

命题赛题目五

企业大数据分析决策平台的设计与实现

一、选题背景：

数据是企业的重要资产，其背后蕴含的巨大商业价值，数据价值的充分挖掘和利用，对企业未来的发展和创新商业模式将有很大的帮助。某企业已上线多个业务系统，既有面向客户的业务系统，例如：网上商城，也有面向公司内部管理的业务系统，例如：财务系统、OA系统、ERP系统、档案管理系统、上网行为监控系统、视频监控系统等，这些业务系统已累积了海量的数据，即包含关系型数据库中的结构化数据，也包含日志、文档、图片、视频等非结构化数据。同时，这些业务系统还在源源不断产生新的业务数据。企业希望建设大数据分析决策平台，对各域数据实现资产化的统一管理，促进企业的业务创新、运营提升、管理优化、开放共享。

二、题目概况：

建设企业大数据分析决策平台，面向海量结构化、半结构化和非结构化数据提供采集、存储、计算、分析挖掘、查询展示等服务，为企业决策提供辅助支持。大数据分析决策平台需具备分布式、高性能、高可用、安全可靠、弹性扩展等特点。

参赛队伍可自主选择某个业务场景，模拟业务数据的产生，经过大数据分析决策平台处理和分析后，得出某个数据模型或某个指标的

改进建议，并通过数据可视化系统进行展示，实现对业务的改进、优化、预警、监控等，解决企业的痛点问题，为企业决策提供数据支撑。

本赛题考核的指标包括但不限于以下：

1. 大数据分析决策平台架构的合理性、可用性、性能、安全性、弹性扩展。
2. 是否可以处理多业务系统的海量数据，包括结构化、半结构化和非结构化数据。
3. 是否可以同时处理离线数据和实时数据。
4. 选取的业务场景是否真实，数据指标是否合理，项目案例是否具有社会意义或商业价值。

三、需求描述：

需求一(必选)：业务场景选取和数据准备

1. 参赛选手自主选择业务场景，构建业务分析指标。
2. 准备静态业务真实数据或仿真数据，业务数据需包含结构化数据（如存储在 MySQL 数据库中的数据）、非结构化数据（如存储在文本中的数据）。
3. 编写业务数据生成工具，模拟业务系统运行，可自动生成业务数据。

需求二(必选)：大数据平台的构建

包括但不限于以下内容：

1. Hadoop 集群的安装、配置和管理。
2. Spark 集群的安装、配置和管理。

3. 数据仓库组件 Hive 的安装、配置和管理。

需求三(必选): 应用开发

在已构建的大数据平台上完成应用开发, 具体内容包括:

1. 根据业务场景建立数据分析模型。
2. 实现数据分析和数据挖掘算法, 可基于 Java、Python 等语言编程实现, 也可基于第三方开源系统。
3. 编程实现数据可视化展示, 可基于第三方开源系统, 编程语言不限。

需求四(必选): 整体联调

系统整体联调, 实现需求 1、需求 2 和需求 3 中各模块的融会贯通, 用于模拟真实业务场景“产生数据→存储和处理数据→分析和挖掘→数据展示”等环节。重点验证以下内容:

1. 结构化数据和非结构化数据是否可以处理。
2. 离线数据和实时数据是否可以处理。
3. 数据模型和分析算法是否正确, 分析结果是否与业务场景相符。

需求五(可选): 工程文档编写

根据项目需求和实施成果, 编写平台的安装部署文档和运维手册。