

# “巴渝工匠杯” 2022 年重庆市职业院校技能大赛 赛项规程

## 一、赛项名称

赛项编号：CQGZ-2022038

赛项名称：大数据技术与应用

英文名称：Big data technology and application

赛项组别：高职组

赛项归属：电子与信息大类

## 二、竞赛目的

为适应大数据产业对高素质技术技能型人才的职业需求，赛项以大数据技术与应用为核心内容和工作基础，重点考查参赛选手基于 Spark、Flink 平台环境下，充分利用 Spark Core、Spark SQL、Flume、Kafka、Flink 等相关技术的特点，基于 Scala、JavaScript 等开发语言，综合软件开发相关技术，解决实际问题的能力，激发学生对大数据相关知识和技术的学习兴趣，提升学生职业素养和职业技能，努力为中国大数据产业的发展储备及输送新鲜血液。

通过举办本赛项，可以搭建校企合作的平台，提升大数据技术与应用专业及其他相关专业毕业生能力素质，满足企业用人需求，促进校企合作协同育人，对接产业发展，实现行业资源、企业资源与教学资源的有机融合，使高职院校在专业建设、课程建设、人才培养方案和人才培养模式等方面，跟踪社会发展的最新需要，缩小人才培养与行业需求差距，引领职业院校专业建设与课程改革。

## 三、竞赛内容

赛项以大数据技术与应用为核心内容和工作基础，重点考查参赛选手基于

Spark、Flink 平台环境下，充分利用 Spark Core、Spark SQL、Flume、Kafka、Flink 等技术的特点，综合软件开发相关技术，解决实际问题的能力，具体包括：

1. 掌握基于 Spark 的离线分析平台、基于 Flink 的实时分析平台，按照项目需求安装相关技术组件并按照需求进行合理配置；
2. 掌握基于 Spark 的离线数据抽取相关技术，完成指定数据的抽取并写入 Hive 分区表中；
3. 综合利用 Spark Core、Spark SQL 等技术，使用 Scala 开发语言，完成某电商系统数据的离线统计任务，包括销量前 5 商品统计、某月的总销售额统计、每个月的销售统计、每个用户在线总时长统计，并将统计结果存入 MySQL 数据库中；
4. 综合利用 Flume、Flink 相关技术，使用 Scala 开发语言，完成将某电商系统的用户操作日志抽取到 Kafka 中，消费 Kafka 中的数据并计算商城在线人数，并统计该电商系统的 UV 与 PV；
5. 综合运用 HTML、CSS、JavaScript 等开发语言，Vue.js 前端技术，结合 Echarts 数据可视化组件，对 MySQL 中的数据进行可视化呈现；
6. 根据数据可视化结果，完成数据分析报告的编写；
7. 竞赛时间 6 小时，竞赛连续进行。

竞赛内容构成如下：

考核环节	考核知识点和技能点
大数据平台环境搭建 (Linux Shell)	Hadoop 完全分布式安装配置
	Spark 安装配置
	Flink 安装配置

	Kafka 安装配置
	Flume 安装配置
离线数据抽取 (Scala)	使用 Spark 抽取 MySQL 指定数据表中的新增的商品数据到 ODS 层的指定的分区表中
	使用 Spark 抽取 MySQL 指定数据表中的新增的用户数据到 ODS 层的指定的分区表中
	使用 Spark 抽取 MySQL 指定数据表中的新增的订单数据到 ODS 层的指定的分区表中
离线数据统计 (Scala)	获取 ODS 层指定分区表中的数据并进行相应的清洗 (缺失字段填充、去重)
	统计销量前 5 的商品并存入 MySQL
	统计某月的总销售额并存入 MySQL
	统计指定几个月的销售额并存入 MySQL
	统计每个用户的在线总时长并存入 MySQL
数据采集与实时计算 (Scala)	使用 Flume 采集某电商系统用户操作日志存入 Kafka 中
	使用 Flink 消费 Kafka 中的数据
	聚合计算出系统在线人数并将结果存入 Redis 中
	统计系统的 UV 与 PV 并存入 MySQL 中
数据可视化 (HTML、CSS、JavaScript、Vue.js)	基于 Vue.js、Echarts 的数据可视化渲染编码 (柱状图、折线图、饼图等)
综合分析报告	文档能力、数据分析能力

竞赛各阶段分值权重和时间分布如下:

阶段	竞赛时间	分值权重
大数据平台环境搭建	6 小时	权重 10%
离线数据抽取		权重 15%

离线数据统计		权重 20%
数据采集与实时计算		权重 20%
数据可视化		权重 20%
综合分析报告		权重 10%
团队协作分工明确合理、操作规范、文明竞赛		权重 5%

#### 四、竞赛方式

1、比赛以团队方式进行，不得跨校组队，同一学校的报名参赛队伍不超过2支。

2、每个参赛队由1名领队（可由指导教师兼任）、2名指导教师、3名选手（其中设队长1名）组成，指导教师须为本校专兼职教师，参赛选手和指导教师报名获得确认后不得随意更换。

