
DataOps

大数据平台

产品白皮书

目录

目录	1
1. 前言	3
2. 产品简介	3
2.1 产品定位	3
2.2 产品价值	4
2.3 产品核心优势	4
2.4 产品核心价值	5
3. 产品架构	6
3.1 功能架构	6
3.2 技术架构	8
4. 产品核心功能	10
4.1 广泛的数据源支持	10
4.1.1 丰富的数据库类型接入	10
4.1.2 支持导入各种类型文件数据	11
4.1.3 支持接口数据源	11
4.2 强大的数据整合	11
4.2.1 可视化的 ETL 过程设计	12
4.2.2 图形化的 ETL 过程流设计	14
4.2.3 可靠的 ETL 过程、过程流发布	16
4.2.4 智能的调度管理	16
4.3 灵活的数据服务管理	17
4.3.1 数据建模	17
4.3.2 数据补录	21
4.3.3 数据服务	22
4.4 一站式数据门户	23
4.5 权限管理	24
4.5.1 功能权限	25
4.5.2 数据权限	25
4.5.3 接口权限	25
4.5.4 开发权限	25
4.6 监控管理	25
4.6.1 调度日志	25

4.6.2	运维监控	26
4.6.3	微服务监控	26
5.	产品设计原则	26
5.1	先进性	26
5.2	可维护性	27
5.3	可靠性	27
5.4	易用性	27
5.5	安全性	28
5.6	扩展性	28
6.	产品部署要求	29
6.1	部署方式	29
6.2	系统要求	29
6.3	浏览器要求	29
6.4	配置要求	29

1. 前言

本白皮书对应产品版本为：DataOps 平台 QDataOps V3..5

最后修订日期：2022 年 6 月

本白皮书将在阐述数据中台、DataOps 及商业智能技术的基础上,详细介绍 DataOps 平台 QDataOpsV3..5 在功能、技术等方面的特点和优势。

2. 产品简介

2.1 产品定位

DataOps 平台是集数据开发、服务、运维、管理为一体的一站式数据开发管理平台软件产品，既可作为数据中台项目实施人员的开发平台，也可作为数据用户、业务人员的一站式数据应用平台。

该产品是我公司在多年数据仓库、数据建模、数据分析的技术经验基础上，运用先进的云原生技术架构，基于微服务软件架构、容器化部署方式，采用湖仓一体存储架构、批量数据和流数据集成的处理模式，遵循 DataOps 的核心理念，经过多年潜心研发而推出的数据开发管理平台产品软件。本产品可最大限度降低数据分析应用的实施门槛，使复杂的工作简单化、便捷化，重复的工作自动化、智能化。

DataOps 平台具备如下特性：

- 部署灵活：微服务架构，容器化部署，满足各类部署要求
- 升级便利：基于 Web 应用，免客户端安装
- 配置方便：集中配置，便捷发布，快速部署
- 功能强大：适配多种主流数据源，支持批流数据处理，集成多种 BI 产品
- 易于使用：低代码、图形化开发，降低用户技术门槛
- 标准开放：提供标准数据接口，快捷提供数据服务

- 性价比高：加速平台搭建速度，提高运维管理效率，降低整体拥有成本

2.2 产品价值

DataOps 平台拥有从丰富的数据源接入、全方位处理数据、多样的数据模型建立、灵活的数据门户设计与数据服务发布，为用户提供一站式数据开发管理，产品价值主要体现在如下几方面：

- 强大的数据集成能力：覆盖主流 RDB 及 NoSql、文件、日志等数据源；
- 灵活的数据建模能力：基于数据对象及数据字典的数据虚拟化方法，提升建模效率；
- 高效的 ETL/ELT 工具：支持自底向上建模，提供数据集市和数仓建模；
- 专有的数字虚拟化技术：采用自主知识产权的数据对象技术实现企业级数据资产建模；
- 易用的数据质量管理：通过数据质量规则及血缘分析提高数据质量，保障数据规范性；
- 低代码化的数据处理：支持低代码方式任务编制及分布式 workflow 任务调度；
- 集成多种 BI 分析工具：为业务用户提供统一数据分析入口，快速构建大数据展示平台，提供炫酷的数据可视化实现，从海量业务数据中挖掘业务知识。

2.3 产品核心优势

(1) 遵循 DataOps 持续集成、持续部署、持续交付的理念，为企业提供基于统一标准和规范的信息化共享服务，提供多个业务环节和系统的共用组件，将以独立的或组合的微服务模式，致力于帮助企业快速搭建数据基础平台，及时通过数据发现问题进而改进业务工作中的数据分析功能。具备操作简单、部署灵活、快速响应等特点。可广泛应用于各行各业，从百亿级数据量的企业到各垂直中小企业，专注为企业提供数字化运

营基础平台，解决各行业的业务数据分析需求。

(2) DataOps 平台提供丰富的数据源接入、数据处理整合、图形化的数据建模、丰富的数据分析方式，可进行一站式数据采集、数据处理、数据分析。帮助企业搭建功能完整的数据服务平台，提供了从系统管理、数据权限等基础功能，到企业多个业务系统的数据建模，任务调度，报表管理，数据服务于一体的完整解决方案。

(3) DataOps 产品主要的功能亮点体现在：为企业提供基于统一标准和规范的信息化共享服务，提供多个业务环节和系统的共用组件，这些组件将以独立的或组合的微服务模式构成，组件包括但不限于：

- 基于组织、角色、用户、权限四大维度的权限管理中心
- 包含数据集成、加工、建模、服务为一体的开发中心
- 包含调度管理、任务调度组件的数据调度中心
- 基于任务管理、实时任务、模板管理的数据集成管理组件
- 第三方的报表、BI 集成组件等

2.4 产品核心价值

(1) 消除数据孤岛，整合业务数据

众所周知，我国的企业信息化建设在近几年逐渐步入正轨，众多企业各个业务系统的信息化建设、IT 规划发展至今，基本搭建完成。在企业当下的信息化建设当中，最大的问题是众多已有的业务系统相对独立，导致数据孤岛长期存在。随着大数据时代的到来，数据来源变得多样化，数据结构也更加复杂化，这使得企业在数据采集、ETL 等方面会耗费很大精力。业务数据的过度分散，让企业很难从全局角度去分析业务的发展情况。针对企业数据孤岛问题，DataOps 提供了完善的解决方案。基于自主研发的数据连接器框架，DataOps 可以对接企业内各个业务系统，包括 ERP、CRM、财务系统、

日志系统等，帮助企业整合所有业务数据，通过 DataOps 系统进行统一管理、认证，在 DataOps 平台上完成众多业务系统报表的共享、统一展现。

(2) 数据资产管理，掌控企业数据脉络

企业数据资产，包括了数据模型、数据字典、数据对象，以及针对这些数据资产的转换处理能力，包括数据服务共享等。DataOps 平台摒除了传统数据管理模式中，复杂，难以理解的数据管理方法，改用业务视角的数据管理模式，从数据模型的增删改查，到数据集的转化处理、业务统建数据服务共享的管理，都让数据管理的复杂度进一步降低。可以使得数据管理人员，甚至业务人员能更好的理解数据。DataOps 平台还让数据血缘关系也得到了重新的设计和梳理，使得业务人员通过简单直接的页面即可方便的查看数据的处理演变过程，为数据业务化提供了进一步的支持。

