

菲斯科白泽产品

需求说明书V1.0



2021-10-13

菲斯科上海（软件）有限公司

上海市延安西路728号华敏翰尊国际4楼J座

V1.0

[1 概要需求 8](#_Toc85121266)

[1.1 产品背景 8](#_Toc85121267)

[1.2 功能概览 8](#_Toc85121268)

[1.3 术语和定义 10](#_Toc85121269)

[2 数据接入 10](#_Toc85121270)

[3 数据建模 10](#_Toc85121271)

[3.1 业务域定义 10](#_Toc85121272)

[3.1.1 故事描述 10](#_Toc85121273)

[3.1.2 规则描述 11](#_Toc85121274)

[3.1.3 验收标准 11](#_Toc85121275)

[3.1.4 设计方案 12](#_Toc85121276)

[3.2 数仓建模 12](#_Toc85121277)

[3.2.1 故事描述 12](#_Toc85121278)

[3.2.2 规则描述 12](#_Toc85121279)

[3.2.3 验收标准 13](#_Toc85121280)

[4 数据调度 0](#_Toc85121281)

[4.1 单表调度 0](#_Toc85121282)

[4.1.1 故事描述 0](#_Toc85121283)

[4.1.2 规则描述 0](#_Toc85121284)

[4.1.3 验收标准 0](#_Toc85121285)

[4.2 数据管道 0](#_Toc85121286)

[4.2.1 故事描述 0](#_Toc85121287)

[4.2.2 规则描述 1](#_Toc85121288)

[4.2.3 验收标准 2](#_Toc85121289)

[4.2.4 设计方案 4](#_Toc85121290)

[5 非功能需求 6](#_Toc85121291)

[5.1 登陆验证方式 6](#_Toc85121292)

[5.2 多语言支持 6](#_Toc85121293)

[5.3 浏览器支持 6](#_Toc85121294)

[5.4 安全性 6](#_Toc85121295)

# 概要需求



## 产品背景

随着传统企业商业智能BI近几十年的不断发展，目前数据在企业发展中的重要性越来越凸显出来，同时越来越多的企业将数据作为资产进行管理的需求也越来越迫在眉睫，企业越来越急切的需要一款好的产品能够对企业数据的资产进行有效的管理，协助企业技术人员完成数据的治理工作。

因此，白泽产品设计的理念应运而生：她是一款为企业数据治理，资产管理提供有效方案的数据产品。

## 功能概览

图示

描述已自动生成

## 术语和定义

|  |  |  |
| --- | --- | --- |
| 术语 | 描述 |  |
| BI | 商业智能 |  |
| DW | 数据仓库 |  |
| Data Lake | 数据湖 |  |

# 数据接入

数据接入的核心理念是将企业的数据进行汇聚，我们可以理解为数据湖（Data Lake）的概念；各个业务系统产生的业务数据，在进入数据建模进行加工之前，我们通过数据湖进行收集存放。

## 数据集成首页

### 用户故事

作为数据开发人员：如果想要在平台中能够得到某一个业务系统的相关业务数据，需要能够在当前平台中完成一个上游应用的配置，并且能够直观的看到应用的负责人信息，以便出现应用无法访问时能够第一时间获取相关帮助；

### 规则描述

#### 名词定义

* 应用：接入白泽系统的上游系统，一个系统称为一个应用。
* 非实时应用：数据的时效性不高，一般是通过白泽平台定义同步时间自动从上游系统抽取数据。
* 实时应用：对数据的失效性要求比较高的上游系统，一般通过和上游系统定义好对接接口实现数据通过消息队列的方式主动推送给白泽产品。

### 验收标准

|  |  |  |
| --- | --- | --- |
| **Given（假如）** | **When（当）** | **Then（然后）** |
| 新建应用 | 应用基本信息添加成功 | 能够在页面呈现一个新的应用 |
| 应用数据源配置成功 | 能够提示测试连接通过 |
| 修改应用 | 配置的数据源等信息修改 | 可以重新测试连接通过。 |

### 设计方案

#### 应用注册首页

电脑软件截图

描述已自动生成

交互规则：

* 添加：能够弹出页面维护进行应用的基本信息维护和应用的数据源配置；
* 导出：能够导出选择的应用到Excel中；
* 过滤：能够进行下拉自定义过滤条件。
* 管理视图：实现过滤条件的保存。

