面向工业制造领域的多模态数据治理平台

软件需求规约

版本 <1.2>

修订历史记录

|  |  |  |  |
| --- | --- | --- | --- |
| **日期** | **版本** | **说明** | **作者** |
| 14/10/23 | 1.0 | 根据立项建议书编写规约文档，修改部分需求明细 | 曾宇欣、陈诺、陈治源 |
| 31/10/23 | 1.2 | 补充并完善非功能需求、整理复核 | 曾宇欣、钱麒丹、张瑞轩 |

目录

1. 简介 5

1.1 目的 5

1.2 定义、首字母缩写词和缩略语 5

1.3 参考资料 5

2. 整体说明 5

2.1 • 产品总体效果 5

2.2 • 产品功能 5

2.3 • 用户特征 6

2.4 • 约束 6

2.5 • 假设与依赖关系 6

2.6 • 需求子集 6

3. 功能需求 7

3.1 <Use case 图> 7

3.2 数据模型定义 7

3.3 标签描述定义 8

3.4 质量规则定义 8

3.5 安全规则定义 8

3.6 数据模型管理 8

3.7 多模态数据源接入 9

3.8 数据源管理 9

3.9 数据资产管理 9

3.10 数据服务管理 9

4. 非功能需求 10

4.1 易用性 10

4.1.1 培训时间 10

4.1.2 易用性需求 10

4.2 可靠性 10

4.2.1 可用性 10

4.2.2 平均故障间隔时间（MTBF） 10

4.2.3 平均修复时间（MTTR） 10

4.2.4 最高错误或缺陷率 10

4.3 性能 10

4.3.1 响应时间 10

4.3.2 吞吐量 10

4.4 可支持性 10

4.4.1 兼容性 10

4.4.2 可维护性 11

4.4.3 编码标准 11

4.4.4 命名约定 11

4.5 设计约束 11

4.5.1 软件语言 11

5. 其它产品需求 11

5.1 联机用户文档和联机帮助的需求 11

5.2 接口需求 11

5.2.1 用户界面 11

5.2.2 硬件接口 11

5.2.3 软件接口 11

5.2.4 通信接口 11

5.3 适用的标准 11

软件需求规约 (简化版)

# 简介

## 目的

此软件需求规约文档用以说明面向工业制造领域的多模态数据治理平台的需求。本文档将详细地说明项目的功能需求、非功能需求以及设计约束，同时提供软件需求说明所需的其他相关要素。

## 定义、首字母缩写词和缩略语

|  |  |
| --- | --- |
| **定义及缩略语** | **含义** |
| 数据治理 | 一种组织内部的管理和控制框架，旨在确保数据的质量、可用性、安全性和合规性，以支持数据的有效管理和利用。 |
| 多模态处理 | 指的是能够处理和整合来自多种不同数据源或数据类型的信息的能力，例如时序、图像或自然文本等。 |
| 大模型 | 在数据治理平台中使用复杂、高性能的机器学习模型，以处理和分析大规模、多模态的工业数据以提取有价值的信息。 |
| ETL | Extract-Transform-Load，描述将数据从数据源端经过抽取、转换、加载到目的端的过程。 |

## 参考资料

1. 《面向工业制造领域的多模态数据治理平台-立项建议书》，高级软件开发与管理课程第1小组。
2. 《数据治理技术》，软件学报。
3. 《人工智能在数据治理中的应用》，信息通信技术与政策。
4. 《智能数据治理平台》，信息安全研究。
5. 《大数据之路 阿里巴巴大数据实践》，电子工业出版社。
6. 《软件工程原理》高等教育出版社，2013年2月第一版。

# 整体说明

## • 产品总体效果

本项目的主要目标是将数据转化为有价值的知识，通过从业务数据中提取关键信息，迅速满足用户需求，并最终生成对企业运营和决策有益的见解。项目的目标是协助企业的各种用户，在业务管理和战略制定方面能够做出及时且准确的决策。这个项目的愿景是为不断发展壮大的企业提供出色性能和高度适配性的数据存储和分析服务。项目还提供数据API，以便上层应用轻松接入，并支持多层次的开发和应用，以构建基于这一数据治理平台的全方位解决方案。

