



(12) 发明专利

(10) 授权公告号 CN 112597207 B

(45) 授权公告日 2022.09.23

(21) 申请号 202011591524.4

G06F 16/28 (2019.01)

(22) 申请日 2020.12.29

G06F 16/215 (2019.01)

(65) 同一申请的已公布的文献号  
申请公布号 CN 112597207 A

(56) 对比文件

CN 111198902 A, 2020.05.26

CN 101515290 A, 2009.08.26

(43) 申请公布日 2021.04.02

US 2020334272 A1, 2020.10.22

(73) 专利权人 科技谷(厦门)信息技术有限公司  
地址 361000 福建省厦门市湖里区安岭二  
路88号A栋404室

彭辛庚, 陈湘涛. 电信企业数据仓库中元数  
据管理的探索与实践.《电信科学》.2009,

彭辛庚, 陈湘涛. 电信企业数据仓库中元数  
据管理的探索与实践.《电信科学》.2009,

(72) 发明人 陈思恩

审查员 林丽香

(74) 专利代理机构 厦门致群财富专利代理事务  
所(普通合伙) 35224  
专利代理师 刘兆庆 邓贵琴

(51) Int. Cl.

G06F 16/2457 (2019.01)

G06F 16/25 (2019.01)

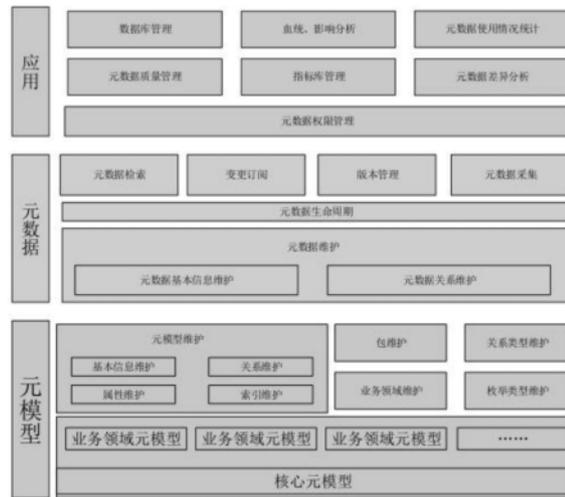
权利要求书1页 说明书3页 附图1页

(54) 发明名称

一种元数据管理系统

(57) 摘要

本发明公开了一种元数据管理系统,包括:应用与分析模块,用于对元数据进行应用和分析,其包括数据库管理子模块、血统或影响分析子模块、元数据使用情况统计子模块及元数据管理子模块;元数据模块,用于对元数据进行维护,其包括元数据检索子模块、变更订阅子模块、版本管理子模块、元数据采集子模块、元数据生命周期子模块及元数据维护子模块;元模型模块,用于对元模型进行维护,其包括元模型维护子模块、包维护子模块、关系类型维护子模块、业务领域维护子模块及枚举类型维护子模块。本发明提供一种元数据管理系统,降低元数据使用难度、提升用户体验,使大数据平台各类用户均能参与到元数据运营维护当中。



1. 一种元数据管理系统,其特征在于,包括:

应用与分析模块,用于对元数据进行应用和分析,其包括数据库管理子模块、血统或影响分析子模块、元数据使用情况统计子模块及元数据管理子模块;

元数据模块,用于对元数据进行维护,其包括元数据检索子模块、变更订阅子模块、版本管理子模块、元数据采集子模块、元数据生命周期子模块及元数据维护子模块;

元模型模块,用于对元模型进行维护,其包括元模型维护子模块、包维护子模块、关系类型维护子模块、业务领域维护子模块及枚举类型维护子模块;

所述元数据模块还包括元数据稽核子模块,用于对元数据采集子模块采集的数据进行稽核,其过程包括:

S1、制定预定义数据稽核和校验规则;

S2、通过ETL 工具对采集到数据做初步的数据清理和预处理;