3、竞赛时间6小时。

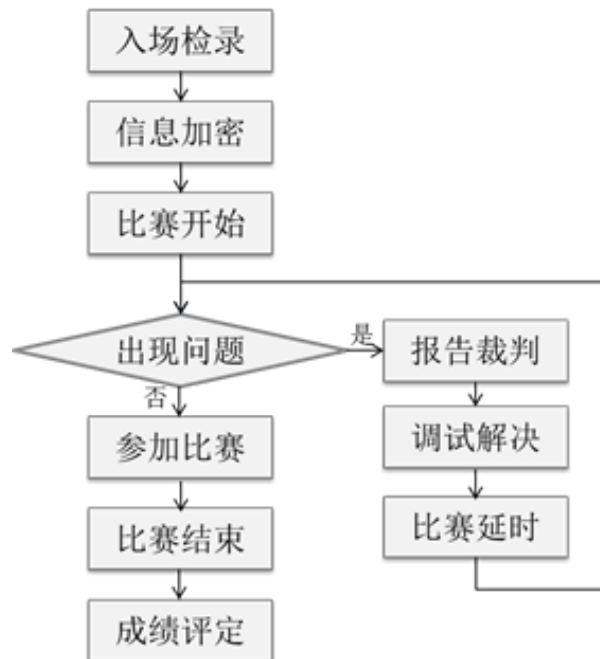
#### 五、竞赛流程

根据竞赛任务要求，参赛队伍在6小时竞赛时间内须完成竞赛任务，每项任务用时可自行掌握。

##### （一）竞赛时间安排

日期	时间	内容
4月14日	14:00-18:00	各参赛队报到
	15:00-16:00	领队会、赛前说明
	16:00-16:30	选手熟悉赛场
4月15日	8:00-9:00	赛场检录，竞赛选手进入赛位
	9:00-15:00	竞赛选手完成竞赛任务
	16:00-19:00	对选手提交的结果文件进行评分

## （二）竞赛流程



## 六、竞赛赛卷

本赛项赛卷样题详见《附件一》。

## 七、竞赛规则

1. 参赛选手须为 2022 年度重庆市高职学校全日制在籍学生。五年制全日制高职四五年级在籍学生可参加竞赛。参赛选手不限性别，年龄须不超过 25 周岁，年龄计算的截止时间以 2022 年 5 月 1 日为准。凡在往届全国职业院校技能大赛中获一等奖的选手，不得再参加同一赛项的比赛。

2. 参赛选手的竞赛工位号采用抽签方式确定。赛题以任务书形式发放，竞赛参考资料在赛前植入参赛选手的计算机，参赛选手根据任务书要求完成竞赛任务。

3. 赛项组委会统一布置竞赛需要的软硬件环境。选手不得私自携带任何移动存储、辅助工具、移动通信等进入赛场。

4. 参赛选手报到当天可预先熟悉比赛场地，但不得进行现场练习。参赛

选手按规定时间到达指定地点，凭参赛证、学生证和身份证（三证必须齐全）进入赛场。选手迟到 10 分钟取消比赛资格。

5. 各<sup>3</sup>领<sup>3</sup>和指导教师，以及观摩人员在赛场指定的观摩区观摩比赛。

6. 参赛选手不得携带通讯工具和其它<sup>1</sup>未经允许的资料、物品进入比赛场地，不得中途退场。如出现较严重的违规、<sup>1</sup>违纪、舞弊等现象，经裁判组裁定取消比赛成绩。

7. 新闻媒体在赛场设定的媒体采访区工作，并且听从现场工作人员的安排和管理，不能影响比赛进行。

8. 在竞赛过程中，参赛选手如有疑问，应举手示意，现场裁判应按要求及时予以答疑。如遇设备或软件等故障，参赛选手应举手示意，现场裁判、技术人员等应及时予以解决。确因计算机软件或硬件故障，致使操作无法继续的，经裁判长确认，予以启用备用设备。

9. 比赛过程中，参赛选手须严格遵守操作标准和规范，保证自身安全，并接受裁判员的监督和警示；若因设备故障导致选手中断或终止比赛，由裁判长视具体情况做出裁决。

10. 参赛选手不得因各种原因提前结束比赛。如确因不可抗因素需要离开赛场的，须向现场裁判举手示意，经裁判长许可并完成记录后，方可离开。凡在竞赛期间内提前离开的选手，不得返回赛场。

11. 现场比赛结束，经裁判长确认后离开赛场。

12. 竞赛由评分裁判现场评分，经裁判长签字确认后予以公布，如有异议请直接向赛项仲裁工作组申请复核。

13. 参赛选手应严格遵守赛场纪律，<sup>1</sup>服从指挥，着装整洁，仪表端庄，讲文明礼貌。各代表<sup>3</sup>之间应团结、友好、协作，避免发生任何形式的矛盾。

14. 其它未尽事宜，将在赛前向各领队做详细说明。

## 八、竞赛环境

1. 竞赛场地。竞赛场地分为：竞赛现场、裁判休息区、指导老师休息区、下务区。其中，竞赛现场又划分为：检录区、场内竞赛区、技术支持区。

2. 竞赛设备。场内竞赛区按照参赛数量准备比赛所需的软硬件平台，为参赛提供统一竞赛设备和备用设备。选手无需自带任何工具及附件。

3. 竞赛工位。竞赛现场各个工作区配备单相 220V/3A 以上交流电源。每个比赛工位上标明编号。

4. 技术支持区。为技术支持人员提供固定工位、电源保障。

5. 下务区。提供医疗等下务保障。

6. 竞赛场地应符合消防安全规定，现场消防器材和消防栓合格有效，应急照明设施状态合格，赛场明显位置张贴紧急疏散图，赛场出入口专人负责。现场临时用电满足《施工现场临时用电安全技术规范》JGJ46-2005 的要求。竞赛现场通风良好、照明需符合教室采光规范。