## • 产品功能

本项目立足于多模态数据处理，致力于构建并实施一款专为工业应用而设计的数据治理平台。该平台旨在将现有工业系统中的各类数据库注册为数据源，实现数据的输入、清洗、存储，并为用户提供强大的数据分析能力和多样的服务接口。其主要特点和功能包括：

1. 数据汇聚：包括对结构化数据源、半结构化数据源、非结构化数据源的配置，以支持数据源直采；定时从数据源采入数据以及实时采集数据。
2. 用户管理与数据安全：包括用户注册登录以及用户权限管理。
3. 数据资产管理：包括自动获得数据源的元数据信息；对数据上下游关系进行可视化；使用成熟的ETL工具，提供跨系统的数据清洗和整合，提供数据完整性；对清洗后的数据进行存储，以支持后续的抽取和分析。
4. 数据模型管理：平台用户查看某数据模型相关信息；更新数据模型信息并点击保存，系统更新数据模型；删除异常或不需要的数据模型。
5. 数据服务：用户可以进行数据相似度查询，并可以通过大模型交互来生成脚本，数据源关系通过数据地图展示。
6. 数据应用：实现数据源关系的可视化以及对数据的分析挖掘。

## • 用户特征

数据中台甲方的运维人员和业务人员。

## • 约束

前端开发语言：HTML、JavaScript、CSS。

服务端编程语言：Java、Python。

Web前端框架：React.JS。

Web UI库：Ant Design。

开发工具：Intellij IDEA、PyCharm、Visual Studio Code。

## • 假设与依赖关系

项目要求自立项起3个月内完成；

多模态数据治理架构的理念契合本系统的工业领域的应用场景；

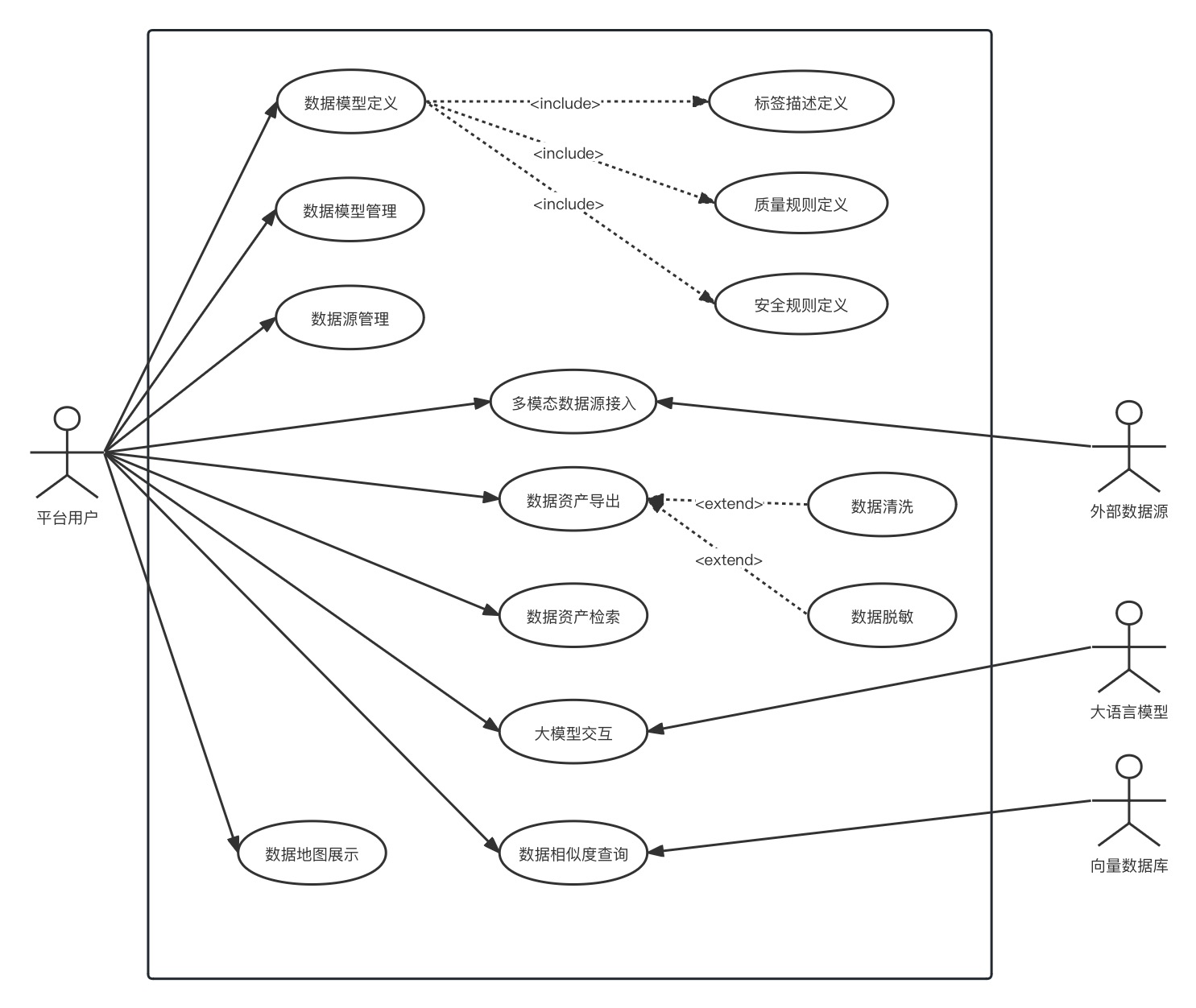
Java具有优秀的生态体系。

## • 需求子集

系统安全稳定，操作简单，方便快捷，用户友好。

# 功能需求

## <Use case 图>



## 数据模型定义

|  |  |  |  |
| --- | --- | --- | --- |
| 用例编号 | UC01 | 用例名称 | 数据模型定义 |
| 描述 | 平台用户定义数据模型 | | |
