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

大数据技术与应用 专业技能考核题库

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、大数据平台部署模块

项目 1：Hadoop 平台部署与运维

1.试题 1-1-1：携程旅游大数据 Hadoop 单机模式部署

（1）任务描述

某青年旅行社根据携程旅游源数据，通过结合大数据和高性能的分析，帮助旅行者提前查找到某地区酒店的酒店名称、酒店地址、酒店评分和酒店价格等数据内容并完成数据展示。

现在要进行大数据分析并完成展示，你作为公司大数据工程师，需安装分布式Hadoop环境，单机模式是Hadoop的默认模式，在该模式下无需任何守护进程，所有程序都在单个JVM上运行，该模式主要用于开发和调试mapreduce的应用逻辑。

表1.1.1单机模式规划

节点角色	虚拟机名	主机名	机器IP
单一节点	master0	node0	192.168.126.100

本环节需要完成Hadoop 平台架设单机模式部署，主要任务如下：

任务一：IP设置及网络互通（10分）

1.1设置并启用IP地址，并提交截图信息，命名1-1。（6分）

1.2内外拼通，并提交截图信息，命名1-2。（4分）

任务二：主机名的设置与映射（10分）

2.1设置主机名，并提交截图信息，命名2-1。（5分）

2.2映射主机名与IP地址，并提交截图信息，命名2-2。（5分）

任务三：正确安装JDK（20分）

3.1上传JDK压缩包至服务器/home/hadoop目录下,并提交截图信息，命名3-1；（5分）

3.2解压JDK至/usr/local/src下并重命名为JDK1.8,提交截图信息，命名3-2。（5分）

3.3配置Java环境变量，并提交截图信息，命名3-3；（5分）

3.4使设置的环境变量生效，验证JDK是否安装成功，并提交截图信息，命名3-4。（5分）

任务四：正确安装Hadoop (20分)

4.1上传Hadoop压缩包（在“F:\BigDataSoft\”文件夹内）至服务器/home/hadoop目录下,并提交截图信息，命名4-1；（5分）

4.2解压Hadoop至/usr/local/src下并重命名为hadoop,并提交截图信息，命名4-2。（5分）

4.3配置Hadoop环境变量，使修改环境变量生效，并提交截图信息，命名4-3；（5分）

4.4验证Hadoop是否安装成功，并提交截图信息，命名4-4。（5分）

任务五：测试 (20分)

5.1 Hadoop自带了一些MapReduce的示例程序，这些程序代码都在hadoop-example.jar包里，找到jar包的安装目录，并提交截图信息，命名5-1；（10分）

5.2单机模式下使用Hadoop计算圆周率，并提交截图信息，命名5-2；（10分）

提交要求：

1)在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南职业技术学院 01 张三。

2)考生文件夹内保存截图：1-1、1-2、2-1、2-2、3-1、3-2、3-2、3-4、4-1、4-2、4-3、4-4、5-1、5-2到一个word文档t1.docx中。

（2）实施条件

测试所需的软硬件设备见表1.1.2。

表1.1.2 考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存8G 以上，WIN7及以上操作系统（64位）	
3	截图工具		系统自带截图工具
4	服务器	安装有Ubuntukylin-16.04-desktop-amd64操作系统	机房/虚拟机
5	JDK安装包	jdk-8u162-linux-x64.tar.gz	Linux版
6	Hadoop安装包	hadoop-2.7.1.tar.gz	Linux版

（3）考核时量

考核时间为3个小时。

（4）评分细则

大数据平台搭建与配置模块考核实行 100 分制，评价内容包括工作任务、职业素养完成情况两个方面。其中，工作任务完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表 1.1.3 所示。

表 1.1.3 评分标准表评价内容

评价内容		分值	评分细则	
工作任务(80分)	IP设置	10分	未设置IP扣2分； 未启用IP扣2分；未能正确查看IP扣2分。	6分
			未正确内外拼通扣4分。	4分
	主机名的设置与映射	10分	未正确设置主机名扣5分。	5分
			未正确映射主机名与IP地址扣5分。	5分
	正确安装JDK	20分	未正确上传JDK压缩包至服务器/home/hadoop目录，扣5分。	5分
			未正确解压JDK至/jncj下，扣3分；未重命名为JDK1.8，扣2分。	5分
			未正确配置Java环境变量，扣3分； 未使设置的环境变量生效，扣2分。	5分
			未正确验证JDK安装成功，扣5分。	5分
	正确安装Hadoop	20分	未正确上传Hadoop压缩包至服务器/home/hadoop目录下，扣5分；	5分
			未正确安装Hadoop至/jncj下，扣5分。	5分
			未正确配置Hadoop环境变量，扣3分；未使设置的环境变量生效，扣2分。	5分
			未正确验证Hadoop安装成功，扣5分。	5分

	测试 Hadoop	20分	未正确找到Hadoop自带MapReduce的示例程序jar包所在目录，扣10分。	10分
			未正确使用Hadoop计算圆周率，扣10分。	10分
职业素养（20分）	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、鼠标等），检查测试所需软件开发环境。 不进行检查操作扣 5 分。	5分
	专业素养	10分	按要求命名文件，截图，答题规范有序得10分。	10分
	职业行为规范	5分	着装干净、整洁。举止文明，遵守考场纪律，按顺序进出考场。	5分

2.试题 1-1-2：千里行旅游大数据 Hadoop 伪分布式模式部署

（1）任务描述

千里行旅游平台需要提供出行点气候查询功能。现通过在天气网站查找天气信息，帮助使用者提前查找到时间、天气状况、气温和风力风向等数据内容并完成数据展示。

现要进行大数据分析并完成展示，你作为公司大数据工程师，需安装分布式Hadoop环境，在伪分布式模式下，Hadoop守护进程运行在一台机器上，模拟一个小规模的集群。该模式在单机模式的基础上增加了代码调试的功能，允许检查NameNode，DataNode，Jobtracker，Tasktracker等模拟节点的运行情况。

表1.2.1伪分布式模式规划

节点角色	虚拟机名	主机名	机器IP
单一节点	master0	node0	192.168.126.150

本环节需要完成Hadoop 平台架设伪分布式模式，主要任务如下：

任务一：克隆一台已安装好Hadoop单机模式的虚拟机并做快照（15分）

1.1当系统需要进行集群分布式部署时，需要多台相同的虚拟机，如果从头安装虚拟机费时费力，大约需要30分钟左右，因此需要克隆，克隆是完整的新建了一台虚拟机，克隆一台新的虚拟机，命名为master0，登录新的虚拟机，并提交截图信息，命名1-1。（5分）

1.2进入克隆后的系统，设置IP地址、映射主机名和IP，截图，并命名1-2。（5分）

1.3“快照”是虚拟机磁盘文件（VMDK）在某个点及时的副本，系统崩溃或系统异常，可以通过使用恢复到快照来保持磁盘文件系统和系统存储，在重要的节点创建虚拟机快照，将有多个还原点可以用于恢复，对上述克隆好的新虚拟机进行快照，命名为“master0快照1”，并提交截图信息，命名1-3。（5分）

任务二：修改Hadoop的5个核心配置文件(30分)

2.1 Hadoop-env.sh文件为Hadoop的运行环境配置文件，Hadoop的运行需要依赖JDK，将其中的export JAVA_HOME的值修改为安装的JDK路径，修改后提交截图信息，命名2-1；（6分）

2.2 core-site.xml文件为Hadoop的核心配置文件，用于定义系统级别的参数，设定namenode的主机名及端口、存放临时文件的目录，修改后提交截图信息，命名2-2；（6分）

2.3 hdfs-site.xml文件为HDFS核心配置文件，如文件副本的个数、块大小及是否使用强制权限等，设定HDFS存储文件的副本个数默认为1、SecondaryNameNode地址和端口，修改后提交截图信息，命名2-3；（6分）

2.4 mapred-site.xml本身这个文件是不存在的，将模版文件mapred-site.xml.template改名为mapred-site.xml，然后进行编辑告诉Hadoop mapreduce运行在yarn，修改后提交截图信息，命名2-4；（6分）

2.5 yarn-site.xml文件为Yarn框架配置文件，指定ResourceManager的地址、指定NodeManager获取数据的方式是shuffle，修改后提交截图信息，命名2-5。（6分）

任务三：格式化、启动和关闭伪分布式Hadoop (15分)

3.1 格式化DFS(Distributed File System)，在格式化的日志中看到succefully format字样，就证明格式化成功，并提交截图信息，命名3-1；（5分）

3.2 启动DFS及Yarn，使用Web 界面查看程序运行结果，并提交截图信息，命名3-2；（5分）

3.3 关闭Hadoop，并提交截图信息，命名3-3；（5分）

任务四：配置SSH免密登录(20分)

4.1 建立密钥对，并提交截图信息，命名4-1；（10分）

4.2 将本机的公钥复制到远程机器的authorized_keys文件中，配置SSH免密登录，并提交截图信息，命名4-2。（10分）

提交要求：

1)在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南职业技术学院 01 张三。

2)考生文件夹内保存截图：1-1、1-2、1-3、2-1、2-2、2-3、2-4、2-5、3-1、3-2、3-3、4-1、4-2到一个word文档t2.docx中。

（2）实施条件

测试所需的软硬件设备见表1.2.2。

表1.2.2 考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存8G 以上，WIN7及以上操作系统（64位）	
3	截图工具		系统自带截图工具
4	服务器	安装有Ubuntukylin-16.04-desktop-amd64操作系统、单机模式Hadoop	机房/虚拟机

、

（3）考核时量

考核时间为3个小时。

（4）评分细则

大数据平台搭建与配置模块考核实行 100 分制，评价内容包括工作任务、职业素养完成情况两个方面。其中，工作任务完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表 1.2.3 所示。

表 1.2.3 评分标准表评价内容

评价内容		分值	评分细则	
工作任务(80分)	克隆虚拟机并做快照	15分	未正确克隆扣5分。	5分
			未正确设置主机名，扣3分； 未正确设置IP地址，扣2分。	5分
			未成功创建快照，扣5分。	5分

	修改Hadoop的5个核心配置文件	30分	未正确配置JDK路径扣6分。	6分
			未正确配置namenode的主机名及端口扣3分；未正确配置namenode存放临时文件的目录扣3分。	6分
			未正确设置定HDFS存储文件的副本个数为1，扣3分；未正确设置SecondaryNameNode地址和端口，扣3分。	6分
			未正确将mapred-site.xml.template改名为mapred-site.xml扣3分；未正确设置Hadoop mapreduce 运行在yarn扣3分。	6分
			未正确配置ResourceManger的地址、扣3分；未正确配置NodeManager获取数据的方式是shuffle，扣3分。	6分
	格式化、启动和关闭伪分布式Hadoop	15分	格式化DFS未成功，扣5分。	5分
			启动DFS及Yarn，未成功使用Web 界面查看程序运行结果，扣5分。	5分
			未成功关闭Hadoop，扣5分。	5分
	配置SSH免密登录	20分	未成功建立密钥对，并提交截图信息,扣10分；	10分
			未成功将本机的公钥复制到远程机器的authorized_keys文件中，扣5分；未成功配置SSH实现免密登录，扣5分。	10分
职业素养（20分）	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、鼠标等），检查测试所需软件开发环境。不进行检查操作扣 5 分。	5分
	专业素养	10分	按要求命名文件，截图，答题规范有序得10分。	10分
	职业行为规范	5分	着装干净、整洁。举止文明，遵守考场纪律，按顺序进出考场。	5分

3.试题 1-1-3：畅游天下旅游大数据 Hadoop 伪分布式模式部署

（1）任务描述

畅游天下旅游平台需要提供出行点气候查询功能。现通过在天气网站查找天气信息，帮助使用者提前查找到时间、天气状况、气温和风力风向等数据内容并完成数据展示。

现要进行大数据分析并完成展示，你作为公司大数据工程师，需安装分布式

Hadoop环境，在伪分布式模式下，Hadoop守护进程运行在一台机器上，模拟一个小规模的集群。该模式在单机模式的基础上增加了代码调试的功能，允许检查NameNode，DataNode，Jobtracker，Tasktracker等模拟节点的运行情况。

表1.3.1伪分布式模式规划

节点角色	虚拟机名	主机名	机器IP
单一节点	master0	node0	192.168.126.150

本环节需要完成Hadoop 平台架设伪分布式模式，主要任务如下：

任务一：克隆一台已安装好Hadoop单机模式的虚拟机（15分）

1.1当系统需要进行集群分布式部署时，需要多台相同的虚拟机，如果从头安装虚拟机费时费力，大约需要30分钟左右，因此需要克隆，克隆是完整的新建了一台虚拟机，克隆一台新的虚拟机，命名为master0，登录新的虚拟机，并提交截图信息，命名1-1。（8分）

1.2进入克隆后的系统，设置IP地址、映射主机名和IP，截图，并命名1-2。（7分）

任务二：修改Hadoop的5个核心配置文件(30分)

2.1 Hadoop-env.sh文件为Hadoop的运行环境配置文件，Hadoop的运行需要依赖JDK，将其中的export JAVA_HOME的值修改为安装的JDK路径，修改后提交截图信息，命名2-1；（6分）

2.2 core-site.xml文件为Hadoop的核心配置文件，用于定义系统级别的参数，设定namenode的主机名及端口、存放临时文件的目录，修改后提交截图信息，命名2-2；（6分）

2.3 hdfs-site.xml文件为HDFS核心配置文件，如文件副本的个数、块大小及是否使用强制权限等，设定HDFS存储文件的副本个数默认为1、SecondaryNameNode地址和端口，修改后提交截图信息，命名2-3；（6分）

2.4 mapred-site.xml本身这个文件是不存在的，将模版文件mapred-site.xml.template改名为mapred-site.xml，然后进行编辑告诉Hadoop mapreduce运行在yarn，修改后提交截图信息，命名2-4；（6分）

2.5 yarn-site.xml文件为Yarn框架配置文件，指定ResourceManger的地址、指定NodeManager获取数据的方式是shuffle，修改后提交截图信息，命名2-5。（6分）

任务三：格式化、启动伪分布式Hadoop (10分)

3.1 格式化DFS(Distributed File System)，在格式化的日志中看到

succefully format字样，就证明格式化成功，并提交截图信息，命名3-1；（5分）

3.2 启动DFS及Yarn，使用命令行命令jps查看运行结果，并提交截图信息，命名3-2。（5分）

任务四：创建HDFS目录，并上传本地文件（10分）

4.1 在HDFS上创建input目录，并提交截图信息，命名4-1；（5分）

4.2在“/home/hadoop/”目录下创建文本文件“note.txt”，内容为“hadoop spark hadoop hive zsf”，使用hdfs shell命令将该文件上传到HDFS的input，并提交截图信息，命名4-2。 （5分）

任务五：伪分布式模式运行Hadoop示例程序(15分)

5.1 使用hadoop jar 命令运行自带示例wordcount程序完成note.txt中单词统计，并提交截图信息，命名5-1；（5分）

5.2 使用命令行查看程序运行结果，并提交截图信息，命名5-2；（5分）

5.3 通过Web 界面查看程序运行结果，并提交截图信息，命名5-3。（5分）

提交要求：

1)在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南职业技术学院 01 张三。

2)考生文件夹内保存截图：1-1、1-2、2-1、2-2、2-3、2-4、2-5、3-1、3-2、4-1、4-2、5-1、5-2、5-3到一个word文档t3.docx中。

（2）实施条件

测试所需的软硬件设备见表1.3.2。

表1.3.2 考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存8G 以上，WIN7及以上操作系统（64位）	
3	截图工具		系统自带截图工具
4	服务器	安装有Ubuntukylin-16.04-desktop-amd64操作系统、单机模式Hadoop	机房/虚拟机

（3）考核时量

考核时间为3个小时。

（4）评分细则

大数据平台搭建与配置模块考核实行 100 分制，评价内容包括工作任务、职业素养完成情况两个方面。其中，工作任务完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表 1.3.3 所示。

表 1.3.3 评分标准表评价内容

评价内容		分值	评分细则	
工作任务（80分）	克隆一台已安装好Hadoop单机模式的虚拟机	15分	未正确克隆扣5分。	8分
			未正确设置主机名，扣3分； 未正确设置IP地址，扣2分。	7分
	修改Hadoop的5个核心配置文件	30分	未正确配置JDK路径扣6分。	6分
			未正确配置namenode的主机名及端口扣3分；未正确配置namenode存放临时文件的目录扣3分。	6分
			未正确设置定HDFS存储文件的副本个数为1，扣3分；未正确设置SecondaryNameNode地址和端口，扣3分。	6分
			未正确将mapred-site.xml.template改名为mapred-site.xml扣3分；未正确设置Hadoop mapreduce 运行在yarn扣3分。	6分
			未正确配置ResourceManger的地址、扣3分；未正确配置NodeManager获取数据的方式是shuffle，扣3分。	6分
	格式化、启动伪分布式Hadoop	10分	格式化DFS未成功，扣5分。	5分
			启动DFS及Yarn，未成功使用命令行命令jps查看程序运行结果，扣5分。	5分
	创建HDFS目录，并上传本地文件	10分	未成功在HDFS上创建input目录，扣5分。	5分
			未成功使用hdfs shell命令将该文件上传到HDFS的input目录，扣5分。	5分

	伪分布式模式运行Hadoop示例程序	15分	未正确使用hadoop jar 命令运行自带示例wordcount程序完成单词统计，扣5分	5分
			未正确使用命令行查看程序运行结果，扣5分； 未正确通过Web 界面查看程序运行结果得，扣5分。	10分
职业素养（20分）	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、鼠标等），检查测试所需软件开发环境。不进行检查操作扣 5 分。	5分
	专业素养	10分	按要求命名文件，截图，答题规范有序得10分。	10分
	职业行为规范	5分	着装干净、整洁。举止文明，遵守考场纪律，按顺序进出考场。	5分

4.试题 1-1-4：鹰眼天下大数据 Hadoop 完全分布式模式部署

（1）任务描述

鹰眼天下作为一家电影服务平台。现在需要通过各个电影网站查询到每部电影的详细信息，帮助使用者查找到电影名、主演、上映时间和评分等数据内容并完成数据展示。

平台现在要进行大数据分析并完成展示，你作为公司大数据工程师，需安装分布式Hadoop环境，企业产品集群一般由多台机器组成，需要在部署于多台机器的Hadoop集群上进行开发。多台机器上运行的Hadoop集群需要部署完全分布式模式的Hadoop。完全分布式模式也叫集群模式，是真正的分布式，由3个及以上的实体机或虚拟机组成的集群。

表1.4.1完全分布式模式Hadoop集群规划

节点角色	虚拟机名	机器IP	主机名	运行进程
主节点	master	192.168.126.200	node	NameNode ResourceManager SecondaryNameNode
从节点	slave1	192.168.126.201	node1	DataNode NodeManager
	slave2	192.168.126.202	node2	DataNode NodeManager

本环节需要完成Hadoop 完全分布式模式部署，主要任务如下：

任务一：已安装好Hadoop伪分布式模式的三台虚拟机、做好完全分布式模式下Hadoop集群规划、并做快照（15分）

1.1 按集群规划修改主机名,并提交截图信息，命名1-1;（5分）

1.2 按集群规划设置IP与主机名映射并提交截图信息，命名1-2;（5分）

1.3 完成以上操作，在主节点上做快照并提交截图信息，命名1-3。（5分）

任务二：配置3台机器两两之间SSH免密登录(20分)

2.1 在node节点上删除原有.ssh目录，然后重新生成密钥对，并提交截图信息，命名2-1；（5分）

2.2 将node节点上的公钥远程拷贝到node、node1、node2的authorized_keys文件,并提交截图信息，命名2-2；（6分）

2.3 查看node节点上的authorized_keys文件，并提交截图信息，命名2-3；（5分）

2.4 验证免密登录，注意查看提示符中主机名称的变化并提交截图信息，命名2-5。（4分）

任务三：修改主节点配置文件并远程拷贝到从节点(35分)

3.1在主节点上修改核心配置文件core-site.xml，并提交截图信息，命名3-1；（5分）

3.2 在主节点上修改HDFS配置文件hdfs-site.xml，并提交截图信息，命名3-2；（5分）

3.3 在主节点上修改MapReduce配置文件mapred-env.sh，并提交截图信息，命名3-3；（5分）

3.4 在主节点上修改hadoop配置文件hadoopp-env.sh，并提交截图信息，命名3-4；（5分）

3.5在主节点上修改Yarn配置文件yarn-site.xml，并提交截图信息，命名3-5；（5分）

3.6 在主节点上修改配置文件slaves文件,并提交截图信息，命名3-6；（5分）

3.7将主节点上的配置文件分发到两个从节点,并提交截图信息，命名3-7。（5分）

任务四：格式化、启动完全分布式Hadoop (10分)

4.1在主节点上格式化HDFS, 在格式化的日志中看到succefully format字样，就证明格式化成功，并提交截图信息，命名4-1；（5分）

4.2在主节点上启动Hadooop，在Web 界面查看程序运行结果，并提交截图信息，命名4-2。（5分）

提交要求：

1)在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南职业技术学院 01 张三。

2)考生文件夹内保存截图：1-1、1-2、1-3、2-1、2-2、2-3、2-4、2-5、3-1、3-2、3-3、3-4、3-5、3-6、3-7、4-1、4-2到一个word文档t4.docx中。

（2）实施条件

测试所需的软硬件设备见表1.4.2。

表1.4.2 考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存8G 以上，WIN7及以上操作系统（64位）	
3	服务器	安装有Ubuntukylin-16.04-desktop-amd64操作系统、伪分布式模式Hadoop	机房/虚拟机

（3）考核时量

考核时间为3个小时。

（4）评分细则

大数据平台搭建与配置模块考核实行 100 分制，评价内容包括工作任务、职业素养完成情况两个方面。其中，工作任务完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表 1.4.3 所示。

表 1.4.3 评分标准表评价内容

评价内容		分值	评分细则	
工作任务(80分)	完全分布式模式下Hadoop集群规划、并做快照	15分	未正确按集群规划修改主机名,扣5分；	5分
			未正确按集群规划设置IP与主机名映射，扣5分；	5分
			未正确在主节点上做快照，扣5分。	5分
	配置3台机器两两之间SSH免密登录	20分	未正确在各节点上生成密钥对，扣5分；	5分
			未正确将node节点上的authorized_keys文件远程拷贝到node1、node2；错一处扣3分；	6分
			未正确查看node节点上的authorized_keys文件，扣5分；	5分
			未成功实现免密登录，扣4分。	4分
	修改主节点配置文件并远程拷贝到从节点	35分	未正确修改核心配置文件core-site.xml，扣5分；	5分
			未正确修改HDFS配置文件hdfs-site.xml，扣5分；	5分
			未正确修改MapReduce配置文件mapred-env.sh，扣5分；	5分
			未正确修改hadoop配置文件hadoop-env.sh，扣5分；	5分
			未正确修改Yarn配置文件yarn-site.xml，扣5分；	5分
			未正确修改配置文件slaves文件，扣5分；	5分
			未将主节点上的配置文件分发到两个从节点,少完成一个节点扣2.5分。	5分
	格式化、	10分	未格式化成功，扣5分；	5分

	启动完全分布式adoop		在主节点上启动Hadooop，不能在Web 界面查看程序运行结果，扣5分。	5分
职业素养（20分）	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、鼠标等），检查测试所需软件开发环境。不进行检查操作扣 5 分。	5分
	专业素养	10分	按要求命名文件，截图，答题规范有序得10分。	10分
	职业行为规范	5分	着装干净、整洁。举止文明，遵守考场纪律，按顺序进出考场。	5分

5.试题 1-1-5：MC 影城大数据 Hadoop 完全分布式模式部署

（1）任务描述

MC影城是一家电影综合服务平台。现在需要通过各个电影网站查询到每部电影的详细信息，帮助使用者查找到电影名、主演、上映时间和评分等数据内容并完成数据展示。

平台现在要进行大数据分析并完成展示，你作为公司大数据工程师，需安装分布式Hadoop环境，企业产品集群一般由多台机器组成，需要在部署于多台机器的Hadoop集群上进行开发。多台机器上运行的Hadoop集群需要部署完全分布式模式的Hadoop。完全分布式模式也叫集群模式，是真正的分布式，由3个及以上的实体机或虚拟机组成的集群。

表1.5.1完全分布式模式Hadoop集群规划

节点角色	虚拟机名	机器IP	主机名	运行进程
主节点	master	192.168.126.200	node	NameNode ResourceManager SecondaryNameNode
从节点	slave1	192.168.126.201	node1	DataNode NodeManager
	slave2	192.168.126.202	node2	DataNode NodeManager

本环节需要完成Hadoop 完全分布式模式部署，主要任务如下：

任务一：配置3台机器两两之间SSH免密登录(20分)

1.1 在各节点上删除原有.ssh目录，然后重新生成密钥对并提交截图信息，命名1-1；(5分)

1.2 将node节点上的公钥远程拷贝到node、node1、node2，并提交截图信息，命名1-2；(6分)

1.3 查看node节点上的authorized_keys文件，并提交截图信息，命名1-3；(5分)

1.4 验证免密登录，注意查看提示符中主机名称的变化，并提交截图信息，命名1-4。(4分)

任务二：修改主节点配置文件并远程拷贝到从节点(25分)

2.1在主节点上修改核心配置文件core-site.xml，并提交截图信息，命名2-1；（5分）

2.2 在主节点上修改HDFS配置文件hdfs-site.xml，并提交截图信息，命名2-2；（5分）

2.3 在主节点上修改MapReduce配置文件mapred-env.sh，并提交截图信息，命名2-3；（5分）

2.4 在主节点上修改配置文件slaves和yarn-site.xml文件,并提交截图信息，命名2-4；（5分）

2.5将主节点上的配置文件分发到两个从节点,并提交截图信息，命名2-5。（5分）

任务三：格式化、启动完全分布式Hadoop (10分)

3.1 在主节点上格式化HDFS, 在格式化的日志中看到succefully format字样，就证明格式化成功，并提交截图信息，命名3-1；（5分）

3.2 在主节点上启动Hadooop，在各节点上使用命令行命令jps查看进程，并提交截图信息，命名3-2。（5分）

任务四：创建HDFS目录，并上传本地文件（10分）

4.1 在HDFS上创建input目录，并提交截图信息，命名4-1；（5分）

4.2 在“/home/hadoop”目录下创建文本文件“note.txt”， 内容为“hadoop spark hadoop hive zsf”，使用hdfs shell命令将该文件上传到HDFS的input目录，并提交截图信息，命名4-2。 （5分）

任务五：在完全分布式模式下使用wordcount示例程序完成单词统计(15分)

5.1 使用hadoop jar 命令运行自带示例wordcount程序完成单词统计，并提交截图信息，命名5-1；（5分）

5.2 使用命令行查看程序运行结果，并提交截图信息，命名5-2；（5分）

5.3 通过Web 界面查看程序运行结果，并提交截图信息，命名5-3。（5分）

提交要求：

1)在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南职业技术学院 01 张三。

2)考生文件夹内保存截图：1-1、1-2、1-3、1-4、1-5、2-1、2-2、2-3、2-4、2-5、3-1、3-2、4-1、4-2、5-1、5-2、5-3到一个word文档t5.docx中。

（2）实施条件

测试所需的软硬件设备见表1.5.2。

表1.5.2 考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存8G 以上，WIN7及以上操作系统（64位）	
3	服务器	安装有Ubuntukylin-16.04-desktop-amd64操作系统、伪分布式模式Hadoop	机房/虚拟机

（3）考核时量

考核时间为3个小时。

（4）评分细则

大数据平台搭建与配置模块考核实行 100 分制，评价内容包括工作任务、职业素养完成情况两个方面。其中，工作任务完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表 1.5.3 所示。

表 1.5.3 评分标准表评价内容

评价内容		分值	评分细则	
工作任务(80分)	配置3台机器两两之间SSH免密登录	20分	未正确在各节点上生成密钥对，扣5分；	5分
			未正确将node节点上的authorized_keys文件远程拷贝到node、node1、node2；错一处扣2分；	6分
			未正确查看node节点上的authorized_keys文件，扣5分；	5分
			未成功实现免密登录，扣4分。	4分
	修改主节点配置文件并远程拷贝到从节点	25分	未正确修改核心配置文件core-site.xml，扣5分；	5分
			未正确修改HDFS配置文件hdfs-site.xml，扣5分；	5分
			未正确修改MapReduce配置文件mapred-env.sh，扣5分；	5分
			未正确修改配置文件slaves文件，扣2分，未正确修改配置文件yarn-site.xml文件，扣3分；	5分
			未将主节点上的配置文件分发到两个从节点,少完成一个节点扣2.5分。	5分
	格式化、启动完全分布式Hadoop	10分	未格式化成功，扣5分；	5分
			在主节点上启动Hadooop，不能在Web 界面查看程序运行结果，扣5分。	5分
	创建HDFS目录，并上传本地文件	10分	未成功在HDFS上创建input目录，扣5分。	5分
			未成功使用hdfs shell命令将指定文件上传到HDFS上的input目录，扣5分。	5分
	wordcount示例程	15分	未正确使用hadoop jar 命令运行自带示例wordcount程序完成单词统计，扣5分	5分

	序完成单词统计		未正确使用命令行查看程序运行结果，扣5分；	5分
			未正确通过Web 界面查看程序运行结果得，扣5分。	5分
职业素养（20分）	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、鼠标等），检查测试所需软件开发环境。不进行检查操作扣 5 分。	5分
	专业素养	10分	按要求命名文件，截图，答题规范有序得10分。	10分
	职业行为规范	5分	着装干净、整洁。举止文明，遵守考场纪律，按顺序进出考场。	5分