#### 应用注册添加页面

待补充

# 数据建模

数据建模的核心理念是将接入数据湖的零零散散的各种数据，通过数据治理的建模方法论，将数据进行重新整合和加工，是数据工厂的核心部分，完成将整理好和清洗好的数据进行仓库存储，并进行提前的预处理和加工生成分析建模，为后续数据消费提供数据服务。

## 业务域定义

### 故事描述

作为企业数据产品负责人：需要站在企业全局的基础上对企业未来的数据按照业务板块的方式进行规划，以便能够将数据湖中零零散散的来源于不同系统的数据按照规划好的业务板块进行数据存储。

### 规则描述

#### 名词定义

* 业务域：根据企业的数据规划，将一类有相关业务含义的数据放在一个整体的数据域内，叫做业务域或者业务板块。

### 验收标准

|  |  |  |
| --- | --- | --- |
| **Given（假如）** | **When（当）** | **Then（然后）** |
| 在页面新建业务域 | 业务域添加成功 | 在业务域页面能够看到该业务域，并看到该业务域的负责人，以及业务域的发布状态（默认为未发布） |
| 当删除业务域 | 业务域删除成功 | 业务域的删除属于风险比较高的操作，需要给予必要的删除提醒；删除之后后台对应的该业务域下的数仓的表（假如有共享维度，则需要把共享维度放到其他业务域下）全部删除，对应的分析模型Doris中的表都要删除，并要提醒用户删除后会影响的调度中的内容，需要用户去手动对调度做出调整； |
| 在业务域下新建了表或者修改了表 | 这些表未发布的时候 | 整个业务域提示待发布状态。 |
| 所有的表都发布了的时候 | 整个业务域提示已发布状态 |

### 设计方案

#### 业务域首页

电脑萤幕截图

描述已自动生成

交互规则：

* 添加：能够弹出页面添加业务域的基础信息，添加完成之后在内容区域多一个新的卡片；
* 导出：能够导出所有的业务域到Excel中；
* 过滤：点击过滤，内容区域向下滑动，展现过滤区域模块，在过滤模块可以自定义过滤条件。
* 管理视图：能够保存当前的过滤条件成一个新的视图，方便下次打开的时候可以选择不同的过滤条件进行查看。
* 内容区域-编辑：可以编辑业务域的基本信息；
* 内容区域-删除：可以删除整个业务域，以及删除该业务域下的所有信息，包含已经创建的数仓的表，doris的表；删除是一个非常危险的动作，要给与客户必要的提醒。
* 内容区域-数仓建模：已发布/待发布状态呈现，当该业务域下的数仓的表只要有一张需要发布，整个状态改为待发布，否则已发布。
* 内容区域-分析建模：已发布/待发布状态呈现，当该业务域下的Doris的表只要有一张需要发布，整个状态改为待发布，否则已发布。

## 数仓建模

### 故事描述

作为数据开发人员：当数据接入模块已经完成数据从上游系统接入到数据湖后，并且也已经规划好了业务域后，我们需要对数据按照数据仓库建设方法论对数据进行数仓建模，以便零散的数据能够按照规定，落地于各个维度表和各个事实表中；

### 规则描述

#### 名词定义

* 数仓建模：创建整个数据仓库的过程即为数仓建模。
* 数据仓库：简称DW，是面向分析创建的数据库，在数据仓库中所有的表大致分为两类：维度表和事实表。
* 维度：维度是业务上的叫法，一个维度后边原则上只能有一张维度表，比如客户维度，在DW中就只有Dim\_Customer 维度表。
* 维度表：维度表是业务上的维度在DW中的一种物理表的反应形式，具体的是一张表。
* 业务过程：业务过程也是业务上的定义，是企业管理中逻辑相关的一组决策和活动的集合。举例：传统企业中的销售出货，HR招聘，财务报销 等都是一个具体的业务过程；
* 事实表：事实表示业务过程在DW中的一种物理表的反应形式，具体也是一张表，一般这样的表会和多个维度表关联，从而从多个维度反应一个局的业务过程，一张事实表一定属于某一个业务过程，一个业务过程可能有多个事实表来反应这个业务。举例：销售出货这个业务过程下，有销售订单事实表，有销售发货事实表，有销售计划事实表，有销售合同事实表，这些事实表同时又需要关联日期维度，产品维度，客户维度等来反应每条数据的具体业务形态和过程。