3. 产品架构

3.1 功能架构



(1) 支持结构化数据、非结构化数据

- 主流的关系型数据库，如：Oracle、Mysql、Sql Server、DB2、HANA 等；

- 分布式数据库，包括： GreenPlum、Hive、SparkSQL、 postgresql 等；
- 数据文件，如： EXCEL、TXT、CSV 等；
- 支持 API 接口。

(2) 数据采集/整合

平台支持各类数据采集接口的接入，可以通过 dataX\kettle\Python shell\sqoop 全量/非实时接入；支持数据增量/实时接入；支持对数据进行数据抽取、数据清洗、数据转换、数据调度、运行监控等。

(3) 数据服务平台

数据工具：平台拥有多个基础组件，贯穿整个系统，包括集成调度、业务建模、数据质量、身份认证、权限管理、元数据管理、运维监控等。

数据模型：平台提供不同方式的主题表创建方式，根据业务性质形诸如财务中心、成本中心、采购中心、销售中心、人力中心、库存中心等多个主题中心，并设置主题之间的关联联系。

数据应用场景：通过对多个业务主题进行数据建模、关联分析，为企业量身打造智能补货、生产调度、智能制造、工艺优化等数据应用场景。

(4) 分析展现

提供各种图表的设计展现，如：固定报表、仪表盘、大屏、看板等，并支持系统管理辅助功能等。

(5) 第三方应用

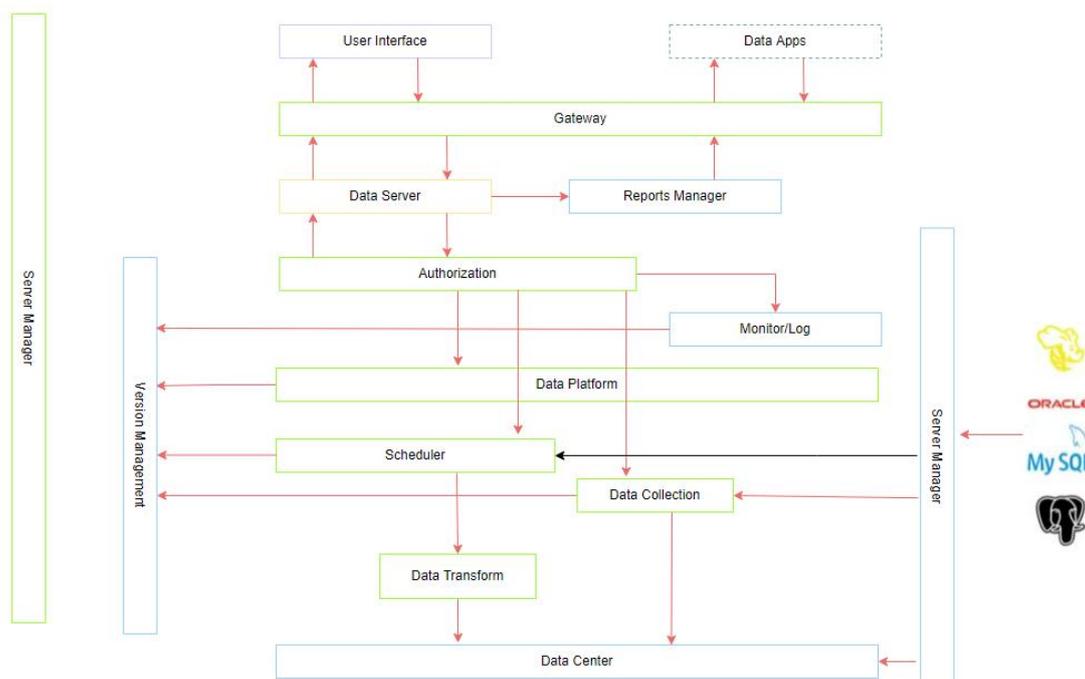
提供丰富的第三方开发接口用于实现项目实施过程中的各种集成需求，包括移动端接口，微信接口等。

(6) 客户端

支持 PC 端、手机、Pad; 支持各种浏览器, 如 IE11 以上、Chrome 等。

3.2 技术架构

DataOps 整体基于 B/S 架构, 业务人员可以通过浏览器, 操控企业数据资产, 并对其进行管理, 加工和输出, 具体架构如下图所示:



DataOps 主要分为 14 个模块, 采用标注的微服务架构, 每个模块都相对独立, 可以单独进行维护升级。同时也可以根据客户情况, 进行选择性的功能模块部署。使整个项目更具灵活性。DataOps 的所有模块都支持集群化部署, 可以动态的根据客户现场的需要, 灵活地增加或减少部分模块的部署数量, 以应付发杂的使用环境。

- **Server Manger 模块:** 用于微服务的注册和管理, 所有的系统服务均需要注册至服务中心, 并通过服务中心进行发现和调用。可以快速获取服务的状态信息, 手动上线/下线需要的服务。
- **User Interface 模块:** DataOps 产品采用前后端分离的模式进行开发, User Interface 属于前端应用模块, 通过 API 服务和后端服务进行通信。
- **Gateway 模块:** DataOps 产品的网关模块, 用于处理前台或其他外部系统的

请求。并通过路由到对应的模块进行功能处理。

- Data Service 模块：数据服务模块。将需要的数据快速进行输入、输出，同时提供对应用户操作 API 接口，对接前台以及其他系统的请求。
- Reports Manger 模块：报表套件管理模块。DataOps 系统可集成 BO, Tableau, 帆软 report、帆软 BI, 永洪, 观远等多种报表套件, Reports Manger 模块能实现对应报表套件的嵌入，单点以及展示。
- Authorization 模块：DataOps 产品拥有一套完善的权限验证系统。支持基于授权码模式，简化模式，密码模式和客户端模式的四种认证模式。同行可以限制单独接口的访问，页面的展示权限，数据的访问权限等。
- Monitor/Log 模块：监控及日志模块能记录用户的操作记录，调度的运行日志。同时支持在服务器宕机，调度任务执行结束后向用户发送邮件。
- Data Platform 模块：数据平台模块，为用户提供了方便快捷的功能操作，允许用户进行中台内数据资产的管理，包括数据模型，数据目录的增删改查，复制，继承 等操作。
- Scheduler 模块：调度模块，能快速方便让用户配置复杂的工作流，支持如 shell, Python, DataX, Kettle 等多种 workflow 节点的配置。
- Data Transform 模块：数据转换处理引擎，可以对数据中台内的数据集进行加工处理，提供标准的数据加工(ETL)算子，可以平滑兼容企业原有的 ETL 过程。
- Data Collection 模块：数据采集模块可以帮助用户将大量的手工数据录入数据中台，同时也支持对已有数据的展示和修改，帮助企业快速搭建数据中台，方便业务人员快速上手操作。