6.试题 1-1-6：途书大数据 Hadoop 完全分布式模式部署

（1）任务描述

途书作为一家网络小说服务平台。现在需要通过各个小说网站查询到每部小说的详细信息，帮助使用者查找到书名、作者、上架时间和评价等数据内容并完成数据展示。

平台现在要进行大数据分析并完成展示，你作为公司大数据工程师，需安装分布式Hadoop环境，企业产品集群一般由多台机器组成，需要在部署于多台机器的Hadoop集群上进行开发。多台机器上运行的Hadoop集群需要部署完全分布式模式的Hadoop。完全分布式模式也叫集群模式，是真正的分布式，由3个及以上的实体机或虚拟机组成的集群。

表1.6.1完全分布式模式Hadoop集群规划

节点角色	虚拟机名	机器IP	主机名	运行进程
主节点	master	192.168.126.200	node	NameNode ResourceManager SecondaryNameNode
从节点	slave1	192.168.126.201	node1	DataNode NodeManager
	slave2	192.168.126.202	node2	DataNode NodeManager

本环节需要完成Hadoop 完全分布式模式部署，主要任务如下：

任务一：已安装好Hadoop伪分布式模式的三台虚拟机、做好完全分布式模式下Hadoop集群规划、并做快照（15分）

1.1 按集群规划修改主机名,并提交截图信息，命名1-1;（5分）

1.2 按集群规划设置IP与主机名映射并提交截图信息，命名1-2;（5分）

1.3 完成以上操作，在主节点上做快照并提交截图信息，命名1-3。（5分）

任务二：配置主节点对自身和 2 个从节点之间SSH免密登录(20分)

2.1 在各节点上删除原有.ssh目录，然后重新生成密钥对，并提交截图信息，命名2-1；（5分）

2.2 将node节点上的公钥远程拷贝到node、node1、node2的authorized_keys文件,并提交截图信息，命名2-2；（6分）

2.3 查看node节点上的authorized_keys文件，并提交截图信息，命名2-3；（5分）

2.4 验证免密登录，注意查看提示符中主机名称的变化并提交截图信息，命名2-4。（4分）

任务三：优化主节点配置文件并远程拷贝到从节点(30分)

3.1 hadoop访问文件的IO操作都需要通过代码库，不论是对硬盘或者是网络操作来讲，较大的缓存都可以提供更高的数据传输，但这也就意味着更大的内存消耗和延迟，在主节点上修改核心配置文件core-site.xml，设置流文件的缓冲区为64K，并提交截图信息，命名3-1；（6分）

3.2 在主节点上设置块复制的最小数量为1，请修改HDFS配置文件hdfs-site.xml，并提交截图信息，命名3-2；（6分）

3.3在主节点上设置块复制的最大数量为64，请修改HDFS配置文件hdfs-site.xml，并提交截图信息，命名3-3；（6分）

3.4在主节点上设置排序文件的内存缓存大小为128M，请修改mapred-site.xml，并提交截图信息，命名3-4；（6分）

3.5将主节点上的配置文件分发到两个从节点,并提交截图信息，命名3-5。（6分）

任务四：格式化、启动完全分布式Hadoop (15分)

4.1在主节点上格式化HDFS, 在格式化的日志中看到succefully format字样，就证明格式化成功，并提交截图信息，命名4-1；（5分）

4.2在主节点上启动Hadooop，在Web 界面查看程序运行结果，并提交截图信息，命名4-2。（10分）

提交要求：

1)在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南职业技术学院 01 张三。

2)考生文件夹内保存截图：1-1、1-2、1-3、2-1、2-2、2-3、2-4、2-5、3-1、3-2、3-3、3-4、3-5、4-1、4-2到一个word文档t6.docx中。

（2）实施条件

测试所需的软硬件设备见表1.6.2。

表1.6.2 考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存8G 以上，WIN7及以上操作系统（64位）	
3	服务器	安装有Ubuntukylin-16.04-desktop-amd64操作系统、伪分布式模式Hadoop	机房/虚拟机

（3）考核时量

考核时间为3个小时。

（4）评分细则

大数据平台搭建与配置模块考核实行 100 分制，评价内容包括工作任务、职业素养完成情况两个方面。其中，工作任务完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表 1.6.3 所示。

表 1.6.3 评分标准表评价内容

评价内容		分值	评分细则	
	完全分布式模式下	15分	未正确按集群规划修改主机名,扣5分；	5分

工作任务（80分）	Hadoop集群规划、并做快照		未正确按集群规划设置IP与主机名映射，扣5分；	5分
			未正确在主节点上做快照，扣5分。	5分
	配置3台机器两两之间SSH免密登录	20分	未正确在各节点上生成密钥对，扣5分；	5分
			未正确将node节点上的authorized_keys文件远程拷贝到node、node1、node2；错一处扣2分；	6分
			未正确查看node节点上的authorized_keys文件，扣5分；	5分
			未成功实现免密登录，扣4分。	4分
	优化主节点配置文件并远程拷贝到从节点	30分	未正确设置流文件的缓冲区为64K，扣6分；	6分
			未正确设置块复制的最小数量为1，扣6分；	6分
			未正确设置块复制的最大数量为64块，扣6分；	6分
			未正确设置排序文件的内存缓存大小为128M，扣6分；	6分
			未将主节点上的配置文件分发到两个从节点,少完成一个节点扣3分。	6分
	格式化、启动完全分布式adoop	15分	未格式化成功，扣5分；	5分
			在主节点上启动Hadooop，不能在Web 界面查看程序运行结果，扣10分。	10分
职业素养（20分）	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、鼠标等），检查测试所需软件开发环境。不进行检查操作扣 5 分。	5分
	专业素养	10分	按要求命名文件，截图，答题规范有序得10分。	10分
	职业行为规范	5分	着装干净、整洁。举止文明，遵守考场纪律，按顺序进出考场。	5分

7.试题 1-1-7：快乐选导购大数据 Hadoop 完全分布式模式部署

（1）任务描述

快乐选导购平台通过各个电商平台网站查询到各类商品的详细信息，帮助使用者查找到品牌、生产厂商、上架时间和评分等数据内容并完成数据展示。

现在要进行大数据分析并完成展示，你作为公司大数据工程师，需安装分布式Hadoop环境，企业产品集群一般由多台机器组成，需要在部署于多台机器的Hadoop集群上进行开发。多台机器上运行的Hadoop集群需要部署完全分布式模式的Hadoop。完全分布式模式也叫集群模式，是真正的分布式，由3个及以上的实体机或虚拟机组成的集群。

表1.7.1完全分布式模式Hadoop集群规划

节点角色	虚拟机名	机器IP	主机名	运行进程
主节点	master	192.168.126.200	node	NameNode ResourceManager SecondaryNameNode
从节点	slave1	192.168.126.201	node1	DataNode NodeManager
	slave2	192.168.126.202	node2	DataNode NodeManager

本环节需要完成Hadoop 完全分布式模式部署，主要任务如下：

任务一：配置3台机器两两之间SSH免密登录(20分)

1.1 在各节点上删除原有.ssh目录，然后重新生成密钥对，并提交截图信息，命名1-1；(5分)

1.2 将node节点上的公钥远程拷贝到node、node1、node2，，并提交截图信息，命名1-2；(6分)

1.3 查看node节点上的authorized_keys文件，，并提交截图信息，命名1-3；(5分)

1.4 验证免密登录，注意查看提示符中主机名称的变化，，并提交截图信息，命名1-4。(4分)

任务二：优化主节点配置文件并远程拷贝到从节点(30分)

2.1 hadoop访问文件的IO操作都需要通过代码库，不论是对硬盘或者是网络操

作来讲，较大的缓存都可以提供更高的数据传输，但这也就意味着更大的内存消耗和延迟，在主节点上修改核心配置文件core-site.xml，设置流文件的缓冲区为128K，并提交截图信息，命名2-1；（6分）

2.2 在主节点上设置磁盘空间统计间隔为5秒，请修改配置文件hdf-site.xml，并提交截图信息，命名2-2；（6分）

2.3 在主节点上设置每个作业缺省的map任务数为3，请修改配置文件mapred-site.xml，并提交截图信息，命名2-3；（6分）

2.4在主节点上设置每个作业缺省的reduce任务数为2，请修改配置文件mapred-site.xml，并提交截图信息，命名2-4；（6分）

2.5将主节点上的配置文件分发到两个从节点,并提交截图信息，命名2-5。（6分）

任务三：格式化、启动完全分布式Hadoop（10分）

3.1 在主节点上格式化HDFS，在格式化的日志中看到succefully format字样，就证明格式化成功，并提交截图信息，命名3-1；（5分）

3.2 在主节点上启动Hadooop，在各节点上使用命令行命令jps查看进程，并提交截图信息，命名3-2。（5分）

任务四：创建HDFS目录，并上传本地文件（8分）

4.1 在HDFS上创建input目录，并提交截图信息，命名4-1；（4分）

4.2在“/home/hadoop”目录下创建文本文件“note.txt”， 内容为“hadoop spark hadoop hive zsf”，使用hdfs shell命令将该文件上传到HDFS的input目录，并提交截图信息，命名4-2。 （4分）

任务五：在完全分布式模式下使用wordcount示例程序完成单词统计(12分)

5.1 使用hadoop jar 命令运行自带示例wordcount程序完成单词统计，并提交截图信息，命名5-1；（4分）

5.2 使用命令行查看程序运行结果，并提交截图信息，命名5-2；（4分）

5.3 通过Web 界面查看程序运行结果，并提交截图信息，命名5-3。（4分）

提交要求：

1)在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南职业技术学院 01 张三。

2)考生文件夹内保存截图：1-1、1-2、1-3、1-4、1-5、2-1、2-2、2-3、2-4、2-5、

3-1、3-2、4-1、4-2、5-1、5-2、5-3到一个word文档t7.docx中。

（2）实施条件

测试所需的软硬件设备见表1.7.2。

表1.7.2 考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存8G 以上，WIN7及以上操作系统（64位）	
3	服务器	安装有Ubuntukylin-16.04-desktop-amd64操作系统、伪分布式模式Hadoop	机房/虚拟机

（3）考核时量

考核时间为3个小时。

（4）评分细则

大数据平台搭建与配置模块考核实行 100 分制，评价内容包括工作任务、职业素养完成情况两个方面。其中，工作任务完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表 1.7.3 所示。

表 1.7.3 评分标准表评价内容

评价内容		分值	评分细则	
工作任务（80分）	配置3台机器两两之间SSH免密登录	20分	未正确在各节点上生成密钥对，扣5分；	5分
			未正确将node节点上的authorized_keys文件远程拷贝到node、node1、node2；错一处扣2分；	6分
			未正确查看node节点上的authorized_keys文件，扣5分；	5分
			未成功实现免密登录，扣4分。	4分
	优化主节	30分	未正确设置流文件的缓冲区为128K，扣6分；	6分

	点配置文件并远程拷贝到从节点		未正确设置设置磁盘空间统计间隔为5秒，扣6分；	6分
			未正确设置每个作业缺省的map任务数为3，扣6分；	6分
			未正确设置每个作业缺省的reduce任务数为2，扣6分；	6分
			未将主节点上的配置文件分发到两个从节点,少完成一个节点扣3分。	6分
	格式化、启动完全分布式Hadoop	10分	未格式化成功，扣5分；	5分
			在主节点上启动Hadooop，不能在Web 界面查看程序运行结果，扣5分。	5分
	创建HDFS目录，并上传本地文件	8分	未成功在HDFS上创建input目录，扣4分。	4分
			未成功使用hdfs shell命令将指定文件上传到HDFS上的input目录，扣4分。	4分
	wordcount示例程序完成单词统计	12分	未正确使用hadoop jar 命令运行自带示例wordcount程序完成单词统计，扣4分	4分
			未正确使用命令行查看程序运行结果，扣4分；	4分
			未正确通过Web 界面查看程序运行结果得，扣4分。	4分
职业素养（20分）	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、鼠标等），检查测试所需软件开发环境。不进行检查操作扣 5 分。	5分
	专业素养	10分	按要求命名文件，截图，答题规范有序得10分。	10分
	职业行为规范	5分	着装干净、整洁。举止文明，遵守考场纪律，按顺序进出考场。	5分

8.试题 1-1-8：Hadoop 高可用(HDFS)

（1）任务描述

某企业Hadoop大数据处理平台由三台服务器组成，采用的是一主两从的架构，

其中一台用作master节点服务器，另两台用作slave节点服务器。因为NameNode是HDFS主从架构中的主节点守护进程，其中存储了HDFS上文件数据的元数据信息，NameNode一旦宕机，其上存储的文件元数据信息就会立刻从内存中丢失 ，HDFS处于瘫痪状态，一切对外提供服务的功能全部丧失，集群无法正常运行。

为了防止出现上述故障，该企业在完全分布式集群上打造了高可用的HDFS集群。作为企业运维工程师，需要对该高可用集群进行日常维护和监控，确保HDFS的高可用HA机制处于正常状态。

表1.8.1高可用集群规划

hostname	IP	NameNode	JournalNode	DataNode	ZFKC	ZooKeeper	ResourceManager	NodeManager
node08	192.168.xxx.180	✓	✓	✓	✓	✓	✓	✓
node1	192.168.xxx.181	✓	✓	✓	✓	✓	✓	✓
node2	192.168.xxx.182		✓	✓		✓		✓

任务一：检查准备设备(25分)

1.1 在VMware Workstation上分别导入主从节点三台虚拟机同时转至HDFS高可用集群预备状态，启动和登录一主（master08）和两从（slave1和slave2）节点，启动成功则截图主节点状态并保存，图片保存到考生文件夹下，并命名为“1-1”。（5分）

1.2 按规划要求分别在各节点上设置IP地址，并通过ping命令，在主节点上检查和slave1及slave2的互连情况，截图并保存，图片保存到考生文件夹下，并命名为“1-2”。（5分）

1.3按规划要求修改主从节点主机名，修改好之后hostname查看，截图并保存，图片保存到考生文件夹下，并命名为“1-3”。（5分）

1.4用Xshell分别登录主从节点，在Xshell上同时显示三台虚拟机的登录状态和查看IP地址截图并保存，图片保存到考生文件夹下，并命名为“1-4”。（5分）

1.5 分别修改一主（master08）和两从（slave1和slave2）节点的/etc/hosts文件，配置IP 地址与主机名之间的映射关系，在Xshell上同时显示三台虚拟机修改后的hosts文件状态截图并保存，图片保存到考生文件夹下，并命名为“1-5”。（5

分）

任务二：审查部署Hadoop HDFS NameNode高可用(25分)

2.1主节点node08切换到hadoop目录下，输入核心配置文件core-site.xml的打开命令，检查/修改两个NameNode的地址是否组装成一个集群mycluster，检查hadoop运行时产生文件的存储目录的指定情况，截图并保存检查和修改情况，图片保存到考生文件夹下，并命名为“2-1”。（5分）

2.2 主节点node08保持在hadoop目录下，输入HDFS配置文件hdfs-site.xml的打开命令，检查/设置dfs副本数为2，确定完全分布式集群名称为mycluster、确定集群中NameNode节点为nn1(主机名为node08)和nn2(主机名为node1) ，截图并保存检查和修改情况，图片保存到考生文件夹下，并命名为“2-2”。（5分）

2.3 slave1从节点node1切换到hadoop目录下，输入HDFS配置文件hdfs-site.xml的打开命令，检查/设置nn1和nn2的RPC通信地址分别为node08:8020,node1:8020; nn1和nn2的http通信地址分别为node08:50070,node1:50070，截图并保存检查和修改情况，图片保存到考生文件夹下，并命名为“2-3”。（5分）

2.4 slave2从节点node2切换到hadoop目录下，输入HDFS配置文件hdfs-site.xml的打开命令，检查/设置权限检查为false关闭状态、指定NameNode元数据在JournalNode上的存放位置分别为node08:8485;node1:8485;node2:8485、配置sshfence隔离机制、设置使用隔离机制时的ssh无秘钥登录、声明Journalnode服务器存储目录、指定mycluster出现故障时负责执行故障切换的类和配置故障转移为自动true，截图并保存检查和修改情况，图片保存到考生文件夹下，并命名为“2-4”。（5分）

2.5把node08、node1、node2三节点同时切换到hadoop目录下，修改DataNode配置文件slaves文件中的主节点名为node08，截图并保存检查和修改情况，图片保存到考生文件夹下，并命名为“2-5”。（5分）

任务三：HDFS高可用集群启动进程查看及验证HA故障自动转移(20分)

3.1用jps查看各个节点上的进程，在Xshell上同时显示三台虚拟机的进程状态截图并保存，图片保存到考生文件夹下，并命名为“3-1”。（5分）。

3.2在主节点node08上分别查看nn1和nn2两个namenode的节点状态，截图并保存，图片保存到考生文件夹下，并命名为“3-2”。（5分）。

3.3手动创造故障即kill掉主节点node08上的namenode进程后再查看nn1和nn2两

个namenode的节点状态，截图并保存，图片保存到考生文件夹下，并命名为“3-3”。（5分）。

3.4在主节点node08上通过hadoop-daemon.sh start重启发生故障后的节点上的namenode进程，再次查看nn1和nn2两个节点的namenode状态，截图并保存，图片保存到考生文件夹下，并命名为“3-4”。（5分）。

任务四：撰写报告(10分)

4.1在考生文件夹下，新建word文档，撰写报告：

（1） 归纳总结HDFS高可用集群的搭建意义。（5分）

（2） 归纳总结HDFS高可用集群的搭建过程及注意事项。（5分）

4.2保存word文档到考生文件夹下，并命名为 “项目报告”。

提交要求：

1) 在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南职业技术学院 01 张三。

2)考生文件夹内保存截图：1-1、1-2、2-1、2-2、2-3、3-1、3-2、3-3、3-4、3-5、4-1、4-2、4-3、4-4到一个word文档t8.docx中。

（2）实施条件

测试所需的软硬件设备见表1.8.2。

表1.8.2 考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存8G 以上，WIN7及以上操作系统（64位）	
3	镜像文件8	安装有Ubuntukylin-16.04-desktop-amd64操作系统、安装好zookeeper的完全分布式Hadoop(安装了sudo apt install psmisc、时间事先同步)	机房/虚拟机

（3）考核时量

考核时间为3个小时。

（4）评分细则

大数据平台搭建与配置模块考核实行 100 分制，评价内容包括工作任务、职业素养完成情况两个方面。其中，工作任务完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表 1.8.3 所示。

表 1.8.3 评分标准表评价内容

评价内容		分值	评分细则	
工作任务（80分）	任务一：检查设备	25分	正确启动和登录主、从节点。未正确启动和登录扣5分。	5分
			正确检查主从节点互联情况。未按集群规划要求完成IP地址修改扣3分，主从节点未能ping通扣2分。	5分
			正确修改主机名。未正确修改主节点主机名扣4分；未能正确查看主节点主机名扣1分。	5分
			远程登录主从节点三台虚拟机。未能在Xshell登录主从机的扣3分，不能在Xshell上查看IP地址扣2分。	5分
			正确修改主从节点的/etc/hosts文件。未能正确配置IP地址与主机名之间的映射关系扣5分。	5分
	任务二：审查部署Hadoop HDFS NameNode高可用	25分	正确查看核心配置文件。未能在主节点node08正确切换到hadoop目录下扣2分；未能按要求正确审核和截图core-site.xml中配置情况扣3分。	5分
			正确查看HDFS配置文件。未能正确打开HDFS配置文件hdfs-site.xml扣2分；未能按要求正确审核和截图集群中NameNode节点为nn1和nn2扣3分。	5分
			正确查看HDFS配置文件。未能按要求正确审核和截图nn1和nn2的RPC通信地址扣2.5分；未能按要求正确审核和截图nn1和nn2的http通信地址扣2.5分；	5分
			正确查看HDFS配置文件。未能正确指定NameNode元数据在JournalNode上存放位置、配置sshfence隔离机制、设置使用隔离机制时的ssh无秘钥登录、声明	5分

			Journalnode服务器存储目录、指定mycluster出现故障时负责执行故障切换的类分别扣1分。	
			正确修改DataNode配置文件slaves文件。未能正确切换到hadoop目录下扣2分；未按要求正确修改主节点名扣3分。	5分
	任务三：验证HA故障自动转移	20分	正确查看各个节点上进程。未能在Xshell上同时显示三台虚拟机的进程状态扣2分；检查HDFS HA集群进程不正常，扣3分。	5分
			正确查看namenode的节点状态。未能正确查看nn1和nn2两个namenode的节点状态，扣5分。	5分
			正确手动创造故障。nn2未能在手动创造nn1故障后状态变为active，扣5分。	5分
			正确重启发生故障后节点的namenode进程。nn1未能正确重启，扣5分。	5分
	任务四：撰写报告	10分	正确撰写报告。未能正确归纳总结HDFS高可用集群的搭建意义，扣5分。未能归纳总结好HDFS高可用集群的搭建过程及注意事项，扣5分。	10分
职业素养（20分）	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、鼠标等），检查测试所需软件开发环境。 不进行检查操作扣 5 分。	5分
	专业素养	10分	电脑等工具使用操作规范、按要求命名和存放文件，操作不规范扣5分，未按要求答题扣5分	10分
	整体形象	5分	着装干净整洁，举止文明。遵守考场纪律，按顺序进出考场，着装不整洁扣2分，举止不文明每次扣1分，扣完3分为止	5分

9.试题 1-1-9：Hadoop 高可用(yarn)

（1）任务描述

某企业Hadoop大数据处理平台由三台服务器组成，采用的是一主两从的架构，其中一台用作master节点服务器，另两台用作slave节点服务器。因为如果一个集群中只存在一个ResourceManager，会存在单点故障而导致一切对外提供服务的功能全部丧失，整个集群无法正常运行。

为了防止出现上述故障，该企业在完全分布式集群上打造了高可用的Yarn集群，作为企业运维工程师，需要对该高可用集群进行日常维护和监控，确保Yarn的HA机

制处于正常状态。

表1.9.1高可用集群规划

hostname	IP	NameNode	JournalNode	DataNode	ZFKC	ZooKeeper	ResourceManager	NodeManager
node09	192.168.xxx.190	✓	✓	✓	✓	✓	✓	✓
node1	192.168.xxx.191	✓	✓	✓	✓	✓	✓	✓
node2	192.168.xxx.192		✓	✓		✓		✓

任务一：检查设备(15分)

1.1 在VMware Workstation上分别导入主从节点三台虚拟机同时转至Yarn高可用集群预备状态，启动和登录一主（master09）和两从（slave1和slave2）节点，启动成功则截图主节点状态并保存，图片保存到考生文件夹下，并命名为“1-1” 。（5分）

1.2 分别在各节点上设置主机名、IP地址和映射文件hosts，并通过测试网络连接量的程序因特网包探索器PING （Packet Internet Groper），在主节点上检查和slave1及slave2的互连情况，截图并保存，图片保存到考生文件夹下，并命名为“1-2”。（5分）

1.3启动HDFS HA集群后，利用jps查看，截图三个节点jps查看结果并保存，图片保存到考生文件夹下，并命名为“1-3”。（5分）。

任务二：审查部署Hadoop Yarn高可用(35分)

2.1关闭HDFS HA集群后，主节点node09切换到hadoop目录下，输入核心配置文件yarn-site.xml的打开命令，检查/修改resourcemanager ha的启用状态为enabled，截图并保存检查和修改情况，图片保存到考生文件夹下，并命名为“2-1”。（7分）

2.2 保持主节点node09切换在hadoop目录下核心配置文件yarn-site.xml的打开，检查/设置resourcemanager的cluster-id为rmCluster，声明两台resourcemanager的名字分别为rm1、rm2，截图并保存检查和修改情况，图片保存到考生文件夹下，并命名为“2-2”。（5分）

2.3保持主节点node09切换在hadoop目录下核心配置文件yarn-site.xml的打开，

检查/设置第一台和第二台resourcemanager的地址分别为node09、node1,指定zookeeper集群的地址为node09:2181,node1:2181,node2:2181，截图并保存检查和修改情况，图片保存到考生文件夹下，并命名为“2-3”。（5分）

2.4保持主节点node09切换在hadoop目录下核心配置文件yarn-site.xml的打开，检查/设置yarn自动恢复启用为true、指定resourcemanager的状态信息存储在zookeeper集群址，截图并保存检查和修改情况，图片保存到考生文件夹下，并命名为“2-4”。（5分）

2.5把node09、node1、node2三节点同时切换到hadoop目录下，查看DataNode配置文件slaves文件中的主从节点名，截图并保存检查情况，图片保存到考生文件夹下，并命名为“2-5”。（3分）

2.6将主节点node09上修改好的配置文件yarn-site.xml分发到两个从节点node1和node2，截图分发到从节点node1、node2的命令及过程，命名为“2-6”。（10分）

任务三：Yarn高可用集群启动进程查看及验证HA故障自动转移（20分）

3.1启动Hadoop HA集群后，用jps查看各个节点上的进程，在Xshell上同时显示三台虚拟机的进程状态截图并保存，图片保存到考生文件夹下，并命名为“3-1”。（10分）。

3.2在主节点node09上分别查看nn1和nn2查看两个ResourceManager的节点状态，截图并保存，图片保存到考生文件夹下，并命名为“3-2”。（5分）。

3.3手动创造故障即kill掉主节点node09上ResourceManager进程后再查看rm1和rm2状态，截图并保存，图片保存到考生文件夹下，并命名为“3-3”。（5分）。

任务四：撰写报告（10分）

4.1在考生文件夹下，word文档t9.docx中，撰写报告：

（1） 归纳总结Yarn高可用集群的搭建意义。（5分）

（2） 归纳总结Yarn高可用集群的搭建过程及注意事项。（5分）

4.2保存word文档到考生文件夹下，并命名为 “项目报告.docx”。

提交要求：

1)在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南职业技术学院 01 张三。

2)考生文件夹内保存截图：1-1、1-2、1-3、2-1、2-2、2-3、2-4、2-5、2-6、3-1、3-2、3-3到一个word文档t9.docx中。

（2）实施条件

测试所需的软硬件设备见表1.9.2。

表1.9.2 考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存8G 以上，WIN7及以上操作系统（64位）	
3	服务器	安装有Ubuntukylin-16.04-desktop-amd64操作系统、完成了HDFS高可用的完全分布式Hadoop集群(yum install psmisc、时间同步)	机房/虚拟机

（3）考核时量

考核时间为3个小时。

（4）评分细则

大数据平台搭建与配置模块考核实行 100 分制，评价内容包括工作任务、职业素养完成情况两个方面。其中，工作任务完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表 1.9.3 所示。

表 1.9.3 评分标准表评价内容

评价内容		分值	评分细则	
工作任务（80分）	任务一：检查设备	15分	正确启动和登录主、从节点。未正确启动和登录扣5分。	5分
			正确检查主从节点互联情况。未正确查看IP地址扣2分，主从节点未能ping通扣3分。	5分
			正确利用jps查看进程。检查HDFS HA集群进程不正常，扣5分。	5分
	任务二：审查部署Hadoop	35分	正确切换目录及检查resourcemanager ha的启用状态。未能正确在主节点node09切换到hadoop目录下，扣3分；未能正确输入核心配置文件yarn-site.xml的打开	7分

	Yarn高可用		命令，扣2分；未能正确检查resourcemanager ha的启用状态，扣2分。	
			正确检查cluster-id和声明resourcemanager。未能正确检查resourcemanager的cluster-id扣2分；未能正确截图两台resourcemanager的声明名字，扣3分。	5分
			正确检查resourcemanager的地址和指定zookeeper集群的地址。未能正确检查第一台和第二台resourcemanager的地址扣2分；未能正确截图zookeeper集群的指定地址，扣3分。	5分
			正确检查yarn自动恢复启用状态及resourcemanager的状态信息存储。未能正确检查yarn自动恢复启用为true状态扣2分；未能正确检查resourcemanager的状态信息存储在zookeeper集群址扣3分。	5分
			正确查看slaves文件中的主从节点名。未能正确打开slaves文件扣2分；未能正确审核DataNode配置文件主从节点名为node09、node1和node2扣1分。	3分
			正确远程分发配置文件信息。未能正确输出远程分发命令扣4分；未能正确将主节点node09上修改好的配置文件yarn-site.xml分发到从节点node1和node2，分别扣3分。	10分
	任务三：验证HA故障自动转移	20分	正确利用jps查看进程。未能在Xshell上同时显示三台虚拟机的进程状态扣3分；检查Hadoop HA集群进程不正常，扣7分。	10分
			正确查看节点状态。未能正确查看两个ResourceManager的节点状态扣5分。	5分
			正确验证HA故障自动转移。手动创造故障查看rm1和rm2故障转移未成功扣 5分。	5分
	任务五：撰写报告	10分	正确撰写报告。未能正确归纳总结Yarn高可用集群的搭建意义，扣5分。未能归纳总结好Yarn高可用集群的搭建过程及注意事项，扣5分。	10分
职业素养（20分）	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、鼠标等），检查测试所需软件开发环境。 不进行检查操作扣 5 分。	5分

	专业素养	10分	电脑等工具使用操作规范、按要求命名和存放文件，操作不规范扣5分，未按要求答题扣5分	10分
	整体形象	5分	着装干净整洁，举止文明。遵守考场纪律，按顺序进出考场，着装不整洁扣2分，举止不文明每次扣1分，扣完3分为止	5分