### 验收标准

|  |  |  |
| --- | --- | --- |
| **Given（假如）** | **When（当）** | **Then（然后）** |
| 新建维度 | 填写维度信息后，选择共享维度 | 在别的业务域下也能看到共享的维度 |
| 在维度下新建维度表的时候 | 只支持创建一个维度表 |
| 新建维度表 | 新建的时候有两种模式1.能够选择从数据接入的某一张表来数据。2，支持自定义SQL脚本拿数据 | 能够保存相关的配置信息到相应的表。 |
| 勾选了从数据接入的表的字段后，或者自定义SQL脚本后，能够自动生成维度属性 | 维度属性可以修改一些描述名称等信息，也可以用默认值直接保存 |
| 新建业务过程 | 填写业务过程的属性信息后 | 能够创建成功业务过程 |
| 在业务过程下创建事实表的时候 | 支持一个业务过程下创建很多事实表 |
| 新建事实表 | 填写事实表基本信息后，并且选择数据来源有两种：1.来源于数据接入的某一张表。2.来源于一个自定义的SQL | 事实表能够创建成功，并且能够自动得到事实表的字段信息。 |
| 添加维度关联的时候，能够填写和维度表关联的字段 | 能够保存这个关联条件，当后台发布的时候能够根据关联字段生成事实表和维度表关联的SQL语句，以便最终生成的事实表中能有相关维度的Key值。 |
| 维度和业务过程的发布 | 点击发布按钮 | 后台会在PG的DW 数据库中生成相应的表结构，以及生成表的处理逻辑的存储过程，同时默认生成一个能够处理该表的NIFI Group |

### 设计方案

#### 数仓模型首页

图形用户界面, 应用程序

描述已自动生成

交互规则：

* 点击数据仓库：整个页面向右滑动进入数据仓库页面；
* 收起操作：默认情况下左右的两个表格的内容都为展开状态，可以针对某一个模块进行收起；
* 维度-新建：可以弹出框维护维度的基本信息；
* 维度-删除：需要选择具体的维度或者维度表进行删除；删除要给与提醒，会同步删除后台相关表。
* 维度-发布：需要选择需要发布的内容，然后进行发布，发布后台会调用相应的逻辑去比较当前数仓中的结构和现在的结构的差异，然后将差异反应到数仓中。

页面所有区域一旦发生字段变化等都将表的状态改为未发布状态（未发布的表表名换个颜色或者用其他方式展现给予明确提醒）

* 添加维度表：点击添加维度表，弹出相应页面进行维度表信息维护。
* 添加字段：待定是否需要。
* 业务过程-新建：可以弹出框创建业务过程基本信息。
* 业务过程-删除：需要选择具体的业务过程或者事实表进行删除，删除要给与提醒，会同步删除后台相关的表。
* 业务过程-发布：需要选择需要发布的内容，然后进行发布，发布后台会调用相应的逻辑去比较当前数仓中的结构和现在的结构的差异，然后将差异反应到数仓中。

页面所有区域一旦发生字段变化等都将表的状态改为未发布状态（未发布的表表名换个颜色或者用其他方式展现给予明确提醒）

* 添加事实表：可以弹出添加事实表的页面：

## 分析建模

### 故事描述

作为数据开发人员：当我们完成数仓建模，实现了将数据湖中零散的数据按照维度和事实表的方式进行归类和整理之后，数据就进入一个可以分析使用的阶段；为了提高数据消费端进行数据消费时的性能，我们在分析建模阶段对经常使用的指标进行预先定义，以便系统能够预先按照我们的定义进行数据聚合计算，从而将聚合好的结果集推送到适合分析使用的Doris数据库中，整个分析建模就是预定义指标，系统根据预定义指标进行聚合存储的过程。

### 规则描述

#### 名词定义

* 分析建模：预定义原子指标，派生指标，根据预定义，在OLAP数据库（Doris）中创建表的整个过程即为分析建模。
* 原子指标：基于DW 中创建的事实表上的度量值，通过SUM，AVG，DISTINCT ，COUNT 等聚合公式计算的指标。
* 派生指标：基于原子指标通过一系列的公式或者一些限定条件生成的指标。
* 业务限定：作为派生指标的输入条件，是对某一张事实表增加的筛选条件的组合。
* 时间周期：作为派生指标的输入条件，时间周期聚合是实现某一个时间段的原子指标数据之和。比如YTD，年初到现在这样一个时间范围。

