



湖南工程职业技术学院
HUNAN VOCATIONAL COLLEGE OF ENGINEERING

大数据技术与应用 专业技能考核标准

专业代码： 610215

所属学院： 信息工程学院

适用年级： 2020级

专业主任： 潘龙英

学院审核人： 龚亮

制（修）订时间： 2020年8月

目 录

一、专业名称及适用对象.....	1
二、考核内容.....	1
三、评价标准.....	3
四、抽考方式.....	5

一、专业名称及适用对象

1. 专业名称

大数据技术与应用(专业代码: 610215)

2. 适用对象

高职高专全日制在籍毕业年级学生。

二、考核内容

根据学院大数据技术与应用专业人才培养方案制订覆盖本专业主要知识点、技能点和职业素养要求的技能考核方案。该方案的主要考核内容包括大数据平台部署技能、数据采集与存储技能、数据分析与可视化技能,包含工信部大数据分析师、大数据分析与应用1+X证书两种职业资格证部分考核内容,专业技能结构关系如下图所示。

模块一	模块二	模块三
大数据平台部署	数据采集与存储	数据分析与可视化
(1) 大数据平台的搭建、配置、优化、测试能力; (2) 符合企业员工的素质要求; (3) 符合大数据平台部署工作师的基本素养要求。	(1) 数据的采集、清洗、预处理、存储能力; (2) 符合企业员工素质要求; (3) 符合大数据采集与存储工程师的基本素养要求。	(1) 大数据预处理、分析统计、可视化能力; (2) 符合大数据分析与可视化工程师的基本素养要求。
项目1	项目1	项目1
Hadoop平台部署与运维	网站数据采集与存储	数据分析与可视化
项目2		
Hadoop生态圈其它组件搭建与配置		

图1 技能抽查模块图

模块一 大数据平台搭建与配置

该模块以大数据平台常见集群部署项目为背景,主要围绕大数据平台所需组件,完成集群安装、配置、优化等工作任务。属于大数据应用开发工程师典型工作流程所需大数据平台部署的基本技能。对Hadoop生态圈组件进行安装、配置、优化和测试,具备大数据平台部署技能。对应核心课程为: Hadoop大数据开发基础, Spark大数据快速运算, Hive技术。

1、Hadoop生态圈组件搭建与配置

基本要求：

- (1) 能在虚拟机的环境下，熟练安装Hadoop生态圈组件并部署Hadoop集群，包括：JDK、Hadoop、Hadoop 高可用的安装与部署。
- (2) 能熟练运用Vi编辑器对Hadoop生态圈组件相应文件进行配置与优化，包括：Hadoop、Hadoop高可用核心文件的配置与优化。
- (3) 能熟练安装ssh服务，完成集群间节点的免密登录：包括：ssh服务开启、密钥创建。
- (4) 能熟练克隆一台已安装好Hadoop单机模式的虚拟机并做快照。
- (5) 能够完成Hadoop大数据平台部署，并完成相应测试。包括：格式化、启动和关闭Hadoop、察看进程、运行Hadoop自带MapReduce示例程序。

2、Hadoop生态圈其它组件搭建与配置

基本要求：

- (1) 能在虚拟机的环境下，熟练安装Hadoop生态圈其它组件并部署集群。包括：Hbase、Hive、Zookeeper、Spark、flum、Sqoop。
- (2) 能熟练运用Vi编辑器对Spark组件相应文件进行配置与优化。包括：Hbase、Hive、Zookeeper、flum、Sqoop、Spark核心文件的配置与优化。
- (3) 能熟练安装ssh服务，完成集群间节点的免密登录：包括：ssh服务开启、密钥创建。
- (4) 能熟练克隆一台已安装好虚拟机并做快照。
- (5) 完成Spark大数据平台部署，并完成相应测试。包括：启动和关闭Spark、察看进程及集群节点情况。

模块二 数据采集与存储

本项目以数据采集与存储应用为背景，完成项目所需数据的采集与存储，基本涵盖了数据采集、数据清洗、数据存储等大数据应用开发工程师典型工作流程所需的基本技能。数据采集与数据存储是后续数据分析与数据可视化的基础，也是大数据学习的重中之重。本模块主要考核学生数据采集、数据清洗及数据存储等相关数据操作等技能。对应专业课程为：Python编程基础，数据采集与网络爬虫，数据库基础，HTML静态网页设计。