S3、对数据进行差异分析和异常分析;

S4、对异常数据发出实时预警和通知,且生成数据稽核统计报表。

2. 如权利要求1所述的一种元数据管理系统,其特征在于:所述元数据管理子模块包括元数据质量管理、指标库管理、元数据差异分析及元数据权限管理。

3. 如权利要求1所述的一种元数据管理系统,其特征在于:所述元数据维护子模块包括元数据基本信息维护和元数据关系维护。

4. 如权利要求1所述的一种元数据管理系统,其特征在于:所述元模型维护子模块包括元模型基本信息维护、元模型关系维护、元模型属性维护及元模型索引维护。

5. 如权利要求1所述的一种元数据管理系统,其特征在于:所述元数据采集子模块包括Web 请求组件、RDBMS 阅读器、HTTP 请求组件、HTTP 响应组件、文件轮询器、文本文件读取器、解压缩组件、二进制文件读取器及XML读取器组件。

6. 如权利要求1所述的一种元数据管理系统,其特征在于:所述变更订阅子模块包括元模型变更订阅和元数据变更订阅。

## 一种元数据管理系统

### 技术领域

[0001] 本发明涉及数据管理技术领域,具体涉及一种元数据管理系统。

### 背景技术

[0002] 元数据是关于数据的数据,元数据主要用于在数据管控平台里描述数据及其环境,它是在数据管控建设过程中所产生的有关数据源定义,目标定义,转换规则等相关的关键数据。因此,元数据的管理对于数据管理平台有着重要的意义。现有元数据使用难度大、管控力度小,提升用户体验差。

### 发明内容

[0003] 本发明提出一种元数据管理系统,降低元数据使用难度、提升用户体验,使大数据平台各类用户均能参与到元数据运营维护当中。

[0004] 为实现上述目的,本发明采用以下技术方案:

[0005] 一种元数据管理系统,包括:

[0006] 应用与分析模块,用于对元数据进行应用和分析,其包括数据库管理子模块、血统或影响分析子模块、元数据使用情况统计子模块及元数据管理子模块;

[0007] 元数据模块,用于对元数据进行维护,其包括元数据检索子模块、变更订阅子模块、版本管理子模块、元数据采集子模块、元数据生命周期子模块及元数据维护子模块;

[0008] 元模型模块,用于对元模型进行维护,其包括元模型维护子模块、包维护子模块、关系类型维护子模块、业务领域维护子模块及枚举类型维护子模块。

[0009] 优选地,所述元数据管理子模块包括元数据质量管理、指标库管理、元数据差异分析及元数据权限管理。

[0010] 优选地,所述元数据维护子模块包括元数据基本信息维护和元数据关系维护。

[0011] 优选地,所述元模型维护子模块包括元模型基本信息维护、元模型关系维护、元模型属性维护及元模型索引维护。

[0012] 优选地,所述元数据采集子模块包括Web请求组件、RDBMS阅读器、HTTP请求组件、HTTP响应组件、文件轮询器、文本文件读取器、解压缩组件、二进制文件读取器及XML读取器组件。

[0013] 优选地,所述元数据模块还包括元数据稽核子模块,用于对元数据采集子模块采集的数据进行稽核,其过程包括:

[0014] S1、制定预定义数据稽核和校验规则;

[0015] S2、通过ETL工具对采集到数据做初步的数据清理和预处理;

[0016] S3、对数据进行差异分析和异常分析;

[0017] S4、对异常数据发出实时预警和通知,且生成数据稽核统计报表。

[0018] 优选地,所述变更订阅子模块包括元模型变更订阅和元数据变更订阅。

[0019] 采用上述技术方案后,本发明与背景技术相比,具有如下优点:

[0020] 本发明提供一种元数据管理系统,各个数据实体来进行实体定义,对实体之间的关系进行约束,实现对数据管理平台里所有数据实体的管理;通过制定规范的数据流程制度,严格控制数据的流向以及数据在各流程中的情况,对数据管理平台中的各项数据流程进行管控;降低元数据使用难度、提升用户体验,使大数据平台各类用户均能参与到元数据运营维护当中。

## 附图说明

[0021] 图1为本发明系统框图。

## 具体实施方式

[0022] 为了使本发明的目的、技术方案及优点更加清楚明白,以下结合附图及实施例,对本发明进行进一步详细说明。应当理解,此处所描述的具体实施例仅仅用以解释本发明,并不用于限定本发明。

[0023] 在本发明中需要说明的是,术语“上”“下”“左”“右”“竖直”“水平”“内”“外”等均为基于附图所示的方位或位置关系,仅仅是为了便于描述本发明和简化描述,而不是指示或暗示本发明的装置或元件必须具有特定的方位,因此不能理解为对本发明的限制。

### [0024] 实施例

[0025] 配合图1所示,本发明公开了一种元数据管理系统,包括:

[0026] 应用与分析模块,用于对元数据进行应用和分析,其包括数据库管理子模块、血统或影响分析子模块、元数据使用情况统计子模块及元数据管理子模块;

[0027] 元数据模块,用于对元数据进行维护,其包括元数据检索子模块、变更订阅子模块、版本管理子模块、元数据采集子模块、元数据生命周期子模块及元数据维护子模块;

[0028] 元模型模块,用于对元模型进行维护,其包括元模型维护子模块、包维护子模块、关系类型维护子模块、业务领域维护子模块及枚举类型维护子模块。

[0029] 元数据管理子模块包括元数据质量管理、指标库管理、元数据差异分析及元数据权限管理。

[0030] 元数据维护子模块包括元数据基本信息维护和元数据关系维护。

[0031] 元模型维护子模块包括元模型基本信息维护、元模型关系维护、元模型属性维护及元模型索引维护。

[0032] 元数据采集子模块支持用户采集或导入元数据和元模型,采集或导入的方式包括自动(或订阅)和手动,支持包括数据库、大数据工具、报表工具等的采集;元数据采集子模块包括Web请求组件、RDBMS阅读器、HTTP请求组件、HTTP响应组件、文件轮询器、文本文件读取器、解压缩组件、二进制文件读取器及XML读取器组件。

[0033] Web请求组件可用于调用WSDL或基于REST的服务,RDBMS阅读器用于通过指定的SQL查询语句从SQL兼容的关系型数据库中读取数据,HTTP请求组件用于定义由服务器托管的RESTful服务,它将接受预期的HTTP请求,HTTP响应组件用于定义由服务器托管的RESTful服务,它将按预期输出HTTP响应,文件轮询器用于查看基于文件的目录结构,并且如果存在与给定文件名称表达式相匹配的单个或多个文件,则向后继组件发送消息。文件轮询器发送的消息的有效载荷内有字符串数组列表,其中的每个文件的文件名与文件轮询

器的文件模式表达式参数匹配,文本文件读取器用于从一个或多个基于文本的文件读取数据,解压缩组件允许提取压缩的归档/压缩文件,以便可以在流程中处理其内容,二进制文件读取器可以读取一个或多个文件,并将读取的数据传递给后继组件,XML读取器组件允许从文件中读取XML数据,并可从标签中的数据作为单独的消息发送。

[0034] 元数据模块还包括元数据稽核子模块,用于对元数据采集子模块采集的数据进行稽核,其过程包括:

[0035] S1、制定预定义数据稽核和校验规则;

[0036] S2、通过ETL工具对采集到数据做初步的数据清理和预处理;

[0037] S3、对数据进行差异分析和异常分析;