### 验收标准

|  |  |  |
| --- | --- | --- |
| **Given（假如）** | **When（当）** | **Then（然后）** |
| 新建原子指标 | 选择了聚合公式，度量字段 | 成功保存该原子指标 |
| 新建派生指标 | 普通指标公式：选择业务限定（非必须），原子指标（必须），时间周期（非必须）， | 能够成功保存该派生指标，并且后续能够在新建派生指标的时候继续引用这个派生指标 |
|  | 自定义公司：可以选择当前业务过程下的所有指标（原子指标和派生指标）进行各类加减乘除的公式计算 | 能够成功保存该指标 |
| 新建业务限定 | 业务限定必须选择在某一张事实表下新建，并且一个业务限定支持多个字段的限定。 | 业务限定能够成功保存，并且可以应用到该事实表上的所有派生指标上。 |
| 发布 | 发布可以只发布某一张维度表和某一个业务过程或者事实表 | 发布操作会在Doris数据库删除该表，并且同时在创建该表，对应的NIFI流程重新生成。 |
| 选择某一个事实表发布时 | 在Doris中生成的事实表需要把配置中的该事实表中的所有原子指标和维度Key放在一起在Doris中创建，只不过每个原子指标的聚合方式按照配置中配置的生成；同时将上述过程生成一个插入Doris的SQL语句进行保存。 |

### 设计方案

#### 分析模型首页

图形用户界面, 应用程序, Teams

描述已自动生成

交互规则：

* 分析建模：点击分析建模，页面向左滑动，进入分析建模页面；
* 收起/展开：默认全部展开，点击收起可以对该模块收起。
* 添加事实表：暂时禁用，待定
* 添加指标：弹出页面选择是创建原子指标还是派生指标，以及基本信息保存，指标会在事实表下边展现；且原子指标和派生指标前边的Icon不同。
* 指标删除：可以删除某一个对应的指标。
* 指标选中：选中指标后右下方区域自动切换到该指标的常规属性，以及指标公式上，且右侧上方区域自动定位对应到该指标所在事实表已经维护的业务限定条件和时间周期。
* 业务限定-添加：点击添加按钮弹出窗口输入业务限定的名称和描述。
* 业务限定-选中：选中之后可以在该业务限定右侧区域看到名称描述，以及限定逻辑，同时可以在此维护限定逻辑。
* 业务限定-逻辑-添加：在该表格区域自动增加一行，可以维护。
* 业务限定-逻辑-保存：对该逻辑整体进行保存。
* 常规：在左侧指标选中之后，常规区域可以对该指标的基础信息进行维护。
* 公式：在左侧指标选中之后，可以切换到该指标的公式页面。

#### 分析模型公式页面

图形用户界面, 应用程序, Teams

描述已自动生成

交互规则：

* 普通：在常规页面可以将鼠标移到时间周期或者业务限定，或者指标处分别拖动到公式的不同位置。
* 自定义：在自定义界面实现一个大的文本框中通过@符号自动获取指标名称的方式完成指标的加减乘除。
* 保存：能够保存当前的公式或者常规配置。
* 清空：清空当前的常规配置或者公式配置。

# 数据调度

数据调度功能主要为IT管理员实现整个平台的所有数据流的调度和配置，

## 单表调度

### 故事描述

作为数据开发人员：当平台上已经完成了有关数据接入，数据建模等的表配置之后，我们需要能够实现在平台中针对每张表设置表的同步时间和同步频率，以便能够实现单表按照设置的同步频率和时间完成数据同步；

### 规则描述

### 验收标准

## 数据管道

### 故事描述

作为数据开发人员：当我们完成了数据接入，数据建模等数据开发操作之后，需要能够将接入的表的处理，数据仓库中表数据的处理，分析建模中表数据的处理通过一定的流程进行串联，才能满足当上游系统数据更新之后，系统能够按照设定的流程从上游系统直接流入到最终的分析建模中，以便报表访问人员能够在报表中直观的看到数据的变化；