## 九、技术规范

本赛项的技术规范将包括：相关专业的教育教学要求、行业、职业技术标准，以及根据高职目录修订后的大数据技术与应用相关专业人才培养标准和规范，适时地修订本赛项遵循的技术规范。

### （一）基础标准

标准	内容
GB/T 11457-2006	信息技术、软件工艺术语
GB/T 8566-2007	信息技术 软件生存周期模型
GB/T 12991.1-2008	信息技术数据库语言 SQL 第 1 部分：框架

GB/T 21025-2007	XML 使用指南
GB/T 20009-2019	信息安全技术数据库管理系统安全评估准则 已发布
GB/T 20273-2019	信息安全技术数据库管理系统安全技术要求
20100383-T-469	信息技术安全技术信息安全管理体系实施指南

## (二) 软件开发标准

标准	内容
GB/T 8566 -2007	信息技术 软件生存周期过程
GB/T 30972-2014	系统与软件工程 软件工程环境 下务 (等同采用 ISO/IEC 15940: 2013 标准)
GB/T 20157-2006	软件维护
GB/T 25000. 51-2010	信息技术 软件包 质量要求和测试

# 十、技术平台

## (一) 竞赛设备

设备类别	数量	设备用途	基本配置
竞赛服务器	每支参赛队伍 1 台。 根据参赛队数量, 配备 10% 的备份机器。	构建大数据平台集群	性能相当于 i5 处理器, 32GB 以上内存, 1TB 以上硬盘, 网卡 (千兆), 显示器要求 1024*768 以上。
竞赛客户机	每支参赛队伍 3 台。 根据参赛团数量, 配备 10% 的备份机器。	竞赛选手比赛使用	性能相当于 i5 处理器, 8GB 以上内存, 1TB 以上硬盘, 显示器要求 1024*768 以上。

## (二) 软件平台

采用四合天地大数据实训管理系统 (北京四合天地科技有限公司研发)。主要功能如下:

1. SHTD 大数据实训管理系统内嵌基于 KVM 的虚拟化私有云平台, 该软件提供镜像上传存储、云主机一键创建、计算存储规格创建及清理、云主机回收站、云主机在线克隆为镜像模板、云主机配置参数修改等功能;

2. 系统基于 Linux 系统部署, 支持多账户管理、云主机在线或远程访问等功能; 支持多角色如 (管理员、教师、学生) 管理、专业管理、班级管理



理、用户管理、操作日志、系统设置、环境池管理、弹性 IP 管理、镜像管理、实训模板管理、课程管理、实训监控、实验环境、在线实训等功能；支持通过 VNC、SSH 等多种模式访问竞赛平台；

3. 支持模拟大数据平台环境搭建、离线数据抽取、离线数据统计、数据采集与实时计算、数据可视化等贯穿大数据技术并符合产业主流技术技能应用的相关知识点，提供大数据竞赛所需的虚拟机，所涉及开发语言包括 Scala、Javascript、HTML 等。

### （三）软件环境

设备类型	软件类别	软件名称、版本号
竞赛服务器	竞赛环境大数据集群操作系统	CentOS 7
	大数据平台组件	Hadoop 2.7.7
		Hive 2.3.4
		Spark 2.1.1
		Kafka 2.0.0
		Redis 4.0.1
		Flume 1.7.0
		Flink 1.10.2
		JDK 1.8
		MySQL 5.7
开发客户端	PC 操作系统	Ubuntu18.04 64 位
	浏览器	Chrome
	开发语言	Scala 2.11
	开发工具	IDEA 2019 (Community Edition)
		Visual Studio Code 1.58
	数据可视化组件	Vue.js 3.0
		Echarts 5.1

	文档编辑器	WPS linux 版
	输入法	搜狗拼音输入法 linux 版

## 十一、成绩评定

### （一）评分标准制定原则

竞赛评分制定严格遵守公平、公正的原则，大数据技术与应用赛项评分采用赛项结果评分方法，始终贯彻落实竞赛一贯坚持的公平、公正和公开原则。

参与竞赛成绩管理的组织机构包括裁判组、监督组和仲裁组等。裁判组实行“裁判长负责制”，设裁判长 1 名，全面负责竞赛的裁判与管理工作，裁判员根据竞赛工作需要分为检录裁判、加密裁判、现场裁判和评分裁判，其中加密裁判 2 人、现场裁判 3 人、评分裁判 6 人。检录裁判负责对参赛队伍（选手）进行点名登记、身份核对等工作；加密裁判负责组织参赛队伍（选手）抽签并对参赛队伍（选手）的信息进行加密、解密；现场裁判按规定做好赛场记录，维护赛场纪律；评分裁判负责对参赛队伍（选手）的技能展示、操作规范和竞赛成果等按赛项评分标准进行评定。