- Data Connector 模块：数据连接器模块，用户可以通过此模块，导入企业内各类异构数据源，包括结构化数据库，NoSQL 数据库，数据文件，API 等数据来源。
- Data Center 模块：DataOps 自有的数据存储引擎，采用先进的 MPP + 内存架构，支持多节点集群部署，可以满足企业大规模数据存储以及快速数据分析响应的要求。
- Version Management 模块：对 DataOps 产品版本管理，方便用户对 DataOps 产品以及数据的版本控制。

4. 产品核心功能

4.1 广泛的数据源支持

4.1.1 丰富的数据库类型接入



平台提供了丰富的数据库类型接入，包括主流关系型数据库以及分布式数据库，能将业务系统数据导入到平台中，提供后续建模分析使用。数据库类型包括：Oracle、Mysql、Sql Server、GreenPlum、Hive、Doris、文本、kafka 等，同时也支持通过接入其他数据库驱动的方式自定义数据库。

4.1.2 支持导入各种类型文件数据

DataOps 平台支持导入以文件形式存储的数据，方便用户提取数据进行分析，支持的类型包括 excel、txt、csv 等文件。

4.1.3 支持接口数据源

互联网上提供了很多对外公开的接口数据，为了方便用户快速提取互联网数据，DataOps 平台支持接口数据源，后续可通过接口获取的数据，将数据转化成规整的格式。

4.2 强大的数据整合能力

(1) 数据集成

在当前信息化服务模式下，越来越多的企业意识到，数据价值的创造往往需要基于正确、一致、完整、结构化的数据，低质量数据将带来运算复杂、数据服务结构偏差等商业风险。然而，在传统 ETL 模式下的数据集成方案下，未能很好的解决“干净”、“准确”的数据基础问题，传统的 ETL+数据仓库的模式，难以实现数据清洗过程的可视化、交付式应用等问题。同时还需要大量的程序编程、SQL 代码，工作量大，效率不高，导致很多企业在数据化应用服务面前犹豫不决。

DataOps 平台提供的强大的数据整合功能可以很好地解决这一痛点。平台提供关系型数据库、大数据存储、半结构化存储、NoSQL 等异构数据源之间的数据共享、数据分发以及数据同步。数据共享包括全量、增量采集模式以及流式数据采集。数据同步的主要的功能有 DataX 和 Sqoop 的同步与批量同步、文本同步以及 Excel 同步。

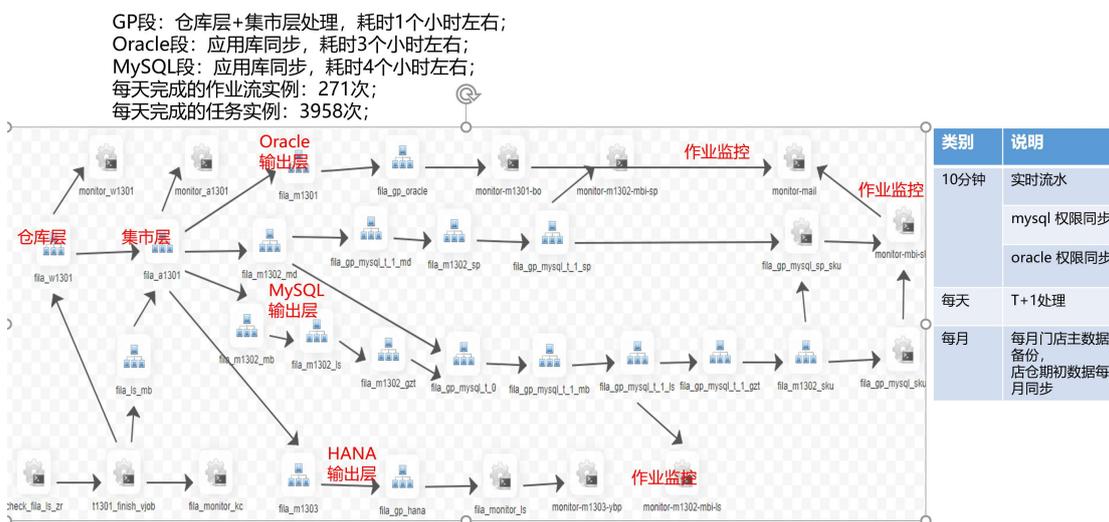
(2) 任务管理

DataOps 将各种数据服务进行任务化的管理，以统一调度的方式，满足数据服务的任务流式管理，通过资源管理、任务创建、任务日志、工作流管理、工作流日志等方式

实现数据服务从集成、清洗、处理、大数据计算到发布等过程完成的程序化或自动化的执行和服务。

用户可以通过调度系统灵活方便地配置和调度大数据 ETL 任务。通过配置任务之间的依赖关系，可以灵活地组织任务流。

支持任务流的定期调度、历史任务重跑等多种调度方式；支持 DS\SQL 等工具进行定制化调度；支持集值判断机制；提供错误跟踪、日志管理，对整个数据服务提供完善的服务功能。结合权限管理，实现数据在受控的模式下进行安全、有效、畅通的各类数据服务。



4.2.1 可视化的 ETL 过程设计

DataOps 平台提供了可视化定义 ETL 作业信息，支持作业的试运行和断点调试等操作，丰富的组件能协助用户完成 ETL 数据加工边调试边预览数据的目的。

4.2.1.1 **任务创建**任务创建页面可以创建放在工作流中运行的任务，如 shell 任务，为图形化的任务调度做准备。

在任务创建页面可以看到已有任务的层级结构⇒点击菜单树中的任务可以看到该

任务的详细信息⇨点击查看流程可以查看该任务被哪些 workflow 引用。

操作	技术名称	描述	类型	创建时间	修改时间
   	asda11		SHELL	2022-06-09 16:47:02	2022-06-09 16:52:27

脚本

```
1 java -jar ./资源文件/load.jar
```

资源

资源文件/load.jar x

操作	技术名称	描述	类型	创建时间	修改时间
   	asda11		SHELL	2022-06-09 16:47:02	2022-06-09 16:52:27

运行标志 正常 禁止执行

任务优先级 Worker分组

失败重试次数 (次) 失败重试间隔 (分)

超时告警

超时策略 超时告警 超时失败

超时时长 分

4.2.1.2 工作流工作流是任务执行的载体。通过图形化、列表化的工作流管理，可以将需要一起执行的任务按照数据采集、数据映射、数据校验、数据深度清洗、基于大数据量计算等方式进行流程化的数据处理以及数据治理。