| 执行者 | 平台用户 | | |
| 前置条件 | 平台用户已登录，并进入数据模型管理页面 | | |
| 后置条件 | 数据模型定义被记录，新增数据模型被保存 | | |
| 基本流 | 1. 点击新增数据模型按钮 2. 平台用户填写数据模型内容 3. 点击确定按钮保存 4. 成功添加新增数据模型 | | |
| 备选流 | 2a. 存在重名数据模型，系统显示相应的提示信息  3a. 用户编辑的信息不合法，系统显示相应的提示信息 | | |

## 标签描述定义

|  |  |  |  |
| --- | --- | --- | --- |
| 用例编号 | UC02 | 用例名称 | 标签描述定义 |
| 描述 | 平台用户对数据模型添加标签描述信息 | | |
| 执行者 | 平台用户 | | |
| 前置条件 | 平台用户已登录，并进入数据模型定义页面 | | |
| 后置条件 | 数据模型的标签描述信息被记录 | | |
| 基本流 | 1. 系统显示推荐的标签和描述 2. 平台用户选择标签描述信息 3. 平台用户添加新的标签描述或修改选择的标签描述 4. 平台用户点击确认，完成标签描述定义 | | |
| 备选流 | 1a. 无推荐的标签描述，系统提示相关信息  4a. 标签描述不合法，系统提示相关信息 | | |

## 质量规则定义

|  |  |  |  |
| --- | --- | --- | --- |
| 用例编号 | UC03 | 用例名称 | 质量规则定义 |
| 描述 | 平台用户对数据模型添加质量规则 | | |
| 执行者 | 平台用户 | | |
| 前置条件 | 平台用户已登录，并进入数据模型定义页面 | | |
| 后置条件 | 数据模型的质量规则被记录 | | |
| 基本流 | 1. 显示质量规则库中本数据模型适用的质量规则 2. 平台用户选择质量规则 3. 平台用户根据需求添加新的质量规则 4. 平台用户点击确认，完成质量规则定义 | | |
| 备选流 | 1a. 无适用的质量规则，系统提示相关信息  3a. 用户添加的质量规则不合法，系统提示相关信息 | | |