项目 2：Hadoop 生态圈其它组件搭建与配置

10.试题 1-2-1：Zookeeper 组件安装

（1）任务描述

某企业产品集群由多台机器组成，采用的是一主多从的架构。因为NameNode是HDFS主从架构中的主节点守护进程，其中存储了HDFS上文件数据的元数据信息，NameNode一旦宕机，其上存储的文件元数据信息就会立刻从内存中丢失 ，HDFS处于瘫痪状态，一切对外提供服务的功能全部丧失。同样，如果一个集群中只存在一个ResourceManager，也会存在单点故障而导致整个集群无法正常运行。ZooKeeper是一个分布式的，开放源码的分布式应用程序协调服务，是Google的Chubby一个开源的实现，是Hadoop和Hbase的重要组件。它是一个为分布式应用提供一致性服务的软件，提供的功能包括：配置维护、域名服务、分布式同步、组服务等。

为了打造高可用的集群，防止出现上述故障，企业需要实现高可用集群的搭建，而要完成高可用集群的搭建，作为企业运维工程师首先需要进行Zookeeper组件的安装，本环节需要完成ZooKeeper安装部署，主要任务如下：

表1.10.1高可用集群规划

节点角色	虚拟机名	机器IP	主机名
主节点	master2-1	192.168.126.210	node2-1
从节点	slave1	192.168.126.211	node1
	slave2	192.168.126.212	node2

任务一：准备和检查设备（30分）

1.1根据集群规划，导入主节点master2-1，在此基础上克隆两台虚拟机作为从

节点，分别命名为slave1和slave2，并在VMware Workstation上新建“Zookeeper集群”文件夹，master2-1、slave1和slave2均放入上述文件夹内，截图并保存，图片保存到考生文件夹下，并命名为“1-1”。（10分）

1.2修改master2-1、slave1和slave2的主机名分别为node2-1、node1和node2,之后用hostname查看主节点master2-1，截图并保存，图片保存到考生文件夹下，并命名为“1-2”。（5分）

1.3 修改主从节点三台虚拟机的IP地址，其余按照“表2.1.1集群规划”的要求，修改完成之后用Xshell分别登录，在Xshell上同时显示三台虚拟机的登录状态和查看IP地址截图并保存，图片保存到考生文件夹下，并命名为“1-3”。（5分）

1.4 分别在各节点上查看IP地址，并通过ping命令，在主节点master2-1上检查和slave1及slave2的互连情况，截图并保存，图片保存到考生文件夹下，并命名为“1-4”。（5分）

1.5 分别修改主（node2-1）和从（node1和node2）节点的/etc/hosts文件，配置IP 地址与主机名之间的映射关系，在Xshell上同时显示三台虚拟机修改后的hosts文件状态截图并保存，图片保存到考生文件夹下，并命名为“1-5”。（5分）

任务二：安装部署ZooKeeper（40分）

2.1使用Xftp远程上传下载工具，往主从三节点上分别上传ZooKeeper压缩包至/home/hadoop下，截图并保存主节点的状态，图片保存到考生文件夹下，并命名为“2-1”。（5分）

2.2在主从三节点上分别解压ZooKeeper压缩包至/usr/local/src下，改名为zookeeper查看主节点/usr/local/src的状态，截图并保存到考生文件夹下，并命名为“2-2”。（5分）

2.3在主节点node2-1中切换至zookeeper下新建文件夹命名为“data”,在data文件夹中echo修改主节点的myid为1，在slave1和slave2中做同样上述操作并分别修改myid为2和3，之后查看主节点的myid，截图并保存，图片保存到考生文件夹下，并命名为“2-3”。（5分）

2.4在主节点node2-1中的zookeeper下切换至conf文件目录，cp备份ZooKeeper配置文件zoo_sample.cfg命名为zoo.cfg，在conf文件目录下查看备份结果，截图并保存，图片保存到考生文件夹下，并命名为“2-4”。（5分）

2.5修改ZooKeeper配置文件zoo.cfg，设置dataDir为usr /local/src文件目录下的/zookeeper/data,主从三节点分别为server.1、server.2、server.3等于主机

名对应2888及3888，截图并保存，图片保存到考生文件夹下，并命名为“2-5”。（5分）

2.6在主从三节点上分别配置ZooKeeper环境变量，并source使得环境变量设置生效，截图主节点环境变量配置的具体内容细节，图片保存到考生文件夹下，并命名为“2-6”。（5分）

2.7远程分发主节点node2-1的zookeeper下的conf文件目录中的zoo.cfg到node1和node2节点的$ZOOKEEPER_HOME/conf/下，截图，图片保存到考生文件夹下，并命名为“2-7.png”。（5分）

2.8 在Xshell上同时启动Zookeeper集群并查看master2-1、slave1和slave2的状态，截图，图片保存到考生文件夹下，并命名为“2-8”。（5分）

任务三：撰写报告（10分）

3.1在考生文件夹下， word文档t10.docx中，撰写报告：

（1） 归纳总结Zookeeper的作用。（5分）

（2） 归纳总结Zookeeper集群的搭建过程及注意事项。（5分）

3.2保存word文档中，并命名为 “项目报告”。

提交要求：

1)在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南职业技术学院 01 张三。

2)考生文件夹内保存截图：1-1、1-2、1-3、1-4、1-5、2-1、2-2、2-3、2-4、2-5、2-6、2-7、2-8到一个word文档t10.docx中。

（2）实施条件

测试所需的软硬件设备见表1.10.2。

表1.10.2 考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存8G 以上，WIN7及以上操作系统（64位）	
3	镜像文件	安装有Ubuntukylin-16.04-desktop-amd64操作系统、完	机房/虚拟机

		全分布式Hadoop集群主节点	
4	远程上传下载工具	Xftp	
5	ZooKeeper压缩包	zookeeper-3.4.6.tar.gz	

（3）考核时量

考核时间为120分钟。

（4）评分细则

大数据平台搭建与配置模块考核实行 100 分制，评价内容包括工作任务、职业素养完成情况两个方面。其中，工作任务完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表1.10.3 所示。

表 1.10.3 评分标准表评价内容

评价内容		分值	评分细则	
工作任务（80分）	任务一：准备和检查设备	30分	正确克隆两台虚拟机。未正确克隆两台虚拟机作为从节点，扣8分；未正确新建“Zookeeper集群”文件夹并放置主从节点，扣2分。	10分
			正确修改主从节点主机名。未正确修改主从节点主机名，每节点扣1分；未正确查看主节点主机名扣2分。	5分
			正确修改主从节点三台虚拟机的IP地址。未按集群规划要求完成IP地址修改扣3分，未能在Xshell上同时查看IP地址扣2分。	5分
			正确检查主从节点互联情况。未正确查看IP地址扣2分，主从节点未能ping通扣3分。	5分
			正确修改主从节点的/etc/hosts文件。未能正确配置IP地址与主机名之间的映射关系扣5分。	5分
	任务二：安装部署ZooKeeper	40分	正确上传ZooKeeper。未正确连接远程上传下载工具扣2分；未往主从三节点上分别上传ZooKeeper压缩包至/home/hadoop下分别扣1分。	5分
			正确解压ZooKeeper压缩包。未正确解压压缩包至/usr/local/src/下扣3分；在/usr/local/src下查看主节点状态不正常，扣2分。	5分

			正确修改节点myid。未能在zookeeper-3.4.8下正确新建文件夹并命名为“data”扣2分,未能修改主节点的myid为1，扣3分。	5分
			正确备份ZooKeeper配置文件。未能在主节点node2-1中的zookeeper-3.4.8下切换至conf文件目录扣2分，未能正确备份ZooKeeper配置文件并将zoo_sample.cfg命名为zoo.cfg扣3分。	5分
			正确修改ZooKeeper配置文件。未能正确设置dataDir为/usr/local/src 下的/zookeeper-3.4.8/data扣2分；未能正确配置server.1、server.2、server.3等于主机名对应2888及3888扣3分。	5分
			正确配置ZooKeeper环境变量。未能在主从三节点上正确配置ZooKeeper环境变量分别扣1分；未能source使得环境变量设置生效扣2分。	5分
			正确远程分发ZooKeeper配置文件。未能正确远程分发主节点node2-1的zookeeper-3.4.8下的conf文件目录中的zoo.cfg到node1和node2节点的$ZOOKEEPER_HOME/conf/下,扣5分。	5分
			正常启动Zookeeper集群。启动Zookeeper集群并查看master2-1、slave1和slave2的状态不正常扣5分。	5分
	任务三：撰写报告	10分	正确撰写报告。未能正确归纳总结Zookeeper的作用，扣5分。未能归纳总结Zookeeper集群的搭建过程及注意事项，扣5分。	10分
职业素养（20分）	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、鼠标等），检查测试所需软件开发环境。 不进行检查操作扣 5 分。	5分
	专业素养	10分	电脑等工具使用操作规范、按要求命名和存放文件，操作不规范扣5分，未按要求答题扣5分	10分
	整体形象	5分	着装干净整洁，举止文明。遵守考场纪律，按顺序进出考场，着装不整洁扣2分，举止不文明每次扣1分，扣完3分为止	5分

11.试题 1-2-2：使用 Flume 采集网络日志

（1）任务描述

某企业正在做一个电子商务网站，需要从消费用户中访问点特定的节点区域来分析消费者的行为或者购买意图. 这样就可以更加快速地将消费者想要的推送到界面上，为了实现这一点，需要将获取到的他访问的页面以及点击的产品数据等日志数据信息收集并移交给Hadoop平台上去分析，而Flume正是有这样的功能。Flume是Cloudera提供的一个高可用的，高可靠的，分布式的海量日志采集、聚合和传输的系统，Flume支持在日志系统中定制各类数据发送方，用于收集数据；同时，Flume提供对数据进行简单处理，并写到各种数据接受方（可定制）的能力。

根据上述需求，作为企业Hadoop开发人员和运维人员，需要把Flume正确安装到企业的Hadoop平台上去，主要任务如下：

任务一：准备和检查设备（15分）

1.1启动镜像集群，修改镜像的主机名为node2-2,之后用hostname查看，截图并保存，图片保存到考生文件夹下，并命名为“1-1”。（5分）

1.2 修改node2-2的IP地址为192.168.xxx.220，和物理机互通，修改完成之后用Xshell登录，在Xshell上显示虚拟机的登录状态和查看IP地址截图并保存，图片保存到考生文件夹下，并命名为“1-2”。（5分）

1.3 修改node2-2的/etc/hosts文件，配置IP 地址与主机名之间的映射关系、hadoop配置文件，在Xshell上显示虚拟机修改后的hosts文件状态截图及进程并保存，图片保存到考生文件夹下，并命名为“1-3”。（5分）

任务二：Flume安装（30分）

2.1使用Xftp远程上传下载工具，往node2-2节点上传Flume压缩包至/home/hadoop/，通过命令行查看，截图并保存节点的状态，图片保存到考生文件夹下，并命名为“2-1”。（5分）

2.2直接解压Flume压缩包至/usr/local/src/下，重命名为flume，查看文件夹/usr/local/src的状态，截图并保存到考生文件夹下，并命名为“2-2”。（5分）

2.3 在/usr/local/src/flume的conf文件夹下，将flume-env.sh.template重命名为flume-env.sh,截图并保存到考生文件夹下，并命名为“2-3”。（5分）

2.4 查看Java的安装目录并复制，修改配置文件flume-env.sh中的JAVA_HOME地址，截图并保存到考生文件夹下，并命名为“2-4”。（5分）

2.5 配置flume环境变量，并source使得环境变量设置生效，截图节点环境变量配置的具体内容细节，图片保存到考生文件夹下，并命名为“2-5”。（5分）

2.6 查看flume-ng的version信息，进行安装验证，截图，图片保存到考生文件

夹下，并命名为“2-6”。（5分）

任务三：Flume实战 (25分)

采集/home/hadoop/flume下的文件，把结果保存到hdfs://node2-2:9000/flume下，生成的文件前缀为ms-

3.1切换到/usr/local/src下的flume中的conf文件目录下，新建并配置Flume2HDFS.conf文件，配置日志采集文件，配置三个组件sources，sinks，channels，截图，图片保存到考生文件夹下，并命名为“3-1”。（5分）

3.2在/home/hadoop目录下创建文件夹flume,查看并截图新创建的文件夹flume，图片保存到考生文件夹下，并命名为“3-2”。（5分）

3.3切换到//usr/local/src/flume/下，利用操作手册中的启动执行命令开启flume,截图启动成功的日志部分，图片保存到考生文件夹下，并命名为“3-3启”。（5分）

3.4 新建同Ip地址的窗口，新建文件a.txt，添加内容“hello flume”,mv移动a.txt到flume/中，回到旧窗口查看收集的结果，截图生成的前缀为ms-的日志部分，图片保存到考生文件夹下，并命名为“3-4”。（5分）

3.5 在web界面中通过IP:50070查看flume文件夹下新生成的文件，截图，图片保存到考生文件夹下，并命名为“3-5”。（5分）

任务四：撰写报告(10分)

4.1在考生文件夹下，新建word文档t11.docx，撰写报告：

（1） 归纳总结Flume的作用。（5分）

（2） 归纳总结Flume的安装过程及注意事项。（5分）

4.2保存word文档到考生文件夹下，并命名为 “项目报告”。

提交要求：

1)在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南职业技术学院 01 张三。

2)考生文件夹内保存截图：1-1、1-2、1-3、2-1、2-2、2-3、2-4、2-5、2-6、3-1、3-2、3-3、3-4、3-5到一个word文档t11.docx中。

（2）实施条件

测试所需的软硬件设备见表1.11.1。

表1.11.1 考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存8G 以上，WIN7及以上操作系统（64位）	
3	镜像文件	安装有Ubuntukylin-16.04-desktop-amd64操作系统、伪分布式模式Hadoop集群（包含了安装好的sqoop,并能实现Mysql和HDFS的导入导出，集群为启动状态）	机房/虚拟机
4	flume安装包	apache-flume-1.7.0-bin.tar.gz	

（3）考核时量

考核时间为120分钟。

（4）评分细则

Hadoop生态圈其它组件搭建与配置考核实行 100 分制，评价内容包括工作任务、职业素养完成情况两个方面。其中，工作任务完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表1.11.2 所示。

表 1.11.2评分标准表评价内容

评价内容		分值	评分细则	
工作任务(80分)	任务一：准备和检查设备	15分	正确修改主机名。未正确修改主机名，扣3分；未正确查看主机名扣2分。	5分
			正确修改IP地址。未按要求完成IP地址修改扣3分，未能在Xshell上显示虚拟机的登录状态和查看IP地址扣2分。	5分
			正确修改/etc/hosts文件。未能正确配置IP 地址与主机名之间的映射关系扣5分。	5分
	任务二：	30分	正确上传Flume。未正确连接远程上传下载工具扣2分；	5分

	Flume安装		未往node2-2节点正确上传Flume压缩包至/home/hadoop/扣3分。	
			正确解压Flume。未正确解压Flume压缩包至/usr/local/src/下扣3分；在/usr/local/src下查看节点状态不正常，扣2分。	5分
			正确重命名flume-env.sh。未正确切换至conf文件夹下扣2分，未正确将flume-env.sh.template重命名为flume-env.sh扣3分。	5分
			正确修改flume-env.sh。未能正确查看Java的安装目录扣2分；未能正确修改配置文件flume-env.sh中的JAVA_HOME地址扣3分。	5分
			正确配置flume环境变量。未能正确配置flume环境变量扣3分；未能source使得环境变量设置生效扣2分。	5分
			正确验证flume-ng的version信息。不能正常查看flume-ng的version信息扣5分。	5分
	任务三：Flume实战	25分	正确配置Flume2HDFS。未能正确切换到flume中的conf文件目录下扣2分；未能正确新建并配置Flume2HDFS.conf文件扣3分。	5分
			正确创建文件夹flume。未能在根目录下正确创建文件夹flume扣5分。	5分
			正确启动执行flume。未能利用操作手册中的启动执行命令正确开启flume扣5分。	5分
			正确新建文件a。未能在同Ip地址的新窗口新建文件a.txt、添加内容“hello flume”、mv移动a.txt到flume/中每一步扣1分；未能在旧窗口查看到收集的结果扣2分。	5分
			在web界面正确查看a文件。未能在web界面中通过IP:50070查看到flume文件夹下新生成的文件扣5分。	5分
	任务四：撰写报告	10分	正确撰写报告。未能正确归纳总结Flume的作用，扣5分。未能归纳总结Flume的安装过程及注意事项，扣5分。	10分

职业素养（20分）	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、鼠标等），检查测试所需软件开发环境。 不进行检查操作扣 5 分。	5分
	专业素养	10分	电脑等工具使用操作规范、按要求命名和存放文件，操作不规范扣5分，未按要求答题扣5分	10分
	整体形象	5分	着装干净整洁，举止文明。遵守考场纪律，按顺序进出考场，着装不整洁扣2分，举止不文明每次扣1分，扣完3分为止	5分

12.试题 1-2-3：Sqoop 的安装

（1）任务描述

在某企业的业务场景中，经常需要Hadoop集群与关系数据库配合完成某些数据处理任务。一般的做法是，用关系数据库存储最新数据，同时使用Hadoop存储老数据。通常业务流需要把数据从一个存储系统移动到另一个存储系统，这就需要使用Sqoop完成这个工作。Sqoop是一款开源的工具，主要用于在Hadoop(Hive)与传统的数据库(mysql、postgresql...)间进行数据的传递，可以将一个关系型数据库（例如 ： MySQL ,Oracle ,Postgres等）中的数据导进到Hadoop的HDFS中，也可以将HDFS的数据导进到关系型数据库中。

根据上述需求，作为企业Hadoop开发人员和运维人员，需掌握能熟练安装Sqoop到企业的Hadoop平台上去，主要任务如下：

任务一：准备和检查设备（20分）

1.1 导入镜像至VMware Workstation中，启动虚拟机master2-3并登录，查看IP地址截图并保存，图片保存到考生文件夹下，并命名为“1-1”。（5分）

1.2修改虚拟机master2-3的IP地址为192.168.xxx.230、主机名为node2-3，和物理机互通，修改完成之后用Xshell登录，在Xshell上显示虚拟机的登录状态和查看IP地址截图并保存，图片保存到考生文件夹下，并命名为“1-2”。（5分）

1.3修改主机node2-3的/etc/hosts文件，配置IP 地址与主机名之间的映射关系，在Xshell上显示修改后的hosts文件状态截图并保存，图片保存到考生文件夹下，并命名为“1-3”。（5分）

1.4 修改伪分布hadoop配置文件，启动hadoop，查看进程，截图并保存，图片保存到考生文件夹下，并命名为“1-4”。（5分）

任务二：Sqoop的安装（50分）

2.1使用Xft远程上传下载工具，上传Sqoop压缩包至/home/hadoop，并截图，图片保存到考生文件夹下，并命名为“2-1”。（5分）

2.2解压Sqoop压缩包至/usr/local/src/下，查看文件夹/usr/local/src/的状态，截图并保存到考生文件夹下，并命名为“2-2”。（5分）

2.3 重新命名sqoop-1.4.7.bin_hadoop-2.6.0/为sqoop，在src文件夹中查看并截图，截图并保存到考生文件夹下，并命名为“2-3”。（5分）

2.4 配置Sqoop环境变量，并source使得环境变量设置生效，截图节点环境变量配置的具体内容细节，图片保存到考生文件夹下，并命名为“2-4”。（5分）

2.5 将数据库驱动mysql-connector-java-5.1.40-bin.jar拷贝到/usr/local/src/sqoop的lib目录下，截图命令，图片保存到考生文件夹下，并命名为“2-5。（8分）

2.6 在/usr/local/src/sqoop的conf文件夹下，重命名模板文件sqoop-env.template.sh为sqoop-env.sh，截图命令，图片保存到考生文件夹下，并命名为“2-6”。（6分）

2.7 修改配置文件sqoop-env.sh，分别给HADOOP_COMMON_HOME及HADOOP_MAPRED_HOME加入hadoop安装路径，截图，图片保存到考生文件夹下，并命名为“2-7”。（10分）

2.8 验证sqoop安装是否成功，截图，图片保存到考生文件夹下，并命名为“2-8”。（6分）

任务三：撰写报告（10分）

3.1在考生文件夹下，t12.docx中，撰写报告：

（1） 归纳总结Sqoop的作用。（5分）

（2） 归纳总结Sqoop的安装过程及注意事项。（5分）

3.2保存word文档到考生文件夹下，并命名为 “项目报告”。

提交要求：

1)在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南职业技术学院 01 张三。

2)考生文件夹内保存截图：1-1、1-2、1-3、2-1、2-2、2-3、2-4、2-5、2-6、2-7、2-8到一个word文档t12.docx中。

（2）实施条件

测试所需的软硬件设备见表1.12.1。

表1.12.1 考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存8G 以上，WIN7及以上操作系统（64位）	
3	服务器	安装有Ubuntukylin-16.04-desktop-amd64操作系统、伪分布式模式Hadoop	机房/虚拟机镜像包
4	mysql	已安装	
5	数据库驱动	mysql-connector-java-5.1.40-bin.jar驱动	
6	Sqoop安装包	sqoop-1.4.7.bin__hadoop-2.6.0.tar.gz	

（3）考核时量

考核时间为90分钟。

（4）评分细则

Hadoop生态圈其它组件搭建与配置考核实行 100 分制，评价内容包括工作任务、职业素养完成情况两个方面。其中，工作任务完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表 1.12.2 所示。

表 1.12.2评分标准表评价内容

评价内容		分值	评分细则	
工作任务（80分）	任务一：准备和检查设备	20分	正确进行镜像导入。未能成功导入镜像至VMware Workstation中扣5分，未能正常启动虚拟机并登录扣3分。	5分
			正确修改IP地址。未能正确修改虚拟机master2-3的IP地址扣3分，未能在Xshell上显示虚拟机的登录状态和查看IP地址扣2分。	5分
			正确进行hosts文件修改。未能打开/etc/hosts文件扣2	5分

			分，未能配置好IP 地址与主机名之间的映射关系扣3分。	
			正确修改伪分布hadoop配置，并启动hadoop,能查看到对应的进程，得满分;否则扣5分。	5分
	任务二：Sqoop的安装	50分	正确上传Sqoop。未能正确连接WinSCP远程上传下载工具扣2分，未能正确上传Sqoop压缩包至/root/扣3分。	5分
			正确解压Sqoop。未能正确解压Sqoop压缩包至/export/server/下，扣5分。	5分
			正确重命名。未能正确重新命名Sqoop压缩包解压文件为sqoop-1.4.7，扣5分。	5分
			正确配置Sqoop环境变量。未能正确配置Sqoop环境变量扣3分，未能source使得环境变量设置生效扣2分。	5分
			正确拷贝数据库驱动。未能正确将数据库驱动mysql-connector-java-5.1.40-bin.jar上传扣3分，未能正确将数据库驱动拷贝到sqoop-1.4.7的lib目录下,扣3分。	6分
			正确重命名模板文件。未能正确切换sqoop-1.4.7的conf文件夹下扣3分，未能正确将模板文件sqoop-env.sh.template重命名为sqoop-env.sh,扣4分。	7分
			正确修改配置文件。未能给HADOOP_COMMON_HOME正确加入hadoop安装路径扣5分，未能正确给HADOOP_MAPRED_HOME加入hadoop安装路径扣5分。	10分
			验证sqoop安装成功。未能正确输入验证sqoop安装命令扣2分，验证sqoop安装未成功，扣5分。	7分
	任务三：撰写报告	10分	正确撰写报告。未能正确归纳总结Sqoop的作用，扣5分。未能归纳总结Sqoop的安装过程及注意事项，扣5分。	10分
职业素养（20分）	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、鼠标等），检查测试所需软件开发环境。 不进行检查操作扣 5 分。	5分
	专业素养	10分	电脑等工具使用操作规范、按要求命名和存放文件，操作不规范扣5分，未按要求答题扣5分	10分
	整体形象	5分	着装干净整洁，举止文明。遵守考场纪律，按顺序进出考场，着装不整洁扣2分，举止不文明每次扣1分，扣完3分为止	5分

13.试题 1-2-4：Hbase 伪分布式部署

（1）任务描述

HBase是一种构建在HDFS之上的分布式、面向列的存储系统。在需要实时读写、随机访问超大规模数据集时，可以使用HBase。应用场景广：如轨迹、气象网格之类，滴滴打车的轨迹数据主要存在HBase之中；另外在技术所有大一点的数据量的车联网企业，数据都是存在HBase之中；在电信领域、银行领域，不少的订单查询底层的存储，另外不少通信、消息同步的应用构建在HBase之上。

你作为某车联网企业的大数据工程师，需在已安装 hadoop 环境下部署伪分布式hbase，满足公司的业务需求。

表1.13.1伪分布式模式Hbase规划

节点角色	虚拟机名	机器IP	主机名	运行进程
主节点	master	192.168.126.200	node	NameNode ResourceManager SecondaryNameNode DataNode NodeManager

本环节需要完成伪分布式Hadoop 平台上架设Hbase 伪分布式模式部署，主要任务如下：

任务一：正确安装Hbase（24分）

1.1导入安装有伪分布hadoop的虚拟机，重新按规划要求设置主机名、IP、映射文件hosts，并提交截图信息，命名1-1；（6分）

1.2 检查修改伪分布hadoop的配置文件，使之与规划的参数一致，删除logs和tmp文件夹里的内容，启动hadoop，用jps查看主节点进程，并提交截图信息，命名1-2；（6分）

1.3上传Hbase 安装包到/home/hadoop，解压到“/usr/local/src”路径，并提交截图信息，命名1-3；（6分）

1.4解压Hbase安装包后文件夹更名为hbase，并提交截图信息，命名1-4。（6分）

任务二：正确配置Hbase（24分）

2.1修改 Hbase 相应配置文件hbase-en.sh,设置java安装位置、设置使用自带

的zookeeper，并提交截图信息，命名2-1；（6分）

2.2修改 Hbase 相应配置文件hbase-site.xml,设置HBase的数据文件存储位置为HDFS的/hbase目录，并提交截图信息，命名2-2；（6分）

2.3修改 Hbase 相应配置文件hbase-site.xml, 设置zk的数据存放目录为/usr/local/src/hbase/data/zookeeper,设置开启hbase分布式。并提交截图信息，命名2-3；（6分）

2.4设置 Hbase 环境变量，使环境变量生效，提交截图信息，命名2-4。（6分）

任务三：运行Hbase并创建数据库表 (32分)

示例：Student 数据表

行键	列族 StuInfo				列族 Grades	
	Name	Age	Sex	Class	BigData	Computer
0001	Tom Green	18	Male		80	90
0002	Amy	19		01	95	

3.1启动Hbase ,用jps查看进程，提交截图信息，命名3-1；（5分）

3.2用hbase shell创建 Hbase 数据库表，提交截图信息，命名3-2；（6分）

3.3将给定数据输入数据库表中，截图并保存结果，提交截图信息，命名3-3；（10分）

3.4查看数据库表结构，提交截图信息，命名3-4；（6分）

3.5查看 Hbase 版本信息，提交截图信息，命名3-5。（5分）

提交要求：

1)在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南职业技术学院 01 张三。

2)考生文件夹内保存截图：1-1、1-2、2-1、2-2、2-3、2-4、3-1、3-2、3-2、3-4、3-5到一个word文档t13.docx中.

