**2023年厦门市职业院校技能大赛**

**高职组“大数据应用开发”赛项规程**

一、赛项名称

赛项名称：大数据应用开发

赛项组别：高职组

赛项归属产业：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，赛项以大数据应用开发为核心内容和工作基础，重点考查参赛选手基于Hadoop、Spark、Flink平台环境下，充分利用Spark Core、Spark SQL、Flume、Kafka、Flink、Hive、HBase、Redis、Maxwell、ClickHouse、MySQL等相关技术的特点，基于Scala、Java、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

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

赛项以大数据应用开发为核心内容和工作基础，重点考查参赛选手基于Hadoop、Spark、Flink平台环境下，充分利用Spark Core、Spark SQL、Flume、Kafka、Flink、Hive、HBase、Redis、Maxwell、ClickHouse、MySQL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Hadoop平台、基于Spark的离线分析平台、基于Flink的实时分析平台，在容器环境下，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的离线数据采集方式方法，完成指定数据的抽取并写入Hive分区表中。掌握基于Flume、Maxwell的实时数据采集，将数据写入Kafka中；

3.综合利用Flink、Kafka、Hive、Redis、HBase、ClickHouse等技术，使用Java开发语言，完成某电商系统的实时数据处理，包括使用Flink处理Kafka中的数据、实时数据仓库、将数据备份至HBase中、建立Hive外表、将数据处理结果存入Redis、ClickHouse中等操作；

4.综合利用Spark、Hive、MySQL、HBase、ClickHouse等相关技术，使用Scala开发语言，完成某电商系统的离线数据处理，包括Hive数据仓库、使用Spark处理离线数据、数据合并、去重、排序、数据类型转换、将数据处理结果存入MySQL、HBase、ClickHouse中等操作；

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，利用后端数据接口完成数据可视化；

6.根据竞赛过程，完成综合分析报告的编写；

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 | Docker基本操作 |
| Hadoop完全分布式安装配置 |
| Spark安装配置 |
| Flink安装配置 |
| Hive安装配置 |
| Kafka安装配置 |
| Flume安装配置 |
| ClickHouse安装配置 |
| HBase安装配置 |
| 数据采集 | 使用Spark抽取MySQL指定数据表中的增量数据到ods层的指定的分区表中 |
| 使用Flume采集某端口的实时数据流并存入Kafka指定的Topic中 |
| 使用Maxwell采集MySQL的binlog日志并存入Kafka指定的Topic中 |
| 实时数据处理 | 使用Flink消费Kafka中的数据并将数据分发至Kafka的dwd层中 |
| 使用Flink消费Kafka中的数据的同时能够将数据备份至HBase中，同时建立Hive外表 |
| 使用Flink对实时数据进行处理并将处理计算结果存入Redis中 |
| 使用Flink对实时数据进行处理并将处理计算结果存入ClickHouse中 |
| 离线数据处理 | 使用Spark对ods层中的离线数据进行清洗，包括数据合并、去重、排序、数据类型转换等操作 |
| 将清洗完的数据存入dwd层中 |
| 根据dwd层的数据使用Spark对数据进行处理计算，并将计算结果存入MySQL中 |
| 根据dwd层的数据使用Spark对数据进行处理计算，并将计算结果存入HBase中 |
| 根据dwd层的数据使用Spark对数据进行处理计算，并将计算结果存入ClickHouse中 |
| 数据可视化 | 根据后端数据接口，基于Vue.js、ECharts的数据可视化编码（柱状图、折线图、饼状图等） |
| 综合分析报告 | 文档能力、综合分析能力 |

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% |
| 数据采集 | 权重15% |
| 实时数据处理 | 权重25% |
| 离线数据处理 | 权重20% |
| 数据可视化 | 权重15% |
| 综合分析报告 | 权重10% |
| 团队分工明确合理、操作规范、文明竞赛 | 权重5% |

四、竞赛方式

比赛以学生比赛方式进行，不得跨校组队，同一学校的报名参赛队伍不超过2支。

每个参赛队由1名领队（可由指导教师兼任）、3名选手（3名学生）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。

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** |
| 10月27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 |
| 10月28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、竞赛命题