4.2.1.3 工作流日志

每次运行的工作流都会有一条工作流日志，方便使用人员快速找到失败的工作流，对问题进行排查。同时支持多种筛选方式，方便快速找到想要的日志信息。

名称	工作流名称	状态	运行类型	开始时间	结束时间	运行时长	运行次数	host	操作
1	flink-holiday-real2-0-16552851 01049	●	调度执行	2022-06-15 17:25:01	2022-06-15 17:25:06	5s	1	192.168.87.128.5678	[操作按钮]
2	dms_god_vt_daily-0-16552850 63693	●	启动工作流	2022-06-15 17:24:24			1	192.168.87.128.5678	[操作按钮]
3	check_routine_load-0-165528 4980802	●	调度执行	2022-06-15 17:23:01	2022-06-15 17:23:32	31s	1	192.168.87.128.5678	[操作按钮]
4	dms_rpt_real_time_xhd-0-1655 28480037	●	调度执行	2022-06-15 17:20:01	2022-06-15 17:23:45	3m 44s	1	192.168.87.128.5678	[操作按钮]
5	flinkTaskCheck-0-1655284800 879	●	调度执行	2022-06-15 17:20:01	2022-06-15 17:20:06	5s	1	192.168.87.128.5678	[操作按钮]
6	wf_ret_dms2dm_10min-0-165 5284741046	●	调度执行	2022-06-15 17:19:01			1	192.168.87.128.5678	[操作按钮]
7	wf_dt_dms2dm_1hour-0-1655 2848082425	●	调度执行	2022-06-15 17:18:00			1	192.168.87.128.5678	[操作按钮]
8	wf_ret_dms2dm_10min-0-165 5284680298	●	调度执行	2022-06-15 17:18:00	2022-06-15 17:18:07	7s	1	192.168.87.128.5678	[操作按钮]
9	wf_ret_dms2dm_30min-0-1655 284829551	●	调度执行	2022-06-15 17:17:01			1	192.168.87.128.5678	[操作按钮]
10	flink-holiday-real2-0-16552845 00302	●	调度执行	2022-06-15 17:15:00	2022-06-15 17:26:23	11m 23s	1	192.168.87.128.5678	[操作按钮]

4.2.1.4 任务日志

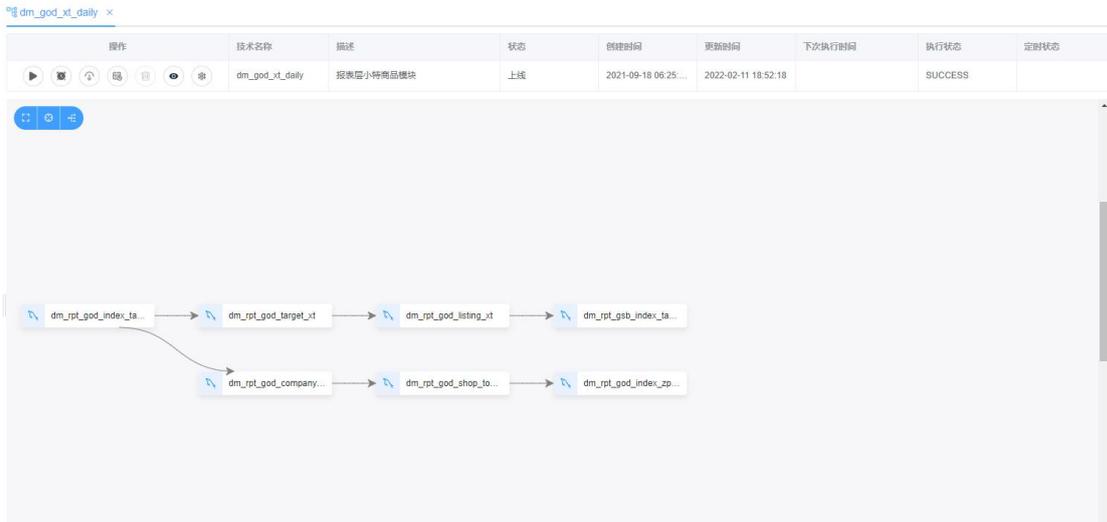
调度系统中的每一个 Task 的运行日志都会被记录，并提供给用户快速查看的可视化界面，方便开发人员快速的定位问题，进行修改和排查。

名称	工作流实例	节点类型	任务	运行时间	开始时间	结束时间	host	运行时长	重试次数	操作
check_routine_load	check_routine_load-0-1 65528480802	EVENT	成功	2022-06-15 17:23:27	2022-06-15 17:23:28	2022-06-15 17:23:30	192.168.87.128.1234	2s	0	[日志] [清除]
routine_load_check	check_routine_load-0-1 65528480802	SHELL	成功	2022-06-15 17:23:04	2022-06-15 17:23:04	2022-06-15 17:23:26	192.168.87.128.1234	22s	0	[日志] [清除]
truncate ods_add_check_routine	check_routine_load-0-1 65528480802	SQL	成功	2022-06-15 17:23:01	2022-06-15 17:23:01	2022-06-15 17:23:03	192.168.87.128.1234	2s	0	[日志] [清除]

4.2.2 图形化的 ETL 过程流设计

通过绘制 DAG 图的方式，可以高效快捷的配置想要的工作流。以图形化的方式完成 ETL 过程的前驱后继关系和调度顺序的定义。ETL 过程依赖关系分为强制依赖和非强制依赖，强制依赖是指作业之间有着业务上的依赖关系，前置作业的输出做为后置作业的输入，只有前置作业执行完成后，才能进行后置作业的调度，而非强制依赖是指作业之间没有任何的业务依赖关系，为了管理的需要将一组作业进行序列化的处理，在前

置作业执行出错时候，可以选择跳过进入到下一个作业的执行中去。



ETL 过程流支持的常用组件包括：

(1) 开始

开始节点是 ETL 过程流的执行入口，执行的引擎从该任务开始逐一的执行后续任务一直到所有的任务执行完毕。

(2) 文件监听任务

文件监听主要用于对文件事件进行监控，是基于文件到达类 ETL 作业的出发条件，文件监听主要通过配置扫描目录、文件名称、文件内容进行判断文件是否到达，是否传输完毕，然后触发后续的进程。