（2）实施条件

测试所需的软硬件设备见表1.13.2

表1.13.2 考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注

1	大数据技术实训机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存8G 以上，WIN7及以上操作系统（64位）,linux 操作系统（ubuntu）。	
3	截图工具		系统自带截图工具
4	Hadoop2.7.1或以上	伪分布，已安装	选用 Hadoop 生产环境稳定版本
5	JDK1.8 及以上	已安装	
6	Hbase-1.1.5或以上		选用与 hadoop 版本兼容的Hbase

（3）考核时量

考核时间为3个小时。

（4）评分细则

大数据平台搭建与配置模块考核实行 100 分制，评价内容包括工作任务、职业素养完成情况两个方面。其中，工作任务完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表 1.13.3 所示。

表 1.13.3 评分标准表评价内容

评价内容		分值	评分细则	
工作任务（80分）	安装Hbase	24分	正确设置主机名、IP、映射文件hosts，每项2分	6分
			按规划修改伪分布hadoop的配置文件，正确启动hadoop得满分，否则不得分	6分
			Hbase正确解压到指定位置得满分，否则不得分；	6分
			解压Hbase安装包后文件夹重命名为hbase得满分，否则不得分。	6分
	配置	24分	正确修改hbase-env.sh文件中配置项得 6分；	6分

	Hbase		正确修改hbase-site.xml文件中配置项得 6分；	6分
			将Hadoop 的相应文件放到 hbase/conf 下,得满分，否则不得分；	6分
			未设置 Hbase 环境变量扣3分，未使环境变量生效扣3分。	6分
	运行Hbase并创建Hbase 数据表	32分	启动 hbase，jps查看启动的进程都有，得 5 分，否则不得分。	5分
			按要求正确创建表结构得分，否则不得分；	6分
			数据录入成功得10分，每少一条数据扣 2 分；	10分
			正确查看数据库表结构得满分，否则不得分；	6分
			正确查看Hbase 版本信息得5分，否则不得分。	5分
职业素养（20分）	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、鼠标等），检查测试所需软件开发环境。不进行检查操作扣 5 分。	5分

14.试题 1-2-5：Hbase 完全分布式部署模块

（1）任务描述

HBase是一种构建在HDFS之上的分布式、面向列的存储系统。在需要实时读写、随机访问超大规模数据集时，可以使用HBase。应用场景广：如轨迹、气象网格之类，滴滴打车的轨迹数据主要存在HBase之中；另外在技术所有大一点的数据量的车联网企业，数据都是存在HBase之中；在电信领域、银行领域，不少的订单查询底层的存储，另外不少通信、消息同步的应用构建在HBase之上。

你作为某车联网企业的大数据工程师，需在已安装 hadoop 环境下部署完全分布式hbase，满足公司的业务需求。

表1.14.1完全分布式模式Hbase规划

节点角色	虚拟机名	机器IP	主机名	运行进程
主节点	master	192.168.126.200	node	NameNode ResourceManager SecondaryNameNode

从节点	slave1	192.168.126.201	node1	DataNode NodeManager
	slave2	192.168.126.202	node2	DataNode NodeManager

本环节需要完成Hbase完全分布式模式部署，主要任务如下：

任务一：正确安装Hbase（22分）

1.1导入安装有分布式hadoop的虚拟机三台，重新按规划要求设置主机名、IP、映射文件hosts，并提交截图信息，命名1-1；（6分）

1.2 检查修改分布式hadoop的配置文件，使之与规划的参数一致，删除logs和tmp文件夹里的内容，启动hadoop，用jps查看主节点进程，并提交截图信息，命名1-2；（6分）

1.3 解压 Hbase 安装包到“/usr/local/src”路径，并修改解压后文件夹名为hbase，并提交截图信息，命名1-3；（5分）

1.4 设置 Hbase 环境变量，并使环境变量只对当前 root 用户生效，并提交截图信息，命名1-4。（5分）

任务二：正确配置Hbase（30分）

2.1修改 Hbase 相应配置文件hbase-env.sh，并提交截图信息，命名2-1;（6分）

2.2修改 Hbase 相应配置文件hbase-site.xml，并提交截图信息，命名2-2;（6分）

2.3修改 Hbase 相应配置文件regionservers，并提交截图信息，命名2-3;（6分）

2.4 分发hbase到其他从节点1中，并提交截图信息，命名2-4; （6分）

2.5 分发hbase到其他从节点2中，并提交截图信息，命名2-5; （6分）

任务三：运行Hbase并创建数据库表 (28分)

示例：Student 数据表

行键	列族 StuInfo				列族 Grades	
	Name	Age	Sex	Class	BigData	Computer
0001	Tom Green	18	Male		80	90
0002	Amy	19		01	95	

3.1 启动Hbase，并命令jps查看主从节点进程，并提交截图信息，命名3-1；

（5分）

3.2 创建 Hbase 数据库表，并提交截图信息，命名3-2； （6分）

3.3 将给定数据导入数据库表中，并提交截图信息，命名3-3；（6分）

3.4查看导入数据库表结构，提交截图信息，命名3-4；（6分）

3.5查看 Hbase 版本信息，提交截图信息，命名3-5。 （5分）

提交要求：

1)在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南职业技术学院 01 张三。

2)考生文件夹内保存截图：1-1、1-2、2-1、2-2、2-3、2-4、2-5、3-1、3-2、3-3、3-4、3-5到一个word文档t14.docx中。

（2）实施条件

测试所需的软硬件设备见表4.14.2

表1.14.2 考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	大数据技术实训机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存8G 以上，WIN7及以上操作系统（64位）,linux 操作系统（ubuntu）。	机房/虚拟机
3	截图工具		系统自带截图工具
4	Hadoop2.6.0或以上	分布式，已安装	选用 Hadoop 生产环境稳定版本
5	JDK1.8 及以上	已安装	
6	Hbase1.3或以上		选用与 hadoop 版本兼容的Hbase

（3）考核时量

考核时间为3个小时。

（4）评分细则

大数据平台搭建与配置模块考核实行 100 分制，评价内容包括工作任务、职业素养完成情况两个方面。其中，工作任务完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表 1.14.3 所示。

表 1.14.3 评分标准表评价内容

评价内容		分值	评分细则	
工作任务（80分）	安装Hbase	22分	按规划修改主机名、IP、映射文件hosts，每项2分	6分
			按规划修改分布式hadoop的配置文件，正确启动hadoop得满分，否则不得分	6分
			Hbase正确解压到指定位置得满分，否则不得分。	5分
			环境变量正确得满分，否则不得分。	5分
	配置Hbase	30分	正确修改hbase-env.sh文件中配置项得满分，否则扣6分；正确修改hbase-site.xml文件中配置项得满分，否则扣6分；正确修改regionservers文件中配置项得满分，否则扣6分。	18分
			正确分发hbase到其他从节点1中得满分，否则扣6分；正确分发hbase到其他从节点2中得满分，否则扣6分。	12分
	运行Hbase并创建Hbase 数据库表	28分	启动 hbase，jps查看启动的进程都有，得 5 分，否则不得分。	5分
			按要求正确创建表结构得分，否则不得分；	6分
			数据导入成功得6分，每少一条数据扣 2 分；	6分
			正确查看导入数据库表结构得满分，否则不得分；	6分
			正确查看Hbase 版本信息得5分，否则不得分。	5分
职业素养（20分）	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、鼠标等），检查测试所需软件开发环境。不进行检查操作扣 5 分。	5分
	专业素养	10分	按要求命名文件，截图，答题规范有序得10分。	10分
	职业行为	5分	着装干净、整洁。举止文明，遵守考场纪律，按顺序进	5分

	规范		出考场。	

15.试题 1-2-6：hadoop 平台架设 Hive 组件部署模块

（1）任务描述

hive是基于Hadoop的一个数据仓库工具，用来进行数据提取、转化、加载，这是一种可以存储、查询和分析存储在Hadoop中的大规模数据的机制。hive数据仓库工具能将结构化的数据文件映射为一张数据库表，并提供SQL查询功能，能将SQL语句转变成MapReduce任务来执行。

某车联网企业随着公司业务量扩展到多个省份，用户数据量越来越大，且难以管理。基于hive十分适合对数据仓库进行统计分析，企业决定部署该组件解决企业用户数据难以管理需求。

你作为某车联网企业的大数据工程师，需在已安装 hadoop 环境下部署伪分布式Hive，满足公司的业务需求。

表1.15.1伪分布式模式Hive规划

节点角色	虚拟机名	机器IP	主机名	运行进程
主节点	master	192.168.126.200	node	NameNode ResourceManager SecondaryNameNode DataNode NodeManager

本环节需要完成Hive伪分布式模式部署，主要任务如下：

任务一：正确安装Hive（24分）

1.1导入镜像至VMware Workstation中，虚拟机并登录，修改主机名、IP、映射文件hosts，修改hadoop配置文件，并启动hadoop，用jps查看进程，截图并保存，图片保存到考生文件夹下，并命名为“1-1”。（6分）

1.2用Xftp上传Hive 安装包到/home/hadoop目录下，解压到“/usr/local/src/”路径，并使用相关命令，修改解压后文件夹名为 Hive，进入 Hive 文件夹，并将查看内容提交截图，命名1-2；（6分）

1.3 设置Hive 环境变量，提交截图信息，命名1-3。（6分）

1.4　使Hive设置后环境变量生效，提交截图信息，命名1-4。（6分）

任务二：正确配置Hive（24分）

2.1 正确将hive-default.xml.template 复制重命名为 hive-site.xml，提交截图信息，命名2-1；（6分）

2.2配置 hive-site.xml 文件，实现“Hive 元存储”的存储位置为 MySQL数据库，提交截图信息，命名2-2；（6分）

2.3、初始化 Hive 元数据（将 MySQL 数据库 JDBC 驱动拷贝到 Hive 安装目录的 lib 下），初始化结果，提交截图信息，命名2-3。（12分）

任务三： 启动并创建Hive数据库表（32分)

test1.txt内容如下：

1,xiaoming,book-TV-code,beijing:chaoyang-shagnhai:pudong

2,lilei,book-code,nanjing:jiangning-taiwan:taibei

3,lihua,music-book,heilongjiang:haerbin

————————————————————

3.1启动 Hive, 检查是否安装成功，提交截图信息，命名3-1；（6分）

3.2 按指定要求(/home/hadoop目录下test1文档描述)创建 Hive 内部表t1，查看内容，提交截图信息，命名3-2；（7分）

3.3 按指定要求(/home/hadoop目录下test1文档描述)创建 Hive 外部表t2，查看内容，提交截图信息，命名3-3；（7分）

3.4实现内外部表转换，把t1转换为外表，t2转换为内表，提交截图信息，命名3-4；（6分）

3.5 按指定要求(/home/hadoop目录下test1文档描述)创建分区表t3，根据性别sex建立分区，test1中的数据全部是男性male，提交截图信息，命名3-5。（6分）

提交要求：

1)在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南职业技术学院 01 张三。

2)考生文件夹内保存截图：1-1、1-2、1-3、2-1、2-2、2-3、3-1、3-2、3-2、3-3、

3-4、3-5到一个word文档t15.docx中。

（2）实施条件

测试所需的软硬件设备见表1.15.2

表1.15.2　考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	大数据技术实训机房	测试场地	保证参考人员有足够间距
2	计算机	CPU　i5以上，内存8G 以上，WIN7及以上操作系统（64位）,linux 操作系统（ubuntu）。	
3	截图工具		系统自带截图工具
4	Hadoop2.7.1或以上	伪分布，已安装	选用 hadoop 生产环境稳定版本
5	JDK1.8 及以上	已安装	
6	mysql	已安装	
7	Hive1.x 及以上		选用与 hadoop 版本兼容的Hive

（3）考核时量

考核时间为3个小时。

（4）评分细则

大数据平台搭建与配置模块考核实行 100 分制，评价内容包括工作任务、职业素养完成情况两个方面。其中，工作任务完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表 1.15.3 所示。

表 1.15.3 评分标准表评价内容

评价内容		分值	评分细则	
	安装Hive	24分	导入镜像，修改配置，启动hadoop，正确查看进程得满分，否则扣6分。	6分
			Hive正确安装到指定位置得满分，否则不得分； 安装后文件夹未重命名扣3分。	6分

工作任务（80分）			环境变量设置正确得满分，否则不得分。环境变量配置错误扣6分；未执行命令使设置后环境变量生效扣6分。	12分
	配置Hive	24分	正确将hive-default.xml.template 复制重命名为 hive-site.xm，未成功扣6分；	6
			正确修改hive-site.xml文件中配置项得满分，否则不得分；	6
			正确将 MySQL 数据库 JDBC 驱动拷贝到 Hive 安装目录的 lib 下得6分，错误扣6分； 初始化 Hive 元数据正确得6分，错误扣6分。	12
	启动Hive及创建Hive 数据库表	32分	Hive启动成功得 6分，否则不得分。	6分
			按要求正确创建 Hive 内部表t1得4分，否则扣4分；正确查看Hive 内部表t1得3分，否则扣3分；	7分
			按要求正确创建 Hive 外部表t2得4分，否则扣4分；正确查看Hive 外部表t2得3分，否则扣3分；	7分
			内外表按要求转换成功得6分，否则不得分。	6分
			按要求正确创建分区表得6分，否则不得分。	6分
职业素养（20分）	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、鼠标等），检查测试所需软件开发环境。不进行检查操作扣 5 分。	5分
	专业素养	10分	按要求命名文件，截图，答题规范有序得10分。	10分
	职业行为规范	5分	着装干净、整洁。举止文明，遵守考场纪律，按顺序进出考场。	5分

16.试题 1-2-7：hadoop 平台架设 Storm 组件部署模块

（1）任务描述

Apache Storm是一个分布式实时大数据处理系统。Storm设计用于在容错和水平可扩展方法中处理大量数据。它是一个流数据框架，具有最高的摄取率。某大型电商企业的电商平台，随着用户量的增大及业务量的上升，企业想实时了解在一些大型商业活动中平台的业务情况，及时作后续的动态调整。现需要对平台数据进行实时的处理。

你作为企业的大数据工程师，需在已安装 hadoop 环境下部署 Storm ，满足公司的业务需求。

表1.16.1 Storm规划

节点角色	虚拟机名	机器IP	主机名	运行进程
主节点	master	192.168.126.200	node	NameNode ResourceManager SecondaryNameNode
从节点	slave1	192.168.126.201	node1	DataNode NodeManager
	slave2	192.168.126.202	node2	DataNode NodeManager

本环节需要完成Storm部署，主要任务如下：

任务一：正确安装Storm（20分）

1.1导入镜像至VMware Workstation中，创建一主两从共三台虚拟机，修改主机名、IP、映射文件hosts，启动hadoop，截图并保存，图片保存到考生文件夹下，并命名为“1-1”。（6分）

1.2 对于前置安装 Zookeeper 集群，修改zookeeper配置文件，能正确启动，提交截图信息，命名1-2；（6分）

1.3 上传Storm 安装包到”/home/hadoop”，解压到“/usr/local/src”路径，并修改解压后文件夹名为storm，提交截图信息，命名1-3。（8分）

任务二：正确配置Storm（30分）

2.1 配置“conf/storm.yaml”文件的storm.zookeeper.servers，提交截图信息，命名2-1；（10分）

2.2 配置 nimbus.seeds，提交截图信息，命名2-2；（10分）

2.3 配置 supervisor.slots.ports，提交截图信息，命名2-3.（10分）

任务三：运行Storm （30分）

3.1 拷贝主节点 Storm 包到从节点，提交截图信息，命名3-1；（10分）

3.2 设置 Storm 环境变量，并使环境变量只对当前 hadoop 用户生效，提交截图信息，命名3-2；（10分）

3.3 在主节点和从节点启动，提交截图信息，命名3-3。（10分）

提交要求：

1)在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南职业技术学院 01 张三。

2)考生文件夹内保存截图：1-1、1-2、2-1、2-2、2-3、3-1、3-2、3-2、3-3到一个word文档t16.docx中。

（2）实施条件

测试所需的软硬件设备见表1.16.2

表1.16.2　考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	大数据技术实训机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存8G 以上，WIN7及以上操作系统（64位）,linux 操作系统（ubuntu）。	机房/虚拟机
3	截图工具		系统自带截图工具
4	Hadoop2.7.1或以上	分布式，已安装	选用 Hadoop 生产环境稳定版本
5	JDK1.8 及以上	已安装	
6	Zookeeper-3.4.6	已安装	选用与 hadoop 版本兼容
7	Storm1.2		选用与 hadoop 版本兼容

（3）考核时量

考核时间为3个小时。

（4）评分细则

大数据平台搭建与配置模块考核实行 100 分制，评价内容包括工作任务、职业素养完成情况两个方面。其中，工作任务完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表 1.16.3 所示。

表 1.16.3 评分标准表评价内容

评价内容		分值	评分细则	
工作任务（80分）	安装 Storm	20分	hadoop启动成功得满分，否则不得分。	6分
			zookeeper启动成功得满分，否则不得分。	6分
			正确解压到指定位置得满分，否则不得分。	8分
	配置 Storm	30分	修改storm.yaml正确得满分，否则不得分。	10分
			修改nimbus.seeds得满分，否则不得分。	10分
			修改supervisor.slots.ports得满分，否则不得分。	10分
	运行 storm	30分	拷贝主节点 Storm 包到所有从节点得满分，否则不得分。	10分
			设置 Storm 环境变量，并使环境变量生效得满分，否则不得分。	10分
			storm启动成功得满分，否则不得分。	10分
职业素养（20分）	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、鼠标等），检查测试所需软件开发环境。不进行检查操作扣 5 分。	5分
	专业素养	10分	按要求命名文件，截图，答题规范有序得10分。	10分
	职业行为规范	5分	着装干净、整洁。举止文明，遵守考场纪律，按顺序进出考场。	5分

17.试题 1-2-8：hadoop 平台架设 Spark 组件部署模块

（1）任务描述

Spark 是一种与 Hadoop 相似的开源集群计算环境，但是两者之间还存在一些不同之处，这些有用的不同之处使 Spark 在某些工作负载方面表现得更加优越。你作为某公司新入职的大数据工程师，因企业业务需求，你需在已安装 hadoop 环境下部署 Spark。

表1.17.1伪分布式模式Spark规划

节点角色	虚拟机名	机器IP	主机名	运行进程
主节点	master	192.168.126.200	node	NameNode

				ResourceManager SecondaryNameNode
从节点	slave1	192.168.126.201	node1	DataNode NodeManager
	slave2	192.168.126.202	node2	DataNode NodeManager

本环节需要完成Spark分布式部署，主要任务如下：

任务一：安装与配置Scala（30分）

1.1导入安装有分布式hadoop的虚拟机三台，重新按规划要求设置主机名、IP、映射文件hosts，并提交截图信息，命名1-1；（6分）

1.2 检查修改分布式hadoop的配置文件，使之与规划的参数一致，删除logs和tmp文件夹里的内容，启动hadoop，用jps查看主节点进程，并提交截图信息，命名1-2；（6分）

1.3 上传scala 安装包到/home/hadoop，并解压到“/usr/local/src”路径下，然后更名为 scala，提交截图信息，命名为1-3；(6分)

1.4 设置 scala 环境变量，并使环境变量只对当前用户生效，提交截图信息，命名为1-4；（6分）

1.5 进入 scala 并截图，截图并保存结果，提交截图信息，命名为1-5。（6分）

任务二：安装与配置 Spark（30分）

2.1 解压 Spark 安装包到“/usr/local/src”路径下，并更名为 spark，提交截图信息，命名为2-1；（6分）

2.2 设置 Spark 环境变量，并使环境变量只对当前用户生效，提交截图信息，命名为2-2；（6分）

2.3配置conf/spark-env.sh，指定SPARK_DIST_CLASSPATH、HADOOP_CONF_DIR、SPARK_MASTER_IP，提交截图信息，命名为2-3。 （6分）

2.4 配置conf/slaves，指定 Spark slave 节点，提交截图信息，命名为2-4。（6分）

2.5 配置sbin/spark-config.sh，指定java安装目录JAVA_HOME，提交截图信息，

命名为2-4。（6分）

任务三：运行 Spark（20分）

3.1 启动 Spark，并使用命令jps查看主从节点进程，提交截图信息，命名为3-1。（10分）

3.2 查看spark webUI 结果，提交截图信息，命名为3-2。（10分）

提交要求：

1)在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南职业技术学院 01 张三。

2)考生文件夹内保存截图：1-1、1-2、1-3、1-4、2-1、2-2、2-3、3-1到一个word文档t17.docx中。

（2）实施条件

测试所需的软硬件设备见表1.17.2

表1.17.2 考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	大数据技术实训机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存8G 以上，WIN7及以上操作系统（64位）,linux 操作系统（ubuntu）。	机房/虚拟机
3	截图工具		系统自带截图工具
4	Hadoop2.7.1	完全分布式，已安装	选用 Hadoop 生产环境稳定版本
5	JDK1.8 及以上	已安装	
6	Scala2.x		选用与 hadoop 版本兼容
7	Spark2.x		选用与 hadoop 版本兼容

（3）考核时量

考核时间为3个小时。

（4）评分细则

大数据平台搭建与配置模块考核实行 100 分制，评价内容包括工作任务、职业素养完成情况两个方面。其中，工作任务完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表 1.17.3 所示。

表 1.17.3 评分标准表评价内容

评价内容		分值	评分细则	
工作任务（80分）	安装与配置Sclal	30分	按规划修改主机名、IP、映射文件hosts，每项2分	6分
			按规划修改分布式hadoop的配置文件，正确启动hadoop得满分，否则不得分	6分
			正确解压到指定位置得3分，否则不得分； 正确修改文件夹名得3分，错误扣3分；	6分
			正确设置Scala环境变量正确得3分，否则扣3分；正确执行命令使环境变量生效得3分，否则扣3分；	6分
			正确进入Scala得满份，否则不得分。	6分
	安装与配置spark	30分	正确解压到指定位置得满分，否则不得分；	6分
			设置spark环境变量并使之生效得满分，未设置环境变量扣4分，未执行命令使之生效扣2分；	6分
			正确配置conf/spark-env.sh得满分，否则不得分	6分
			正确配置conf/slaves得满分，否则不得分	6分
			正确配置sbin/spark-config.sh得满分，否则不得分。	6分
	启动spark	20分	spark 启动成功，正确查看主从节点进程	10分
			spark 启动成功，查看spark webUI 结果，得满分，否则不得分。	10分
职业素养（20分）	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、鼠标等），检查测试所需软件开发环境。不进行检查操作扣 5 分。	5分
	专业素养	10分	按要求命名文件，截图，答题规范有序得10分。	10分
	职业行为规范	5分	着装干净、整洁。举止文明，遵守考场纪律，按顺序进出考场。	5分

二、数据采集与存储模块

1.试题 2-1-1：7 天天气数据采集与存储

（1）任务描述

天气预报和人们的工作、生活息息相关，人们一般都会在天气网站查找天气信息。下面以天气预报网站http://www.weather.com.cn/weather/101250101.shtml为例（对应局域网地址：http://172.16.7.152/sevenDaysWeather.html，其中IP地址根据实际局域网配置修改）。请根据天气预报网站源数据，综合利用大数据采集工具和相关技术对网站的数据进行数据采集，完成数据采集任务。利用数据库技术，根据采集的有效数据信息，创建数据库表结构，并将采集的有效数据存储到Mysql数据库中，且能正确查看数据结果，完成数据存储任务。帮助使用者提前查找到7天天气状况、气温和风力风向等数据内容并完成数据展示与存储。

采集数据样式：

日期	天气	温度	风向风力
15日（今天）	雷阵雨	33/22℃	南风 南风<3级
16日（明天）	晴	33/22℃	西风 南风<3级

任务一：项目创建（10分）

1.1创建项目。（10分）

2.1截图项目结构，命名1-1。

任务二：解析网页（10分）

2.1解析网页，获取html文档。（10分）

2.2截图代码，命名2-1。

任务三：提取数据（20分）

3.1根据待采集数据样式，实现数据采集，能够在控制台输出所有待采集数据。（20分）

3.2截图输出结果，命名3-1。

任务四：创建数据库表（10分）

4.1根据采集的数据样式，创建对应的表结构。（10分）

4.2截图表结构，命名4-1。

任务五：数据库连接（10分）

5.1编写代码，利用“数据库资源参数”进行数据库连接。（10分）

5.2截图代码，命名5-1。

任务六：批量插入数据（15分）

6.1 编写代码，遍历数据采集结果集合，依次将采集结果存储至数据库表中。（12分）

6.2释放资源。（3分）

6.3截图代码，命名6-1。

任务七：查看存储结果（5分）

7.1打开数据库，查看数据库表中数据。（3分）

7.2截图数据库表中的数据，命名7-1。

7.3将完整的项目提交到考生文件夹中。（2分）

提交要求：

1)在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南工程职业技术学院 01 张三。

2)考生文件夹内共创建：项目文件，截图文件夹（保存截图）。

（2）实施条件

测试所需的软硬件设备见下表2.1.1。

表 2.1.1考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存 8G 以上， windows操作系统	用于软件开发和软件部署，每人一台。
3	截图工具		系统自带截图工具
4	pycharm		用于数据采集。
5	MySQL5.5或以上		用于数据存储。
6	Navicat10或以上		用于数据存储。

（3）考核时量

考核时间为3小时

（4）评分细则

数据采集与存储模块的考核实行 100 分制，评价内容包括技能要求、职业素养完成情况两个方面。其中，技能要求完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表 2.1.2 所示。

表 2.1.2 评分标准表评价内容

评价内容		分值	评分细则	
工作任务(80分)	项目创建	10分	没有正确导入urllib库或requests库扣5分。	5分
			没有正确导入pymysql扣5分。	5分
	解析网页	10分	没有设置headers扣4分。	4分
			没有正确返回Document对象扣6分。	6分
	提取数据	20分	数据数量不足每一个扣1分，扣完为止。	5分
			数据少一个字段扣2分，扣完为止。	5分
			结果没有正确显示在控制台上，扣10分。	10分
	创建数据库表	10分	数量数据不足扣5分，每缺少1个字段扣3分，扣完为止。	10分
	数据库连接	10分	没有获取数据库连接，扣10分； 使用参数不正确，每处扣1分； 没有正确调用对应方法，每处扣2分。	10分
	批量插入数	15分	没有释放资源，扣2分；	15分

	据		使用参数不正确，每处扣1分； 没有正确调用对应方法，每处扣2分； 异常处理不正确，扣3分； 没有释放资源，扣2分。	
	查看存储结果	5分	数据库表中数据不正确，扣3分；	3分
			没有提交项目文件，扣2分。	2分
职业素养(20分)	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、鼠标等），检查测试所需软件开发环境。不进行检查操作扣5分。	5分
	专业素养	10分	代码符合软件开发规范，命名规范，能做到见名知意；缩进统一，方便阅读；注释规范正确。整个项目命名不规范每一处扣1分；完全没有注释扣2分；有注释，但注释不规范每一处扣1分。	10分
	职业行为规范	5分	着装干净整洁，举止文明。遵守考场纪律，按顺序进出考场。	5分

2.试题 2-1-2：8-15 天天气数据采集与存储

（1）任务描述

天气预报和人们的工作、生活息息相关，人们一般都会在天气网站查找天气信息。下面以天气预报网站http://www.weather.com.cn/weather15d/101250101.shtml/为例（对应局域网地址：http://172.16.7.152/eightDaysWeather.html，其中IP地址根据实际局域网配置修改）。请根据天气预报网站源数据，综合利用大数据采集工具和相关技术对网站的数据进行数据采集，完成数据采集任务。利用数据库技术，根据采集的有效数据信息，创建数据库表结构，并将采集的有效数据存储到Mysql数据库中，且能正确查看数据结果，完成数据存储任务。帮助使用者提前查找到8-15天天气状况、气温和风力风向等数据内容并完成数据展示与存储。

采集数据样式：

日期	天气	温度	风向风力
6日（周二）	多云转阴	36/27℃	南风<3级
7日（周二）	阴	36/27℃	东南风转南风<3级

任务一：项目创建（10分）

1.1创建项目。（10分）

2.1截图项目结构，命名1-1。

任务二：解析网页（10分）

2.1解析网页，获取html文档。（10分）

2.2截图代码，命名2-1。

任务三：提取数据（20分）

3.1根据待采集数据样式，实现数据采集，能够在控制台输出所有待采集数据。（20分）

3.2截图输出结果，命名3-1。

任务四：创建数据库表（10分）

4.1根据采集的数据样式，创建对应的表结构。（10分）

4.2截图表结构，命名4-1。

任务五：数据库连接（10分）

5.1编写代码，利用“数据库资源参数”进行数据库连接。（10分）

5.2截图代码，命名5-1。

任务六：批量插入数据（15分）

6.1编写代码，遍历数据采集结果集合，依次将采集结果存储至数据库表中。（12分）

6.2释放资源。（3分）

6.3截图代码，命名6-1。

任务七：查看存储结果（5分）

7.1 打开数据库，查看数据库表中数据。（3分）

7.2截图数据库表中的数据，命名7-1。

7.3将完整的项目提交到考生文件夹中。（2分）

提交要求：

1)在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南工程职业技术学院 01 张三。

2)考生文件夹内共创建：项目文件，截图文件夹（保存截图）。

（2）实施条件

测试所需的软硬件设备见下表2.2.1。

表 2.2.1考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存 8G 以上， windows操作系统	用于软件开发和软件部署，每人一台。
3	截图工具		系统自带截图工具
4	pycharm		用于数据采集。
5	MySQL5.5或以上		用于数据存储。
6	Navicat10或以上		用于数据存储。

（3）考核时量

考核时间为3小时

（4）评分细则

数据采集与存储模块的考核实行 100 分制，评价内容包括技能要求、职业素养完成情况两个方面。其中，技能要求完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表 2.2.2 所示。

表 2.2.2 评分标准表评价内容

评价内容		分值	评分细则	
工作任务(80分)	项目创建	10分	没有正确导入urllib库或requests库扣5分。	5分
			没有正确导入pymysql扣5分。	5分
	解析网页	10分	没有设置headers扣4分。	4分
			没有正确返回Document对象扣6分。	6分
	提取数据	20分	数据数量不足每一个扣1分，扣完为止。	5分
			数据少一个字段扣2分，扣完为止。	5分
			结果没有正确显示在控制台上，扣10分。	10分
	创建数据库表	10分	数量数据不足扣5分，每缺少1个字段扣3分，扣完为止。	10分
	数据库连接	10分	没有获取数据库连接，扣10分； 使用参数不正确，每处扣1分；	10分

			没有正确调用对应方法，每处扣2分。	
	批量插入数据	15分	没有释放资源，扣2分； 使用参数不正确，每处扣1分； 没有正确调用对应方法，每处扣2分； 异常处理不正确，扣3分； 没有释放资源，扣2分。	15分
	查看存储结果	5分	数据库表中数据不正确，扣3分；	3分
			没有提交项目文件，扣2分。	2分
职业素养(20分)	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、鼠标等），检查测试所需软件开发环境。不进行检查操作扣5分。	5分
	专业素养	10分	代码符合软件开发规范，命名规范，能做到见名知意；缩进统一，方便阅读；注释规范正确。整个项目命名不规范每一处扣1分；完全没有注释扣2分；有注释，但注释不规范每一处扣1分。	10分
	职业行为规范	5分	着装干净整洁，举止文明。遵守考场纪律，按顺序进出考场。	5分

3.试题 2-1-3：常用电话号码数据采集与存储

（1）任务描述

日常生活中，我们可能会遇到一些紧急情况，人们一般都会在网站查找电话信息。下面以便民查询网站http://changyongdianhuahaoma.51240.com/为例（对应局域网地址：http://172.16.7.152/commonTelephonNumbers.html，其中IP地址根据实际局域网配置修改）。请根据便民查询网站源数据，综合利用大数据采集工具和相关技术对网站的数据进行数据采集，完成数据采集任务。利用数据库技术，根据采集的有效数据信息，创建数据库表结构，并将采集的有效数据存储到Mysql数据库中，且能正确查看数据结果，完成数据存储任务。帮助使用者提前查找到名称、电话等数据内容并完成数据展示与存储。

