



湖南网络工程职业学院

大数据技术专业 专业技能考核标准

2021 级

2021 年 8 月

目 录

一、专业名称及适用对象	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
（一）专业基本技能	1
（二）专业核心技能	2
四、评价标准	5
1. 评价方式	5
2. 分值分配	5
3. 技能评价要点	5
五、考核方式	7
1. 参考模块确定	7
2. 考题抽选	7
六、附录	7
1. 相关法律法规	7
2. 相关规范与标准	7

湖南网络工程职业学院

大数据技术专业学生技能考核标准

一、专业名称及适用对象

1. 专业名称

专业名称：大数据技术专业

专业代码：510205

专业大类：电子信息

隶属专业群：数字媒体技术专业群（湖南省高水平专业群）

2. 适用对象

高职全日制在籍毕业年级学生。

二、考核目标

依据本专业的人才培养方案,本技能考核标准通过设置数据管理技术、大数据平台部署、Python 数据采集技术、数据处理与可视化 4 个技能考核模块,测试学生的数据库管理、大数据平台运维、数据采集、数据处理与数据可视化展现等能力,检测学生程序编写规范、技术文档编写、操作规范等职业素养。

三、考核内容

（一）专业基本技能

模块一：数据管理技术

本模块以大数据应用系统中数据管理为背景,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、数据库访问等工作内容。基本涵盖了大数据平台运维工程师、数据采集工程师、大数据可视化工程师岗位中关于数据管理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工具;
- (2) 能使用主流的数据库图形化工具;

(3) 能完成数据库管理系统的连接和使用。

2. 数据库及数据表的创建与管理

(1) 能通过 E-R 图理解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；

(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，并具有较好的可读性。

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 语句实现单表数据的查询

(3) 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；

(4) 能使用 SQL 语句实现视图的创建、修改或删除操作；

5. 完成时间：180 分钟

(二) 专业核心技能

模块二：大数据平台部署

本模块以企业大数据平台部署为背景，完成大数据技术应用中大数据平台应用环境的搭建、配置、使用、维护、调优。基本涵盖了大数据平台运维岗位所需的基本技能。

1. Hadoop 平台环境的配置基本要求：

(1) 能够搭建和管理虚拟机环境，为部署 Hadoop 做准备；

- (2) 能够配置和管理虚拟机；
- (3) 能够完成操作系统的安装和配置；
- (4) 能够完成操作系统的网络设置并进行测试；
- (5) 能够完成大数据平台的日常维护。

2. 大数据处理工具 Hadoop 的部署基本要求：

- (1) 能根据业务需求，选择 Hadoop 正确模式进行部署；
- (2) 能在 Hadoop 伪分布式模式基础上继续部署完全分布式模式；
- (3) 能根据需处理的数据量大小对 Hadoop 集群进行优化；
- (4) 能根据业务需求修改相关配置文件，提高 Hadoop 集群的性能；
- (5) 能够合理调整 Hadoop 集群节点数，以满足数据处理要求；
- (6) 能够利用 Hadoop 集群对海量数据进行存储、查询、分析。

3. 分布式文件系统的使用基本要求：

- (1) 能根据需求设计 HDFS 副本存放策略；
- (2) 能根据各节点硬件条件，对 HDFS 集群进行负载均衡设置；
- (3) 能使用 HDFS SHELL 命令对数据进行上传、下载、查看、删除等操作；
- (4) 能通过 HDFS 集群的 WEB 界面掌握集群的情况；
- (5) 熟悉相关规范和标准，具备团队协作能力，能识读相关软件技术文档；

4、完成时间：180 分钟

模块三：Python 数据采集技术

本模块以企、事业单位应用项目为背景，完成 Python 项目开发平台的使用、项目模型的设计与建立、程序代码的编写与运行、网络数据采集等工作内容，基本涵盖了数据采集工程师、大数据可视化工程师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. 项目的设计与建模基本要求

(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、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) 能使用列表和字典等数据结构,存放数据及读取数据,分析数据;

(3) 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;

(4) 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。

4. 网络数据采集

(1) 能够使用常看网页源代码并掌握浏览器开发者工具的使用方法;

(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) 能够使用正则表达式解析数据;

(6) 能够使用 Xpath 解析数据;

(7) 能够使用 BeautifulSoup 解析数据;