(3) ETL 过程

ETL 过程组件是用来调用 ETL 过程的。

(4) ETL 过程流

ETL 过程流中的任务用来定义 ETL 过程的执行顺序。

(5) 分支任务

分支任务主要用于控制流程的条件分支，只能拥有一个输入，但是有多个输出，当输入条件满足任意分支时，才能进行后续操作。

(6) 存储过程任务

在选择连接池后，存储过程任务组件会根据不同的连接池刷新存储过程列表，选择存储过程，参数列表会相应字段加载。

(7) SQL 任务

SQL 任务组件用来执行 SQL 语句，SQL 语句支持引用参数，SQL 组件大多数来执行清理表数据等。

(8) Shell 任务

Shell 任务是用来调用系统脚本的，来完成 ETL 流程后续的触发。

(9) 参数赋值

参数赋值任务用于根据输入的记录来更新 ETL 编辑器中定义参数值。

(10) 结束节点

结束节点是所有流程的尾部，是一个空节点，无输出，但是必须有输入节点与之相连接，用于标识流程已经结束。

4.2.3 可靠的 ETL 过程、过程流发布

平台提供了设计区和运行区隔离的机制，来保证运行环境的稳定和可靠，所有作业的修订和更改通过在设计区完成调试和试运行后，通过发布机制，发布到稳定的运行环境。

4.2.4 智能的调度管理

调度设置主要用于配置作业的计划执行时间。调度设置能指定到月份、日期、小时、分钟的粒度，也支持按照星期的方式来进行计划任务的执行。同时能设置调度的时间窗

口，重调时间间隔。作业只有在配置的调度时间达到后才会出发作业的执行，在规定开始时间没能启动作业执行时，会根据配置的重调时间间隔和重调次数来进行作业的重新启动，在操作时间窗口和重调次数时，会进行后续的作业运行和处理。并且调度可设置邮件发送执行情况，当 ETL 流执行出错时，相关人员可收到通知和相关信息，并根据信息判断做出处理。调度支持图形监控和列表监控，图形监控提供 ETL 运行的总览图，能直接的展现各 ETL 的执行状态和执行时长信息，列表监控提供明细的运行监控，粒度细化到 ETL 作业内部的组件单元。

4.3 灵活的数据服务管理

DataOps 平台支持不同业务系统的数据源接入，在接入不同业务系统的数据源之后，从业务数据应用场景出发，采用业务数据集的方式，对数据模型进行管理，业务用户可以直观的通过业务模型视角，观察不同数据集之间的关系。也可以对数据模型进行增删改查操作，或者进行复制，修改等操作。让企业数据真正意义上打破数据孤岛，为企业提供全方位、全业务流程、全生命周期的数据分析平台。

4.3.1 数据建模

4.3.1.1 数据字典

DataOps 平台通过数据字典功能对数据的数据项、数据结构、数据流、数据存储、处理逻辑等进行定义和描述，其目的是对数据流程图中的各个元素做出详细的说明，使用数据字典为简单的建模项目。

数据字典页面可以定义数据字段，如主数据字段，用于在表定义中直接选择。在字典定义中，可设置字段类型、最大长度、整数精度、小数精度。考虑到数据字典的复用性以及表字典的统一管理，本系统中所有用在表定义中的字段，都必须是在数据字典中

创建的字段。

名称	编码	类型	创建时间	修改时间
国家键值	LAND1	字符类型	2021-09-16 10:30:59	2021-09-16 10:30:59
省份	REGIO	字符类型	2021-09-26 11:01:03	2021-09-26 11:01:03
省份文本	REGIO_TXT	字符类型	2021-09-26 11:01:03	2021-09-26 11:01:03
城市	ORT01	字符类型	2021-09-26 11:01:03	2021-09-26 11:01:03
城市文本	CITY_TXT	字符类型	2021-10-23 11:23:40	2021-10-23 11:23:54

点击查看可以看到该字典被哪些表引用过。

模板定义详情(模板字段必须指定列排序)

列文本: 列名: 搜索:

<input type="checkbox"/>	列排序	列文本	列名	数据类型	最大长度	整数精度	小数精度	主键标志	是否必填	是否唯一	权限绑定	操作
<input type="checkbox"/>	1	公司主数据*	BUKRS	varchar	5	0	0	否 <input checked="" type="checkbox"/> 是	否 <input checked="" type="checkbox"/> 是	否 <input type="checkbox"/> 是	146	<input type="button" value="白"/>
<input type="checkbox"/>	2	年月*	YEARMONT H	varchar	6	0	0	否 <input checked="" type="checkbox"/> 是	否 <input checked="" type="checkbox"/> 是	否 <input type="checkbox"/> 是	请选择字典	<input type="button" value="白"/>
<input type="checkbox"/>	3	产品线*	PRODUCTT YPE	varchar	40	0	0	否 <input checked="" type="checkbox"/> 是	否 <input checked="" type="checkbox"/> 是	否 <input type="checkbox"/> 是	请选择字典	<input type="button" value="白"/>
<input type="checkbox"/>	4	公司描述	BUKRS_TE XT	varchar	40	0	0	否 <input type="checkbox"/> 是	否 <input type="checkbox"/> 是	否 <input type="checkbox"/> 是	请选择字典	<input type="button" value="白"/>
<input type="checkbox"/>	5	整车收入 (万元)	ZC_INCOM E	decimal	10	0	4	否 <input type="checkbox"/> 是	否 <input type="checkbox"/> 是	否 <input type="checkbox"/> 是	请选择字典	<input type="button" value="白"/>
<input type="checkbox"/>	6	配件收入 (万元)	PJ_INCOM E	decimal	10	0	4	否 <input type="checkbox"/> 是	否 <input type="checkbox"/> 是	否 <input type="checkbox"/> 是	请选择字典	<input type="button" value="白"/>
<input type="checkbox"/>	7	整车销量	ZC_SALE	decimal	10	0	4	否 <input type="checkbox"/> 是	否 <input type="checkbox"/> 是	否 <input type="checkbox"/> 是	请选择字典	<input type="button" value="白"/>
<input type="checkbox"/>	8	销售旧车公 司编码*	BUKRS_YX	varchar	30	0	0	否 <input checked="" type="checkbox"/> 是	否 <input checked="" type="checkbox"/> 是	否 <input type="checkbox"/> 是	请选择字典	<input type="button" value="白"/>

共 10 条 10条/页 < 1 > 前往 1 页

4.3.1.2 表定义

根据已有的数据字典，支持主题表全部手动定义：表定义页面可以根据数据字典中的字段定义出主数据表，元数据表.....通过设置表名称、数据源、SchemeName、数据库表，并从数据字典中选择字段，创建新表，点击发布后可在所选的数据源中找到这张表。

<input type="checkbox"/>	模板名称	数据源	SchemeName	数据库表	处理模式	是否发布	操作
<input type="checkbox"/>	1 营销年度计划	140	data_collection	sd_income_target	替换	未发布 <input checked="" type="checkbox"/> 已发布	<input type="button" value="详情"/> <input type="button" value="删除"/> <input type="button" value="白"/>
<input type="checkbox"/>	2 营销-经销商目标	140	data_collection	sd_ccsale_target	替换	未发布 <input checked="" type="checkbox"/> 已发布	<input type="button" value="详情"/> <input type="button" value="删除"/> <input type="button" value="白"/>
<input type="checkbox"/>	3 营销-经销商	140	data_collection	site_manage_level02	替换	未发布 <input checked="" type="checkbox"/> 已发布	<input type="button" value="详情"/> <input type="button" value="删除"/> <input type="button" value="白"/>
<input type="checkbox"/>	4 公司部门营销归属	140	data_collection	bulrs_owm_department	替换	未发布 <input checked="" type="checkbox"/> 已发布	<input type="button" value="详情"/> <input type="button" value="删除"/> <input type="button" value="白"/>