采集数据样式：

名称	电话
匪警	110
火警	119

任务一：项目创建（10分）

1.1创建项目。（10分）

2.1截图项目结构，命名1-1。

任务二：解析网页（10分）

2.1解析网页，获取html文档。（10分）

2.2截图代码，命名2-1。

任务三：提取数据（20分）

3.1根据待采集数据样式，实现数据采集，能够在控制台输出所有待采集数据。（20分）

3.2截图输出结果，命名3-1。

任务四：创建数据库表（10分）

4.1根据采集的数据样式，创建对应的表结构。（10分）

4.2截图表结构，命名4-1。

任务五：数据库连接（10分）

5.1编写代码，利用“数据库资源参数”进行数据库连接。（10分）

5.2截图代码，命名5-1。

任务六：批量插入数据（15分）

6.1编写代码，遍历数据采集结果集合，依次将采集结果存储至数据库表中。（12分）

6.2释放资源。（3分）

6.3截图代码，命名6-1。

任务七：查看存储结果（5分）

7.1 打开数据库，查看数据库表中数据。（3分）

7.2截图数据库表中的数据，命名7-1。

7.3将完整的项目提交到考生文件夹中。（2分）

提交要求：

1)在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南工程职业技术学院 01 张三。

2)考生文件夹内共创建：项目文件，截图文件夹（保存截图）。

（2）实施条件

测试所需的软硬件设备见下表2.3.1。

表 2.3.1考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存 8G 以上， windows操作系统	用于软件开发和软件部署，每人一台。
3	截图工具		系统自带截图工具
4	pycharm		用于数据采集。
5	MySQL5.5或以上		用于数据存储。
6	Navicat10或以上		用于数据存储。

（3）考核时量

考核时间为3小时

（4）评分细则

数据采集与存储模块的考核实行 100 分制，评价内容包括技能要求、职业素养完成情况两个方面。其中，技能要求完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表 2.3.2 所示。

表 2.3.2 评分标准表评价内容

评价内容		分值	评分细则	
工作任务(80分)	项目创建	10分	没有正确导入urllib库或requests库扣5分。	5分
			没有正确导入pymysql扣5分。	5分
	解析网页	10分	没有设置headers扣4分。	4分
			没有正确返回Document对象扣6分。	6分
	提取数据	20分	数据数量不足每一个扣1分，扣完为止。	5分
			数据少一个字段扣3分，扣完为止。	5分
			结果没有正确显示在控制台上，扣10分。	10分
	创建数据库表	10分	数量数据不足扣5分，每缺少1个字段扣3分，扣完为止。	10分
	数据库连接	10分	没有获取数据库连接，扣10分； 使用参数不正确，每处扣1分； 没有正确调用对应方法，每处扣2分。	10分
	批量插入数	15分	没有释放资源，扣2分；	15分

	据		使用参数不正确，每处扣1分； 没有正确调用对应方法，每处扣2分； 异常处理不正确，扣3分； 没有释放资源，扣2分。	
	查看存储结果	5分	数据库表中数据不正确，扣3分；	3分
			没有提交项目文件，扣2分。	2分
职业素养 (20分)	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、鼠标等），检查测试所需软件开发环境。不进行检查操作扣5分。	5分
	专业素养	10分	代码符合软件开发规范，命名规范，能做到见名知意；缩进统一，方便阅读；注释规范正确。整个项目命名不规范每一处扣1分；完全没有注释扣2分；有注释，但注释不规范每一处扣1分。	10分
	职业行为规范	5分	着装干净整洁，举止文明。遵守考场纪律，按顺序进出考场。	5分

4.试题 2-1-4：国家域名缩写和电话代码数据采集与存储

（1）任务描述

作为外贸员必须收集世界各国的国家名称，对应英文名，对应国际域名，对应电话代码等信息。下面以便民查询网站https://yumingsuoxie.bmcx.com/为例（对应局域网地址：http://172.16.7.152/abbreviationForCountries.html，其中IP地址根据实际局域网配置修改）。请根据便民查询网站源数据，综合利用大数据采集工具和相关技术对网站的数据进行数据采集，完成数据采集任务。利用数据库技术，根据采集的有效数据信息，创建数据库表结构，并将采集的有效数据存储到Mysql数据库中，且能正确查看数据结果，完成数据存储任务。帮助使用者提前查找到国家域名缩写、国家或地区、英文名、电话代码等数据内容并完成数据展示与存储。

采集数据样式：

国际域名缩写	国家或地区	英文名	电话代码
AD	安道尔共和国	Andorra	376
AE	阿拉伯联合酋长国	United Arab Emirates	971
AF	阿富汗	Afghanistan	93

任务一：项目创建（10分）

1.1创建项目。（10分）

2.1截图项目结构，命名1-1。

任务二：解析网页（10分）

2.1解析网页，获取html文档。（10分）

2.2截图代码，命名2-1。

任务三：提取数据（20分）

3.1根据待采集数据样式，实现数据采集，能够在控制台输出所有待采集数据。（20分）

3.2截图输出结果，命名3-1。

任务四：创建数据库表（10分）

4.1根据采集的数据样式，创建对应的表结构。（10分）

4.2截图表结构，命名4-1。

任务五：数据库连接（10分）

5.1编写代码，利用“数据库资源参数”进行数据库连接。（10分）

5.2截图代码，命名5-1。

任务六：批量插入数据（15分）

6.1 编写代码，遍历数据采集结果集合，依次将采集结果存储至数据库表中。（12分）

6.2释放资源。（3分）

6.3截图代码，命名6-1。

任务七：查看存储结果（5分）

7.1 打开数据库，查看数据库表中数据。（3分）

7.2截图数据库表中的数据，命名7-1。

7.3将完整的项目提交到考生文件夹中。（2分）

提交要求：

1)在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南工程职业技术学院 01 张三。

2)考生文件夹内共创建：项目文件，截图文件夹（保存截图）。

（2）实施条件

测试所需的软硬件设备见下表2.4.1。

表 2.4.1考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存 8G 以上，windows操作系统	用于软件开发和软件部署，每人一台。
3	截图工具		系统自带截图工具
4	pycharm		用于数据采集。
5	MySQL5.5或以上		用于数据存储。
6	Navicat10或以上		用于数据存储。

（3）考核时量

考核时间为3小时

（4）评分细则

数据采集与存储模块的考核实行 100 分制，评价内容包括技能要求、职业素养完成情况两个方面。其中，技能要求完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表 2.4.2 所示。

表 2.4.2 评分标准表评价内容

评价内容		分值	评分细则	
工作任务(80分)	项目创建	10分	没有正确导入urllib库或requests库扣5分。	5分
			没有正确导入pymysql扣5分。	5分
	解析网页	10分	没有设置headers扣4分。	4分
			没有正确返回Document对象扣6分。	6分
	提取数据	20分	数据数量不足每一个扣1分，扣完为止。	5分
			数据少一个字段扣3分，扣完为止。	5分
			结果没有正确显示在控制台上，扣10分。	10分
	创建数据库表	10分	数量数据不足扣5分，每缺少1个字段扣3分，扣完为止。	10分
	数据库连接	10分	没有获取数据库连接，扣10分； 使用参数不正确，每处扣1分； 没有正确调用对应方法，每处扣2分。	10分
	批量插入数	15分	没有释放资源，扣2分；	15分

	据		使用参数不正确，每处扣1分； 没有正确调用对应方法，每处扣2分； 异常处理不正确，扣3分； 没有释放资源，扣2分。	
	查看存储结果	5分	数据库表中数据不正确，扣3分；	3分
			没有提交项目文件，扣2分。	2分
职业素养(20分)	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、鼠标等），检查测试所需软件开发环境。不进行检查操作扣5分。	5分
	专业素养	10分	代码符合软件开发规范，命名规范，能做到见名知意；缩进统一，方便阅读；注释规范正确。整个项目命名不规范每一处扣1分；完全没有注释扣2分；有注释，但注释不规范每一处扣1分。	10分
	职业行为规范	5分	着装干净整洁，举止文明。遵守考场纪律，按顺序进出考场。	5分

5.试题 2-1-5：各国人口数量数据采集与存储

（1）任务描述

人口是一个内容复杂、综合多种社会关系的社会实体，具有性别和年龄及自然构成，多种社会构成和社会关系、经济构成和经济关系。在当前信息时代人们可以通过网站查询到各国（地区）的大概人口数量和各国人口所占世界人口百分比。下面以查询者在快易网站https://www.kylc.com/查询人口信息为例（对应局域网地址：http://172.16.7.152/PopulationOfEachCountry.html，其中IP地址根据实际局域网配置修改）。请根据快易网站源数据，综合利用大数据采集工具和相关技术对网站数据的进行数据采集，完成数据采集任务。利用数据库技术，根据采集的有效数据信息，创建数据库表结构，并将采集的有效数据存储到Mysql数据库中，且能正确查看数据结果，完成数据存储任务。帮助查询者查找到排名、国家（地区）、所在洲、人口和所占百分比等数据内容并完成数据展示与存储。

采集数据样式：

排名	国家/地区	所在洲	人口	占世界%
1	全世界		75.92亿 (7,591,932,906)	
2	中国	亚洲	13.93亿 (1,392,730,000)	18.3449%
3	印度	亚洲	13.53亿 (1,352,617,328)	17.8165%

任务一：项目创建（10分）

1.1创建项目。（10分）

2.1截图项目结构，命名1-1。

任务二：解析网页（10分）

2.1解析网页，获取html文档。（10分）

2.2截图代码，命名2-1。

任务三：提取数据（20分）

3.1根据待采集数据样式，实现数据采集，能够在控制台输出所有待采集数据。（20分）

3.2截图输出结果，命名3-1。

任务四：创建数据库表（10分）

4.1根据采集的数据样式，创建对应的表结构。（10分）

4.2截图表结构，命名4-1。

任务五：数据库连接（10分）

5.1编写代码，利用“数据库资源参数”进行数据库连接。（10分）

5.2截图代码，命名5-1。

任务六：批量插入数据（15分）

6.1编写代码，遍历数据采集结果集合，依次将采集结果存储至数据库表中。（12分）

6.2释放资源。（3分）

6.3截图代码，命名6-1。

任务七：查看存储结果（5分）

7.1打开数据库，查看数据库表中数据。（3分）

7.2截图数据库表中的数据，命名7-1。

7.3将完整的项目提交到考生文件夹中。（2分）

提交要求：

1)在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南工程职业技术学院 01 张三。

2)考生文件夹内共创建：项目文件，截图文件夹（保存截图）。

（2）实施条件

测试所需的软硬件设备见下表2.5.1。

表2.5.1考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存 8G 以上， windows操作系统	用于软件开发和软件部署，每人一台。
3	截图工具		系统自带截图工具
4	pycharm		用于数据采集。
5	MySQL5.5或以上		用于数据存储。
6	Navicat10或以上		用于数据存储。

（3）考核时量

考核时间为3小时

（4）评分细则

数据采集与存储模块的考核实行 100 分制，评价内容包括技能要求、职业素养完成情况两个方面。其中，技能要求完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表 2.5.2 所示。

表 2.5.2 评分标准表评价内容

评价内容		分值	评分细则	
工作任务(80分)	项目创建	10分	没有正确导入urllib库或requests库扣5分。	5分
			没有正确导入pymysql扣5分。	5分
	解析网页	10分	没有设置headers扣4分。	4分
			没有正确返回Document对象扣6分。	6分
	提取数据	20分	数据数量不足每一个扣1分，扣完为止。	5分
			数据少一个字段扣3分，扣完为止。	5分

			结果没有正确显示在控制台上，扣10分。	10分
	创建数据库表	10分	数量数据不足扣5分，每缺少1个字段扣3分，扣完为止。	10分
	数据库连接	10分	没有获取数据库连接，扣10分； 使用参数不正确，每处扣1分； 没有正确调用对应方法，每处扣2分。	10分
	批量插入数据	15分	没有释放资源，扣2分； 使用参数不正确，每处扣1分； 没有正确调用对应方法，每处扣2分； 异常处理不正确，扣3分； 没有释放资源，扣2分。	15分
	查看存储结果	5分	数据库表中数据不正确，扣3分；	3分
			没有提交项目文件，扣2分。	2分
职业素养(20分)	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、鼠标等），检查测试所需软件开发环境。不进行检查操作扣5分。	5分
	专业素养	10分	代码符合软件开发规范，命名规范，能做到见名知意；缩进统一，方便阅读；注释规范正确。整个项目命名不规范每一处扣1分；完全没有注释扣2分；有注释，但注释不规范每一处扣1分。	10分
	职业行为规范	5分	着装干净整洁，举止文明。遵守考场纪律，按顺序进出考场。	5分

6.试题 2-1-6：各国 GNP 数据采集与存储

（1）任务描述

国民生产总值(Gross National Product，简称GNP)是最重要的宏观经济指标，指一个国家(地区)所有常驻机构单位在一定时期内(年或季)收入初次分配的最终成果，是一国所拥有的生产要素所生产的最终产品价值，是一个国民概念。下面以快易理财网站https://www.kylc.com查询国民生产总值为例（对应局域网地址：http://172.16.7.152/GNPDataForEachCountry.html，其中IP地址根据实际局域网配置修改）。请根据快易网站源数据，综合利用大数据采集工具和相关技术对网站数据的进行数据采集，完成数据采集任务。利用数据库技术，根据采集的有效数据

信息，创建数据库表结构，并将采集的有效数据存储到Mysql数据库中，且能正确查看数据结果，完成数据存储任务。帮助使用者查找到世界排名、国家、地区、年份、GNP值和百分比等数据内容并完成数据展示与存储。

采集数据样式

世界排名	国家	地区	年份	GNP	占世界%
	全世界		2019	88.76万亿 （88,775,236,041,873）	
1	美国	美洲	2019	21.62万亿 （21,615,817,525,647）	24.3544%
	欧盟地区		2019	16.03万亿 （16,026,517,409,439）	18.0570%
2	中国	亚洲	2019	14.52万亿 （14,519,055,044,099）	16.3585%

任务一：项目创建（10分）

1.1创建项目。（10分）

2.1截图项目结构，命名1-1。

任务二：解析网页（10分）

2.1解析网页，获取html文档。（10分）

2.2截图代码，命名2-1。

任务三：提取数据（20分）

3.1根据待采集数据样式，实现数据采集，能够在控制台输出所有待采集数据。（20分）

3.2截图输出结果，命名3-1。

任务四：创建数据库表（10分）

4.1根据采集的数据样式，创建对应的表结构。（10分）

4.2截图表结构，命名4-1。

任务五：数据库连接（10分）

5.1编写代码，利用“数据库资源参数”进行数据库连接。（10分）

5.2截图代码，命名5-1。

任务六：批量插入数据（15分）

6.1编写代码，遍历数据采集结果集合，依次将采集结果存储至数据库表中。（12分）

6.2释放资源。（3分）

6.3截图代码，命名6-1。

任务七：查看存储结果（5分）

7.1打开数据库，查看数据库表中数据。（3分）

7.2截图数据库表中的数据，命名7-1。

7.3将完整的项目提交到考生文件夹中。（2分）

提交要求：

1)在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南工程职业技术学院 01 张三。

2)考生文件夹内共创建：项目文件，截图文件夹（保存截图）。

（2）实施条件

测试所需的软硬件设备见下表2.6.1。

表 2.6.1考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存 8G 以上，windows操作系统	用于软件开发和软件部署，每人一台。
3	截图工具		系统自带截图工具
4	pycharm		用于数据采集。
5	MySQL5.5或以上		用于数据存储。
6	Navicat10或以上		用于数据存储。

（3）考核时量

考核时间为3小时

（4）评分细则

数据采集与存储模块的考核实行 100 分制，评价内容包括技能要求、职业素养完成情况两个方面。其中，技能要求完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表 2.6.2 所示。

表2.6.2 评分标准表评价内容

评价内容		分值	评分细则	
工作任务(80分)	项目创建	10分	没有正确导入urllib库或requests库扣5分。	5分
			没有正确导入pymysql扣5分。	5分
	解析网页	10分	没有设置headers扣4分。	4分
			没有正确返回Document对象扣6分。	6分
	提取数据	20分	数据数量不足每一个扣1分，扣完为止。	5分
			数据少一个字段扣3分，扣完为止。	5分
			结果没有正确显示在控制台上，扣10分。	10分
	创建数据库表	10分	数量数据不足扣5分，每缺少1个字段扣3分，扣完为止。	10分
	数据库连接	10分	没有获取数据库连接，扣10分； 使用参数不正确，每处扣1分； 没有正确调用对应方法，每处扣2分。	10分
	批量插入数据	15分	没有释放资源，扣2分； 使用参数不正确，每处扣1分； 没有正确调用对应方法，每处扣2分； 异常处理不正确，扣3分； 没有释放资源，扣2分。	15分
	查看存储结果	5分	数据库表中数据不正确，扣3分；	3分
			没有提交项目文件，扣2分。	2分
职业素养(20分)	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、鼠标等），检查测试所需软件开发环境。不进行检查操作扣5分。	5分
	专业素养	10分	代码符合软件开发规范，命名规范，能做到见名知意；缩进统一，方便阅读；注释规范正确。整个项目命名不规范每一处扣1分；完全没有注释扣2分；有注释，但注释不规范每一处扣1分。	10分
	职业行为规范	5分	着装干净整洁，举止文明。遵守考场纪律，按顺序进出考场。	5分

7.试题 2-1-7：各国 GDP 数据采集与存储

（1）任务描述

GDP显示一个国家（或地区）所有常驻单位在一定时期内生产活动的最终成果，常被公认为是衡量国家经济状况的最佳指标。下面以快易理财网站https://www.kylc.com为例（对应局域网地址：http://172.16.7.152/GDPOfEachCounty.html，其中IP地址根据实际局域网配置修改）。请根据快易网站源数据，综合利用大数据采集工具和相关技术对网站数据的进行数据采集，完成数据采集任务。利用数据库技术，根据采集的有效数据信息，创建数据库表结构，并将采集的有效数据存储到Mysql数据库中，且能正确查看数据结果，完成数据存储任务。帮助使用者查找到世界排名、国家（或地区）、所在洲、GDP值和百分比等数据内容并完成数据展示与存储。

采集数据样式：

世界排名	国家/地区	所在洲	GDP（美元）	占世界%
	全世界		86.41万亿 (86,408,955,453,299)	
1	美国	美洲	20.58万亿 (20,580,223,000,000)	23.8172%
	欧盟地区		15.93万亿 (15,931,983,317,841)	18.4379%
2	中国	亚洲	13.89万亿 (13,894,817,110,036)	16.0803%

任务一：项目创建（10分）

1.1创建项目。（10分）

2.1截图项目结构，命名1-1。

任务二：解析网页（10分）

2.1解析网页，获取html文档。（10分）

2.2截图代码，命名2-1。

任务三：提取数据（20分）

3.1根据待采集数据样式，实现数据采集，能够在控制台输出所有待采集数据。（20分）

3.2截图输出结果，命名3-1。

任务四：创建数据库表（10分）

4.1根据采集的数据样式，创建对应的表结构。（10分）

4.2截图表结构，命名4-1。

任务五：数据库连接（10分）

5.1编写代码，利用“数据库资源参数”进行数据库连接。（10分）

5.2截图代码，命名5-1。

任务六：批量插入数据（15分）

6.1编写代码，遍历数据采集结果集合，依次将采集结果存储至数据库表中。（12分）

6.2释放资源。（3分）

6.3截图代码，命名6-1。

任务七：查看存储结果（5分）

7.1打开数据库，查看数据库表中数据。（3分）

7.2截图数据库表中的数据，命名7-1。

7.3将完整的项目提交到考生文件夹中。（2分）

提交要求：

1)在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南工程职业技术学院 01 张三。

2)考生文件夹内共创建：项目文件，截图文件夹（保存截图）。

（2）实施条件

测试所需的软硬件设备见下表2.7.1。

表 2.7.1考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存 8G 以上， windows操作系统	用于软件开发和软件部署，每人一台。
3	截图工具		系统自带截图工具
4	Pycharm		用于数据采集。
5	MySQL5.5或以上		用于数据存储。
6	Navicat10或以上		用于数据存储。

（3）考核时量

考核时间为3小时

（4）评分细则

数据采集与存储模块的考核实行 100 分制，评价内容包括技能要求、职业素养完成情况两个方面。其中，技能要求完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表 2.7.2 所示。

表2.7.2 评分标准表评价内容

评价内容		分值	评分细则	
工作任务（80分）	项目创建	10分	没有正确导入urllib库或requests库扣5分。	5分
			没有正确导入pymysql扣5分。	5分
	解析网页	10分	没有设置headers扣4分。	4分
			没有正确返回Document对象扣6分。	6分
	提取数据	20分	数据数量不足每一个扣1分，扣完为止。	5分
			数据少一个字段扣3分，扣完为止。	5分
			结果没有正确显示在控制台上，扣10分。	10分
	创建数据库表	10分	数量数据不足扣5分，每缺少1个字段扣3分，扣完为止。	10分
	数据库连接	10分	没有获取数据库连接，扣10分； 使用参数不正确，每处扣1分； 没有正确调用对应方法，每处扣2分。	10分
	批量插入数据	15分	没有释放资源，扣2分； 使用参数不正确，每处扣1分； 没有正确调用对应方法，每处扣2分； 异常处理不正确，扣3分； 没有释放资源，扣2分。	15分
	查看存储结果	5分	数据库表中数据不正确，扣3分；	3分
			没有提交项目文件，扣2分。	2分
职业素养（20分）	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、 鼠标等），检查测试所需软件开发环境。 不进行检查操作扣 5 分。	5分
	专业素养	10分	代码符合软件开发规范，命名规范，能做到见名知意；缩进统一，方便阅读；注释规范正确。整个项目命名不规范每一处扣 1 分；完全没有注释	10分

			扣 2 分；有注释，但注释不规范每一处扣 1 分。	
	职业行为规范	5分	着装干净整洁，举止文明。遵守考场纪律，按顺序进出考场。	5分

8.试题 2-1-8：各国国土面积数据采集与存储

（1）任务描述

世界各国领土面积排名是指各国（不含水域、殖民地）国土面积排行榜，这些国家可以分为超大型国家、大型国家、中型国家、小型国家、超小型国家和微型国家。下面以快易理财网站https://www.kylc.com为例（对应局域网地址：http://172.16.7.152/landAreaOfEeachCountryhtml.html，其中IP地址根据实际局域网配置修改）。请根据快易网站源数据，综合利用大数据采集工具和相关技术对网站数据的进行数据采集，完成数据采集任务。利用数据库技术，根据采集的有效数据信息，创建数据库表结构，并将采集的有效数据存储到Mysql数据库中，且能正确查看数据结果，完成数据存储任务。帮助使用者查找到世界排名、国家、所在洲、年份、面积值和占世界百分比等数据内容并完成数据展示与存储。

采集数据样式：

排名	国家	所在洲	年份	面积	百分比
	全世界		2018	1.35亿（134,542，704）	
1	俄罗斯	欧洲	2018	1710万（17,098,250）	12.7084%
2	加拿大	美洲	2018	988万（9,879,750）	7.3432%

任务一：项目创建（10分）

1.1创建项目。（10分）

2.1截图项目结构，命名1-1。

任务二：解析网页（10分）

2.1解析网页，获取html文档。（10分）

2.2截图代码，命名2-1。

任务三：提取数据（20分）

3.1根据待采集数据样式，实现数据采集，能够在控制台输出所有待采集数据。（20分）

3.2截图输出结果，命名3-1。

任务四：创建数据库表（10分）

4.1根据采集的数据样式，创建对应的表结构。（10分）

4.2截图表结构，命名4-1。

任务五：数据库连接（10分）

5.1编写代码，利用“数据库资源参数”进行数据库连接。（10分）

5.2截图代码，命名5-1。

任务六：批量插入数据（15分）

6.1编写代码，遍历数据采集结果集合，依次将采集结果存储至数据库表中。（12分）

6.2释放资源。（3分）

6.3截图代码，命名6-1。

任务七：查看存储结果（5分）

7.1打开数据库，查看数据库表中数据。（3分）

7.2截图数据库表中的数据，命名7-1。

7.3将完整的项目提交到考生文件夹中。（2分）

提交要求：

1)在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南工程职业技术学院 01 张三。

2)考生文件夹内共创建：项目文件，截图文件夹（保存截图）。

（2）实施条件

测试所需的软硬件设备见下表2.8.1。

表 2.8.1考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存 8G 以上， windows操作系统	用于软件开发和软件部署，每人一台。
3	截图工具		系统自带截图工具
4	Pycharm		用于数据采集。
5	MySQL5.5或以上		用于数据存储。
6	Navicat10或以上		用于数据存储。

（3）考核时量

考核时间为3小时

（4）评分细则

数据采集与存储模块的考核实行 100 分制，评价内容包括技能要求、职业素养完成情况两个方面。其中，技能要求完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表 2.8.2 所示。

表 2.8.2 评分标准表评价内容

评价内容		分值	评分细则	
工作任务（80分）	项目创建	10分	没有正确导入urllib库或requests库扣5分。	5分
			没有正确导入pymysql扣5分。	5分
	解析网页	10分	没有设置headers扣4分。	4分
			没有正确返回Document对象扣6分。	6分
	提取数据	20分	数据数量不足每一个扣1分，扣完为止。	5分
			数据少一个字段扣3分，扣完为止。	5分
			结果没有正确显示在控制台上，扣10分。	10分
	创建数据库表	10分	数量数据不足扣5分，每缺少1个字段扣3分，扣完为止。	10分
	数据库连接	10分	没有获取数据库连接，扣10分； 使用参数不正确，每处扣1分； 没有正确调用对应方法，每处扣2分。	10分
	批量插入数据	15分	没有释放资源，扣2分； 使用参数不正确，每处扣1分； 没有正确调用对应方法，每处扣2分； 异常处理不正确，扣3分； 没有释放资源，扣2分。	15分
	查看存储结果	5分	数据库表中数据不正确，扣3分；	3分
			没有提交项目文件，扣2分。	2分
职业素养（20分）	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、鼠标等），检查测试所需软件开发环境。不进行检查操作扣 5 分。	5分
	专业素养	10分	代码符合软件开发规范，命名规范，能做到见名知	10分

			意；缩进统一，方便阅读；注释规范正确。整个项目命名不规范每一处扣 1 分；完全没有注释扣 2 分；有注释，但注释不规范每一处扣 1 分。	
	职业行为规范	5分	着装干净整洁，举止文明。遵守考场纪律，按顺序进出考场。	5分

9.试题 2-1-9：野生动物图片采集与存储

（1）任务描述

工作或生活中，常常要下载图片素材，人们一般都会在图片网站查找图片并进行下载。下面以昵图网站https://www.nipic.com/为例（对应局域网地址：http://172.16.7.152/picturesForWildlife.html，其中IP地址根据实际局域网配置修改），请根据昵图网站源数据，综合利用大数据采集工具和相关技术对昵图网摄影类中的野生动物图片第1页进行采集，并将图片存储至本地的文件夹中。

采集数据样式：

任务一：项目创建（10分）

1.1创建项目。（10分）

2.1截图项目结构，命名1-1。

任务二：解析网页（35分）

2.1找出网页编码的规律，设置正确的url。（5分）

2.2解析网页，获取html文档。（10分）

2.3将所有图片的标题合并到列表中。（10分）

2.4将所有图片的文件地址合并到列表中。（10分）

2.5截图代码，命名2-1。

任务三：提取数据（30分）

3.1为待采集图片编号，编号从数字1开始。（5分）

3.2为待采集图片命名，名称为编号+标题。（5分）

3.3下载图片至本地文件夹。（10分）

3.4在控制台输出正在下载的图片的提示信息。（5分）

3.3下载结束后，输出提示信息。（5分）

3.3截图输出结果，命名3-1。

任务四：查看存储结果（5分）

4.1打开文件夹，查看图片的数量和名称。（3分）

4.2截图文件夹中的文件，命名4-1。

4.3将完整的项目提交到考生文件夹中。（2分）

提交要求：

1)在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南工程职业技术学院 01 张三。

2)考生文件夹内共创建：项目文件，截图文件夹（保存截图）。

（2）实施条件

测试所需的软硬件设备见下表2.9.1。

表 2.9.1考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存 8G 以上， windows操作系统	用于软件开发和软件部署，每人一台。
3	截图工具		系统自带截图工具
4	Pycharm		用于数据采集。

（3）考核时量

考核时间为3小时

（4）评分细则

数据采集与存储模块的考核实行 100 分制，评价内容包括技能要求、职业素养完成情况两个方面。其中，技能要求完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表 2.9.2 所示。

表 2.9.2 评分标准表评价内容

评价内容		分值	评分细则	
工作任务(80分)	项目创建	10分	没有正确导入urllib库或requests库扣5分。	5分
			没有正确导入pymysql扣5分。	5分
	解析网页	35分	没有设置headers扣4分。	4分
			没有设置正确的url，少一页，扣1分。	5分
			没有正确返回Document对象扣6分。	6分
			没有将所有图片的标题合并到列表中，少1页，扣2分。	10分
			没有将所有图片的文件地址合并到列表中，少1页，扣2分。	10分
	提取数据	30分	未用数字编号，扣5分；数字编号未从1开始编号，扣1分。	5分
			图片命名未加标题，扣5分。	5分
			图片未下载至本地文件夹，扣10分。	10分
			未输出正在下载图片的提示信息，扣10分。	10分
			下载结束后，未输出提示信息，扣5分。	5分
	查看存储结果	5分	文件夹中图片编号和数量不正确，扣3分；	3分
			没有提交项目文件，扣2分。	2分
职业素养(20分)	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、鼠标等），检查测试所需软件开发环境。不进行检查操作扣 5 分。	5分
	专业素养	10分	代码符合软件开发规范，命名规范，能做到见名知意；缩进统一，方便阅读；注释规范正确。整个项目命名不规范每一处扣 1 分；完全没有注释扣 2 分；有注释，但注释不规范每一处扣 1 分。	10分
	职业行为规范	5分	着装干净整洁，举止文明。遵守考场纪律，按顺序进出考场。	5分