技能大赛的命题工作由赛项执委会指定的命题专家组负责。

本赛项制定样题一套，并与本规程同步发布。具体详见附件1。

正式赛卷在比赛前一天由裁判长在监督组的监督下从命题专家组负责的命题库中随机抽取一套赛题，在保密室中打印并存放。保密室全程监控，并安排专人看守。比赛前1小时由两名裁判及比赛监督员将赛题从保密室运往赛场。

七、竞赛规则

1. 学生须为高等职业学校专科、高等职业学校本科、技师学院全日制在籍学生，五年制高职四、五年级学生也可报名参赛。凡在往届全国职业院校技能大赛中获一等奖的学生，不能再参加同一项目同一组别的比赛。参赛选手由省教育厅、省人力资源社会保障厅负责。

2. 竞赛前1日安排各参赛队领队、参赛选手熟悉赛场。

3. 严禁参赛选手、赛项裁判、工作人员私自携带通讯、摄录设备进入比赛场地。

4. 参赛选手所需的硬件、软件和辅助工具统一提供，参赛队不得使用自带的任何有存储功能的设备，如硬盘、光盘、U盘、手机、平板电脑等。

5. 所有参赛选手都必须携带参赛证件进行检录。

6. 参赛队在赛前领取比赛任务并进入比赛工位，比赛正式开始后方可进行相关操作。

7. 比赛过程中，选手须严格遵守操作规程，确保人身及设备安全，并接受裁判员的监督和指示。因选手原因造成设备故障或损坏而无法继续比赛的，裁判长有权决定中止该队比赛；非因选手个人原因造成设备故障的，由裁判长视具体情况作出裁决。

8. 竞赛开始时统一发放本阶段赛卷，竞赛结束后，参赛选手要确认已成功提交竞赛要求的配置文件和文档，裁判员与参赛选手一起签字确认，参赛选手在确认后不得再进行任何操作。

9. 赛项成绩解密后，在指定地点，以纸质形式向全体参赛队进行公布，成绩无异议后，在闭赛式上予以宣布。

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、服务区。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队数量准备比赛所需的软硬件平台，为参赛队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. 服务区。提供医疗等服务保障。

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应用开发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、软件工程术语  |
| GB8566-88 | 计算机软件开发规范 |
| GB/T 12991-2008 | 信息技术数据库语言SQL第1部分：框架 |
| GB/T 21025-2007 | XML使用指南 |
| GB/T 20009-2005 | 信息安全技术数据库管理系统安全评估准则 已发布 |
| GB/T 20273-2006 | 信息安全技术数据库管理系统安全技术要求 |
| 20100383-T-469 | 信息技术安全技术信息安全管理体系实施指南 |

**（二）软件开发标准**

|  |  |
| --- | --- |
| **标 准** | **内 容** |
| GB/T 8566 -2001 | 信息技术 软件生存周期过程  |
| GB/T 15853 -1995 | 软件支持环境  |
| GB/T 14079 -1993 | 软件维护指南 |
| GB/T 17544-1998 | 信息技术 软件包 质量要求和测试 |

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，64GB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。根据参赛团队数量，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，16GB以上内存，1TB以上硬盘，显示器要求1024\*768以上。 |

**（二）软件平台**

由2022年全国职业院校技能大赛（高职组）大数据应用开发赛项合作企业——北京四合天地科技有限公司提供四合天地大数据实训管理系统。

系统基于自研Docker容器编排管理引擎，运用云原生和容器技术构建训练环境，支持快速创建训练环境。实现每个学生环境互相隔离、训练过程互不干扰，教师可以一键操作即可创建一套全新的环境供学生进行备赛训练，方便学生高效的完成训练操作的同时，大幅降低了教师组织训练的难度和成本。具体功能如下：

