

“大数据应用开发”赛项大赛训练基地

大数据应用开发项目为团队项目，每个参赛队由 3 位选手组成，要求在规定的时间内相互配合并完成赛题的任务。赛题包含掌握基于 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；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；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服务，包括销量前 5 商品统计、某月的总销售额统计、每个月的销售额统计、每个用户在线总时长统计，并将统计结果存入 MySQL 数据库中；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；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；根据数据可视化结果，完成数据分析报告的编写。训练基地如下：



首次参赛时间：2019年5月