

2024 年四川省职业院校技能大赛

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 4 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	4 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

- 1、比赛以学生赛方式进行，不得跨校组队，同一学校的报名参赛队伍不超过 3 支。
- 2、每个参赛队由 2 名指导教师和 2 名选手组成，参赛选手报名获得确认后不得随意更换。
- 3、竞赛时间 4 小时。

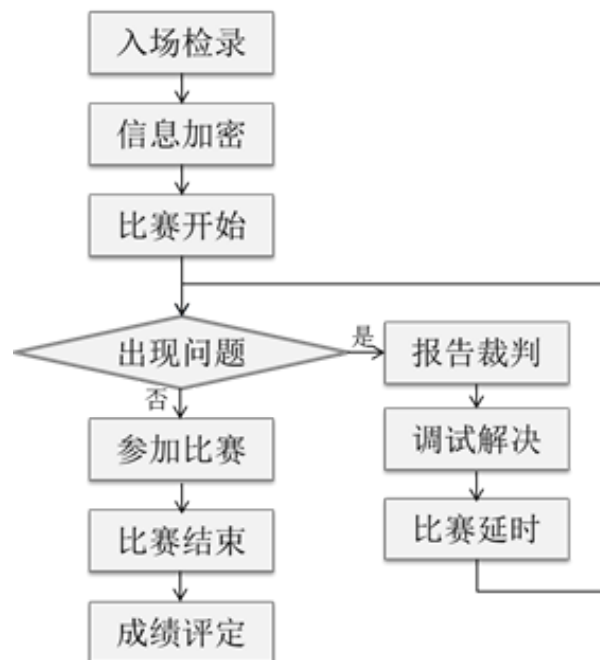
五、竞赛流程

根据竞赛任务要求，参赛队伍在 4 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
12 月 24 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
12 月 25 日	8: 00-9: 00	赛场检录，竞赛选手进入赛位
	9: 00-13: 00	竞赛选手完成竞赛任务
	14: 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。

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国赛中获一等奖和 2024 年世界职业院校技能大赛总决赛争夺赛获金奖的选手，不得参加本届大赛同一专业大类赛项的比赛。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，经裁判长等签字确认后在群内向参赛队进行公布。

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 10 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	Intel 银牌 4316, 硬盘容量 ≥4TB*2 SATA 接口，内存 640GB
竞赛客户机	每支参赛队伍 4 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
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%。

（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“裁判员负责制”。

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

选手在完成比赛任务之后，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	大数据相关平	在指定的宿主机上，基于 Docker 环境完成 Hadoop 完全分布式、	主要评分点包括 Hadoop 完全分布式安装配置、	10

台环境搭建 (10分)	台组件安装配置	Spark、Flink、Hive、Kafka、Flume、ClickHouse、HBase 等的安装配置。	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 (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

各参赛队伍对设施配备、比赛执裁、赛场管理、成绩评定，以及工作人员的不规范行为等持有异议时，由各参赛队伍领队向赛项仲裁工作组提出书面申诉。提出申诉时间应在选手比赛结束后 2 小时内，超过时效不予受理。申诉书应对申诉事件的基本情况、发生时间、涉及人员、申诉依据等进行充分、实事求是的叙述。申诉书应由参赛队领队签字递交。

(二) 仲裁

1. 2024 年四川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

本次竞赛不能进入赛场进行公开观摩，采用视频观看方式。

(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2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3支；

3. 各参赛院校应指定1名负责人任赛项领队（可由参赛教师兼任），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。

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. 领队务必带好有效身份证件，在活动过程中佩戴“领队证”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）

一、竞赛时间、内容及总成绩

（一）竞赛时间

竞赛时间共为 4 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。

二、 任务须知

1. 每组参赛队分配一台线上云服务器、两台客户机，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、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）。