



湖南现代物流职业技术学院
HUNAN MODERN LOGISTICS COLLEGE

2021 级大数据技术专业 技能考核标准

大数据技术与应用专业技能考核标准

一. 专业名称及适用对象

1. 专业名称

大数据技术专业（专业代码：510205）。

2. 适用对象

高等职业院校三年一期全日制在籍学生。

二. 考核目标

本专业技能考核标准通过程序设计、数据库技术、数据处理、分布式应用等 4 个技能考核模块，测试学生的编程能力、数据采集能力、数据清洗能力、数据分析能力、数据可视化技能和分布式系统安装、配置、应用与维护能力以及从事大数据开发工作的程序编写规范、技术文档编写、交流与沟通等职业素养。引导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与应用高素质技术、技能人才。

三. 考核内容

模块一：程序设计（Python）

本模块以企、事业单位应用项目为背景，完成 Python 项目开发平台的配置与使用、项目模型的设计与建立、程序代码的编写与运行等工作内容，基本涵

盖了数据采集、数据处理与分析等岗位设计与开发工作所需语言开发的基本技能。

1. 开发平台的配置与使用基本要求

- (1) 能熟练使用主流的软件开发平台，并进行相关参数的配置；
- (2) 能使用平台进行项目的创建、开发、编译、运行及调试；
- (3) 具有较强的分析与解决问题的能力。

2. 项目的设计与建模

- (1) 能使用面向对象思想对数据处理分析项目进行建模与设计；
- (2) 能将编程任务以流程图的形式描述出来；
- (3) 具有较强的分析问题的能力、发散思维和创新意识。

3. 程序的编写与实现基本要求

- (1) 能使用数据类型、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的业务逻辑单元；
- (2) 能使用封装、继承、多态、类、对象、重载等语言机制，进行面向对象程序的编写，实现代码的可重用性；
- (3) 能使用列表、元组和字典等数据结构，存放数据及读取数据，分析数据；
- (4) 能实现文件的读取和写入，能将处理完的数据写入到 CSV、JSON 等格式的文件中；
- (5) 具有良好的编程习惯、较强的逻辑思维能力及综合运用知识的能力；
- (6) 具备大数据应用程序员的严谨认真、规范的工作态度和正确的价值观。

模块二：数据库设计技术（MySQL）

本模块以企业事业单位大数据应用系统项目开发为背景，完成应用系统中 MySQL 数据库开发环境的配置与使用、数据库及数据表的设计、创建与管理、数据表的约束与关系、数据库访问和数据库备份管理等工作内容。基本涵盖了大数据应用开发工程师、ETL 工程师岗位从事应用系统数据库的设计与开发工作所需的基本技能。本模块主要掌握 MySQL 数据库设计技术。

1. 数据库管理系统的配置与使用基本要求：

(1) 能使用主流的数据库管理系统如 MySQL，在常用操作系统上进行相关参数的配置，完成数据库管理系统的安装；

(2) 能使用主流的数据库管理工具，完成数据库管理系统的连接和使用；

(3) 能在 Linux 环境中用命令完成数据库服务器的启动与停止。

2. 数据库及数据表的创建与管理

(1) 能使用 E-R 图构建应用系统数据库模型；

(2) 能使用 SQL 语句实现数据库及数据表的创建；

(3) 能使用 SQL 语句实现数据库及数据表的重命名；

(4) 能使用 SQL 语句实现数据库及数据表的修改或删除；

(5) 在应用系统数据库设计过程中，具有数据库管理员、程序员必备的数据库操作和管理习惯，数据表结构设计合理，SQL 语句执行效率高，表名、字段名命名规范具较好的可读性和可维护性。

3. 创建数据表的约束和关系基本要求：

(1) 能使用 E-R 图设计数据表关系；

(2) 能使用 SQL 语句实现数据表的主、外键约束的添加、修改或删除操作；

(3) 能使用 SQL 语句实现唯一约束、检查约束、默认约束的添加、修改或删除；

(4) 能根据业务系统需求对数据表的约束及关系进行规范命名, 并具有良好的可读性。

4. 数据表的访问基本要求:

(1) 能使用 SQL 语句实现数据表记录的插入、修改和删除操作;

(2) 能使用 SQL 语句实现联合查询、嵌套查询, 并能实现对查询结果集进行筛选、排序、统计操作;

(3) 能使用 SQL 语句实现视图的创建、修改或删除操作;

(4) 能使用 SQL 语句实现存储过程、数据表触发器的创建、修改或删除操作;

(5) 能使用批处理、流程控制语句进行 SQL 编程。

5. 数据库的备份与恢复基本要求:

(1) 能使用数据库管理工具或 SQL 语句完成数据库的备份;

(2) 能使用数据库管理工具或 SQL 语句完成数据库的恢复。

模块三：数据处理

本模块以企业事业单位对数据人才需求为背景, 涉及网络数据的爬取和采集、数据处理和分析、数据可视化等操作。基本涵盖了大数据相关爬虫工程师、ETL 工程师、数据分析师等岗位从事数据采集、处理和分析工作所需的主要技能。