10. 试题 2-1-10：PPT 背景图片采集与存储

（1）任务描述

工作或生活中，常常要下载图片素材，人们一般都会在图片网站查找图片并进行下载。下面以优品PPT网站https://www.ypppt.com/beijing/为例（对应局域网地址：http://172.16.7.152/backGroundForPPT.html，其中IP地址根据实际局域网配置修改），请根据优品PPT网站源数据，综合利用大数据采集工具和相关技术对优品PPT背景图片第1页进行采集，并将图片存储至本地的文件夹中。

采集数据样式：

任务一：项目创建（10分）

1.1创建项目。（10分）

2.1截图项目结构，命名1-1。

任务二：解析网页（35分）

2.1找出网页编码的规律，设置正确的url。（5分）

2.2　解析网页，获取html文档。（10分）

2.3将所有图片的标题合并到列表中。（10分）

2.4将所有图片的文件地址合并到列表中。（10分）

2.5截图代码，命名2-1。

任务三：提取数据（30分）

3.1为待采集图片编号，编号从数字1开始。（5分）

3.2为待采集图片命名，名称为编号+标题。（5分）

3.3下载图片至本地文件夹。（10分）

3.4在控制台输出正在下载的图片的提示信息。（5分）

3.3下载结束后，输出提示信息。（5分）

3.3截图输出结果，命名3-1。

任务四：查看存储结果（5分）

4.1打开文件夹，查看图片的数量和名称。（3分）

4.2截图文件夹中的文件，命名4-1。

4.3将完整的项目提交到考生文件夹中。（2分）

提交要求：

1)在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南工程职业技术学院 01 张三。

2)考生文件夹内共创建：项目文件，截图文件夹（保存截图）。

（2）实施条件

测试所需的软硬件设备见下表2.10.1。

表 2.10.1考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存 8G 以上，windows操作系统	用于软件开发和软件部署，每人一台。
3	截图工具		系统自带截图工具
4	Pycharm		用于数据采集。

（3）考核时量

考核时间为3小时

（4）评分细则

数据采集与存储模块的考核实行 100 分制，评价内容包括技能要求、职业素养完成情况两个方面。其中，技能要求完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表 2.10.2 所示。

表 2.10.2 评分标准表评价内容

评价内容		分值	评分细则	
工作任务(80分)	项目创建	10分	没有正确导入urllib库或requests库扣5分。	5分
			没有正确导入pymysql扣5分。	5分
	解析网页	35分	没有设置headers扣4分。	4分
			没有设置正确的url，少一页，扣1分。	5分
			没有正确返回Document对象扣6分。	6分
			没有将所有图片的标题合并到列表中，少1页，扣2	10分

			分。	
			没有将所有图片的文件地址合并到列表中，少1页，扣2分。	10分
	提取数据	30分	未用数字编号，扣5分；数字编号未从1开始编号，扣1分。	5分
			图片命名未加标题，扣5分。	5分
			图片未下载至本地文件夹，扣10分。	10分
			未输出正在下载图片的提示信息，扣10分。	10分
			下载结束后，未输出提示信息，扣5分。	5分
	查看存储结果	5分	文件夹中图片编号和数量不正确，扣3分；	3分
			没有提交项目文件，扣2分。	2分
职业素养(20分)	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、鼠标等），检查测试所需软件开发环境。不进行检查操作扣5分。	5分
	专业素养	10分	代码符合软件开发规范，命名规范，能做到见名知意；缩进统一，方便阅读；注释规范正确。整个项目命名不规范每一处扣1分；完全没有注释扣2分；有注释，但注释不规范每一处扣1分。	10分
	职业行为规范	5分	着装干净整洁，举止文明。遵守考场纪律，按顺序进出考场。	5分

11.试题 2-1-11：招聘信息采集与存储

（1）任务描述

随着网络的普及，在网上找工作的人越来越多，人们可以通过网站查询招聘信息，实现就业。下面以查询者在猎聘网https://www.liepin.com/zhaopin/?inputFrom=head_navigation&scene=init&workYearCode=0&ckId=k80cz868nqygmoltydpeu577i2ua4cp9查询职位信息为例（对应局域网地址：http://172.16.7.152/RecruitmentInformation.html，其中IP地址根据实际局域网配置修改）。请综合利用大数据采集工具和相关技术对爬虫岗位数据的第1页进行采集，完成数据采集任务。利用数据库技术，根据采集的有效数据信息，创建数据库表结构，并将采集的有效数据存储到Mysql数据库中，且能正确查看数据

结果，完成数据存储任务。帮助查询者查找到职位名称、工作地点、薪资待遇等数据内容并完成数据展示与存储。

采集数据样式：

职位名称	公司名称	工作地点	薪资
网络爬虫工程师	成都金榜路教育科技有限公司	成都-武侯区	6-10k
python爬虫工程师	深圳市中海通物流股份有限公司	深圳	10-20k·13薪

任务一：项目创建（10分）

1.1创建项目。（10分）

2.1截图项目结构，命名1-1。

任务二：解析网页（10分）

2.1解析网页，获取html文档。（10分）

2.2截图代码，命名2-1。

任务三：提取数据（20分）

3.1根据待采集数据样式，实现数据采集，能够在控制台输出所有待采集数据。（20分）

3.2截图输出结果，命名3-1。

任务四：创建数据库表（10分）

4.1根据采集的数据样式，创建对应的表结构。（10分）

4.2截图表结构，命名4-1。

任务五：数据库连接（10分）

5.1编写代码，利用“数据库资源参数”进行数据库连接。（10分）

5.2截图代码，命名5-1。

任务六：批量插入数据（15分）

6.1编写代码，遍历数据采集结果集合，依次将采集结果存储至数据库表中。（12分）

6.2释放资源。（3分）

6.3截图代码，命名6-1。

任务七：查看存储结果（5分）

7.1打开数据库，查看数据库表中数据。（3分）

7.2截图数据库表中的数据，命名7-1。

7.3将完整的项目提交到考生文件夹中。（2分）

提交要求：

1)在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南工程职业技术学院 01 张三。

2)考生文件夹内共创建：项目文件，截图文件夹（保存截图）。

（2）实施条件

测试所需的软硬件设备见下表2.11.1。

表 2.11.1考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存 8G 以上， windows操作系统	用于软件开发和软件部署，每人一台。
3	截图工具		系统自带截图工具
4	pycharm		用于数据采集。
5	MySQL5.5或以上		用于数据存储。
6	Navicat10或以上		用于数据存储。

（3）考核时量

考核时间为3小时

（4）评分细则

数据采集与存储模块的考核实行 100 分制，评价内容包括技能要求、职业素养完成情况两个方面。其中，技能要求完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表 2.11.2 所示。

表2.11.2 评分标准表评价内容

评价内容		分值	评分细则	
工作任务(80分)	项目创建	10分	没有正确导入urllib库或requests库扣5分。	5分
			没有正确导入pymysql扣5分。	5分
	解析网页	10分	没有设置headers扣4分。	4分
			没有正确返回Document对象扣6分。	6分
	提取数据	20分	数据数量不足每一个扣1分，扣完为止。	5分
			数据少一个字段扣3分，扣完为止。	5分

			结果没有正确显示在控制台上，扣10分。	10分
	创建数据库表	10分	数量数据不足扣5分，每缺少1个字段扣3分，扣完为止。	10分
	数据库连接	10分	没有获取数据库连接，扣10分； 使用参数不正确，每处扣1分； 没有正确调用对应方法，每处扣2分。	10分
	批量插入数据	15分	没有释放资源，扣2分； 使用参数不正确，每处扣1分； 没有正确调用对应方法，每处扣2分； 异常处理不正确，扣3分； 没有释放资源，扣2分。	15分
	查看存储结果	5分	数据库表中数据不正确，扣3分；	3分
			没有提交项目文件，扣2分。	2分
职业素养(20分)	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、鼠标等），检查测试所需软件开发环境。不进行检查操作扣5分。	5分
	专业素养	10分	代码符合软件开发规范，命名规范，能做到见名知意；缩进统一，方便阅读；注释规范正确。整个项目命名不规范每一处扣1分；完全没有注释扣2分；有注释，但注释不规范每一处扣1分。	10分
	职业行为规范	5分	着装干净整洁，举止文明。遵守考场纪律，按顺序进出考场。	5分

12.试题 2-1-12：房产销售数据采集与存储

（1）任务描述

信息时代人们可以通过房地产网站随时随地查询到房屋的出租、出售等各项信息。下面以购房者在链家网站https://cs.lianjia.com/查询梅溪湖周边二手房房屋信息为例（对应局域网地址：http://172.16.7.152/houseSales.html，其中IP地址根据实际局域网配置修改）。请综合利用大数据采集工具和相关技术对网站第1页数据进行数据采集，完成数据采集任务。帮助购房者查找到标题、位置、房屋信息、总价和单价等数据内容并完成数据展示。利用数据库技术，根据采集的有效数据信息，创建数据库表结构，并将采集的有效数据存储到Mysql数据库中，且能正确查看

数据结果，完成数据存储任务。帮助查询者查找到标题、位置、房屋信息、总价和单价等数据内容并完成数据展示与存储。

采集数据样式：

标题	位置	房屋信息	总价	单价
浅水湾精装复式公寓，看湖景红本在手 可拎包入住！	卓越浅水湾-梅溪湖南岸	1室1厅 | 34.9平米 | 南 北 | 精装 | 中楼层(共45层) | 板塔结合	58.8万	单价16849元/平米

任务一：项目创建（10分）

1.1创建项目。（10分）

2.1截图项目结构，命名1-1。

任务二：解析网页（10分）

2.1解析网页，获取html文档。（10分）

2.2截图代码，命名2-1。

任务三：提取数据（20分）

3.1根据待采集数据样式，实现数据采集，能够在控制台输出所有待采集数据。（20分）

3.2截图输出结果，命名3-1。

任务四：创建数据库表（10分）

4.1根据采集的数据样式，创建对应的表结构。（10分）

4.2截图表结构，命名4-1。

任务五：数据库连接（10分）

5.1编写代码，利用“数据库资源参数”进行数据库连接。（10分）

5.2截图代码，命名5-1。

任务六：批量插入数据（15分）

6.1编写代码，遍历数据采集结果集合，依次将采集结果存储至数据库表中。（12分）

6.2释放资源。（3分）

6.3截图代码，命名6-1。

任务七：查看存储结果（5分）

7.1打开数据库，查看数据库表中数据。（3分）

7.2截图数据库表中的数据，命名7-1。

7.3将完整的项目提交到考生文件夹中。（2分）

提交要求：

1)在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南工程职业技术学院 01 张三。

2)考生文件夹内共创建：项目文件，截图文件夹（保存截图）。

（2）实施条件

测试所需的软硬件设备见下表2.12.1。

表 2.12.1考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存 8G 以上， windows操作系统	用于软件开发和软件部署，每人一台。
3	截图工具		系统自带截图工具
4	pycharm		用于数据采集。
5	MySQL5.5或以上		用于数据存储。
6	Navicat10或以上		用于数据存储。

（3）考核时量

考核时间为3小时

（4）评分细则

数据采集与存储模块的考核实行 100 分制，评价内容包括技能要求、职业素养完成情况两个方面。其中，技能要求完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表 2.12.2 所示。

表 2.12.2 评分标准表评价内容

评价内容		分值	评分细则	
	项目创建	10分	没有正确导入urllib库或requests库扣5分。	5分
			没有正确导入pymysql扣5分。	5分
	解析网页	10分	没有设置headers扣4分。	4分
			没有正确返回Document对象扣6分。	6分

工作任务(80分)	提取数据	20分	数据数量不足每一个扣1分，扣完为止。	5分
			数据少一个字段扣3分，扣完为止。	5分
			结果没有正确显示在控制台上，扣10分。	10分
	创建数据库表	10分	数量数据不足扣5分，每缺少1个字段扣3分，扣完为止。	10分
	数据库连接	10分	没有获取数据库连接，扣10分； 使用参数不正确，每处扣1分； 没有正确调用对应方法，每处扣2分。	10分
	批量插入数据	15分	没有释放资源，扣2分； 使用参数不正确，每处扣1分； 没有正确调用对应方法，每处扣2分； 异常处理不正确，扣3分； 没有释放资源，扣2分。	15分
	查看存储结果	5分	数据库表中数据不正确，扣3分；	3分
			没有提交项目文件，扣2分。	2分
职业素养(20分)	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、鼠标等），检查测试所需软件开发环境。不进行检查操作扣5分。	5分
	专业素养	10分	代码符合软件开发规范，命名规范，能做到见名知意；缩进统一，方便阅读；注释规范正确。整个项目命名不规范每一处扣1分；完全没有注释扣2分；有注释，但注释不规范每一处扣1分。	10分
	职业行为规范	5分	着装干净整洁，举止文明。遵守考场纪律，按顺序进出考场。	5分

13.试题 2-1-13：世界各国服务业增加值数据采集与存储

（1）任务描述

服务业与通常所指的第三产业，大体上可以理解为一个概念。服务业是国际通行的产业分类概念，指那些提供非实物产品为主的行业。下面以查询者在快易网站https://www.kylc.com/查询世界各国服务业增加值为例（对应局域网地址：http://172.16.7.152/serviceIndustryInTheWorld.html，其中IP地址根据实际局域网配置修改）。请综合利用大数据采集工具和相关技术对网站数据的进行数据采

集，完成数据采集任务。利用数据库技术，根据采集的有效数据信息，创建数据库表结构，并将采集的有效数据存储到Mysql数据库中，且能正确查看数据结果，完成数据存储任务。帮助查询者查找到排名、国家（地区）、所在洲、年份和服务业增加值(美元)等数据内容并完成数据展示与存储。

采集数据样式：

排名	国家/地区	所在洲	年份	服务业增加值(美元)
	全世界	美洲	2019年	56.65万亿 (56,650,602,909,964)
1	美国		2019年	16.57万亿 (16,571,067,139,000)
	欧盟地区		2020年	10.02万亿 (10,018,539,020,954)
2	中国	亚洲	2020年	8.03万亿 (8,027,718,525,388)

任务一：项目创建（10分）

1.1创建项目。（10分）

2.1截图项目结构，命名1-1。

任务二：解析网页（10分）

2.1解析网页，获取html文档。（10分）

2.2截图代码，命名2-1。

任务三：提取数据（20分）

3.1根据待采集数据样式，实现数据采集，能够在控制台输出所有待采集数据。（20分）

3.2截图输出结果，命名3-1。

任务四：创建数据库表（10分）

4.1根据采集的数据样式，创建对应的表结构。（10分）

4.2截图表结构，命名4-1。

任务五：数据库连接（10分）

5.1编写代码，利用“数据库资源参数”进行数据库连接。（10分）

5.2截图代码，命名5-1。

任务六：批量插入数据（15分）

6.1编写代码，遍历数据采集结果集合，依次将采集结果存储至数据库表中。（12分）

6.2释放资源。（3分）

6.3截图代码，命名6-1。

任务七：查看存储结果（5分）

7.1打开数据库，查看数据库表中数据。（3分）

7.2截图数据库表中的数据，命名7-1。

7.3将完整的项目提交到考生文件夹中。（2分）

提交要求：

1)在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南工程职业技术学院 01 张三。

2)考生文件夹内共创建：项目文件，截图文件夹（保存截图）。

（2）实施条件

测试所需的软硬件设备见下表2.13.1。

表2.13.1考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存 8G 以上， windows操作系统	用于软件开发和软件部署，每人一台。
3	截图工具		系统自带截图工具
4	pycharm		用于数据采集。
5	MySQL5.5或以上		用于数据存储。
6	Navicat10或以上		用于数据存储。

（3）考核时量

考核时间为3小时

（4）评分细则

数据采集与存储模块的考核实行 100 分制，评价内容包括技能要求、职业素养完成情况两个方面。其中，技能要求完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表 2.13.2 所示。

表 2.13.2 评分标准表评价内容

评价内容		分值	评分细则	
	项目创建	10分	没有正确导入urllib库或requests库扣5分。	5分
			没有正确导入pymysql扣5分。	5分

工作任务（80分）	解析网页	10分	没有设置headers扣4分。	4分
			没有正确返回Document对象扣6分。	6分
	提取数据	20分	数据数量不足每一个扣1分，扣完为止。	5分
			数据少一个字段扣3分，扣完为止。	5分
			结果没有正确显示在控制台上，扣10分。	10分
	创建数据库表	10分	数量数据不足扣5分，每缺少1个字段扣3分，扣完为止。	10分
	数据库连接	10分	没有获取数据库连接，扣10分； 使用参数不正确，每处扣1分； 没有正确调用对应方法，每处扣2分。	10分
	批量插入数据	15分	没有释放资源，扣2分； 使用参数不正确，每处扣1分； 没有正确调用对应方法，每处扣2分； 异常处理不正确，扣3分； 没有释放资源，扣2分。	15分
	查看存储结果	5分	数据库表中数据不正确，扣3分；	3分
			没有提交项目文件，扣2分。	2分
职业素养（20分）	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、鼠标等），检查测试所需软件开发环境。不进行检查操作扣5分。	5分
	专业素养	10分	代码符合软件开发规范，命名规范，能做到见名知意；缩进统一，方便阅读；注释规范正确。整个项目命名不规范每一处扣1分；完全没有注释扣2分；有注释，但注释不规范每一处扣1分。	10分
	职业行为规范	5分	着装干净整洁，举止文明。遵守考场纪律，按顺序进出考场。	5分

14.试题 2-1-14：世界各国国民总储蓄数据采集与存储

（1）任务描述

总储蓄指可支配总收入用于最终消费后的余额。国民总储蓄指各部门的总储蓄之和。总储蓄和居民储蓄的概念不同，总储蓄是宏观的概念，是可支配收入减去最终消费后，用于一个国家或地区投资资金的主要来源。下面以查询者在快易网站

https://www.kylc.com/查询世界各国总储蓄为例（对应局域网地址：http://172.16.7.152/totalSavings.html，其中IP地址根据实际局域网配置修改）。请综合利用大数据采集工具和相关技术对网站数据的进行数据采集，完成数据采集任务。利用数据库技术，根据采集的有效数据信息，创建数据库表结构，并将采集的有效数据存储到Mysql数据库中，且能正确查看数据结果，完成数据存储任务。帮助查询者查找到排名、国家（地区）、所在洲、年份和总储蓄(US$)等数据内容并完成数据展示与存储。

采集数据样式：

排名	国家/地区	所在洲	年份	总储蓄(US$)
1	中国	亚洲	2019	6.26万亿 (6,256,953,481,218)
2	美国	美洲	2020	4.01万亿 (4,014,933,312,453)
3	日本	亚洲	2019	1.42万亿 (1,415,514,616,627)

任务一：项目创建（10分）

1.1创建项目。（10分）

2.1截图项目结构，命名1-1。

任务二：解析网页（10分）

2.1解析网页，获取html文档。（10分）

2.2截图代码，命名2-1。

任务三：提取数据（20分）

3.1根据待采集数据样式，实现数据采集，能够在控制台输出所有待采集数据。（20分）

3.2截图输出结果，命名3-1。

任务四：创建数据库表（10分）

4.1根据采集的数据样式，创建对应的表结构。（10分）

4.2截图表结构，命名4-1。

任务五：数据库连接（10分）

5.1编写代码，利用“数据库资源参数”进行数据库连接。（10分）

5.2截图代码，命名5-1。

任务六：批量插入数据（15分）

6.1编写代码，遍历数据采集结果集合，依次将采集结果存储至数据库表中。（12分）

6.2释放资源。（3分）

6.3截图代码，命名6-1。

任务七：查看存储结果（5分）

7.1打开数据库，查看数据库表中数据。（3分）

7.2截图数据库表中的数据，命名7-1。

7.3将完整的项目提交到考生文件夹中。（2分）

提交要求：

1)在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南工程职业技术学院 01 张三。

2)考生文件夹内共创建：项目文件，截图文件夹（保存截图）。

（2）实施条件

测试所需的软硬件设备见下表2.14.1。

表 2.14.1考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存 8G 以上， windows操作系统	用于软件开发和软件部署，每人一台。
3	截图工具		系统自带截图工具
4	pycharm		用于数据采集。
5	MySQL5.5或以上		用于数据存储。
6	Navicat10或以上		用于数据存储。

（3）考核时量

考核时间为3小时

（4）评分细则

数据采集与存储模块的考核实行 100 分制，评价内容包括技能要求、职业素养完成情况两个方面。其中，技能要求完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表2.14.2 所示。

表 2.14.2 评分标准表评价内容

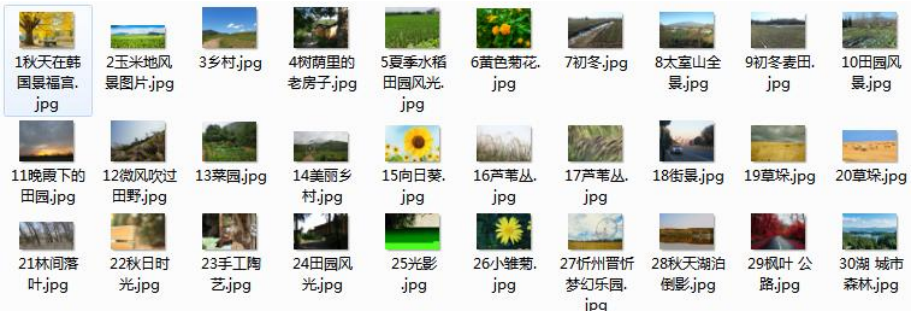
评价内容		分值	评分细则	
工作任务(80分)	项目创建	10分	没有正确导入urllib库或requests库扣5分。	5分
			没有正确导入pymysql扣5分。	5分
	解析网页	10分	没有设置headers扣4分。	4分
			没有正确返回Document对象扣6分。	6分
	提取数据	20分	数据数量不足每一个扣1分，扣完为止。	5分
			数据少一个字段扣3分，扣完为止。	5分
			结果没有正确显示在控制台上，扣10分。	10分
	创建数据库表	10分	数量数据不足扣5分，每缺少1个字段扣3分，扣完为止。	10分
	数据库连接	10分	没有获取数据库连接，扣10分； 使用参数不正确，每处扣1分； 没有正确调用对应方法，每处扣2分。	10分
	批量插入数据	15分	没有释放资源，扣2分； 使用参数不正确，每处扣1分； 没有正确调用对应方法，每处扣2分； 异常处理不正确，扣3分； 没有释放资源，扣2分。	15分
	查看存储结果	5分	数据库表中数据不正确，扣3分；	3分
			没有提交项目文件，扣2分。	2分
职业素养(20分)	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、鼠标等），检查测试所需软件开发环境。不进行检查操作扣5分。	5分
	专业素养	10分	代码符合软件开发规范，命名规范，能做到见名知意；缩进统一，方便阅读；注释规范正确。整个项目命名不规范每一处扣1分；完全没有注释扣2分；有注释，但注释不规范每一处扣1分。	10分
	职业行为规范	5分	着装干净整洁，举止文明。遵守考场纪律，按顺序进出考场。	5分

15.试题 2-1-15：自然景观图片采集与存储

（1）任务描述

工作或生活中，常常要下载图片素材，人们一般都会在图片网站查找图片并进行下载。下面以昵图网站https://www.nipic.com/为例（对应局域网地址：http://172.16.7.152/NaturalLandscapePictures.html，其中IP地址根据实际局域网配置修改）。请根据昵图网站源数据，综合利用大数据采集工具和相关技术对昵图网摄影类中的自然景观图片第1页进行采集，并将图片存储至本地的文件夹中。

采集数据样式：



任务一：项目创建（10分）

1.1创建项目。（10分）

2.1截图项目结构，命名1-1。

任务二：解析网页（35分）

2.1找出网页编码的规律，设置正确的url。（5分）

2.2解析网页，获取html文档。（10分）

2.3将所有图片的标题合并到列表中。（10分）

2.4将所有图片的文件地址合并到列表中。（10分）

2.5截图代码，命名2-1。

任务三：提取数据（30分）

3.1为待采集图片编号，编号从数字1开始。（5分）

3.2为待采集图片命名，名称为编号+标题。（5分）

3.3下载图片至本地文件夹。（10分）

3.4在控制台输出正在下载的图片的提示信息。（5分）

3.3下载结束后，输出提示信息。（5分）

3.3截图输出结果，命名3-1。

任务四：查看存储结果（5分）

4.1打开文件夹，查看图片的数量和名称。（3分）

4.2截图文件夹中的文件，命名4-1。

4.3将完整的项目提交到考生文件夹中。（2分）

提交要求：

1)在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南工程职业技术学院 01 张三。

2)考生文件夹内共创建：项目文件，截图文件夹（保存截图）。

（2）实施条件

测试所需的软硬件设备见下表2.15.1。

表 2.15.1考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存 8G 以上， windows操作系统	用于软件开发和软件部署，每人一台。
3	截图工具		系统自带截图工具
4	Pycharm		用于数据采集。

（3）考核时量

考核时间为3小时

（4）评分细则

数据采集与存储模块的考核实行 100 分制，评价内容包括技能要求、职业素养完成情况两个方面。其中，技能要求完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表 2.15.2 所示。

表 2.15.2 评分标准表评价内容

评价内容		分值	评分细则	
工作任务	项目创建	10分	没有正确导入urllib库或requests库扣5分。	5分
			没有正确导入pymysql扣5分。	5分
	解析网页	35分	没有设置headers扣4分。	4分
			没有设置正确的url，少一页，扣1分。	5分
			没有正确返回Document对象扣6分。	6分

(80分)			没有将所有图片的标题合并到列表中，少1页，扣2分。	10分
			没有将所有图片的文件地址合并到列表中，少1页，扣2分。	10分
	提取数据	30分	未用数字编号，扣5分；数字编号未从1开始编号，扣1分。	5分
			图片命名未加标题，扣5分。	5分
			图片未下载至本地文件夹，扣10分。	10分
			未输出正在下载图片的提示信息，扣10分。	10分
			下载结束后，未输出提示信息，扣5分。	5分
	查看存储结果	5分	文件夹中图片编号和数量不正确，扣3分；	3分
			没有提交项目文件，扣2分。	2分
职业素养(20分)	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、鼠标等），检查测试所需软件开发环境。不进行检查操作扣5分。	5分
	专业素养	10分	代码符合软件开发规范，命名规范，能做到见名知意；缩进统一，方便阅读；注释规范正确。整个项目命名不规范每一处扣1分；完全没有注释扣2分；有注释，但注释不规范每一处扣1分。	10分
	职业行为规范	5分	着装干净整洁，举止文明。遵守考场纪律，按顺序进出考场。	5分

16.试题 2-1-16：字体库信息采集与存储

（1）任务描述

工作或生活中，常常要下载图片素材，人们一般都会在图片网站查找图片并进行下载。下面以优品PPT网站https://www.ypppt.com/ziti/为例（对应局域网地址：http://172.16.7.152/FontLibraryInformation.html，其中IP地址根据实际局域网配置修改）。请根据优品PPT网站源数据，综合利用大数据采集工具和相关技术对字体库第1页的字体名称和下载地址信息进行采集，完成数据展示与存储。

采集数据样式：

字体	下载地址
默陌月芽体	https://www.ypppt.com/article/2022/13981.html
博洋柳体3500	https://www.ypppt.com/article/2022/13980.html
造字工房风舞体(非商用)	https://www.ypppt.com/article/2022/13866.html
微软雅黑	https://www.ypppt.com/article/2022/13865.html

任务一：项目创建（10分）

1.1创建项目。（10分）

2.1截图项目结构，命名1-1。

任务二：解析网页（10分）

2.1解析网页，获取html文档。（10分）

2.2截图代码，命名2-1。

任务三：提取数据（20分）

3.1根据待采集数据样式，实现数据采集，能够在控制台输出所有待采集数据。（20分）

3.2截图输出结果，命名3-1。

任务四：创建数据库表（10分）

4.1根据采集的数据样式，创建对应的表结构。（10分）

4.2截图表结构，命名4-1。

任务五：数据库连接（10分）

5.1编写代码，利用“数据库资源参数”进行数据库连接。（10分）

5.2截图代码，命名5-1。

任务六：批量插入数据（15分）

6.1编写代码，遍历数据采集结果集合，依次将采集结果存储至数据库表中。（12分）

6.2释放资源。（3分）

6.3截图代码，命名6-1。

任务七：查看存储结果（5分）

7.1打开数据库，查看数据库表中数据。（3分）

7.2截图数据库表中的数据，命名7-1。

7.3将完整的项目提交到考生文件夹中。（2分）

提交要求：

1)在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南工程职业技术学院 01 张三。

2)考生文件夹内共创建：项目文件，截图文件夹（保存截图）。

（2）实施条件

测试所需的软硬件设备见下表2.16.1。

表 2.16.1考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存 8G 以上，windows操作系统	用于软件开发和软件部署，每人一台。
3	截图工具		系统自带截图工具
4	pycharm		用于数据采集。
5	MySQL5.5或以上		用于数据存储。
6	Navicat10或以上		用于数据存储。

（3）考核时量

考核时间为3小时

（4）评分细则

数据采集与存储模块的考核实行 100 分制，评价内容包括技能要求、职业素养完成情况两个方面。其中，技能要求完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表 2.16.2 所示。

表 2.16.2 评分标准表评价内容

评价内容		分值	评分细则	
	项目创建	10分	没有正确导入urllib库或requests库扣5分。	5分
			没有正确导入pymysql扣5分。	5分
	解析网页	10分	没有设置headers扣4分。	4分
			没有正确返回Document对象扣6分。	6分