1. 本系统运行在开源操作系统Linux平台下，应使用浏览器/服务器模式提供服务,用户使用最新版本的谷歌浏览器访问系统。
2. 系统支持包括管理员、教师、学生三种角色。管理员负责系统配置维护、镜像环境维护等工作；教师负责小组维护、训练任务维护等工作；学生参与并完成训练。
3. 管理员功能：
4. 专业管理：管理员可自主创建专业，编辑专业相关内容，如：新增（编辑）专业，删除专业。
5. 班级管理：管理员可自主创建班级，编辑班级相关内容，如：新增（编辑）班级，删除班级。
6. 用户管理：管理员可自主创建用户，编辑用户相关内容，如：新增（编辑）用户的账号、姓名、手机号、角色、状态、密码等，同时也可对用户进行删除操作。
7. 镜像环境：管理员可维护镜像环境，可通过网页上传Dockerfile文件来完成镜像环境的创建，也可从本地镜像仓库进行镜像同步，并可对镜像环境执行编辑等操作。
8. 环境配置：管理员依据镜像环境的具体应用场景，通过添加环境配置来完善该镜像所需的CPU、内存、磁盘空间等配置信息，并可对环境配置信息进行编辑和删除。
9. 服务器配置：管理员可配置计算节点服务器相关信息，包括新增服务器名称、服务器IP地址、CPU核数、内存大小、磁盘空间，也可对服务器配置信息进行编辑和删除。
10. 训练资源监控：管理员可监控正在进行的训练资源，监控正在进行的训练状态、使用人数、占用资源等信息，并可对相关的环境执行启动、挂起等操作。
11. 角色管理：管理员可自主创建角色，编辑角色相关内容，如：添加（编辑）角色名称、角色备注（描述）、状态、权限分配（权限分配依据角色名称分配相应权限）等，同时也可对角色进行删除操作。
12. 日志管理：管理员可查看当前系统操作日志和登录日志。操作日志主要记录操作的账户、操作模块、时间、IP等信息。登录日志主要记录登录的用户、IP地址、登录状态等信息。
13. 系统设置：可自定义系统名称，如浏览器标题、浏览器LOGO、登录页标题、首页标题、首页LOGO等。
14. 教师功能：
15. 训练模块：教师可自主创建训练模块，编辑相应的训练内容。内容包括训练的名称、介绍、使用状态、该训练指定的一个或多个实训环境、可供下载的资源包、任务设置的具体形式包括上传任务说明、任务参考等。
16. 小组管理：教师可自主创建小组，并对小组进行编辑、启用、禁用等操作。已启用的小组可以进行添加成员、设置任务参考操作。在编辑具体小组时，通过选择该小组所包含的成员，内容上选择所用到的训练模块，完成小组的整体设置。
17. 训练监控：教师可查看各小组的训练情况，选择启动、挂起、继续或结束训练环境。通过“进入”按钮，可进入小组的某个环境进行指导操作，也可对小组的某个训练环境执行启动、停止、重置或重启。
18. 训练报告：对已启用小组，记录每个小组中，学生具体的训练情况，查阅学生提交上传的报告、附件等信息，编辑每个学生对应的训练成绩。
19. 个人实验环境：教师可在个人实验环境中选择镜像环境进行个人实验或练习等操作。
20. 消息通知：当学生完成训练报告的提交或附件上传时，教师会收到消息通知并可通过消息通知直接找到训练报告或附件。
21. 个人资料：教师可修改个人头像及密码。
22. 学生功能：
23. 学生登录平台，可进入对应的训练模块，在线查看任务说明、任务参考，并进入训练环境进行操作，可在浏览器中访问自己的训练环境，也可通过SSH工具直连相应训练环境容器的ip地址进入该环境，支持通过环境的ip地址访问环境中部署的相应服务，也可对环境进行重置、重启、上传、下载等操作，同时学生可以下载相关的资源文件，提交个人的训练报告，上传相关附件完成训练。
24. 个人中心：学生可修改个人头像及密码，也可查看自己的训练记录。

**（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、Docker-CE 20.10 |
| 大数据平台组件 | Hadoop 3.1.3 |
| Hive 3.1.2 |
| HBase 2.2.3 |
| Spark 3.1.1 |
| Kafka 2.4.1 |
| Redis 6.2.6 |
| Flume 1.9.0 |
| Maxwell 1.29.0 |
| Flink 1.14.0 |
| ClickHouse 21.9.4 |
| 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 2.12 |
| Java 8 |
| 开发工具 | IDEA 2022（Community Edition） |
| Visual Studio Code 1.69 |
| 数据库连接工具 | MySQL Workbench |
| SSH工具 | Asbru-cm或Ubuntu SSH客户端 |
| API测试工具 | Postman API Platform |
| 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）奖项设定**