1、网络数据的爬取和采集

(1) 能够使用 HTML 脚本语言书写基本网页

(2) 能够使用 Python 主流模块爬取网页

(3) 能够爬取简单静态网页

(4) 能够爬取常规动态网页

(5) 能够模拟登陆爬取网页

(6) 能够使用 Scrapy 框架

2、数据处理和分析

- (1) 能够使用 Numpy 进行数值计算
- (2) 能够使用 Pandas 进行统计分析
- (3) 能够使用 Pandas 进行数据预处理
- (4) 能够使用常见的机器学习算法进行数据分析

3、数据可视化

- (1) 能够使用 Matplotlib 进行数据可视化
- (2) 能够绘制并定制化图表
- (3) 能够绘制点图、条形图和 3D 图表等常见形状图表
- (4) 能够进行交互式绘图
- (5) 能够使用图像和地图绘制图表

模块四：分布式处理系统（Hadoop）

本模块以企业事业单位对大数据人才需求为背景，涉及大数据的存储和管理、数据处理和分析、数据应用等操作。基本涵盖了大数据相关大数据工程师、ETL 工程师、系统运维工程师等岗位从事大数据存储、管理和分析应用工作所需的主要技能。

1、虚拟机安装与使用

- (1) 虚拟机下载与安装
- (2) 虚拟机的配置
- (3) 虚拟机硬件的分配
- (4) 虚拟机网络连接

2、Linux 操作系统安装与管理

- (1) Linux 操作系统的安装

- (2) Linux 操作系统的网络配置
 - (3) Linux 系统的软件安装
 - (4) Linux 操作系统的文件管理
 - (5) JDK 软件的安装
- 3、Hadoop 分布式系统的安装与应用
- (1) Hadoop 分布式系统的安装与配置
 - (2) HDFS 的文件管理
 - (3) Hadoop 集群的应用
- 4、Hbase 安装与应用
- (1) Hbase 安装与配置
 - (2) ZooKeeper 安装配置
 - (3) Hbase Shell 操作
 - (4) Hbase 的应用
- 5、Hive 安装与应用
- (1) Hive 安装与配置
 - (2) Hive Shell 操作
 - (3) Hive 的应用

四. 评价标准

1. 评价方式

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。

2. 分值分配

本专业技能考核满分为 100 分，其中专业技能占 90 分，职业素养占 10 分。抽取的学生按抽取学生人数的 30%抽程序设计模块、30%抽数据库技术模块、20%抽数据处理模块、20%抽分布式应用模块。

3. 技能评价要点

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。模块三题目亦可提供下载的网页供爬取。各模块和项目的技能评价要点内容如表 1 所示。

表 1 大数据技术与应用专业技能考核评价要点

模块	项目	评价要点
程序设计 (Python)	开发平台的配置与使用	正确配置软件开发环境； 开发环境配置过程符合职业规范。
	项目的设计与建模	项目的设计步骤清晰、方法科学合理； 正确将面向对象的思想运用于项目设计中，有效降低代码的冗余度，提高代码的复用性； 正确运用各种图例画出程序流程图； 设计过程符合职业规范。
	程序的编写与实现	正确定义变量、常量，名称符合命名规范；正确使用运算符、表达式、函数进行编程； 正确使用顺序、分支、循环三种控制结构实现项目的业务逻辑单元； 正确使用数组等基本数据结构进行编程； 正确使用封装、继承、多态、类、对象等面向对象语言机制，实现代码的复用； 正确使用列表、元组和字典、Numpy 和 Pandas 等数据结构存储数据； 程序书写结构良好，注释清晰，可维护性好；程序设计合理、语法正确、功能正确完备； 开发过程遵循软件开发的规范。
数据库设计技术	数据库管理系统的配置与使用	正确根据项目的需求选取数据库系统产品； 正确使用数据库管理工具，实现系统用户的管理与配置； 正确使用命令实现数据库服务器的启动与停止。