[0038] S4、对异常数据发出实时预警和通知,且生成数据稽核统计报表。

[0039] 变更订阅子模块包括元模型变更订阅和元数据变更订阅,二者均采用EMAIL的形式发送给用户,且订阅元数据变更订阅的同时可以订阅元数据的下级节点。

[0040] 血统分析子模块表现为:当用户配置了同步任务,并通过SQL任务进行多个步骤的清洗、转化处理之后最终会将结果数据输出,在整个处理链路中,数据的血缘关系就隐含在同步任务和SQL代码中,数据血缘表示每个统计指标是如何从原始数据得到的过程。

[0041] 影响分析子模块表现为:动分析并展示该操作对数据产生的影响,提供图形可视化的影响分析功能,如接口文件、库表、SQL程序、报表、指标对其他数据或程序的影响。

[0042] 以上所述,仅为本发明较佳的具体实施方式,但本发明的保护范围并不局限于此,任何熟悉本技术领域的技术人员在本发明揭露的技术范围内,可轻易想到的变化或替换,都应涵盖在本发明的保护范围之内。因此,本发明的保护范围应该以权利要求的保护范围为准。

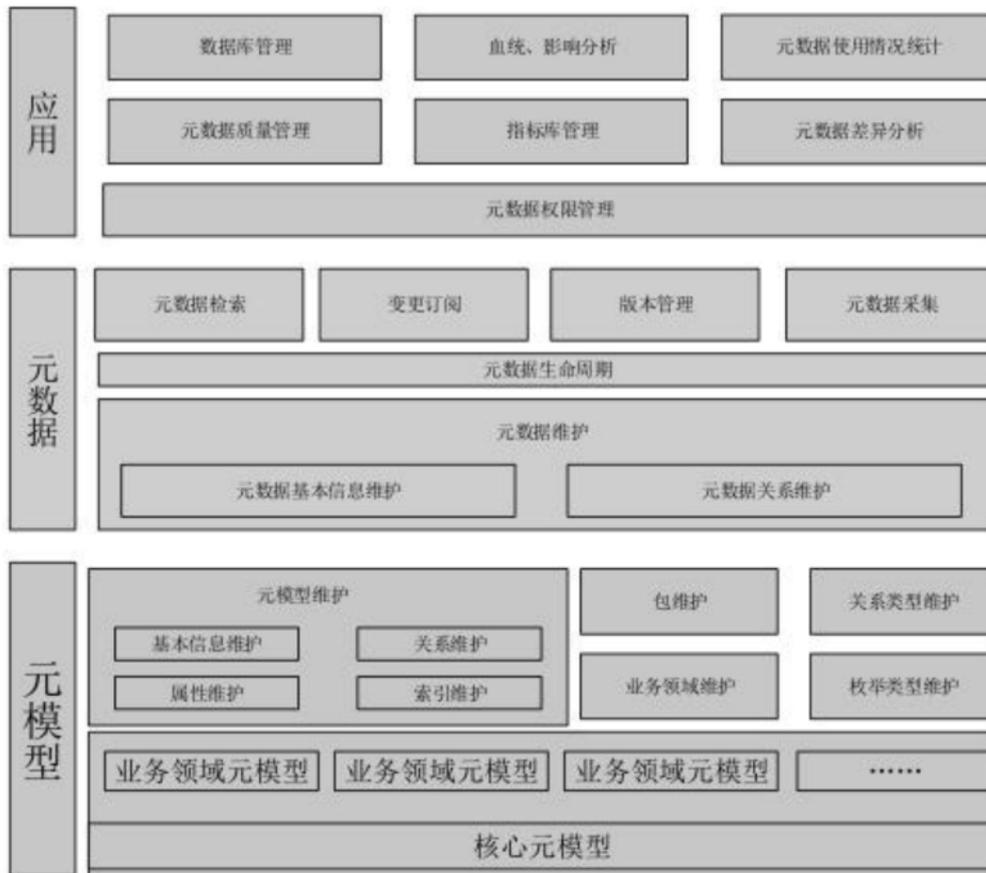


图1