竞赛设参赛选手团体奖。奖项设置按参赛队数量确定，其中一等奖10%，二等奖20%，三等奖30%。为团队一等奖选手的指导教师设优秀指导教师奖。

**（二）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，大数据应用开发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、加密裁判2名、现场裁判4名、评分裁判6名，共13名裁判。

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。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。

**（三）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。

参赛队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，给参赛队记警告一次。

在竞赛过程中，选手如有不服从裁判判决、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，队员退出比赛现场。

**（四）评分标准**

| **任务** | **考查点** | **描述** | **评分标准** | **分值（分）** |
| --- | --- | --- | --- | --- |
| 大数据平台环境搭建（10分） | 大数据相关平台组件安装配置 | 在指定的宿主机上，基于Docker环境完成Hadoop完全分布式、Spark、Flink、Hive、Kafka、Flume、ClickHouse、HBase等的安装配置。 | 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、Hive安装配置、Kafka安装配置、Flume安装配置、ClickHouse安装配置、HBase安装配置。 | 10 |
| 数据采集（15分） | 离线数据采集、实时数据采集 | 按照要求基于Scala语言完成特定函数的编写，使用Spark完成离线数据采集；按照要求使用Linux命令，利用Flume、Maxwell、Kafka等工具完成实时数据采集。 | 主要评分点包括Spark数据读取、数据存储、Flume数据采集、Maxwell数据采集、Kafka等操作。 | 15 |
| 实时数据处理（25分） | 实时数据处理计算代码编写 | 使用Java语言基于Flink完成Kafka中的数据消费，将数据分发至Kafka的dwd层中，并在HBase中进行备份同时建立Hive外表，基于Flink完成相关的数据指标计算并将计算结果存入Redis、ClickHouse中。 | 主要评分点包括Flink数据处理、数据指标计算、HBase、Hive、ClickHouse、Redis等相关操作。 | 25 |
| 离线数据处理（20分） | 离线数据处理计算代码编写 | 使用Scala语言基于Spark完成离线数据清洗、处理、计算，包括数据的合并、去重、排序、数据类型转换等并将计算结果存入MySQL、HBase、ClickHouse中。 | 主要评分点包括基于Spark的数据清洗、数据指标计算、HBase、Hive、ClickHouse、MySQL等相关操作。 | 20 |
| 数据可视化（15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。 | 15 |
| 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 |
| 职业素养（5分） | 职业素养 | 团队分工明确合理、操作规范、文明竞赛。 | 主要评分点包括：竞赛团队分工明确合理、操作规范、文明竞赛。 | 5 |

**（五）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。

**（六）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、赛场预案

赛场备用工位：赛场提供占总参赛队伍10%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：现场提供占总参赛队伍10%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。

十三、申诉与仲裁

**（一）申诉**

1. 参赛队对不符合竞赛规定的设备、工具、软件，有失公正的评判、奖励，以及对工作人员的违规行为等，均可提出申诉。

2. 申诉应在竞赛结束后2小时内提出，超过时效将不予受理。申诉时，应按照规定的程序由参赛队领队向相应赛项裁判委员会递交书面申诉报告。报告应对申诉事件的现象、发生的时间、涉及到的人员、申诉依据与理由等进行充分、实事求是的叙述。事实依据不充分、仅凭主观臆断的申诉将不予受理。申诉报告须有申诉的参赛选手、领队签名。

3. 赛项裁判委员会收到申诉报告后，应根据申诉事由进行审查，2小时内书面通知申诉方，告知申诉处理结果。如受理申诉，要通知申诉方举办听证会的时间和地点；如不受理申诉，要说明理由。

4. 申诉人不得无故拒不接受处理结果，不允许采取过激行为刁难、攻击工作人员，否则视为放弃申诉。申诉人不满意赛项裁委会的处理结果的，可向赛项仲裁工作组提出复议申请。