## 安全规则定义

|  |  |  |  |
| --- | --- | --- | --- |
| 用例编号 | UC04 | 用例名称 | 安全规则定义 |
| 描述 | 平台用户对数据模型添加安全规则 | | |
| 执行者 | 平台用户 | | |
| 前置条件 | 平台用户已登录，并进入数据模型定义页面 | | |
| 后置条件 | 数据模型的安全规则被记录 | | |
| 基本流 | 1. 平台用户选择数据模型的安全密级 2. 平台用户选择数据导出时是否脱敏 3. 平台用户选择数据存储时是否加密 4. 平台用户点击确认，完成安全规则定义 | | |
| 备选流 | 4a. 用户添加的安全规则不合法，系统提示相关信息 | | |

## 数据模型管理

|  |  |  |  |
| --- | --- | --- | --- |
| 用例编号 | UC05 | 用例名称 | 数据模型定义 |
| 描述 | 平台用户对已有数据模型进行管理，包括更新、删除、查看详情等 | | |
| 执行者 | 平台用户 | | |
| 前置条件 | 平台用户已登录，并进入数据模型管理页面 | | |
| 后置条件 | 平台用户操作被记录，系统根据操作进行信息变更，并由前端显示 | | |
| 基本流 | 1. 平台用户查看某数据模型相关信息 2. 平台用户更新数据模型信息并点击保存，系统更新数据模型 3. 平台用户删除异常或不需要的数据模型 | | |
| 备选流 | 1a. 用户更新的数据模型信息不合法，系统提示相关信息  3a. 数据模型删除失败，系统提示相关信息 | | |

## 多模态数据源接入

|  |  |  |  |
| --- | --- | --- | --- |
| 用例编号 | UC06 | 用例名称 | 多模态数据源接入 |
| 描述 | 平台用户注册数据源信息，包含数据源访问ip账号密码或文件位置等 | | |
| 执行者 | 平台用户 | | |
| 前置条件 | 平台用户已登录，并进入数据源接入页面 | | |
| 后置条件 | 数据源信息被记录 | | |
| 基本流 | 1. 系统显示数据源接入表单 2. 平台用户填写数据源相关信息并点击提交 3. 系统调取元数据信息并在验证页面显示 4. 平台用户填写数据源描述信息并点击确认，完成注册 | | |
| 备选流 | 3a. 未连接成功数据源，系统提示相关信息  4a. 平台用户确认连接错误，返回数据源连接表单页重新填写 | | |

## 数据源管理

|  |  |  |  |
| --- | --- | --- | --- |
| 用例编号 | UC07 | 用例名称 | 数据源管理 |
| 描述 | 平台用户对已有数据源进行管理，包括更新、删除、查看详情等 | | |
| 执行者 | 平台用户 | | |
| 前置条件 | 平台用户已登录，并进入数据模型管理页面 | | |
| 后置条件 | 平台用户操作被记录，系统根据操作进行信息变更，并由前端显示 | | |
| 基本流 | 1. 系统显示数据模型已接入的所有数据源 2. 平台用户查看某数据源的详细信息 3. 平台用户更新数据源信息并点击保存，系统更新数据源相关信息 4. 平台用户删除异常或不需要的数据源 | | |
| 备选流 | 2a. 未连接成功数据源，系统提示相关信息  3a. 数据源删除失败，系统给出提示 | | |

## 数据资产管理

|  |  |  |  |
| --- | --- | --- | --- |
| 用例编号 | UC08 | 用例名称 | 数据资产管理 |
| 描述 | 平台用户对已有数据资产进行管理，包括导出、检索、查看详情等 | | |
| 执行者 | 平台用户 | | |
| 前置条件 | 平台用户已登录，并进入数据资产管理页面 | | |
| 后置条件 | 平台用户操作被记录，系统根据操作进行信息变更，并由前端显示 | | |
| 基本流 | 1. 系统显示已有的数据资产及其检索界面 2. 平台用户查看某数据资产目录的详细信息 3. 平台用户进行相关数据资产的检索，系统展示相关数据资产信息 4. 平台用户删除异常或不需要的数据源 | | |
| 备选流 | 2a. 未检索成功数据资产，系统提示相关信息  3a. 数据源导出失败，系统给出提示 | | |