监督组对裁判组的工作进行全程监督，并对竞赛成绩抽检复核。

仲裁组负责接受由参赛队领队提出的对裁判结果的申诉，组织复议并及时反馈复议结果。

### （二）评分方法

选手在完成任务之后，将任务完成结果拷贝至 U 盘中，由参赛选手队长签字确认（签工位号）。

评分采取分步得分、累计总分的计分方式。

不计参赛选手的个人得分，只记录团体得分。

参赛<sup>3</sup>提交比赛任务结束请求或者在比赛时间终止后，不得再进行任何操作。否则，视为比赛作弊，给参赛<sup>3</sup>记警告一次。

在竞赛过程中，选手如有不<sup>3</sup>从裁判判决、扰乱赛场秩序、舞弊等不文明行为，由裁判长按照规定扣减相应分数并且给予警告，情节严重的取消竞赛资格，竞赛成绩记 0 分，<sup>3</sup>员退出比赛现场。

### (三) 评分标准

任务	考查点	描述	评分标准	分值(分)
模块 A: 大数据平台环境搭建 (10 分)	大数据相关平台组件安装配置	在指定的主机上完成 Hadoop 完全分布式、Spark、Flink、Kafka、Flume、Hive 的安装配置。	主要评分点包括 Hadoop 完全分布式安装配置、Spark 安装配置、Flink 安装配置、Kafka 安装配置、Flume 安装配置、Hive 安装配置。	10
模块 B: 离线数据抽取与统计 (35 分)	离线数据抽取与统计代码编写	按照要求使用 Scala 语言完成特定函数的编写，使用 Spark 抽取 MySQL 指定数据表中的新增的数据到 ODS 层的指定的分区表中，获取 ODS 层指定分区表中的数据进行清洗，并完成销量前 5 的商品统计、某月的总销售额统计、指定月份的销售额统计、各用户在线总时长统计，并将统计后的数据存入 MySQL 数据库中。	主要评分点包括 Spark 数据读取、数据存储、数据清洗、数据统计相关操作。	35
模块 C: 数据采集与实时计算 (20 分)	数据采集与实时计算代码编写	启动业务系统，按照要求使用 Flume 将用户操作日志采集并存入 Kafka 中并使用 Flink、Scala 消费 Kafka 中的数据将其进行聚合计算出商城在线人数，将结果存入 Redis 中，并统计该系统的 UV 与 PV 存入 MySQL 中。	主要评分点包括 Flume、Kafka、Flink、Redis 相关操作。	20

模块 D: 数据可视化 (20 分)	数据可视化代码编写	编写前端 Web 界面, 调用后台数据接口, 使用 Vue.js、ECharts 完成数据可视化。	主要评分点包括可视化前端代码开发、前端展示。	20
模块 E: 综合分析报告 (10 分)	文档编写	根据项目要求, 完成综合分析报告编写。	主要评分点包括能够按照赛项要求进行综合分析。	10
模块 F: 职业素养 (5 分)	职业素养	团队协作分工明确合理、操作规范、文明竞赛	主要评分点包括: 竞赛团队协作分工明确合理、操作规范、文明竞赛。	5

如出现参赛队总分相同情况, 按照模块分值权重顺序的得分高低排序, 即总成绩相同的情况下比较模块 B 的成绩, 模块 B 成绩高的排名优先, 如果模块 B 成绩也相同, 则按模块 C、模块 E、模块 A 的成绩进行排名, 以此类推完成相同成绩的排序。

#### (四) 成绩审核方法

竞赛结束后, 由裁判长向裁判员核实竞赛过程中有无异常。如无异常, 成绩单由裁判长签字确认并封存直至公布成绩时开启。如有异常, 在裁判长主持下, 由专家组成员、裁判员、仲裁员和监督员共同处理。

#### (五) 成绩公布方法

竞赛成绩经复核无误后, 经裁判长、监督人员审核签字后, 以赛项组委会最终公布结果为准

竞赛结束后, 如参赛队对比赛成绩有异议, 提出异议申诉或仲裁, 可按照相关规定进行申诉和仲裁, 按照仲裁结果公布竞赛成绩。

## 十二、奖项设定

本赛项设参赛选手团体一、二、三等奖。以赛项实际参赛队总数为基数, 一、二、三等奖获奖比例分别为 10%、20%、30% (小数点后四舍五

入)。为赛项获得一等奖的参赛选手的指导教师颁发“优秀指导教师”证书。

### **十三、赛场预案**

赛场备用工位：赛场提供占总参赛选手 10%的备用工位。

竞赛系统可靠性：竞赛系统使用的服务器应进行冗余，数据库、存储应使用高可用架构。提前开始运行，经过多次压力测试，由学校组织的真实竞赛环境测试。

竞赛备用服务器：现场提供占总参赛选手 10%的备用服务器。

现场应急预案详情，如下：

#### **（一）服务器问题预案**

若服务器在比赛过程中出现卡顿、死机等情况，参赛选手举手示意裁判，在裁判与技术支持人员确定情况后，可更换服务器。更换服务器的等待时间，可在比赛结束后延时。

#### **（二）交换机问题预案**

若交换机在比赛过程中出现传输速度慢或无故中断等情况，参赛选手举手示意裁判，在裁判与技术支持人员确定情况后，可更换交换机。更换交换机的等待时间，可在比赛结束后延时。