**（二）仲裁**

1. 2023年XX省职业院校技能大赛（高职组）“大数据应用开发”赛项裁判委员会设仲裁工作组，负责受理竞赛中出现的申诉复议并进行仲裁，以保证竞赛的顺利进行和竞赛结果公平、公正。

2. 仲裁工作组的裁决为最终裁决，参赛队不得因对仲裁处理意见不服而停止比赛或滋事，否则按弃权处理。

十四、竞赛观摩

**（一）视频观摩**

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。

十五、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、开赛式和闭赛式、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赛式、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。

十六、竞赛须知

**（一）参赛队须知**

1. 参赛队名称：统一使用规定的学校代表队名称，不使用其他组织、团体的名称；

2. 参赛队组成：每支参赛队由3名参赛选手组成，须为同校在籍学生，其中队长1名。每支参赛队可配2名指导教师，指导教师须为本校专兼职教师。不接受跨校组队，同一学校的报名参赛队伍不超过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在报名获得确认后，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，须由其所在学校供职部门于赛项开赛前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. 参赛队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. 参赛队选手、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，并由参赛队长确认签字（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. 参赛队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队比赛。

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，并由参赛队长签字确认（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，监督检查参赛队安全有序竞赛。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。

5.参赛队进入赛场，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，交由参赛队随行人员保管，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。

# 附件一：大数据应用开发赛项竞赛试题（样卷）

# 竞赛时间、内容及总成绩

**（一）竞赛时间**

竞赛时间共为6小时，参赛队自行安排任务进度，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 | 按照任务书要求，需要基于Docker环境完成Hadoop完全分布式、Spark安装配置、Flink安装配置、Hive安装配置、Kafka安装配置、Flume安装配置、ClickHouse安装配置、HBase安装配置等中的任意三个组件的安装配置 |
| 任务二 | 数据采集 | 按照任务书要求基于Scala语言基于Spark完成离线数据采集，将数据存入Hive的ods层中；按照要求使用Linux命令，利用Flume、Maxwell、Kafka等工具完成实时数据采集 |
| 任务三 | 实时数据处理 | 按照任务书要求使用Java语言基于Flink完成Kafka中的数据消费，将数据分发至Kafka的dwd层中，并在HBase中进行备份同时建立Hive外表，基于Flink完成相关的数据指标计算并将计算结果存入Redis、ClickHouse中 |
| 任务四 | 离线数据处理 | 按照任务书要求使用Scala语言基于Spark完成离线数据清洗、处理、计算，包括数据的合并、去重、排序、数据类型转换等并将计算结果存入MySQL、HBase、ClickHouse中 |
| 任务五 | 数据可视化 | 按照任务书要求编写前端代码，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**

“大数据应用开发”赛项竞赛总成绩为100分，其中包含赛场职业素养5分。

# 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。

# 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、数据采集、实时数据处理、离线数据处理、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. 电商相关脱敏业务数据
3. 大数据分析集群环境
4. 数据采集开发环境
5. 实时数据处理开发环境
6. 离线数据处理开发环境
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

按照任务书要求，需要基于Docker环境完成Hadoop完全分布式、Spark安装配置、Flink安装配置、Hive安装配置、Kafka安装配置、Flume安装配置、ClickHouse安装配置、HBase安装配置等中的任意三个组件的安装配置。

任务二：数据采集

按照任务书要求基于Scala语言基于Spark完成离线数据采集，将数据存入Hive的ods层中；按照任务书要求使用Linux命令，利用Flume、Maxwell等工具完成实时数据采集，将数据存入Kafka指定的Topic中。

任务三：实时数据处理

按照任务书要求使用Java语言基于Flink完成Kafka中的数据消费，将数据分发至Kafka的dwd层中，并在HBase中进行备份同时建立Hive外表，基于Flink完成相关的数据指标计算并将计算结果存入Redis、ClickHouse中。

任务四：离线数据处理

按照任务书要求使用Scala语言基于Spark完成离线数据清洗、处理、计算，包括数据的合并、去重、排序、数据类型转换等并将计算结果存入MySQL、HBase、ClickHouse中。

任务五：数据可视化

按照任务书要求编写前端代码，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

按照任务书要求，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队信息和参赛选手信息，竞赛文档需要填写参赛队信息时以工位号代替（XX代表工位号）。