## 数据服务管理

|  |  |  |  |
| --- | --- | --- | --- |
| 用例编号 | UC09 | 用例名称 | 数据服务管理 |
| 描述 | 平台用户进行数据相似度查询、数据地图展示等 | | |
| 执行者 | 平台用户 | | |
| 前置条件 | 平台用户已登录，并进入数据服务页面 | | |
| 后置条件 | 平台用户操作被记录，系统根据操作进行信息变更，并由前端显示 | | |
| 基本流 | 1. 系统显示数据模型已接入的所有数据源 2. 平台用户进行数据的相似度查询 3. 平台用户查看数据地图，系统展示数据地图 4. 平台用户进行大模型交互，系统生成相关处理脚本，并由前端展示，可在系统内运行。 | | |
| 备选流 | 2a. 未查询成功数据相似度，系统提示相关信息  3a. 与大模型交互的脚本生成失败，系统给出提示 | | |

# 非功能需求

## 易用性

### 培训时间

管理人员：本项目要求管理人员能够完成部分专业数据库技术操作，涉及部分较为复杂的功能模块，其培训时间不超过2小时。

操作人员：主要负责使用产品基本功能，其培训时间不应超过1小时。

### 易用性需求

系统应确定统一的用户界面设计规范和友好的人机交互方式，提供明确、美观、友好的前端用户界面，并为用户提供易于操作的相关功能项。

## 可靠性

### 可用性

为了保证软件系统的无故障执行水平、可恢复性及准确性，本项目要求一年内的平均正常运行时间达到95%。

### 平均故障间隔时间（MTBF）

系统的平均故障间隔时间应大于30×24小时。

### 平均修复时间（MTTR）

数据在处理流程中的故障为严重故障，其修复时间不超过2小时；

由前端界面引发的用户交互故障为普通故障，其修复时间不超过1小时。

### 最高错误或缺陷率

根据能力成熟度模型集成（Capability Maturity Model Integration，CMMI），要求千行代码出错率不超过11.95‰（CMMI 1级）。

## 性能

### 响应时间

对于不涉及实时数据处理的一般用户操作，要求95%情况下的响应时间不超过1.5秒（网络不稳定情况下为3秒）。

### 吞吐量

在开发环境的约束下，系统最多可容纳同时在线用户数不超过500个，系统最多可支持上千并发请求。

## 可支持性

### 兼容性

Web前端支持Chrome58、Edge14、Firefox54、safari10、opera58及以上版本的浏览器。前端支持的分辨率：1920\*1080，1440\*900，1366\*768。

可支持的配置数据库包括但不限于Postgresql等关系型数据库以及MongoDB等非关系型数据库。

### 可维护性

以基于UML的面向对象方法作为软件开发方法，以实现整个系统的灵活可扩展与高可维护性。

### 编码标准

前端：ESLint。

后端：Google Java Style。

### 命名约定

前端：Airbnb JavaScript。

## 设计约束

### 软件语言

中文。

# 其它产品需求

## 联机用户文档和联机帮助的需求

用户手册需要提供详细的用户使用帮助说明，包括系统的基本介绍、功能。软件安装界面要求。

## 接口需求

### 用户界面

基于浏览器/服务器架构，用户采用浏览器作为客户端访问功能界面。

### 硬件接口

无。

### 软件接口

Restful API。

### 通信接口

无。

## 适用的标准

本系统遵守《中华人民共和国保密法》、《计算机信息系统国际联网保密管理规定》、《中华人民共和国计算机信息系统安全保护条例》、《计算机信息网络国际联网安全保护管理办法》、《中华人民共和国计算机信息网络国际联网管理暂行规定》及其实施办法等相关法律法规的任何及所有的规定，并对用户以任何方式使用服务的任何行为及其结果承担全部责任。