#### **（三）PC 机问题预案**

若 PC 机在比赛过程中出现死机、蓝屏等现象（重启后无法解决），参赛选手举手示意裁判，在裁判与技术支持人员确定情况后，可更换备用工位或更换 PC 机进行答题。

### **十四、赛项安全**

赛事安全是技能竞赛一切工作顺利开展的先决条件，是赛事筹备和运行工作必须考虑的核心问题。赛项执委会采取切实有效措施保证竞赛期间参赛

选手、指导教师、裁判员、工作人员及观众的人身安全。

### （一）比赛环境

1. 赛项执委会须在赛前组织专人对比赛现场、住宿场所和交通保障进行考察，并对安全工作提出明确要求。赛场的布置，赛场内的器材、设备，应符合国家有关安全规定。如有必要，也可进行赛场仿真模拟测试，以发现可能出现的问题。承办单位赛前须按照执委会要求排除安全隐患。

2. 严格控制与参赛无关的易燃易爆以及各类危险品进入比赛场地，不许随便携带书包进入赛场。

3. 配备先进的仪器，防止有人利用电磁波干扰比赛秩序。竞赛现场需对赛场进行网络安全控制，以免场内外信息交互，充分体现竞赛的严肃、公平和公正性。

4. 竞赛期间，承办单位须在赛场管理的关键岗位，增加力量，建立安全管理日志。

### （二）生活条件

1. 竞赛期间，原则上由执委会统一安排参赛选手和指导教师食宿。承办单位须尊重少数民族的信仰及文化，根据国家相关的民族政策，安排好少数民族选手和教师的饮食起居。

2. 竞赛期间安排的住宿地应具有宾馆/住宿经营许可资质。以学校宿舍作为住宿地的，竞赛期间的住宿、卫生、饮食安全等由赛项执委会和提供宿舍的学校共同负责。

3. 赛项的安全管理，除了可以采取必要的安全隔离措施外，应严格遵守国家相关法律法规，保护个人隐私和人身自由。

### （三）组织责任

1. 各学校组织代表<sup>队</sup>时，须安排为参赛选手购买竞赛期间的人身意外伤害保险。

2. 各学校代表<sup>队</sup>组成后，须制定相关管理制度，并对所有选手、指导教师进行安全教育。

3. 各参赛<sup>队</sup>伍须加强对参与竞赛人员的安全管理，实现与赛场安全管理的对接。

#### （四）应急处理

竞赛期间发生意外事故，发现者应第一时间报告赛项执委会，同时采取措施避免事态扩大。赛项执委会应立即启动预案予以解决。赛项出现重大安全问题可以停赛。

#### （五）处罚措施

1. 因参赛<sup>队</sup>伍原因造成重大安全事故的，取消其获奖资格。

2. 参赛<sup>队</sup>伍有发生重大安全事故隐患，经赛场工作人员提示、警告无效的，可取消其继续比赛的资格。

3. 赛事工作人员违规的，按照相应的制度追究责任。情节恶劣并造成重大安全事故的，由司法机关追究相应法律责任。

## 十五、竞赛须知

### （一）参赛<sup>队</sup>须知

1. 参赛<sup>队</sup>名称：统一使用规定的学校代表<sup>队</sup>名称，不使用其他组织、团体的名称；

2. 参赛<sup>队</sup>组成：每支参赛<sup>队</sup>由3名参赛选手组成，须为同校在籍学生，其中<sup>队</sup>长1名。每支参赛<sup>队</sup>可配2名指导教师，指导教师须为本校专任教师。不接受跨校组<sup>队</sup>，同一学校的报名参赛<sup>队</sup>伍不超过2支；

3. 各参赛院校应指定1名负责人任赛项领队，全权负责该校参赛事务的组织、协调和领导工作。

4. 参赛选手及指导教师报名获得确认后，原则上不再更换。如在筹备过程中，参赛选手和指导教师因故不能参赛，须由其所在学校职能部门于赛项开赛前10个工作日之前出具书面说明，经赛项执委会办公室核实后予以更换。允许队员缺席比赛；允许指导教师缺席比赛。