点击详情可以维护这张表的字段，定义该字段是否是主键、是否唯一、是否必输。

模板定义详情(模板字段必须指定列排序)

列排序	列文本	列名	数据类型	最大长度	整数精度	小数精度	主键标志	是否必输	是否唯一	权限绑定	操作
1	公司主数据*	BUKRS	varchar	5	0	0	否 <input checked="" type="checkbox"/>	是 <input checked="" type="checkbox"/>	否 <input type="checkbox"/>	146	自
2	年月*	YEARMONTH	varchar	6	0	0	否 <input checked="" type="checkbox"/>	是 <input checked="" type="checkbox"/>	否 <input type="checkbox"/>	请选择字典	自
3	产品线*	PRODUCTTYPE	varchar	40	0	0	否 <input checked="" type="checkbox"/>	是 <input checked="" type="checkbox"/>	否 <input type="checkbox"/>	请选择字典	自
4	公司描述	BUKRS_T EXT	varchar	40	0	0	否 <input type="checkbox"/>	是 <input type="checkbox"/>	否 <input type="checkbox"/>	请选择字典	自
5	整年收入(万元)	ZC_INCOME	decimal	10	0	4	否 <input type="checkbox"/>	是 <input type="checkbox"/>	否 <input type="checkbox"/>	请选择字典	自
6	配件收入(万元)	PJ_INCOME	decimal	10	0	4	否 <input type="checkbox"/>	是 <input type="checkbox"/>	否 <input type="checkbox"/>	请选择字典	自
7	整年销量	ZC_SALE	decimal	10	0	4	否 <input type="checkbox"/>	是 <input type="checkbox"/>	否 <input type="checkbox"/>	请选择字典	自
8	销售归属公	BUKRS_Y	varchar	30	0	0	否 <input type="checkbox"/>	是 <input type="checkbox"/>	否 <input type="checkbox"/>	请选择字典	自

共 10 条 10 条/页 < 1 > 前往 1 页

4.3.1.3 模板定义

平台提供主题表模板定义功能，用户可以对自己关心以及常用的主题表进行模板定义、订阅和收藏。

在 DataOps 中，可以方便的将一张数据库中的表映射成一个模板，新建模板时，可定义模板名称、模板类型、数据源、SchemeName、数据库表、处理模式（替换、覆盖、追加），通过模板对数据表进行增、删、改、查。

模板定义

模板名称	数据源	处理模式	是否发布	操作
1 营销考核计划	140	替换	未发布 <input type="checkbox"/> 已发布 <input checked="" type="checkbox"/>	详情 授权 删除
2 营销经销商目标	140	替换	未发布 <input type="checkbox"/> 已发布 <input checked="" type="checkbox"/>	详情 授权 删除
3 营销二网经销商	140	替换	未发布 <input type="checkbox"/> 已发布 <input checked="" type="checkbox"/>	详情 授权 删除
4 公司部门营销归集	140	替换	未发布 <input type="checkbox"/> 已发布 <input checked="" type="checkbox"/>	详情 授权 删除
5 工作计划公司	140	替换	未发布 <input type="checkbox"/> 已发布 <input checked="" type="checkbox"/>	详情 授权 删除
6 生产生产发货目标	140	替换	未发布 <input type="checkbox"/> 已发布 <input checked="" type="checkbox"/>	详情 授权 删除
7 生产生产事故	140	替换	未发布 <input type="checkbox"/> 已发布 <input checked="" type="checkbox"/>	详情 授权 删除
8 生产库存	140	替换	未发布 <input type="checkbox"/> 已发布 <input checked="" type="checkbox"/>	详情 授权 删除
9 生产质量问题	140	替换	未发布 <input type="checkbox"/> 已发布 <input checked="" type="checkbox"/>	详情 授权 删除
10 生产质量问题	140	替换	未发布 <input type="checkbox"/> 已发布 <input checked="" type="checkbox"/>	详情 授权 删除

新增模板定义

模板名称: ods

数据源: 140

SchemeName: qsd

数据库表: act_cmmn_casedef

处理模式: 替换

点击详情可以维护这张表的字段，定义该字段是否是主键、是否唯一、是否必输。

模板定义详情(模板字段必须指定列排序)

列文本 列名

<input type="checkbox"/>	列排序	列文本	列名	数据类型	最大长度	整数精度	小数精度	主键标志	是否必填	是否唯一	权限绑定	操作
<input type="checkbox"/>	1	公司主数据*	BUKRS	varchar	5	0	0	<input checked="" type="checkbox"/> 是	<input checked="" type="checkbox"/> 是	<input type="checkbox"/> 否	146	<input type="button" value="白"/>
<input type="checkbox"/>	2	年月*	YEARMONTH	varchar	6	0	0	<input checked="" type="checkbox"/> 是	<input checked="" type="checkbox"/> 是	<input type="checkbox"/> 否	请选择字典	<input type="button" value="白"/>
<input type="checkbox"/>	3	产品线*	PRODUCTTYPE	varchar	40	0	0	<input checked="" type="checkbox"/> 是	<input checked="" type="checkbox"/> 是	<input type="checkbox"/> 否	请选择字典	<input type="button" value="白"/>
<input type="checkbox"/>	4	公司描述	BUKRS_TEXT	varchar	40	0	0	<input type="checkbox"/> 否	<input type="checkbox"/> 否	<input type="checkbox"/> 否	请选择字典	<input type="button" value="白"/>
<input type="checkbox"/>	5	整车收入 (万元)	ZC_INCOME	decimal	10	0	4	<input type="checkbox"/> 否	<input type="checkbox"/> 否	<input type="checkbox"/> 否	请选择字典	<input type="button" value="白"/>
<input type="checkbox"/>	6	配件收入 (万元)	PJ_INCOME	decimal	10	0	4	<input type="checkbox"/> 否	<input type="checkbox"/> 否	<input type="checkbox"/> 否	请选择字典	<input type="button" value="白"/>
<input type="checkbox"/>	7	整车销量	ZC_SALE	decimal	10	0	4	<input type="checkbox"/> 否	<input type="checkbox"/> 否	<input type="checkbox"/> 否	请选择字典	<input type="button" value="白"/>
<input type="checkbox"/>	8	销售白票公司编码*	BUKRS_YX	varchar	30	0	0	<input checked="" type="checkbox"/> 是	<input checked="" type="checkbox"/> 是	<input type="checkbox"/> 否	请选择字典	<input type="button" value="白"/>

共 10 条 前往 页

4.3.1.4 模板展示

模板展示页面，可以根据当前用户所拥有权限，展示对应数据。支持增删改数据、导出数据、导入数据、导出本页、导出模板。

模板展示

过滤条件: =

<input type="checkbox"/>	公司主数据*	年月*	产品线*	公司描述	整车收入 (万元)	配件收入 (万元)	整车销量	销售白票公司编码*	预留字段1	预留字段2
1	<input type="checkbox"/>	<input type="text"/>								
2	<input type="checkbox"/>	<input type="text"/>								
3	<input type="checkbox"/>	<input type="text"/>								
4	<input type="checkbox"/>	<input type="text"/>								
5	<input type="checkbox"/>	<input type="text"/>								
6	<input type="checkbox"/>	<input type="text"/>								
7	<input type="checkbox"/>	<input type="text"/>								
8	<input type="checkbox"/>	<input type="text"/>								
9	<input type="checkbox"/>	<input type="text"/>								

