 1 | ISO/IEC23000-5:2011 | 多媒体应用格式(MPEG-A) |
| 2 | ISO/IEC10646:2011 | 通用编码字符集(UCS) |
| 3 | ISO/IEC 27005:2011 | 信息安全风险管理 |
| 4 | ISO/IEC25010:2011 | 系统和软件质量模型 |
| 5 | ISO/IEC 26512:2010 | 对用户文档需方和供方的要求 |
| 6 | ISO/IEC TR19791:2010 | 操作系统的安全评价 |
| 7 | ISO/IEC 25000:2005 | 软件产品质量要求与评价指南 |
| 8 | GB/T 11457-2006 | 软件工程术语 |
| 9 | GB/T 8566-2007 | 软件生存周期过程 |
| 10 | GB/T 8567-2006 | 计算机软件文档编制规范 |
| 11 | GB50311-2007 | 综合布线工程设计规范 |
| 12 | GB50312-2007 | 综合布线工程验收规范 |
| 13 | GB/T 15532-2008 | 计算机软件测试规范 |
| 14 | GB/T 9385-2008 | 计算机软件需求规格说明规范 |
| 15 | GB/T 19668.5-2007 | 信息化工程监理规范 |
| 16 | GB/T 20270—2006 | 网络基础安全技术要求 |

| | | |
|----|-----------------|--------------|
| 17 | SJ/T 11372-2007 | 中文办公软件用户界面要求 |
| 18 | SJ/T 11409-2009 | 软件构件模型 |