5. 参赛选手按照赛项竞赛规程安排，凭赛项执委会颁发的参赛证和有效身份证件参加比赛及相关活动。

6. 赛项执委会统一安排各参赛队在比赛前一天进入赛场熟悉环境和设施情况。

7. 参赛选手、领队和指导教师要有良好的职业道德，严格遵守比赛规则和比赛纪律，服从裁判，尊重裁判和赛场工作人员，自觉维护赛场秩序。

8. 领队应负责赛事活动期间本队所有选手的人身及财产安全，如发现意外事故，应及时向赛项执委会报告。

9. 各学校组织代表队时，须为参赛选手购买竞赛期间的人身意外伤害保险。

## （二）领队和指导教师须知

1. 严格遵守赛场的各项规定，服从裁判，文明竞赛。如发现弄虚作假者，取消参赛资格，名次无效。

2. 领队和指导教师务必带好有效身份证件，在活动过程中佩戴“指导教师证”参加竞赛相关活动。

3. 各代表队领队要坚决执行竞赛的各项规定，加强对参赛人员的管理，做好赛前准备工作，督促选手带好证件等竞赛相关材料。



4. 在比赛期间要严格遵守比赛规则，不得私自接触裁判人员。

5. 竞赛过程中，未经裁判许可，领队、指导教师及其他人员一律不得进入竞赛现场。

6. 如对竞赛过程有疑议，由领队和指导教师负责以书面形式向赛项仲裁委员会反映，但不得影响竞赛进行。

7. 对申诉的仲裁结果，领队要带头服从和执行，并做好选手工作。参赛选手不得因申诉或对处理意见不服而停止竞赛，否则以弃权处理。

8. 领队和指导老师应及时查看有关赛项的通知和内容，认真研究和掌握本赛项竞赛的规程、技术规范和赛场要求，指导选手做好赛前的一切技术准备和竞赛准备。

### （三）参赛选手须知

1. 参赛选手应严格遵守赛场规章、操作规程和工艺准则，保证人身及设备安全，接受裁判员的监督和警示，文明竞赛。

2. 参赛选手应按照规定时间抵达赛场，凭身份证、学生证，以及统一发放的参赛证，完成入场检录、抽签确定竞赛工位号，不得迟到早退。

3. 参赛选手凭竞赛工位号进入赛场，不允许携带任何电子设备及其他资料、用品。

4. 参赛选手应在规定的时间段进入赛场，认真核对竞赛工位号，在指定位置就座。

5. 参赛选手入场后，迅速确认竞赛设备状况，填写相关确认文件，并由参赛组长确认签字（竞赛工位号）。

6. 参赛选手在收到开赛信号前不得启动操作。在竞赛过程中，确因计算机软件或硬件故障，致使操作无法继续的，经裁判长确认，予以启用备用计算

机。

7. 参赛选手应在竞赛规定时间内完成任务书内容，并按照规定要求，将相应文档拷贝到U盘。

8. 参赛选手需及时保存工作记录。对于因各种原因造成的数据丢失，由参赛选手自行负责。

9. 参赛选手所提交的答卷采用竞赛工位号进行标识，不得出现地名、校名、姓名、参赛证编号等信息，否则取消竞赛成绩。

10. 竞赛过程中，因严重操作失误或安全事故不能进行比赛的（例如因操作原因发生短路导致赛场断电的、造成设备不能正常工作的），现场裁判有权中止该场比赛。

11. 在比赛中如遇非人为因素造成的设备故障，经裁判确认后，可向裁判长申请补足排除故障的时间。

12. 参赛选手不得因各种原因提前结束比赛。如确因不可抗因素需要离开赛场的，须向现场裁判举手示意，经裁判长许可并完成记录后，方可离开。凡在竞赛期间内提前离开的选手，不得返回赛场。

13. 竞赛操作结束后，参赛选手需要根据任务书要求，将相关成果文件拷贝至U盘，填写结束比赛相关确认文件，并由参赛选手签字确认（竞赛工位号）。因参赛选手未能按要求，将相应的文档等拷贝至U盘的，竞赛成绩计为零分。

14. 竞赛时间结束，选手应全体起立，停止操作。将资料和工具整齐摆放在操作平台上，经工作人员清点后可离开赛场，离开赛场时不得带走任何资料。

15. 在竞赛期间，未经执委会批准，参赛选手不得接受其他单位和个人进行的与竞赛内容相关的采访。参赛选手不得将竞赛的相关信息私自公布。

16. 符合下列情形之一的参赛选手，经裁判组裁定后中止其竞赛：

(1) 不服从裁判员/监考员管理、扰乱赛场秩序、干扰其他参赛选手比赛，裁判员应提出警告，二次警告后无效，或情节特别严重，造成竞赛中止的，经裁判长确认，中止比赛，并取消竞赛资格和竞赛成绩。

(2) 竞赛过程中，由于选手人为造成计算机、仪器设备及工具等严重损坏，负责赔偿其损失，并由裁判组裁定其竞赛结束与否、是否保留竞赛资格、是否累计其有效竞赛成绩。

(3) 竞赛过程中，产生重大安全事故、或有产生重大安全事故隐患，经裁判员提示没有采取措施的，裁判员可暂停其竞赛，由裁判组裁定其竞赛结束，保留竞赛资格和有效竞赛成绩。

#### (四) 工作人员须知

1. 竞赛现场设现场裁判组，裁判长 1 名，现场裁判若干名。裁判要秉公裁判，监督检查参赛选手安全有序竞赛。如遇疑问或争议，须请示裁判长裁决，裁判长的决定为现场最终裁定。