(8) 能够按要求进行数据存储。

4. 完成时间: 180 分钟

模块四: 数据处理与可视化

本模块以数字内容服务行业、社会服务行业和资讯服务行业的数据处理与可视化需求为背景,主要涉及数据处理分析、数据可视化等操作。基本涵盖了大数据可视化工程师等岗位从事处理和可视化与分析工作所需的主要技能。

1. 数据处理和分析

(1) 能够读取、写入文件

(2) 能够正确使用 Pandas 进行数据预处理

(3) 能够对数据进行清洗、合并、重塑、聚合和分组运算

2. 数据可视化

- (1) 能够使用合适的工具进行数据可视化；
- (2) 能够绘制并定制化图表；
- (3) 能够绘制折线图、条形图、饼图等常见形状图表；
- (4) 能够进行交互式绘图。

3. 完成时间：180 分钟

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。

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。模块四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下表所示。

类型	模块	项目	评价要点
专业基本技能	数据管理技术	数据库管理系统的配置与使用	能够正确根据项目的需求选取数据库系统产品； 能够正确使用数据库管理工具，实现系统用户的管理与配置； 能使用主流的数据库图形化工具； 能完成数据库管理系统的连接和使用。
		数据库及数据表的创建与管理	能够正确使用 E-R 图完成数据表的结构设计； 能够正确使用 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	能够正确使用 E-R 图实现数据表的关系设计； 能够正确使用 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	能使用 SQL 语句实现数据表记录的插入、修改和删除操作； 能使用 SQL 语句实现单表数据的查询 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； 能使用 SQL 语句实现视图的创建、修改或删除操作。

专业 核心 技能	大数据 平台部 署	Hadoop 平台环 境的配置	能够搭建和管理虚拟机环境，为部署 Hadoop 做准备； 能够配置和管理虚拟机； 能够完成操作系统的安装和配置； 能够完成操作系统的网络设置并进行测试； 能够完成大数据平台的日常维护。
		大数据处理工 具 Hadoop 的部 署	能根据业务需求，选择 Hadoop 正确模式进行部署； 能在 Hadoop 伪分布式模式基础上继续部署完全分布式模式； 能根据需处理的数据量大小对 Hadoop 集群进行优化； 能根据业务需求修改相关配置文件，提高 Hadoop 集群的性能； 能够合理调整 Hadoop 集群节点数，以满足数据处理要求； 能够利用 Hadoop 集群对海量数据进行存储、查询、分析。
		分布式文件系 统的使用	能根据需求设计 HDFS 副本存放策略； 能根据各节点硬件条件，对 HDFS 集群进行负载均衡设置； 能使用 HDFS SHELL 命令对数据进行上传、下载、查看、删除等操作； 能通过 HDFS 集群的 WEB 界面掌握集群的情况； 熟悉相关规范和标准，具备团队协作能力，能识读相关软件技术文档。
	Python 数据采 集技术	开发平台的配 置与使用	能够正确使用软件开发环境，符合职业规范。
		项目的设计与 建模	设计完成项目时步骤清晰、方法科学合理； 能够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； 能够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	能够正确定义变量、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； 能够正确使用顺序、分支、循环三种控制结构实现项目的业务逻辑单元； 能够正确使用字符串、列表和字典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。
		网络数据采集	能够认识 HTML 等脚本语言 能够正确使用 Request 模块、BeautifulSoup 模块 能够爬取简单的静态网页 能够爬取常规的动态网页 能够模拟登陆爬取网页
	数 据 处 理 与 可 视 化	数据处理和分 析	能够读取、写入文件 能够正确使用 Pandas 进行数据预处理 能够对数据进行清洗、合并、重塑、聚合和分组运算
		数据可视化	能够使用 Matplotlib、echarts、pyecharts 等技术中的一项进行数据可视 化展示 能够绘制并定制化图表 能够绘制折线图、条形图、饼图等常见形状图表 能够进行交互式绘图

五、考核方式

技能考核为现场操作考核，成绩评定采用过程考核与结果考核相结合。具体方式如下：

1.参考模块确定

专业基本技能和专业核心技能所涉及的四个模块都为考试模块，参考学生按规定比例随机抽取模块。其中，25%的考生选择模块一，25%的考生选择模块二，25%的考生选择模块三，25%的考生选择模块四。考试过程中可以根据实际情况对参考学生比例进行适当调整，各模块考生人数按四舍五入计算。

2.考题抽选

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

《中华人民共和国网络安全法》

《中国互联网网络版权自律公约》

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 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

T/31SCTA 003-2017 工业大数据平台技术规范 数据处理

T/31SCTA 002-2017 工业大数据平台技术规范 数据存储