共 144 条 前往 页

当执行一次增删改操作后，可以在导入日志中查看当前进度。

导入日志 ×

刷新

	Job名称	文件名称	开始时间	结束时间	状态	错误信息
1	界面录入-insert		2022-05-27 08:24:29	2022-05-27 08:24:30	正式表插入完成	
2	界面录入-insert		2022-05-27 08:24:10	2022-05-27 08:24:11	正式表插入完成	
3	界面录入-delete		2022-05-27 08:23:57	2022-05-27 08:23:58	正式表插入完成	
4	界面录入-insert		2022-05-27 08:23:42	2022-05-27 08:23:43	正式表插入完成	
5	界面录入-insert		2022-05-27 08:22:48	2022-05-27 08:22:48	插入正式表失败	<pre> ### Error updating database. Cause: java.sql.SQLException: Duplicate entry '1800-总装车' for key 'PRIMARY' ### The error may involve com.open.capacity.data.mapper.CustomSqlMapper.update-inline ### The error occurred while setting parameters ### SQL: UPDATE `data_collection`.`safety_accident` SET `PRODUCTTYPE`='总装车',`YEARMONTH`='202202`,`LOSS_MONEY`=0,`WERKS`='1800`,`WERKS_TXT`='天津洪申公司`,`WORKSHOP_CODE`='TJ01' WHERE `PRODUCTTYPE`='总装车' </pre>

共 1807 条 10条/页 < 1 2 3 4 5 6 ... 181 > 前往 1 页

4.3.2 数据补录

DataOps 平台对于线下数据或者手工整理的的数据，通过 Excel 或者其他文件批量导入系统或者页面手工进行新增补录，并在数据补录流程中穿插数据校验和审核，以确保正确补录数据以供使用。用户在通过数据补录将数据导入到正式表后，如果后期发现数据存在问题，需要对数据进行修改，那么可以通过补录编辑页面进行数据的查询，修改，删除。

数据补录主要包含模板定义、数据补录、补录审批以及补录查看。模版定义的主要功能是定义补录模板，包括目标数据库，补录列字段名称、类型、长度、校验方式、默认值等信息；

数据补录是整个功能模块的重点，他根据补录模版动态生成不同补录页面，支持 excel 批量录入、校验、提交；补录查看则是查看、下载已补录数据，并支持对已导入数据编辑、修改操作。

4.3.3 数据服务

DataOps 提供了便捷的数据分享功能，可以根据企业各业务用户的需求灵活配置需要分发的数据服务，同时支持自定义分发规则，多时间维度定义分发频次。自动定时分发与手动选择分发结合满足企业多条件分发要求。并且提供了易于用户监控数据的分发结果展示界面。。

4.3.3.1 分发权限管理

- 可以配目标组织，目标系统代码，文件服务器（发送的服务器）
- 权限组分配可以配相应的权限组：权限组里面还有一个页面可以配相应的权限值（例如
- 权限组可以配置省级别的权限，权限值则细分到要加该省下面的哪个地市的权限)

4.3.1.2 分发执行管理

分发平台可以选择立即执行或者定时执行的方式。分发组织信息：查看定义的权限是否一致。

4.3.1.3 流程化的管理模式

支持异常复杂业务模式下的数据分发服务，集成任务管理和任务调度管理，实现 DataOps 平台到目标系统的便捷、高效、可控的数据集成服务模式。满足大数据业务模型下，大数据量、数据加密、数据压缩、数据分发的业务目的。

DataOps 的这一数据服务能力，尽可能的满足了，企业数据中心服务模式更多的商业应用场景和业务模式，为企业提供信息化基础服务设置，实现数据服务业务的赋能。

4.4 一站式数据门户

DataOps 报表管理平台支持与成熟的数据展示工具进行集成，让用户通过配置的方式能够轻松构建出理想的可视化报表模型，实现各种图像化类型业务的展示。

DataOps 报表管理平台可集成多种第三方报表工具，包括 SAP BO, Tableau, 永洪, 帆软 BI、帆软 Report、观远等。用户只需要配置好对应的报表工具的服务器地址等信息，就可以快速的将已经开发好的报表挂载进 DataOps 系统，使用户访问的入口统一的同时，更为方便的控制用户的权限。报表管理平台具有以下优势：

- 快速集成成熟工具。如 包括 SAP BO, Tableau, 永洪, 帆软等市面上成熟的 BI 工具均可快速集成于 DataOps 平台，进行统一管控。
- 报表统一系统访问入口，在 DataOps 中直接展示图形化报表
- 快速的实现各种类型的图像化业务的展示
- 实现高并发模式下报表数据的并发控制
- 安全可控-报表访问监控

除了对报表本身的可见性控制以外，报表管理平台还可以与 DataOps 平台权限管理系统结合，通过对报表内数据进行权限管控，实现对同一报表，不同用户/租户/角色查看不同数据。同时，报表管理平台还提供报表访问监控，访问日志管理，可以分析用户访问内容和频率分析，保证所有操作均可追溯，从源头保证数据安全。

- 支持移动应用，支持与移动端设备的集成，提供移动端的报表数据展示

4.5 权限管理

针对数据权限，DataOps 提供一套完整的权限体系。包括：角色组、权限对象、权限值、权限聚合等管理方式。如：数据采集时的权限控制，可视化集成报表查看的权限控制、数据服务分享的数据访问权限控制。满足云端化服务完整权限管理，并具有极强的灵活性和配置操作便捷性。

➤ 角色组

角色组是一系列具有相同查看权限、填报权限的用户的组合。通过对权限进行组别管理，实现具有业务意义的权限汇聚，并可进行灵活的统一授权管理。

➤ 权限对象及权限值管理

(1) 权限对象：可以理解为不同维度内的权限分类。比如：公司权限、省份权限；

(2) 权限值：可以理解为权限对象下的一个个实体，比如：公司权限对象，其下对应着一个个公司。

通过设定父级参数，将权限对象、权限值进行值集关系管理。如，省份为公司的父级参数，公司 1 的父级参数为省份 1，则只有在省份参数选择为“省份 1”时，方可看到“公司 1”。

➤ 权限聚合管理

权限聚合管理囊括角色组、权限对象、权限值，三者汇聚于一处，组、对象、值三者的关联关系直观明了。聚合管理提供了角色组、对象、值的便利的分配操作。

4.5.1 功能权限 DataOps 系统可实现对报表本身的可见性控制，及在同一页面中，不同用户的操作区域控制。另外，DataOps 权限管理系统通过与第三方账号配置结合，实现用户对报表的仅查看/编辑权限控制。