2. 赛场工作人员由赛项执委会统一聘用并进行工作分工，进入竞赛现场须佩戴赛项执委会统一提供的胸牌。

3. 赛场工作人员需服从赛项执委会的管理，严格执行赛项各项比赛规则，执行各项工作安排，积极维护好赛场秩序，坚守岗位，为赛场提供有序的服务。

4. 赛场工作人员进入现场，不得携带任何通讯工具或与竞赛无关的物品。

5. 参赛选手进入赛场，现场裁判应按规定审查参赛选手带入赛场的物品，如发现不允许带入赛场的物品，交由参赛选手随行人员保管，赛场不提供保管服务。

6. 赛场工作人员在竞赛过程中不回答选手提出的任何有关比赛技术问题，如遇争议问题，应及时报告裁判长。

## 十六、申诉与仲裁

(一) 各参赛队对不符合赛项规程规定的设备、工具、材料、计算机软硬件、竞赛执裁、赛场管理及工作人员的不规范行为等，可向赛项仲裁工作组提出申诉。

(二) 申诉主体为参赛队领队。

(三) 申诉启动时，参赛队以该队领队亲笔签字同意的书面报告的形式递交赛项仲裁工作组。报告应对申诉事件的现象、发生时间、涉及人员、申诉依据等进行充分、实事求是的叙述。非书面申诉不予受理。

(四) 提出申诉应在赛项比赛结束后 2 小时内提出。超过 2 小时不予受理。

(五) 赛项仲裁工作组在接到申诉报告后的 2 小时内组织复议，并及时将复议结果以书面形式告知申诉方。申诉方对复议结果仍有异议，可由领队向大赛组委会办公室提出申诉。大赛组委会办公室的仲裁结果为最终结果。

(六) 申诉方不得以任何理由拒绝接收仲裁结果；不得以任何理由采取过激行为扰乱赛场秩序；仲裁结果由申诉人签收，不能代收；如在约定时间和地点申诉人离开，视为自行放弃申诉。

## 十七、竞赛观摩

(一) 视频观摩

为预防控制新冠肺炎疫情，不能进入赛场进行公开观摩，采用视频观看方式。赛场外设置开放式观摩区，向媒体、企业代表、院校师生等社会公众开放，通过室外大屏幕对赛场进行直播，同时还可以通过竞赛系统进度监控图实时观看选手答题进度。

(二) 组织安排

在竞赛开始 1 个小时之后，由承办校组织并派人带领媒体、专家、企业代表、院校师生等进入赛场外的开放式观摩区，按照指定路线进行观摩。

### （三）纪律要求

为保证大赛顺利进行，在观摩期间应遵循以下纪律要求：

1. 除与竞赛直接有关工作人员、裁判员、参赛选手外，其余人员均为观摩观众。
2. 不得违反职业院校技能大赛规定的各项纪律。
3. 观摩人员需批准，佩戴观摩证件，遵循观摩区的工作人员指挥。
4. 文明观摩，保持观摩区清洁，不得大声喧哗，杜绝各种违反观摩秩序的不文明行为。

## 十八、竞赛直播

### （一）直播方式

1. 赛场内部署无盲点录像设备，能实时录制并播送赛场情况。
2. 赛场外有大屏幕或投影，同步显示赛场内竞赛状况。
3. 赛场外通过大屏幕或投影，实时展示竞赛系统选手答题进度。

### （二）直播安排

1. 对赛项赛场准备、开赛式和闭幕式、比赛期间进行录像。
2. 从选手进入赛场开始，全程进行赛场实时录像直播。
3. 从比赛开始时到比赛结束，全程进行竞赛系统的进度监控直播。

### （三）直播内容

1. 赛项执行委员会安排专人对赛项开闭幕式、比赛过程进行全程直播和录像。
2. 制作参赛选手、指导教师采访实录，裁判专家点评和企业人士采访视

频资料，突出赛项的技能重点与优势特色。为宣传、仲裁、资源转化提供全面的信息资料。

## 十九、资源转化

依照有关要求，赛项资源转化方案按要求与规程同时公布；承办方于赛后30日内向大赛执委会办公室提交资源转化实施方案，在三个月完成新资源的补充与完善，半年内完成全部资源的转化工作。

### （一）竞赛过程中获得的主要资源

1. 竞赛样题、试题库；
2. 竞赛赛题的评分标准；
3. 考核环境描述；
4. 竞赛过程音视频记录；
5. 评委、裁判、专家点评和技术分析报告；
6. 优秀选手、指导教师访谈。

### （二）资源转化基本方案与呈现形式

资源转化成果按照行业标准、契合课程标准、突出技能特色、展现竞赛优势，形成满足职业教育教学需求、体现先进教学模式、反映职业教育先进水平的共享性职业教育教学资源。资源转化成果包含基本资源和拓展资源，充分体现本赛项技能考核特点：资源转化成果应包含基本资源和拓展资源。

#### 1. 基本资源

基本资源按照风采展示、技能概要、教学资源三大模块设置。

（1）风采展示。赛后即时制作时长15分钟左右的赛项宣传片，以及时长10分钟左右的获奖代表<sup>3)</sup>（选手）的风采展示片。供专业媒体进行宣传播放。

（2）技能概要。包括技能介绍、技能操作要点、评价指标等。

（3）教学资源。教学资源充分涵盖赛项内容。包括教学方案、训练指导、作业/任务、实验/实训/实习资源等，其呈现形式主要是演示文稿、图片、操作流程演示视频、动画及相关微课程、微资源等。

