

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）。