(MySQL)	数据库及数据表的创建与管理	正确使用 E-R 图完成数据表的结构设计； 正确使用 SQL 语句完成数据库及数据表的修改和删除。
	创建数据表的约束和关系	正确使用 E-R 图实现数据表的关系设计； 正确使用 SQL 语句实现主键约束、外键约束、唯一约束、检查约束及默认约束的添加、修改或删除。
	数据表的访问	正确书写 SQL 语句执行增、删、改、查等数据操作； 正确书写 SQL 语句创建视图、触发器及存储过程并执行； 正确书写 SQL 语句管理数据库对象； 遵守相关职业规范。
	数据库的备份和恢复	正确使用数据库管理工具实现数据库的备份； 正确使用数据库管理工具实现数据库的恢复。
数据处理技术	网络数据的爬取和采集	认识 HTML 等脚本语言 Request 模块、BeautifulSoup 模块的使用 爬取简单静态网页 爬取常规动态网页 模拟登陆爬取网页 使用 Scrapy 框架
	数据处理和分析	使用 Numpy 进行数值计算 使用 Pandas 进行统计分析 使用 Pandas 进行数据预处理 使用常见的机器学习算法进行数据分析
	数据可视化	使用 Matplotlib 进行数据可视化 绘制并定制化图表 绘制点图、条形图和 3D 图表等常见形状图表 进行交互式绘图 使用图像和地图绘制图表
分布式系统 (Hadoop)	虚拟机安装与使用	正确下载的 ISO 映像文件 正确选择安装过程中使用的语言、选择安装位置 正确创建密码和创建用户、设置 root 密码 成功重启 CentOS7 操作系统，输入用户名和密码验证登录 正确运用指令自动获取 IP 并查看 正确修改好静态 IP 模式后指定 IP 地址、子网掩码以及网关、配置 DNS 的 IP
	LINUX 操作系统安装与管理	正确安装 Linux 工具包 正确远程登录工具软件的下载 正确远程登录工具软件的安装与配置 正确用远程登录工具，操作 ls 命令查看文件 正确安装 JDK 环境

<p>Hadoop 分布式系统的安装与应用</p>	<p>正确解压 Hadoop 安装包 正确解压后的 Hadoop 文件夹移动到/opt/hadoop 正确修改 hadoop-env.sh 文件，配置 jdk 路径 正确修改核心组件文件是 core-site.xml 正确修改文件系统配置文件是 hdfs-site.xml 正确修改 Yarn 的站点配置文件是 yarn-site.xml 正确修改.bash_profile 文件 正确格式化文件系统 正确启动 Hadoop 集群 正确运行 Hadoop 集群程序</p>
<p>Hbase 安装与应用</p>	<p>正确下载解压 zookeeper 软件，并移动到/opt/zookeeper 正确修改配置文件 zoo.cfg 正确给每个节点做好标识文件 myid 正确给每个节点启动 ZooKeeper，并查看节点运行情况 正确下载解压指定的 hbase 软件，并移动到/opt/hbase 路径 正确修改 hbase-env.sh 中 JDK 环境变量 正确修改 hbase-env.sh 中 HBASE_MANAGES_ZK=false 正确修改配置文件 hbase-site.xml 正确设置 regionservers 列表 正确启动 HBase，正确操作 Hbase shell 正确查看当前用户 正确查看已有数据表 正确创建表 yg_info 正确给表 yg_info 添加记录 正确扫描整张表 正确统计表中记录数</p>
<p>Hive 安装与应用</p>	<p>正确下载解压 Hive 软件，并移动到/opt/hive 路径 正确修改配置文件 hive-default.xml 正确修改 hive-env.sh 正确启动 hive shell 正确启动 hadoop、zookeeper、hbase 和 hive 正确查看当前数据库中的数据表 show tables 正确创建店铺信息表 正确导入信息到 shop_info 表中 正确查询出商圈 (Business) 是“左家庄”的所有店铺</p>

五. 抽考方式

1. 考题抽选

试题总分 100 分，学生在模块一至模块四中任选一题 100 分（含职业素养 10 分）。

2. 测试时间

90 分钟

3. 场地设备

技能抽查考点由教育厅指定，考点提供计算机、网络设备、软件等资源，但不允许自带存储介质和软件等相关材料。

4. 评分方式

根据现场测评专家的记录和测试结果，由湖南省职业院校职业能力考试委员会指定的考评员集体评判成绩。

5. 成绩公布

由湖南省教育厅在网上统一公布抽查学校的平均成绩和合格率及排名等，不公布个人成绩。

6. 抽考方式

按该年级注册人数随机抽取 10%，抽取样本不足 10 人，取 10 人，超过 30 人取 30 人。技能抽查前一天通知被抽查学校，抽查当天随机确定被抽查学生，抽查对象一经确定，不得随意更改。

六. 附录

1. 相关法律法规

2002 年 《计算机软件著作权登记办法》
2001 年 《计算机软件保护条例》
2000 年 《计算机病毒防治管理办法》
1997 年 《计算机信息系统安全专用产品检测和销售许可证管
法》

2. 相关规范与标准

GB/T 8566 -2001 信息技术软件生存周期过程
GB/T 15853 -1995 软件支持环境
GB/T 14079 -1993 软件维护指南
GB/T 17544-1998 信息技术软件包质量要求和测试
GB/T 11457-2006 信息技术、软件工程术语
GB8566-88 计算机软件开发规范
GB/T 12991-2008 信息技术数据库语言 SQL 第 1 部分：框架
20120567-T-469 信息技术云数据存储和管理第 1 部分：总则
GB/T 21025-2007 XML 使用指南
GB/T 20009-2005 信息安全技术数据库管理系统安全评估准则
GB/T 20273-2006 信息安全技术数据库管理系统安全技术要求
20100383-T-469 信息技术安全技术信息安全管理体系实施指南
软件开发标准
GB/T 17544-1998 信息技术软件包质量要求和测试
T/31SCTA 002-2017 工业大数据平台技术规范 数据存储