工作任务(80分)	提取数据	20分	数据数量不足每一个扣1分，扣完为止。	5分
			数据少一个字段扣3分，扣完为止。	5分
			结果没有正确显示在控制台上，扣10分。	10分
	创建数据库表	10分	数量数据不足扣5分，每缺少1个字段扣3分，扣完为止。	10分
	数据库连接	10分	没有获取数据库连接，扣10分； 使用参数不正确，每处扣1分； 没有正确调用对应方法，每处扣2分。	10分
	批量插入数据	15分	没有释放资源，扣2分； 使用参数不正确，每处扣1分； 没有正确调用对应方法，每处扣2分； 异常处理不正确，扣3分； 没有释放资源，扣2分。	15分
	查看存储结果	5分	数据库表中数据不正确，扣3分；	3分
			没有提交项目文件，扣2分。	2分
职业素养(20分)	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、鼠标等），检查测试所需软件开发环境。不进行检查操作扣5分。	5分
	专业素养	10分	代码符合软件开发规范，命名规范，能做到见名知意；缩进统一，方便阅读；注释规范正确。整个项目命名不规范每一处扣1分；完全没有注释扣2分；有注释，但注释不规范每一处扣1分。	10分
	职业行为规范	5分	着装干净整洁，举止文明。遵守考场纪律，按顺序进出考场。	5分

三、数据分析与可视化模块

1.试题 3-1-1：51JOB 网站大数据岗位数据分析与可视化

（1）任务描述

51job网站是中国具有广泛影响力的人力资源服务供应商。求职者一般都会在该网站上查找求职信息。对相关招聘数据进行数据分析与可视化，能帮助求职者更好地求职。

下面以51job网站为例，请根据“51job.csv”源数据，利用 Python分析知识对大数据岗位数据的清洗和整理，完成数据处理和分析任务。使用Python语言完成数据可视化，帮助求职者提前了解岗位的现状。

任务一：数据处理和分析（50分）

1.1打开Jupyter Notebook编辑器新建Python文件。（10分）

1.2读取“51job.csv”文件，显示前5条数据，截图1-1。（20分）

1.3读取csv文件，利用Python语言分析各城市大数据相关岗位的职位数量并显示所有数据，截图1-2。（20分）

任务二:数据可视化（30分）

2.1在上一步骤的Jupyter Notebook文件中，使用matplotlib库绘出大数据岗位招聘对比柱状图。X轴表示地区名称，Y轴表示岗位数量。将代码和结果截图，截图2-1。

要求（每个要求5分，共30分）：

1）自定义尺寸，宽为10英寸，高为8英寸；

2）含有标题“大数据岗位招聘对比柱状图”；

3）设置x轴标签为“地区名称”；

4）设置y轴标签为“岗位数量”；

5）柱体为红色；

6）要求中文能正常显示。

提交要求：

1） 在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南工程职业技术学院 01 张三。

2）“技能抽查提交资料”文件夹内共创建：项目文件，截图文件夹（包含截图1-1、1-2、2-1）。

（2）实施条件

测试所需的软硬件设备见下表 3.1.1

表3.1.1 考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存8G 以上，	

		windows操作系统	
3	截图工具		系统自带截图工具
4	Python3.7或以上		用于数据分析
5	Anaconda3（自带Jupter Notebook）		用于数据可视化

（3）考核时量

考核时长 90 分钟。

（4）评分细则

数据分析与可视化模块的考核实行 100 分制，评价内容包括工作任务、职业素养完成情况两个方面。其中，工作任务完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表 3.1.2 所示。

表 3.1.2 评分标准表评价内容

评价内容		分值	评分细则	
工作任务（80分）	数据处理和分析	50分	新建Jupyter Notebook项目正确。	10分
			正确编写Python分析代码。代码不正确每处扣 1 分。	30分
			运行程序结果正确，运行结果进行截图。结果不正确扣8分，截图不正确扣2分。	10分
	数据可视化	30分	导入模块正确，模块导入错误每次扣1分。	4分
			正确使用函数读取文件，处理后显示所有数据。代码不正确每处扣 1 分。	6分
			编写Python代码，绘制图形，图形能正确显示。X轴和Y轴数据不正确，每处扣2分，代码函数使用不正确每处扣 1 分。	10分
			图形正确按照题目要求绘制，运行结果和代码进行截图。代码截图不正确扣2分，图形截图不正确扣2分。	10分
职业素养（20分）	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、 鼠标等），检查测试所需软件开发环境。 不进行检查操作扣 5 分。	5分

	专业素养	10分	代码符合软件开发规范，命名规范，能做到见名知意；缩进统一，方便阅读；注释规范正确。整个项目命名不规范每一处扣 1 分；完全没有注释扣 2 分；有注释，但注释不规范每一处扣 1 分。	10分
	职业行为规范	5分	着装干净整洁，举止文明。遵守考场纪律，按顺序进出考场	5分

2.试题 3-1-2：51JOB 网站地区平均薪资数据分析与可视化

（1）任务描述

51job网站是中国具有广泛影响力的人力资源服务供应商。求职者一般都会在该网站上查找求职信息。对相关招聘数据进行数据分析与可视化，能帮助求职者更好地求职。

下面以51job网站为例，请根据“51job.csv”源数据，利用 Python分析知识对大数据薪资数据的清洗和整理，完成数据处理和分析任务。使用Python语言完成数据可视化，帮助求职者提前了解薪资的现状。

任务一：数据处理和分析（50分）

1.1打开Jupyter Notebook编辑器新建Python文件。（10分）

1.2读取“51job.csv”文件，显示前5条数据，截图1-1。（20分）

1.3读取csv文件，利用Python语言提取大数据每个岗位的最低薪资，然后按照薪资进行分组统计并显示所有数据，截图1-2。（20分）

任务二:数据可视化（30分）

2.1在上一步骤的Jupyter Notebook文件中，使用matplotlib库绘出大数据相同薪资岗位数量对比的柱状图。X轴为薪资，单位为元，Y轴表示薪资对应的工作数量。将代码和结果截图，截图2-1。（30分）

要求（每个要求5分，共30分）：

1）自定义尺寸，宽为10英寸，高为8英寸；

2）含有标题“各城市大数据相关岗位的平均薪资对比柱状图”；

3）设置x轴标签为“薪资/元”；

4）设置y轴标签为“工作数量”；

5）柱体为红色；

6）要求中文能正常显示。

提交要求：

1）在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南工程职业技术学院 01 张三。

2）“技能抽查提交资料”文件夹内共创建：项目文件，截图文件夹（包含截图1-1、1-2、2-1）。

（2）实施条件

测试所需的软硬件设备见下表 3.2.1

表3.2.1 考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存8G 以上，windows操作系统	
3	截图工具		系统自带截图工具
4	Python3.7或以上		用于数据分析
5	Anaconda3（自带Jupter Notebook）		用于数据可视化

（3）考核时量

考核时长 90 分钟。

（4）评分细则

数据分析与可视化模块的考核实行 100 分制，评价内容包括工作任务、职业素养完成情况两个方面。其中，工作任务完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表3.2.2 所示。

表3.2.2 评分标准表评价内容

评价内容		分值	评分细则	
工作任务（80分）	数据处理和分析	50分	新建Jupyter Notebook项目正确。	10分
			正确编写Python分析代码。代码不正确每处扣 1分。	30分
			运行程序结果正确，运行结果进行截图。结果不正	10分

			确扣8分，截图不正确扣2分。	
	数据可视化	30分	导入模块正确，模块导入错误每次扣1分。	4分
			正确使用函数读取文件，处理后显示所有数据。代码不正确每处扣 1 分。	6分
			编写Python代码，绘制图形，图形能正确显示。X轴和Y轴数据不正确，每处扣2分，代码函数使用不正确每处扣 1 分。	10分
			图形正确按照题目要求绘制，运行结果和代码进行截图。代码截图不正确扣2分，图形截图不正确扣2分。	10分
职业素养（20分）	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、 鼠标等），检查测试所需软件开发环境。 不进行检查操作扣 5 分。	5分
	专业素养	10分	代码符合软件开发规范，命名规范，能做到见名知意；缩进统一，方便阅读；注释规范正确。整个项目命名不规范每一处扣 1 分；完全没有注释扣 2 分；有注释，但注释不规范每一处扣 1 分。	10分
	职业行为规范	5分	着装干净整洁，举止文明。遵守考场纪律，按顺序进出考场	5分

3.试题 3-1-3： 51JOB 网站岗位分类数据分析与可视化

（1）任务描述

51job网站是中国具有广泛影响力的人力资源服务供应商。求职者一般都会在该网查找求职信息。下面以51job网站为例，请根据“51job.csv”源数据，综合利用Python分析知识对数据的清洗和整理，完成数据处理和分析任务。使用Python语言完成数据可视化，帮助求职者提前了解大数据岗位对应的职位数。

任务一：数据处理和分析（50分）

1.1打开Jupyter Notebook编辑器新建Python文件。（10分）

1.2读取“51job.csv”文件，显示前5条数据，截图1-1。（20分）

1.3读取csv文件，利用matplotlib数据分析所有地区岗位数分类：职位名称中包含“开发”、“工程师”的归类“开发工程师”；职位名称中包含“分析”、“数据”的归类“数据分析师”；职位名称中包含“运维”的归类“运维工程师”；

职位名称中包含“测试”的归类“测试人员”；职位名称中包含“销售”、“售前”、“营销”的归类“销售”；职位名称中包含“运营”的归类“运营人员”；其他职位归类“其他”。统计每个分类所包含的工作数量并显示所有数据，截图1-2。（20分）

任务二:数据可视化（30分）

2.1使用matplotlib库绘出大数据岗位招聘情况饼图。需要显示招聘职位和所占百分比。将代码和结果截图，截图2-1。

要求（每个要求6分，共30分）：

1）自定义尺寸，宽为10英寸，高为8英寸；

2）标题为“大数据岗位招聘情况饼图”；

3）显示百分比，格式化输出百分比为“%.2f%%”；

4）显示标签为对应的职位归类名称；

5）要求中文能正常显示。

提交要求：

1）在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南工程职业技术学院-01-张三。

2）“技能抽查提交资料”文件夹内共创建：项目文件，输出结果，截图文件夹（包含截图1-1、1-2、2-1）。

（2）实施条件

测试所需的软硬件设备见下表 3.3.1

表3.3.1 考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存8G 以上，windows操作系统	
3	截图工具		系统自带截图工具
4	Python3.7或以上		用于数据分析
5	Anaconda3（自带Jupter Notebook）		用于数据可视化

（3）考核时量

考核时长 90 分钟。

（4）评分细则

数据分析与可视化模块的考核实行 100 分制，评价内容包括工作任务、职业素养完成情况两个方面。其中，工作任务完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表3.3.2 所示。

表 3.3.2评分标准表评价内容

评价内容		分值	评分细则	
工作任务（80分）	数据处理和分析	50分	新建Jupyter Notebook项目正确。	10分
			正确编写Python分析代码。代码不正确每处扣 1 分。	30分
			运行程序结果正确，运行结果进行截图。结果不正确扣8分，截图不正确扣2分。	10分
	数据可视化	30分	导入模块正确，模块导入错误每次扣1分。	4分
			正确使用函数读取文件，处理后显示所有数据。代码不正确每处扣 1 分。	6分
			编写Python代码，绘制图形，图形能正确显示。X轴和Y轴数据不正确，每处扣2分，代码函数使用不正确每处扣 1 分。	10分
			图形正确按照题目要求绘制，运行结果和代码进行截图。代码截图不正确扣2分，图形截图不正确扣2分。	10分
职业素养（20分）	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、 鼠标等），检查测试所需软件开发环境。 不进行检查操作扣 5 分。	5分
	专业素养	10分	代码符合软件开发规范，命名规范，能做到见名知意；缩进统一，方便阅读；注释规范正确。整个项目命名不规范每一处扣 1 分；完全没有注释扣 2 分；有注释，但注释不规范每一处扣 1 分。	10分
	职业行为规范	5分	着装干净整洁，举止文明。遵守考场纪律，按顺序进出考场	5分

4.试题 3-1-4： AppleStore 平台上 App 收费与免费数量分析与可视化

（1）任务描述

AppleStore是苹果手机端软件下载市场，为苹果手机用户提供了App的主要来源。下面以AppleStore上的App为例，请根据其源数据，综合利用Python分析知识对数据的清洗和整理，完成数据处理和分析任务。使用Python语言完成数据可视化，帮助用户了解App的使用情况，直观了解市面上App的使用趋势。

任务一：数据处理和分析（50分）

1.1打开Jupyter Notebook编辑器新建Python文件。（10分）

1.2读取“applestore.csv”文件，利用Python数据分析显示前5条数据，截图1-1。（20分）

1.3 利用Python数据分析清理Uname 0这个变量并显示数据，截图1-2。（10分）

1.4 对价格数据进行分析，新增“价格”列，对原价格大于0的记为1，其他记为0。显示前5条数据并截图1-3。（10分）

任务二:数据可视化（30分）

2.1在上一步骤的Jupyter Notebook文件中，使用matplotlib库绘出App收费和不收费的对比柱状图。X轴为收费和不收费两种情况，Y轴表示对应的App数量，单位为数量。将代码和结果截图，截图2-1。（30分）

要求（每个要求5分，共30分）：

1）自定义尺寸，宽为10英寸，高为8英寸；

2）含有标题“App收费和不收费的对比柱状图”；

3）设置X轴标签为“收费情况”；

4）设置Y轴标签为“App数量”；

5）设置柱体颜色为红色；

6）要求中文能正常显示。

提交要求：

1）在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南工程职业技术学院 01 张三。

2）“技能抽查提交资料”文件夹内共创建：项目文件，截图文件夹（包含截图1-1、1-2、1-3、2-1）。

（2）实施条件

测试所需的软硬件设备见下表 3.4.1

表3.4.1 考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存8G 以上，windows操作系统	
3	截图工具		系统自带截图工具
4	Python3.7或以上		用于数据分析
5	Anaconda3 （自带Jupter Notebook）		用于数据可视化

（3）考核时量

考核时长 90 分钟。

（4）评分细则

数据分析与可视化模块的考核实行 100 分制，评价内容包括工作任务、职业素养完成情况两个方面。其中，工作任务完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表 3.4.2所示。

表 3.4.2评分标准表评价内容

评价内容		分值	评分细则	
工作任务（80分）	数据处理和分析	50分	新建Jupyter Notebook项目正确。	10分
			正确编写Python分析代码。代码不正确每处扣 1 分。	30分
			运行程序结果正确，运行结果进行截图。结果不正确扣8分，截图不正确扣2分。	10分
	数据可视化	30分	导入模块正确，模块导入错误每次扣1分。	4分
			正确使用函数读取文件，处理后显示所有数据。代码不正确每处扣 1 分。	6分
			编写Python代码，绘制图形，图形能正确显示。X轴和Y轴数据不正确，每处扣2分，代码函数使用不正确每处扣 1 分。	10分
			图形正确按照题目要求绘制，运行结果和代码进行	10分

			截图。代码截图不正确扣2分，图形截图不正确扣2分。	
职业素养（20分）	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、 鼠标等），检查测试所需软件开发环境。 不进行检查操作扣 5 分。	5分
	专业素养	10分	代码符合软件开发规范，命名规范，能做到见名知意；缩进统一，方便阅读；注释规范正确。整个项目命名不规范每一处扣 1 分；完全没有注释扣 2 分；有注释，但注释不规范每一处扣 1 分。	10分
	职业行为规范	5分	着装干净整洁，举止文明。遵守考场纪律，按顺序进出考场	5分

5.试题 3-1-5： AppleStore 平台上 App 评分数分析与可视化

（1）任务描述

AppleStore是苹果手机端软件下载市场，为苹果手机用户提供了App的主要来源。下面以AppleStore上的App为例，请根据其源数据，综合利用Python分析知识对数据的清洗和整理，完成数据处理和分析任务。使用Python语言完成数据可视化，帮助用户了解App的使用情况，直观了解市面上App的使用趋势。

任务一：数据处理和分析（50分）

1.1打开Jupyter Notebook编辑器新建Python文件。（10分）

1.2读取“applestore.csv”文件，利用Python数据分析显示前5条数据，截图1-1。（20分）

1.3 利用Python数据分析清理Uname 0这个变量并显示数据，截图1-2。（10分）

1.4 将数据按照用户评分分组并显示数据，截图1-3（10分）

任务二:数据可视化（30分）

2.1在上一步骤的Jupyter Notebook文件中，使用matplotlib库绘出用户对App评分对比情况柱状图。X轴为分数段，Y轴表示对应的人的数量，单位为数量。将代码和结果截图，截图2-1。（30分）

要求（每个要求5分，共30分）：

1）自定义尺寸，宽为10英寸，高为8英寸；

2）含有标题“用户评分对比情况柱状图”；

3）设置X轴标签为“分数段”；

4）设置Y轴标签为“人数”；

5）设置柱体颜色为红色；

6）要求中文能正常显示。

提交要求：

1）在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南工程职业技术学院 01 张三。

2）“技能抽查提交资料”文件夹内共创建：项目文件，截图文件夹（包含截图1-1、1-2、1-3、2-1）。

（2）实施条件

测试所需的软硬件设备见下表 3.5.1

表3.5.1 考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存8G 以上，windows操作系统	
3	截图工具		系统自带截图工具
4	Python3.7或以上		用于数据分析
5	Anaconda3（自带Jupter Notebook）		用于数据可视化

（3）考核时量

考核时长 90 分钟。

（4）评分细则

数据分析与可视化模块的考核实行 100 分制，评价内容包括工作任务、职业素养完成情况两个方面。其中，工作任务完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表 3.5.2 所示。

表 3.5.2评分标准表评价内容

评价内容		分值	评分细则	
工作任务（80分）	数据处理和分析	50分	新建Jupyter Notebook项目正确。	10分
			正确编写Python分析代码。代码不正确每处扣 1 分。	30分
			运行程序结果正确，运行结果进行截图。结果不正确扣8分，截图不正确扣2分。	10分
	数据可视化	30分	导入模块正确，模块导入错误每次扣1分。	4分
			正确使用函数读取文件，处理后显示所有数据。代码不正确每处扣 1 分。	6分
			编写Python代码，绘制图形，图形能正确显示。X轴和Y轴数据不正确，每处扣2分，代码函数使用不正确每处扣 1 分。	10分
			图形正确按照题目要求绘制，运行结果和代码进行截图。代码截图不正确扣2分，图形截图不正确扣2分。	10分
职业素养（20分）	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、 鼠标等），检查测试所需软件开发环境。 不进行检查操作扣 5 分。	5分
	专业素养	10分	代码符合软件开发规范，命名规范，能做到见名知意；缩进统一，方便阅读；注释规范正确。整个项目命名不规范每一处扣 1 分；完全没有注释扣 2 分；有注释，但注释不规范每一处扣 1 分。	10分
	职业行为规范	5分	着装干净整洁，举止文明。遵守考场纪律，按顺序进出考场	5分

6.试题 3-1-6： AppleStore 平台上 App 类别分析与可视化

（1）任务描述

AppleStore是苹果手机端软件下载市场，为苹果手机用户提供了App的主要来源。下面以AppleStore上的App为例，请根据其源数据，综合利用Python分析知识对数据的清洗和整理，完成数据处理和分析任务。使用Python语言完成数据可视化，帮助用户了解App的使用情况，直观了解市面上App的使用趋势。

任务一：数据处理和分析（50分）

打开Jupyter Notebook编辑器新建Python文件。（10分）

1.1读取“applestore.csv”文件，利用Python数据分析显示前5条数据，截图1-1。（20分）

1.2 利用Python数据分析清理Uname 0这个变量并显示前5条数据，截图1-2。（10分）

1.3 将数据按照APP分类分组统计每组数据并显示所有数据，截图1-3。（10分）

任务二:数据可视化（30分）

2.1在上一步骤的Jupyter Notebook文件中，使用matplotlib库绘出App分类对比情况柱状图。x轴为App类别名称，Y轴表示对应的App数量。将代码和结果截图，截图2-1。（30分）

要求（每个要求5分，共30分）：

1）自定义尺寸，宽为10英寸，高为8英寸；

2）含有标题“App分类对比情况柱状图”；

3）设置X轴标签为“App分类名称”；

4）设置Y轴标签为“数量”；

5）设置柱体颜色为红色；

6）要求中文能正常显示。

提交要求：

1）在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南工程职业技术学院 01 张三。

2）“技能抽查提交资料”文件夹内共创建：项目文件，截图文件夹（包含截图1-1、1-2、1-3、2-1）。

（2）实施条件

测试所需的软硬件设备见下表3.6.1

表3.6.1 考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存8G 以上，windows操作系统	
3	截图工具		系统自带截图工具
4	Python3.7或以上		用于数据分析
5	Anaconda3 （自带Jupter Notebook）		用于数据可视化

（3）考核时量

考核时长 90 分钟。

（4）评分细则

数据分析与可视化模块的考核实行 100 分制，评价内容包括工作任务、职业素养完成情况两个方面。其中，工作任务完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表3.6.2 所示。

表 3.6.2评分标准表评价内容

评价内容		分值	评分细则	
工作任务（80分）	数据处理和分析	50分	新建Jupyter Notebook项目正确。	10分
			正确编写Python分析代码。代码不正确每处扣 1 分。	30分
			运行程序结果正确，运行结果进行截图。结果不正确扣8分，截图不正确扣2分。	10分
	数据可视化	30分	导入模块正确，模块导入错误每次扣1分。	4分
			正确使用函数读取文件，处理后显示所有数据。代码不正确每处扣 1 分。	6分
			编写Python代码，绘制图形，图形能正确显示。X轴和Y轴数据不正确，每处扣2分，代码函数使用不正确每处扣 1 分。	10分
			图形正确按照题目要求绘制，运行结果和代码进行	10分

			截图。代码截图不正确扣2分，图形截图不正确扣2分。	
职业素养（20分）	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、 鼠标等），检查测试所需软件开发环境。 不进行检查操作扣 5 分。	5分
	专业素养	10分	代码符合软件开发规范，命名规范，能做到见名知意；缩进统一，方便阅读；注释规范正确。整个项目命名不规范每一处扣 1 分；完全没有注释扣 2 分；有注释，但注释不规范每一处扣 1 分。	10分
	职业行为规范	5分	着装干净整洁，举止文明。遵守考场纪律，按顺序进出考场	5分

7.试题 3-1-7: 猫眼电影网站各类型电影评分数分析与可视化

（1）任务描述

猫眼电影是国内观众喜爱的观影平台,为您提供了在线购票服务。 下面以猫眼电影网站为例，请根据其源数据，利用 Python知识对数据的清洗和整理，完成数据处理和分析任务。使用Python语言完成数据可视化，帮助观众了解各类型电影的评分均值的趋势。

任务一：数据处理和分析（50分）

1.1 打开Jupyter Notebook编辑器新建Python文件。（10分）

1.2读取“maoyan.csv”文件，利用Python数据分析显示前5条数据，截图1-1。（20分）

1.3 将数据按照用户评分分组统计并显示所有数据，截图1-2。（20分）

任务二:数据可视化（30分）

2.1在上一步骤的Jupyter Notebook文件中，使用matplotlib库绘出猫眼电影用户评分情况柱状图。X轴为用户评分情况，Y轴表示对应的用户数量。将代码和结果截图，截图2-1。（30分）

要求（每个要求5分，共30分）：

1）自定义尺寸，宽为10英寸，高为8英寸；

2）含有标题“猫眼电影用户评分情况柱状图”；

3）设置X轴标签为“用户评分情况”；

4）设置Y轴标签为“数量”；

5）设置柱体颜色为红色；

6）要求中文能正常显示。

提交要求：

1）在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南工程职业技术学院 01 张三。

2）“技能抽查提交资料”文件夹内共创建：项目文件，截图文件夹（包含截图1-1、1-2、2-1）。

（2）实施条件

测试所需的软硬件设备见下表3.7.1

表3.7.1 考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存8G 以上，windows操作系统	
3	截图工具		系统自带截图工具
4	Python3.7或以上		用于数据分析
5	Anaconda3（自带Jupter Notebook）		用于数据可视化

（3）考核时量

考核时长 90 分钟。

（4）评分细则

数据分析与可视化模块的考核实行 100 分制，评价内容包括工作任务、职业素养完成情况两个方面。其中，工作任务完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表 3.7.2 所示。

表 3.7.2 评分标准表评价内容

评价内容		分值	评分细则	
工作任务（80分）	数据处理和分析	50分	新建Jupyter Notebook项目正确。	10分
			正确编写Python分析代码。代码不正确每处扣 1 分。	30分
			运行程序结果正确，运行结果进行截图。结果不正确扣8分，截图不正确扣2分。	10分
	数据可视化	30分	导入模块正确，模块导入错误每次扣1分。	4分
			正确使用函数读取文件，处理后显示所有数据。代码不正确每处扣 1 分。	6分
			编写Python代码，绘制图形，图形能正确显示。X轴和Y轴数据不正确，每处扣2分，代码函数使用不正确每处扣 1 分。	10分
			图形正确按照题目要求绘制，运行结果和代码进行截图。代码截图不正确扣2分，图形截图不正确扣2分。	10分
职业素养（20分）	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、 鼠标等），检查测试所需软件开发环境。 不进行检查操作扣 5 分。	5分
	专业素养	10分	代码符合软件开发规范，命名规范，能做到见名知意；缩进统一，方便阅读；注释规范正确。整个项目命名不规范每一处扣 1 分；完全没有注释扣 2 分；有注释，但注释不规范每一处扣 1 分。	10分
	职业行为规范	5分	着装干净整洁，举止文明。遵守考场纪律，按顺序进出考场	5分

8.试题 3-1-8：猫眼电影网站电影时长数据分析与可视化

（1）任务描述

猫眼电影是国内观众喜爱的观影平台,为您提供了在线购票服务。 下面以猫眼电影网站为例，请根据其源数据，利用Python知识对数据的清洗和整理，完成数据

处理和分析任务。使用Python语言完成数据可视化，帮助观众了解各类型电影的评分均值的趋势。

任务一：数据处理和分析（50分）

1.1 打开Jupyter Notebook编辑器新建Python文件。（10分）

1.2读取“maoyan.csv”文件，利用Python数据分析显示前5条数据，截图1-1。（20分）

1.3 将数据按照上映时间分组统计并显示所有数据，截图1-2。（20分）

任务二:数据可视化（30分）

2.1在上一步骤的Jupyter Notebook文件中，使用matplotlib库绘出电影上映时间饼图。将代码和结果截图，截图2-1。（30分）

要求（每个要求6分，共30分）：

1）自定义尺寸，宽为10英寸，高为8英寸；

2）标题为“电影上映时间饼图”；

3）显示百分比，格式化输出百分比为“%.2f%%”；

4）显示标签为对应的上映时间；

5）要求中文能正常显示。

提交要求：

1）在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南工程职业技术学院 01 张三。

2）“技能抽查提交资料”文件夹内共创建：项目文件，截图文件夹（包含截图1-1、1-2、2-1）。

（2）实施条件

测试所需的软硬件设备见下表3.8.1

表3.8.1 考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存8G 以上，windows操作系统	
3	截图工具		系统自带截图工具
4	Python3.7或以上		用于数据分析
5	Anaconda3（自带Jupter Notebook）		用于数据可视化

（3）考核时量

考核时长 90 分钟。

（4）评分细则

数据分析与可视化模块的考核实行 100 分制，评价内容包括工作任务、职业素养完成情况两个方面。其中，工作任务完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表3.8.2所示。

表3.8.2 评分标准表评价内容

评价内容		分值	评分细则	
工作任务（80分）	数据处理和分析	50分	新建Jupyter Notebook项目正确。	10分
			正确编写Python分析代码。代码不正确每处扣 1 分。	30分
			运行程序结果正确，运行结果进行截图。结果不正确扣8分，截图不正确扣2分。	10分
	数据可视化	30分	导入模块正确，模块导入错误每次扣1分。	4分
			正确使用函数读取文件，处理后显示所有数据。代码不正确每处扣 1 分。	6分
			编写Python代码，绘制图形，图形能正确显示。X轴和Y轴数据不正确，每处扣2分，代码函数使用不正确每处扣 1 分。	10分
			图形正确按照题目要求绘制，运行结果和代码进行截图。代码截图不正确扣2分，图形截图不正确扣2分。	10分
职业素养（20分）	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、 鼠标等），检查测试所需软件开发环境。 不进行检查操作扣 5 分。	5分
	专业素养	10分	代码符合软件开发规范，命名规范，能做到见名知意；缩进统一，方便阅读；注释规范正确。整个项目命名不规范每一处扣 1 分；完全没有注释扣 2 分；有注释，但注释不规范每一处扣 1 分。	10分
	职业行为规范	5分	着装干净整洁，举止文明。遵守考场纪律，按顺序进出考场	5分

9.试题 3-1-9：猫眼电影网站各类型电影数据分析与可视化

（1）任务描述

猫眼电影是国内观众喜爱的观影平台,为您提供了在线购票服务。 下面以猫眼电影网站为例，请根据其源数据，利用Python知识对数据的清洗和整理，完成数据处理和分析任务。使用Python语言完成数据可视化，帮助观众了解各类型电影的评分均值的趋势。

任务一：数据处理和分析（50分）

1.1 打开Jupyter Notebook编辑器新建Python文件。（10分）

1.2读取“maoyan.csv”文件，利用Python数据分析显示前5条数据，截图1-1。（20分）

1.3 将数据按照类型进行分类，“剧情或爱情”归类为“剧情”，“喜剧”类归类为“喜剧”，“动作”类归类为“动作”，“恐怖，惊悚”归类为“恐怖”，“动画，家庭”类归类为“家庭”。统计每个分类所包含的电影数量并显示所有数据，截图1-2。（20分）