网站数据采集与存储基本要求：

- (1) 能够编写代码解析指定网页。
- (2) 能够查看网页源码，分析网页结构，按照要求采集指定数据或文件。
- (3) 能够将采集的数据或文件信息输出到控制台。
- (4) 能够根据给定数据源设计数据库表结构。
- (5) 能够编写代码完成数据库连接，完成数据批量插入到MySQL或者Hbase。
- (6) 能够查看数据库数据。
- (7) 能将采集到的文件存储到本地文件夹。

模块三 数据分析与可视化

本项目以企、事业单位应用项目为背景，完成项目所需数据的分析，基本涵盖了数据分析、数据可视化等岗位从事大数据技术与应用项目设计与开发工作所需的基本技能。数据分析与可视化将直接影响到数据分析报告的质量。主要考核学生数据分析、数据可视化等相关数据操作等技能。对应核心课程为：Python数据分析与应用，数据可视化技术，Python编程基础。

基本要求：

- (1) 能够理解数据分析目标及分析解决方案。
- (2) 能够选择最佳分析及可视化图表。
- (3) 使用 Python语言进行目标数据读取、数据探索、数据预处理、数据特征构造等，并按要求分析样例数据并输出到指定文件中。
- (4) 能够对数据进行计算、分析，并完成数据可视化。
- (5) 能够展示分析结果。

三、评价标准

(一)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；根据运行结果、测试结果和提交的文档及报告质量等因素评价结果成绩。

(二)分值分配

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分。

(三)评价标准

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。各项目的技能考核内容及要求如表1所示。

表 1 大数据技术与应用专业技能考核评价要点

序号	模块	项目	评价内容	分值(分)	评价要点
	大数据平台搭建与配置	Hadoop生态圈组件搭建与配置	集群搭建	50	正确配置与管理Hadoop集群；管理过程符合职业规范；能搭建hadoop集群部署；
			集群优化	30	能对hadoop集群及其它组件进行适当优化。
			职业素养	20	按要求命名文件，截图，答题规范有序；遵守相关职业规范，操作过程体现较高的职业素养；
		Hadoop生态圈其它组件搭建与配置	集群搭建	50	正确配置与管理Hadoop生态圈其它组件；管理过程符合职业规范；能搭建生态圈其它组件部署；
			集群优化	30	能对Hadoop生态圈其它组件进行适当优化。
			职业素养	20	按要求命名文件，截图，答题规范有序；遵守相关职业规范，操作过程体现较高的职业素养；
数据采集与存储	网站数据采集与存储	数据采集	40	能够使用 Python或者Java 语言编写程序，爬取指定网页数据，并将采集到的数据按要求存放到本地文件中；	
		数据存储	40	能够将采集到的数据文件，用Python或者Java 语言编写程序，根据应用场景，编写代码完成数据库连接，并对数据库进行相应设计；能将采集到的数据存放到本地数据库Mysql或Hbase中。	
		职业素养	20	按要求命名文件，截图，答题规范有序；遵守相关职业规范，操作过程体现较高的职业素养；	
数据分析与可视化	数据分析与可视化	数据处理与分析	50	采用Python语言完成离线数据读取、清洗、整理、分组和分析。	
		数据可视化	30	能够使用Python语言编写程序，根据要求使用相关数据绘制出各种图形。	
		职业素养	20	按要求命名文件，截图，答题规范有序；遵守相关职业规范，操作过程体现较高的职业素养。	

四、抽考方式

按照专业培养方案，根据岗位专业技能基本要求，本专业技能考核标准共设置三个考核模块，每个考核模块均设多个考核项目。本专业技能考核题库共50套题，其中大数据平台搭建与配置模块共17套题，数据采集与存储模块共16套题，数据分析与可视化模块共17套题。

(一) 学生样本抽取

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。

(二) 模块选取

本专业技能考核标准设置的三个模块均为必考模块。其中，30%考生考模块一，30%的考生考模块二，40%的考生考模块三。各模块考生人数按四舍五入计算，剩余的尾数考生随机在三个模块中抽取应试模块。模块抽取可以根据实际情况适当调整参考学生比例。

(三) 项目抽取

考生根据抽取的考核模块，从对应模块中随机抽取考核项目。

(四) 试题抽取

学生在相应项目题库中随机抽取1套试题进行测试。