4.5.2 数据权限

DataOps 权限管理系统与 DataOps 可视化平台结合，通过对报表内数据进行权限管控，实现对同一报表，不同用户/租户/角色查看不同数据。

4.5.3 接口权限

对 hdfs、hive 等维护策略，包括表访问权限、表策略设置、列访问权限、数据脱敏策略等，控制用户的列级、行级、表级访问权限，对数据进行脱敏。

4.5.4 开发权限

由权限对象、权限值，生成权限表，通过权限聚合管理，控制当前登录用户及其拥有的报表开发调整权限，确保创建者的报表不被其他用户修改。

4.6 监控管理

运维监控平台提供集群服务、硬件、调度等内容一站式监控与告警服务，方便企业快速发现定位问题原因，对运营维护自身系统提供监控应用与服务能力。

4.6.1 调度日志

监控流程执行情况，可查看执行到哪个节点，每个任务节点的执行情况等。并可查询流程情况，进行中断、跳转、暂停等操作。同时可以查看所有 ETL 执行过程中产生的日志以及节点执行状态。

4.6.2 运维监控

DataOps 提供报表访问监控，访问日志管理，可以分析用户访问内容和频率分析，保证所有操作均可追溯，从源头保证数据安全。

4.6.3 微服务监控

分层式监控，通过获取系统运行参数如延迟时间、请求量、错误率，掌握系统运行情况，助力企业快速发现问题、分析问题、解决问题和优化问题，高效保障系统的稳定性与持续服务能力。

5. 产品设计原则

5.1 先进性

- SQL 解析技术
- 遵循事实结构的星型数据仓库模型
- 用户 WEB 界面采用 VUE2 技术
- 使用国际化编码 UTF - 8
- 采用 J2EE 架构，适应各类平台部署
- 易于使用：简洁的界面，便于用户理解、操作
- 标准开放：与其它业务系统方便地集成
- 遵循维度事实结构的星型数据仓库模型
- 遵循多维 (MD) 和 ROLAP 理论
- 不但支持标准的 OLAP 分析，同样也支持 OLTP 报表
- 支持深入的数理统计分析功能
- 遵循 J2EE、VUE2 等多种业界主流的、先进的技术标准

- 采用多层应用体系架构，将各种应用作为被共享的服务的集合

5.2 可维护性

- 提供集成化的系统管理模块，通过图形化界面管理、配置所有系统对象
- 数据标准可方便进行备份和恢复
- 支持自定义数据标准属性集，可适应不同项目对数据标准属性项的定义
- 纯 web 化，报表分析展示无需任何插件，客户端零安装，大大方便系统升级维护
- 所有系统对象和元数据全部持久保存在数据库中，服务器迁移十分方便
- 报表模板、对象结构等元数据可方便的进行备份和恢复

5.3 可靠性

- 合理的 JAVA 内存回收机制，用户操作结束后，系统不允许仍然占有该连接的内存
- 有完善的错误诊断和恢复机制，不会出现因用户误操作而导致系统崩溃
- 所有系统对象和元数据全部持久保存在数据库中，不会因应用服务器崩溃或病毒入侵等损坏系统
- 数据整合支持数据缓存、端点续传和错误通知等技术，保证工具使用的稳定性和可靠性。

5.4 易用性

- 系统安装配置简单，基于 B/S 结构，客户端零安装
- 界面操作简单，流程清晰
- 面向业务人员提供零编程、交互式分析功能，而非定制的分析报表
- 广泛采用 VUE2 技术，为用户提供了十分友好的交互式 WEB 操作界面

- 敏捷开发模式，集数据建模和 ETL 设计于一体，能快速响应开发，灵活应对需求变动
- 大量批量创建的方式完成数据的落地和抽取、转换、装载的数据处理过程
- 开发者与应用者视角隔离，支持多人协同定义，共享资料库
- 可视化模型定义和 ETL 过程定义
- ETL 定义支持数据抽样，转换过程支持预警设置等多样化参数

5.5 安全性

- 自动限制每级用户只能访问数据权限对应的层级数据
- 管理员可详细配置每个用户的操作权限和可访问对象
- 采用严格的技术手段，杜绝非法用户绕过安全认证直接调阅数据标准
- 记录详细的操作日志以备审查
- 防暴力攻击
- 采用全方位、严密的安全编码来加强产品安全性。
- 采用 Base64 密码加密

5.6 扩展性

- 所有功能均根据可扩展通用的原则来设计的，可以适应未来未知的需求变化
- 系统的模块化设计，一个模块发生变化不会影响整个系统
- 提供各种定制接口和调用接口 Web Services 服务，满足第三方服务需要
- 提供二次开发平台设计，满足未来各种未知需求
- 内置脚本支持，满足未来各种未知需求
- 数据整合内置了多种函数库，提供跨数据库类型和版本的算法兼容性，同时支持自定义，提供了丰富的脚本组件、转换组件、清洗组件、存储过程组件、系统命令

组件的方式来保证工具的扩展性。

6. 产品部署要求

6.1 部署方式

DataOps 支持多种部署方案，可以独立部署至企业内网，也可以采用混合云架构，部署至企业私有云。

同时 DataOps 支持企业内部 Windows 域、VLAN、VPN 等网络环境，从而确保企业在不同的网络环境下都可以使用 DataOps，满足企业的实际应用需求。

6.2 系统要求

DataOps 支持市面上几乎所有的主流操作系统，包括 Windows Server，Linux，Unix 等。针对 Linux 的众多发行版，DataOps 也都有很好的支持，如 CentOS、Ubuntu 等。

针对国产化需求，DataOps 产品同样适应各种国产 Linux 操作系统：深度、红旗，麒麟等。

6.3 浏览器要求

DataOps 属于标准的 BS 结构产品，在开发中使用了前后端分离的技术，其主要通过浏览器进行访问。

DataOps 目前支持主流的浏览器，包括：Chrome、Firefox、IE edge 等多种常用浏览器。大多数情况下，我们建议您使用 Chrome 浏览器。

6.4 配置要求

DataOps 产品中，由于采用了微服务架构，所有的模块均可以根据实际业务情况进行独立部署，同时随着业务量，数据量的增长，也对任意模块进行集群化扩展。

DataOps 支持最小化的部署方案，支持在最低 8 核 CPU，16G 内存配置下单机运

行。或者每个模块可以在 1 个 CPU 核心，2G 内存的服务器中单独运行。

推荐硬件配置方案：

基本配置	高级配置
2 x Intel® Xeon® E5-2670, v1~v3 128~256GB Memory DDR3/4	2 x Intel® Xeon® E5-2690, v1~v3 128~256GB Memory DDR3/4
2 x 300Gb 10K SAS for OS	2 x 300Gb 10K SAS for OS
14 x 600GB 10K SAS for data	22/24 x 1.2TB 10K SAS for data

具体实施部署场景的硬件需求，请咨询 DataOps 售前工程师。