任务二:数据可视化（30分）

2.1在上一步骤的Jupyter Notebook文件中，对归类后的新的类型进行统计，并使用matplotlib中的柱状图进行显示。X轴为电影归类后的类型，Y轴表示对应的影片的数量。将代码和结果截图，截图2-1。（30分）

要求（每个要求5分，共30分）：

1）自定义尺寸，宽为10英寸，高为8英寸；

2）含有标题“猫眼电影分类对比柱状图”；

3）设置X轴标签为“统计类型”；

4）设置Y轴标签为“电影数量”；

5）柱体为红色；

6）要求中文能正常显示。

提交要求：

1）在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南工程职业技术学院 01 张三。

2）“技能抽查提交资料”文件夹内共创建：项目文件，截图文件夹（包含截图1-1、1-2、2-1）。

（2）实施条件

测试所需的软硬件设备见下表3.9.1

表3.9.1 考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存8G 以上，windows操作系统	
3	截图工具		系统自带截图工具
4	Python3.7或以上		用于数据分析
5	Anaconda3（自带Jupter Notebook）		用于数据可视化

（3）考核时量

考核时长 90 分钟。

（4）评分细则

数据分析与可视化模块的考核实行 100 分制，评价内容包括工作任务、职业素养完成情况两个方面。其中，工作任务完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表 3.9.2 所示。

表3.9.2 评分标准表评价内容

评价内容		分值	评分细则	
工作任务（80分）	数据处理和分析	50分	新建Jupyter Notebook项目正确。	10分
			正确编写Python分析代码。代码不正确每处扣 1 分。	30分
			运行程序结果正确，运行结果进行截图。结果不正确扣8分，截图不正确扣2分。	10分
	数据可视化	30分	导入模块正确，模块导入错误每次扣1分。	4分
			正确使用函数读取文件，处理后显示所有数据。代码不正确每处扣 1 分。	6分

			编写Python代码，绘制图形，图形能正确显示。X轴和Y轴数据不正确，每处扣2分，代码函数使用不正确每处扣 1 分。	10分
			图形正确按照题目要求绘制，运行结果和代码进行截图。代码截图不正确扣2分，图形截图不正确扣2分。	10分
职业素养（20分）	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、 鼠标等），检查测试所需软件开发环境。 不进行检查操作扣 5 分。	5分
	专业素养	10分	代码符合软件开发规范，命名规范，能做到见名知意；缩进统一，方便阅读；注释规范正确。整个项目命名不规范每一处扣 1 分；完全没有注释扣 2 分；有注释，但注释不规范每一处扣 1 分。	10分
	职业行为规范	5分	着装干净整洁，举止文明。遵守考场纪律，按顺序进出考场	5分

10.试题 3-1-10：2021 年东京奥运会数据分析与可视化

（1）任务描述

2021年东京奥运会是竞技场，也是大舞台，为全世界人民呈现了一场体育盛事的同时，也为大家重新定义了美的概念。在这场奥运会中我们前所未有地接近世界舞台的中心，前所未有地接近实现中华民族伟大复兴的目标。下面以奥运会中运动员的数据为例，请根据其源数据，利用Python分析知识对数据的清洗和整理，完成数据处理和分析任务。使用Python语言完成数据可视化，帮助用户了解各国运动健儿的相关数据。

任务一：数据处理和分析（50分）

1.1 打开Jupyter Notebook编辑器新建Python文件。（10分）

1.2读取“athletes.csv”文件，利用Python数据分析显示前5条数据，截图1-1。（20分）

1.3 对数据按照国家进行分组统计总人数并按照降序排序，显示前10条数据，截图1-2。（20分）

任务二:数据可视化（30分）

2.1在上一步骤的Jupyter Notebook文件中，使用matplotlib库绘出各国参赛运动员人数情况条形图。X轴为参赛人数分数，Y轴表示对应的国家，将代码和结果截图，截图2-1。（30分）

要求（每个要求5分，共30分）：

1）自定义尺寸，宽为10英寸，高为8英寸；

2）含有标题“各国参赛运动员人数情况条形图”；

3）设置X轴标签为“参赛人数”；

4）设置Y轴标签为“国家”；

5）柱体为红色；

6）要求中文能正常显示。

提交要求：

1）在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南工程职业技术学院 01 张三。

2）“技能抽查提交资料”文件夹内共创建：项目文件，截图文件夹（包含截图1-1、1-2、2-1）。

（2）实施条件

测试所需的软硬件设备见下表3.10.1

表3.10.1 考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存8G 以上，windows操作系统	
3	截图工具		系统自带截图工具
4	Python3.7或以上		用于数据分析
5	Anaconda3（自带Jupter Notebook）		用于数据可视化

（3）考核时量

考核时长 90 分钟。

（4）评分细则

数据分析与可视化模块的考核实行 100 分制，评价内容包括工作任务、职业素养完成情况两个方面。其中，工作任务完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表 3.10.2 所示。

表 3.10.2 评分标准表评价内容

评价内容		分值	评分细则	
工作任务（80分）	数据处理和分析	50分	新建Jupyter Notebook项目正确。	10分
			正确编写Python分析代码。代码不正确每处扣 1 分。	30分
			运行程序结果正确，运行结果进行截图。结果不正确扣8分，截图不正确扣2分。	10分
	数据可视化	30分	导入模块正确，模块导入错误每次扣1分。	4分
			正确使用函数读取文件，处理后显示所有数据。代码不正确每处扣 1 分。	6分
			编写Python代码，绘制图形，图形能正确显示。X轴和Y轴数据不正确，每处扣2分，代码函数使用不正确每处扣 1 分。	10分
			图形正确按照题目要求绘制，运行结果和代码进行截图。代码截图不正确扣2分，图形截图不正确扣2分。	10分
职业素养（20分）	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、 鼠标等），检查测试所需软件开发环境。 不进行检查操作扣 5 分。	5分
	专业素养	10分	代码符合软件开发规范，命名规范，能做到见名知意；缩进统一，方便阅读；注释规范正确。整个项目命名不规范每一处扣 1 分；完全没有注释扣 2 分；有注释，但注释不规范每一处扣 1 分。	10分
	职业行为规范	5分	着装干净整洁，举止文明。遵守考场纪律，按顺序进出考场	5分

11.试题 3-1-11：2021 年东京奥运会数据分析与可视化

（1）任务描述

2021年东京奥运会是竞技场，也是大舞台，为全世界人民呈现了一场体育盛事的同时，也为大家重新定义了美的概念。在这场奥运会中我们前所未有地接近世界舞台的中心，前所未有地接近实现中华民族伟大复兴的目标。下面以奥运会中运动员的数据为例，请根据其源数据，利用Python分析知识对数据的清洗和整理，完成数据处理和分析任务。使用Python语言完成数据可视化，帮助用户了解各国获奖情况。

任务一：数据处理和分析（50分）

1.1 打开Jupyter Notebook编辑器新建Python文件。（10分）

1.2读取“medal.csv”文件，利用Python数据分析显示前5条数据，截图1-1。（20分）

1.3对奖牌数据进行降序排序，显示前10条数据，截图1-2。（20分）

任务二:数据可视化（30分）

2.1在上一步骤的Jupyter Notebook文件中，使用matplotlib库绘出前10名奖牌获取情况条形图。X轴为奖牌数，Y轴表示对应的国家。将代码和结果截图，截图2-1。（30分）

要求（每个要求5分，共30分）：

1）自定义尺寸，宽为10英寸，高为8英寸；

2）含有标题“前10名奖牌获取情况条形图”；

3）设置X轴标签为“奖牌数”；

4）设置Y轴标签为“国家”；

5）柱体为红色；

6）要求中文能正常显示。

提交要求：

1）在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南工程职业技术学院 01 张三。

2）“技能抽查提交资料”文件夹内共创建：项目文件，截图文件夹（包含截图1-1、1-2、2-1）。

（2）实施条件

测试所需的软硬件设备见下表 3.11.1

表3.11.1 考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存8G 以上，windows操作系统	
3	截图工具		系统自带截图工具
4	Python3.7或以上		用于数据分析
5	Anaconda3（自带Jupter Notebook）		用于数据可视化

（3）考核时量

考核时长 90 分钟。

（4）评分细则

数据分析与可视化模块的考核实行 100 分制，评价内容包括工作任务、职业素养完成情况两个方面。其中，工作任务完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表 3.11.2 所示。

表 3.11.2评分标准表评价内容

评价内容		分值	评分细则	
工作任务（80分）	数据处理和分析	50分	新建Jupyter Notebook项目正确。	10分
			正确编写Python分析代码。代码不正确每处扣 1 分。	30分
			运行程序结果正确，运行结果进行截图。结果不正确扣8分，截图不正确扣2分。	10分
	数据可视化	30分	导入模块正确，模块导入错误每次扣1分。	4分
			正确使用函数读取文件，处理后显示所有数据。代码不正确每处扣 1 分。	6分
			编写Python代码，绘制图形，图形能正确显示。X轴和Y轴数据不正确，每处扣2分，代码函数使用不正确每处扣 1 分。	10分
			图形正确按照题目要求绘制，运行结果和代码进行	10分

			截图。代码截图不正确扣2分，图形截图不正确扣2分。	
职业素养（20分）	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、 鼠标等），检查测试所需软件开发环境。 不进行检查操作扣 5 分。	5分
	专业素养	10分	代码符合软件开发规范，命名规范，能做到见名知意；缩进统一，方便阅读；注释规范正确。整个项目命名不规范每一处扣 1 分；完全没有注释扣 2 分；有注释，但注释不规范每一处扣 1 分。	10分
	职业行为规范	5分	着装干净整洁，举止文明。遵守考场纪律，按顺序进出考场	5分

12.试题 3-1-12：2021 年东京奥运会数据分析与可视化

（1）任务描述

2021年东京奥运会是竞技场，也是大舞台，为全世界人民呈现了一场体育盛事的同时，也为大家重新定义了美的概念。在这场奥运会中我们前所未有地接近世界舞台的中心，前所未有地接近实现中华民族伟大复兴的目标。下面以奥运会中运动员的数据为例，请根据其源数据，利用Python分析知识对数据的清洗和整理，完成数据处理和分析任务。使用Python语言完成数据可视化，帮助用户了解各国获奖情况。

任务一：数据处理和分析（50分）

1.1 打开Jupyter Notebook编辑器新建Python文件。（10分）

1.2读取“medal.csv”文件，利用Python数据分析显示前5条数据，截图1-1。（20分）

1.3获取中国各项奖牌数据并显示所有数据，截图1-2（20分）

任务二:数据可视化（30分）

2.1在上一步骤的Jupyter Notebook文件中，使用matplotlib库绘出中国各项奖牌获取情况饼图。将代码和结果截图，截图2-1。（30分）

要求（每个要求6分，共30分）：

1）自定义尺寸，宽为10英寸，高为8英寸；

2）标题为“中国各项奖牌获取情况饼图”；

3）显示百分比，格式化输出百分比为“%.2f%%”；

4）显示标签为对应的奖牌名称；

5）要求中文能正常显示。

提交要求：

1）在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南工程职业技术学院 01 张三。

2）“技能抽查提交资料”文件夹内共创建：项目文件，截图文件夹（包含截图1-1、1-2、2-1）。

（2）实施条件

测试所需的软硬件设备见下表3.12.1

表3.12.1 考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存8G 以上，windows操作系统	
3	截图工具		系统自带截图工具
4	Python3.7或以上		用于数据分析
5	Anaconda3（自带Jupter Notebook）		用于数据可视化

（3）考核时量

考核时长 90 分钟。

（4）评分细则

数据分析与可视化模块的考核实行 100 分制，评价内容包括工作任务、职业素养完成情况两个方面。其中，工作任务完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表3.12.2 所示。

表 3.12.2 评分标准表评价内容

评价内容		分值	评分细则	
工作任务（80分）	数据处理和分析	50分	新建Jupyter Notebook项目正确。	10分
			正确编写Python分析代码。代码不正确每处扣 1	30分

			分。	
			运行程序结果正确，运行结果进行截图。结果不正确扣8分，截图不正确扣2分。	10分
	数据可视化	30分	导入模块正确，模块导入错误每次扣1分。	4分
			正确使用函数读取文件，处理后显示所有数据。代码不正确每处扣 1 分。	6分
			编写Python代码，绘制图形，图形能正确显示。X轴和Y轴数据不正确，每处扣2分，代码函数使用不正确每处扣 1 分。	10分
			图形正确按照题目要求绘制，运行结果和代码进行截图。代码截图不正确扣2分，图形截图不正确扣2分。	10分
职业素养（20分）	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、 鼠标等），检查测试所需软件开发环境。 不进行检查操作扣 5 分。	5分
	专业素养	10分	代码符合软件开发规范，命名规范，能做到见名知意；缩进统一，方便阅读；注释规范正确。整个项目命名不规范每一处扣 1 分；完全没有注释扣 2 分；有注释，但注释不规范每一处扣 1 分。	10分
	职业行为规范	5分	着装干净整洁，举止文明。遵守考场纪律，按顺序进出考场	5分

13.试题 3-1-13：COVID-19 世界范围内疫苗接种进度情况分析与可视化

（1）任务描述

下面以新冠肺炎疫苗接种进度情况为例，请根据其源数据，利用Python分析知识对数据的清洗和整理，完成数据处理和分析任务。使用Python语言完成数据可视化，帮助用户了解各国新冠肺炎疫苗接种的相关数据。

任务一：数据处理和分析（50分）

1.1 打开Jupyter Notebook编辑器新建Python文件。（10分）

1.2读取“country_vaccinations_by_manufacturer.csv”文件，利用Python数据分析显示前5条数据，截图1-1。（20分）

1.3 利用Python对疫苗生产厂商进行分组统计并显示，截图1-2。（20分）

任务二:数据可视化（30分）

2.1在上一步骤的Jupyter Notebook文件中，使用matplotlib库绘出各疫苗生产厂商在全世界所占的比例饼图。将代码和结果截图，截图2-1。（30分）

要求（每个要求6分，共30分）：

1）自定义尺寸，宽为10英寸，高为8英寸；

2）标题为“各疫苗生产厂商在全世界所占的比例饼图”；

3）显示百分比，格式化输出百分比为“%.2f%%”；

4）显示标签为对应的疫苗生产厂商名称；

5）要求中文能正常显示。

提交要求：

1）在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南工程职业技术学院 01 张三。

2）“技能抽查提交资料”文件夹内共创建：项目文件，截图文件夹（包含截图1-1、1-2、2-1）。

（2）实施条件

测试所需的软硬件设备见下表3.13.1

表3.13.1 考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存8G 以上，windows操作系统	
3	截图工具		系统自带截图工具
4	Python3.7或以上		用于数据分析
5	Anaconda3 （自带Jupter Notebook）		用于数据可视化

（3）考核时量

考核时长 90 分钟。

（4）评分细则

数据分析与可视化模块的考核实行 100 分制，评价内容包括工作任务、职业素养完成情况两个方面。其中，工作任务完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表3.13.2 所示。

表 3.13.2 评分标准表评价内容

评价内容		分值	评分细则	
工作任务（80分）	数据处理和分析	50分	新建Jupyter Notebook项目正确。	10分
			正确编写Python分析代码。代码不正确每处扣 1 分。	30分
			运行程序结果正确，运行结果进行截图。结果不正确扣8分，截图不正确扣2分。	10分
	数据可视化	30分	导入模块正确，模块导入错误每次扣1分。	4分
			正确使用函数读取文件，处理后显示所有数据。代码不正确每处扣 1 分。	6分
			编写Python代码，绘制图形，图形能正确显示。X轴和Y轴数据不正确，每处扣2分，代码函数使用不正确每处扣 1 分。	10分
			图形正确按照题目要求绘制，运行结果和代码进行截图。代码截图不正确扣2分，图形截图不正确扣2分。	10分
职业素养（20分）	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、 鼠标等），检查测试所需软件开发环境。 不进行检查操作扣 5 分。	5分
	专业素养	10分	代码符合软件开发规范，命名规范，能做到见名知意；缩进统一，方便阅读；注释规范正确。整个项目命名不规范每一处扣 1 分；完全没有注释扣 2 分；有注释，但注释不规范每一处扣 1 分。	10分
	职业行为规范	5分	着装干净整洁，举止文明。遵守考场纪律，按顺序进出考场	5分

14.试题 3-1-14：COVID-19 世界范围内疫苗接种进度情况分析与可视化

（1）任务描述

下面以新冠肺炎疫苗接种进度情况为例，请根据其源数据，利用Python分析知识对数据的清洗和整理，完成数据处理和分析任务。使用Python语言完成数据可视化，帮助用户了解各国新冠肺炎疫苗接种的相关数据。

任务一：数据处理和分析（50分）

1.1 打开Jupyter Notebook编辑器新建Python文件。（10分）

1.2读取“country_vaccinations_by_manufacturer.csv”文件，利用Python数据分析显示前5条数据，截图1-1。（20分）

1.3 利用Python对每个国家接种疫苗人数进行分组求和并显示，截图1-2。（20分）

任务二:数据可视化（30分）

2.1在上一步骤的Jupyter Notebook文件中，使用matplotlib库绘出各个地区接种疫苗人数的情况柱状图。X轴为地区名称，Y轴为人数。将代码和结果截图，截图2-1。（30分）

要求（每个要求5分，共30分）：

1）自定义尺寸，宽为10英寸，高为8英寸；

2）含有标题“各个地区接种疫苗人数的情况柱状图”；

3）设置X轴标签为“地区名称”；

4）设置Y轴标签为“国家名称”；

5）柱体为红色；

6）要求中文能正常显示。

提交要求：

1）在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南工程职业技术学院 01 张三。

2）“技能抽查提交资料”文件夹内共创建：项目文件，截图文件夹（包含截图1-1、1-2、2-1）。

（2）实施条件

测试所需的软硬件设备见下表 3.14.1

表3.14.1 考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存8G 以上，windows操作系统	
3	截图工具		系统自带截图工具

4	Python3.7或以上		用于数据分析
5	Anaconda3（自带Jupter Notebook）		用于数据可视化

（3）考核时量

考核时长 90 分钟。

（4）评分细则

数据分析与可视化模块的考核实行 100 分制，评价内容包括工作任务、职业素养完成情况两个方面。其中，工作任务完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表 3.14.2 所示。

表 3.14.2 评分标准表评价内容

评价内容		分值	评分细则	
工作任务（80分）	数据处理和分析	50分	新建Jupyter Notebook项目正确。	10分
			正确编写Python分析代码。代码不正确每处扣 1 分。	30分
			运行程序结果正确，运行结果进行截图。结果不正确扣8分，截图不正确扣2分。	10分
	数据可视化	30分	导入模块正确，模块导入错误每次扣1分。	4分
			正确使用函数读取文件，处理后显示所有数据。代码不正确每处扣 1 分。	6分
			编写Python代码，绘制图形，图形能正确显示。X轴和Y轴数据不正确，每处扣2分，代码函数使用不正确每处扣 1 分。	10分
			图形正确按照题目要求绘制，运行结果和代码进行截图。代码截图不正确扣2分，图形截图不正确扣2分。	10分
职业素养（20分）	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、 鼠标等），检查测试所需软件开发环境。 不进行检查操作扣 5 分。	5分
	专业素养	10分	代码符合软件开发规范，命名规范，能做到见名知意；缩进统一，方便阅读；注释规范正确。整个项目命名不规范每一处扣 1 分；完全没有注释扣 2	10分

			分；有注释，但注释不规范每一处扣 1 分。	
	职业行为规范	5分	着装干净整洁，举止文明。遵守考场纪律，按顺序进出考场	5分

15.试题 3-1-15：超市销售情况分析与可视化

（1）任务描述

下面以超市销售情况为例，请根据其源数据，利用Python分析知识对数据的清洗和整理，完成数据处理和分析任务。使用Python语言完成数据可视化，帮助用户了解某超市销售的相关数据。

任务一：数据处理和分析（50分）

1.1 打开Jupyter Notebook编辑器新建Python文件。（10分）

1.2读取“sales.csv”文件，利用Python数据分析显示前5条数据，截图1-1。（20分）

1.3 利用Python对每个月的全部订单数进行分组计数，截图1-2。（20分）

任务二:数据可视化（30分）

2.1在上一步骤的Jupyter Notebook文件中，使用matplotlib库绘出每个月全部订单数的柱状图。X轴为年月，Y轴为订单数。将代码和结果截图，截图2-1。（30分）

要求（每个要求5分，共30分）：

1）自定义尺寸，宽为10英寸，高为8英寸；

2）含有标题“每个月全部订单数的柱状图”；

3）设置X轴标签为“年月”；

4）设置Y轴标签为“订单数”；

5）柱体为红色；

6）要求中文能正常显示。

提交要求：

1）在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南工程职业技术学院 01 张三。

2）“技能抽查提交资料”文件夹内共创建：项目文件，截图文件夹（包含截图1-1、1-2、2-1）。

（2）实施条件

测试所需的软硬件设备见下表 3.15.1

表3.15.1 考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存8G 以上，windows操作系统	
3	截图工具		系统自带截图工具
4	Python3.7或以上		用于数据分析
5	Anaconda3 （自带Jupter Notebook）		用于数据可视化

（3）考核时量

考核时长 90 分钟。

（4）评分细则

数据分析与可视化模块的考核实行 100 分制，评价内容包括工作任务、职业素养完成情况两个方面。其中，工作任务完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表 3.15.2 所示。

表 3.15.2 评分标准表评价内容

评价内容		分值	评分细则	
工作任务（80分）	数据处理和分析	50分	新建Jupyter Notebook项目正确。	10分
			正确编写Python分析代码。代码不正确每处扣 1 分。	30分
			运行程序结果正确，运行结果进行截图。结果不正确扣8分，截图不正确扣2分。	10分
	数据可视化	30分	导入模块正确，模块导入错误每次扣1分。	4分
			正确使用函数读取文件，处理后显示所有数据。代码不正确每处扣 1 分。	6分
			编写Python代码，绘制图形，图形能正确显示。X轴和Y轴数据不正确，每处扣2分，代码函数使用不正确每处扣 1 分。	10分
			图形正确按照题目要求绘制，运行结果和	10分

			代码进行截图。代码截图不正确扣2分，图形截图不正确扣2分。	
职业素养（20分）	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、鼠标等），检查测试所需软件开发环境。不进行检查操作扣 5 分。	5分
	专业素养	10分	代码符合软件开发规范，命名规范，能做到见名知意；缩进统一，方便阅读；注释规范正确。整个项目命名不规范每一处扣 1 分；完全没有注释扣 2 分；有注释，但注释不规范每一处扣 1 分。	10分
	职业行为规范	5分	着装干净整洁，举止文明。遵守考场纪律，按顺序进出考场	5分

16.试题 3-1-16：超市销售情况分析与可视化

（1）任务描述

下面以超市销售情况为例，请根据其源数据，利用Python分析知识对数据的清洗和整理，完成数据处理和分析任务。使用Python语言完成数据可视化，帮助用户了解某超市销售的相关数据。

任务一：数据处理和分析（50分）

1.1 打开Jupyter Notebook编辑器新建Python文件。（10分）

1.2读取“sales.csv”文件，利用Python数据分析显示前5条数据，截图1-1。（20分）

1.3 利用Python对每个地区的全部订单数进行分组统计，截图1-2。（20分）

任务二:数据可视化（30分）

2.1在上一步骤的Jupyter Notebook文件中，使用matplotlib库绘出每个地区全部订单数的柱状图。X轴为地区代码，Y轴为订单数量。将代码和结果截图，截图2-1。（30分）

要求（每个要求5分，共30分）：

1）自定义尺寸，宽为10英寸，高为8英寸；

2）含有标题“每个地区全部订单数的柱状图”；

3）设置X轴标签为“地区代码”；

4）设置Y轴标签为“订单数”；

5）柱体为红色；

6）要求中文能正常显示。

提交要求：

1）在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南工程职业技术学院 01 张三。

2）“技能抽查提交资料”文件夹内共创建：项目文件，截图文件夹（包含截图1-1、1-2、2-1）。

（2）实施条件

测试所需的软硬件设备见下表 3.16.1

表3.16.1 考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存8G 以上，windows操作系统	
3	截图工具		系统自带截图工具
4	Python3.7或以上		用于数据分析
5	Anaconda3（自带Jupter Notebook）		用于数据可视化

（3）考核时量

考核时长 90 分钟。

（4）评分细则

数据分析与可视化模块的考核实行 100 分制，评价内容包括工作任务、职业素养完成情况两个方面。其中，工作任务完成质量占该项目总分的 80%，职业素养占该项目总分的 20%。

具体评价标准见表 3.16.2 所示。

表 3.16.2 评分标准表评价内容

评价内容		分值	评分细则	
工作任务（80分）	数据处理和分析	50分	新建Jupyter Notebook项目正确。	10分
			正确编写Python分析代码。代码不正确每处扣 1 分。	30分
			运行程序结果正确，运行结果进行截图。结果不正确扣8分，截图不正确扣2分。	10分
	数据可视化	30分	导入模块正确，模块导入错误每次扣1分。	4分
			正确使用函数读取文件，处理后显示所有数据。代码不正确每处扣 1 分。	6分
			编写Python代码，绘制图形，图形能正确显示。X轴和Y轴数据不正确，每处扣2分，代码函数使用不正确每处扣 1 分。	10分
			图形正确按照题目要求绘制，运行结果和代码进行截图。代码截图不正确扣2分，图形截图不正确扣2分。	10分
职业素养（20分）	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、 鼠标等），检查测试所需软件开发环境。 不进行检查操作扣 5 分。	5分
	专业素养	10分	代码符合软件开发规范，命名规范，能做到见名知意；缩进统一，方便阅读；注释规范正确。整个项目命名不规范每一处扣 1 分；完全没有注释扣 2 分；有注释，但注释不规范每一处扣 1 分。	10分
	职业行为规范	5分	着装干净整洁，举止文明。遵守考场纪律，按顺序进出考场	5分

17.试题 3-1-17：超市销售情况分析与可视化

（1）任务描述

下面以超市销售情况为例，请根据其源数据，利用Python分析知识对数据的清洗和整理，完成数据处理和分析任务。使用Python语言完成数据可视化，帮助用户了解某超市销售的相关数据。

任务一：数据处理和分析（50分）

1.1 打开Jupyter Notebook编辑器新建Python文件。（10分）

1.2读取“sales.csv”文件，利用Python数据分析显示前5条数据，截图1-1。（20分）

1.3 利用Python对所有店铺的全部订单数进行求和并排序，显示前10条数据，截图1-2。（20分）

任务二:数据可视化（30分）

2.1在上一步骤的Jupyter Notebook文件中，使用matplotlib库绘出前10名店铺订单数的饼图。将代码和结果截图，截图2-1。（30分）

要求（每个要求6分，共30分）：

1）自定义尺寸，宽为10英寸，高为8英寸；

2）标题为“前10名店铺订单数的饼图”；

3）显示百分比，格式化输出百分比为“%.2f%%”；

4）显示标签为对应的店铺代码；

5）要求中文能正常显示。

提交要求：

1）在“e:\技能抽查提交资料\”文件夹内创建考生文件夹，考生文件夹的命名规则：考生学校+考生号+考生姓名，示例：湖南工程职业技术学院 01 张三。

2）“技能抽查提交资料”文件夹内共创建：项目文件，截图文件夹（包含截图1-1、1-2、2-1）。

（2）实施条件

测试所需的软硬件设备见下表 3.17.1

表3.17.1 考点提供的主要设备及软件

序号	设备、软件名称	规格/技术参数、用途	备注
1	机房	测试场地	保证参考人员有足够间距
2	计算机	CPU i5以上，内存8G 以上，windows操作系统	
3	截图工具		系统自带截图工具
4	Python3.7或以上		用于数据分析
5	Anaconda3（自带Jupter Notebook）		用于数据可视化

（3）考核时量

考核时长 90 分钟。

（4）评分细则

数据分析与可视化模块的考核实行 100 分制，评价内容包括工作任务、职业素养完成情况两个方面。其中，工作任务完成质量占该项目总分的 80%，职业素养占该项目总分的20%。

具体评价标准见表 3.17.2 所示。

表 3.17.2 评分标准表评价内容

评价内容		分值	评分细则	
工作任务（80分）	数据处理和分析	50分	新建Jupyter Notebook项目正确。	10分
			正确编写Python分析代码。代码不正确每处扣 1 分。	30分
			运行程序结果正确，运行结果进行截图。结果不正确扣8分，截图不正确扣2分。	10分
	数据可视化	30分	导入模块正确，模块导入错误每次扣1分。	4分
			正确使用函数读取文件，处理后显示所有数据。代码不正确每处扣 1 分。	6分
			编写Python代码，绘制图形，图形能正确显示。X轴和Y轴数据不正确，每处扣2分，代码函数使用不正确每处扣 1 分。	10分
			图形正确按照题目要求绘制，运行结果和代码进行截图。代码截图不正确扣2分，图形截图不正确扣2分。	10分
职业素养（20分）	工作前准备	5分	做好工作前准备，检查电脑硬件（键盘、 鼠标等），检查测试所需软件开发环境。 不进行检查操作扣 5 分。	5分
	专业素养	10分	代码符合软件开发规范，命名规范，能做到见名知意；缩进统一，方便阅读；注释规范正确。整个项目命名不规范每一处扣 1 分；完全没有注释扣 2 分；有注释，但注释不规范每一处扣 1 分。	10分
	职业行为规范	5分	着装干净整洁，举止文明。遵守考场纪律，按顺序进出考场	5分