整个数据流程确保了数据的流的连续性；

### 规则描述

#### 名词定义

* 数据管道：定义整个数据流程按照一定的数据流进行处理的整个过程称为数据管道；
* 管道活动：每一个活动代表这一种任务类型；

|  |  |  |
| --- | --- | --- |
| **活动名称** | **实现功能** | **备注** |
| 调度活动 | 这个活动只是用来设置时间调度的，只能在这个活动上设置调度时间 | 在Nifi中是一个空的Processor |
| 任务组活动 | 这个活动是用来解决并行等问题的，如果使用了任务组，则任务组支持打开，在任务组中继续添加其他的活动 | 对应在NIFI中是一个Group |
| 数据湖表活动 | 当用户使用数据湖表任务这个活动后，需要在数据湖任务上配置数据湖中对应的表，从而实现对该表的数据处理 | 在NIFI中对应生成一组处理数据湖单表的Porcessor，可以放在一个大的Group中 |
| 数仓维度表活动 | 当用户使用数仓维度表任务这个活动后，需要在该任务上配置数仓维度表，从而实现对该数据仓库表的处理。 | 在NIFI中生成的是一个Processor，处理该表 |
| 数仓事实表活动 | 当用户使用该活动，并配置该活动指定某一个事实表，从而实现对数仓事实表的处理 | 在NIFI中生成的是一个Processor，处理该表 |
| 分析模型维度活动 | 当用户使用该活动，并配置分析模型中指定的维度表，从而实现对分析模型维度表的处理 | 在NIFI中生成的是一个Processor，处理该表 |
| 分析模型事实活动 | 当用户使用该活动，并配置分析模型中指定的维度表，从而实现对分析模型事实表的处理 | 在NIFI中生成的是一个Processor，处理该表 |

### 验收标准

|  |  |  |
| --- | --- | --- |
| **Given（假如）** | **When（当）** | **Then（然后）** |
| 页面添加管道并为管道设置活动 | 管道添加成功，并且设置完成每个管道内的活动之后，并发布管道后 | 再NIFI中能够生成一个新的Nifi流程，并且管道中设置的各个节点都能够成功的反应到Nifi中 |
| 页面删除管道 | 当管道删除成功，并发布成功后 | NIFI中的整个管道对应的流程需要全部删除 |
| 在管道中的对调度任务的活动设置调度时间 | 当设置时间成功后，并发布成功后 | NIFI流程能够按照设置的时间在这个调度任务之后的所有活动自动按照顺序完成整个数据流程的同步 |
| 在管道中删除活动 | 活动删除成功，并发布成功后 | 该活动后续的活动都将删除，nifi中也将删除 |

### 设计方案

#### 管道首页

图形用户界面, 应用程序

描述已自动生成

交互规则：

* 添加：点击添加按钮能够弹出添加管道页面，完成管道基本信息的表单维护；
* 导出：能够将页面的管道信息导出到Excel中；
* 过滤：点击过滤能够下拉出过滤的筛选条件界面，这里可以自定义过滤内容；
* 管道区域：每个模块展现管道的名称，负责人，发布状态等信息；
* 管道编辑：点击每个管道右上角的编辑按钮能够实现管道基本信息的编辑；
* 管道删除：点击删除按钮能够完成管道删除，管道删除后，后台对应的Nifi流程也要删除；
* 管道进入：进入分为两个视图，进入可视化视图和进入表格视图：

#### 管道表格视图

图形用户界面, 表格

描述已自动生成

交互规则：

* 添加：点击添加，能够弹出添加界面完成下表中信息的维护；
* 编辑：能够弹出编辑页面完成下表中信息的维护；
* 删除：删除时请判断如果当前活动是一个节点的输入的时候，删除的时候要给与提醒，当前活动删除，该活动之后的活动将都被删除；
* 一般任务组：如果活动类型是一般任务组，则需要在该活动名称增加下划线连接的方式，点击连接能够进入这个一般任务组中，同时在上边的页面当前位置上增增加一个该活动名称，方便用户在回退到上一个页面；在该任务组页面中依然展现的是类似的这样一张表格。

#### 管道可视化视图

待补充

# 数据治理

# 非功能需求

## 登陆验证方式

白泽产品采用form表单验证方式，管理员提供账号密码，用户输入账号密码登录即可。

## 多语言支持

1.0版本只支持中文

## 浏览器支持

Windows系统的IE8以上浏览器，谷歌浏览器，Microsoft Edge浏览器等

## 安全性

当服务器因为某些不确定因素受到破坏时，可以根据以前备份的数据库和程序进行简单的恢复即可运行。

数据在网络传输过程中采用https协议进行加密。