#### 2. 拓展资源

拓展资源是指反映技能特色、可应用于各教学与训练环节、支持技能教学和学习过程的较为成熟的多样性辅助资源。

### (三) 资源的技术标准

资源转化成果以文本文档、演示文稿、视频文件、Flash 文件、图形/图片素材和网页型资源等。

### (四) 资源的提交方式与版权

赛项资源转化成果的版权由大赛执委会和赛项执委会共享。

### (五) 资源的使用与管理

资源转化成果的使用与管理由大赛执委会统一使用与管理,会同赛项承办单位、赛项有关专家,联系出版社编辑出版有关赛项实训教材等精品资源。

## 附件一：大数据技术与应用赛项竞赛试题（样卷）

### 一、 竞赛时间、内容及总成绩

#### (一) 竞赛时间

竞赛时间共为 6 小时,参赛选手自行安排任务进度,休息、饮水、如厕等不设专门用时,统一含在竞赛时间内。

#### (二) 竞赛内容概述

序号	任务名称	具体内容
任务一	大数据平台环境搭建	按照大数据分析平台需求,需要完成 Hadoop 完全分布式、Spark 安装配置、Flink 安装配置、Kafka 安装配置、Flume 安装配置
任务二	离线数据抽取	按照要求使用 Scala 语言完成特定函数的编写,使用 Spark 抽取 MySQL 指定数据表中的新增的数据到 ODS 层的指定的分区表中
任务三	离线数据统计	使用 Scala 语言编写程序获取 ODS 层指定分区表中的数据进行清洗,并完成销量前 5 的商品统计、某月的总销售额统计、指定月份的销售统计、各用户在线总时长统计,并将统计后的数据存入 MySQL 数据库中
任务四	数据采集与实时计算	启动业务系统,按照要求使用 Flume 将用户操作日志采集并存入 Kafka 中并使用 Flink、

		Scala 消费 Kafka 中的数据将其进行聚合计算出商城在线人数，将结果存入 Redis 中，并统计该系统的 UV 与 PV 将结果存入 MySQL 中
任务五	数据可视化	编写前端 Web 界面，调用后台数据接口，使用 Vue.js、Echarts 完成数据可视化
任务六	综合分析报告	根据要求编写综合分析报告

### （三）竞赛总成绩

“大数据技术与应用”赛项竞赛总成绩为 100 分，其中包含赛场职业素养 5 分。

## 二、任务须知

1. 每组参赛队分配一台竞赛服务器、三台客户机，拥有独立 IP 组。
2. 本次比赛采用统一网络环境比赛，请不要随意更改客户端的网络地址信息，对于更改客户端信息造成的问题，由参赛选手自行承担比赛损失；
3. 请不要恶意破坏竞赛环境，对于恶意破坏竞赛环境的参赛者，组委会根据其行为予以处罚直至取消比赛资格。
4. 比赛过程中及时保存相关文档。
5. 比赛相关文档中不能出现参赛学校名称和参赛选手名称，以赛位号（工位号）代替。
6. 参赛选手请勿删除模板内容，若因删除导致任何问题后果自负。
7. 若同一文档由不同选手完成，须将文档合并后作为最终结果提交到 U 盘中。
8. 比赛中出现各种问题及时向现场裁判举手示意，不要影响其他参赛队比赛。



### 三、 任务说明

本项目要求完成离线电商数据统计分析，完成大数据平台环境搭建、离线数据抽取、离线数据统计、数据采集与实时计算、数据可视化及综合分析报告编写等工作。

提供的相关资源包括：

1. 大数据环境搭建中需要用到的组件安装包
2. 电商相关数据
3. 大数据分析集群环境
4. 离线数据抽取开发环境
5. 离线数据统计开发环境
6. 数据采集与实时计算开发环境
7. 数据可视化开发环境
8. 综合分析报告文档模板

任务一：大数据平台环境搭建

按照大数据分析平台需求，需要完成 Hadoop 完全分布式、Spark 安装配置、Flink 安装配置、Kafka 安装配置、Flume 安装配置。

任务二：离线数据抽取

按照要求使用 Scala 语言完成特定函数的编写，使用 Spark 抽取 MySQL 指定数据表中的新增的数据到 ODS 层的指定的分区表中。

任务三：离线数据统计

使用 Scala 语言编写程序获取 ODS 层指定分区表中的数据进行清洗，并完成销量前 5 的商品统计、某月的总销售额统计、指定月份的销售额统计、各用户在线总时长统计，并将统计后的数据存入 MySQL 数据库中。

#### 任务四：数据采集与实时计算

启动业务系统，按照要求使用 Flume 将用户操作日志采集并存入 Kafka 中并使用 Flink、Scala 消费 Kafka 中的数据将其进行聚合计算出商城在线人数，将结果存入 Redis 中，并统计该系统的 UV 与 PV 将结果存入 MySQL 中。

#### 任务五：数据可视化

编写前端 Web 界面，调用后台数据接口，用 Vue.js、Echarts 完成数据可视化。

#### 任务六：综合分析报告

根据项目要求，完成综合分析报告编写。

### 四、竞赛结果提交要求

#### （一）提交方式

任务成果需拷贝至提供的 U 盘中。在 U 盘中以 XX 工位号建一个文件夹（例如 01），将所有任务成果文档保存至该文件夹中。

#### （二）文档要求

竞赛提交的所有文档中不能出现参赛选手信息和参赛选手信息，竞赛文档需要填写参赛选手信息时以工位号代替（XX 代表工位号）。