

# 数据共享交换管理

## 用户操作手册

## 目录

1 产品定位.....	1
2 产品概述.....	1
3 客户价值.....	3
4 关键特性.....	4
4.1 支持结构化数据、大数据、文件数据的共享发布.....	4
4.2 支持单表、结果集等形式的实时接口服务发布.....	5
4.3 支持在线、离线、模板等多种开发形式.....	5
4.4 提供多视角、全链路数据资源使用状况管控.....	6
4.5 元数据驱动实现数据资源目录.....	6
4.6 支持大文件传输.....	7
4.7 统一的作业调度中心.....	7
4.8 数据服务安全管理.....	8
5 核心架构.....	8
5.1 应用架构.....	8
5.2 数据架构.....	10
5.3 部署架构.....	10
5.4 技术架构.....	11
5.4.1 技术视图.....	11
5.4.2 技术实现.....	12
6 产品组成.....	13
7 应用功能概述.....	15
7.1 数据资源目录管理.....	16
7.1.1 节点管理.....	16
7.1.2 数据库资源管理.....	17
7.1.3 文件资源管理.....	17
7.1.4 接口服务管理.....	18
7.1.5 大数据资源管理.....	19
7.1.6 资源分类管理.....	19

7.2 数据使用 .....	20
7.2.1 消费方注册 .....	20
7.2.2 数据服务使用 .....	21
7.2.2.1 数据资源申请 .....	21
7.2.2.2 数据资源订阅 .....	22
7.2.3 数据服务审批 .....	23
7.3 数据服务开发 .....	24
7.3.1 实时数据服务 .....	24
7.3.2 批量数据服务 .....	25
7.3.2.1 结构化 .....	25
7.3.2.2 非结构化 .....	27
7.3.3 订阅任务开发 .....	28
7.3.4 脱敏算法配置 .....	28
7.3.5 作业模板开发 .....	29
7.4 服务控制 .....	31
7.4.1 系统调用关系管理 .....	31
7.4.2 服务状态管理 .....	31
7.4.3 系统 IP 地址管理 .....	32
7.5 调度管理 .....	33
7.5.1 作业策略管理 .....	33
7.5.2 作业计划场景 .....	33
7.5.3 作业图形编排 .....	34
7.5.4 事件触发 .....	35
7.6 运行监控 .....	36
7.6.1 交换监控 .....	36
7.6.2 消费者监控 .....	36
7.6.3 计划、作业流监控 .....	37
7.6.4 故障告警监控 .....	38
7.6.5 元数据采集监控 .....	39
7.6.6 历史记录查询 .....	39

7.6.7 物理资源监控 .....	40
7.6.8 服务引擎监控 .....	41
7.7 统计分析 .....	42
7.7.1 交换总量 .....	42
7.7.2 作业统计 .....	43
7.7.3 文件统计 .....	43
7.7.4 消费方数据统计 .....	44
7.7.5 Web 服务统计 .....	45

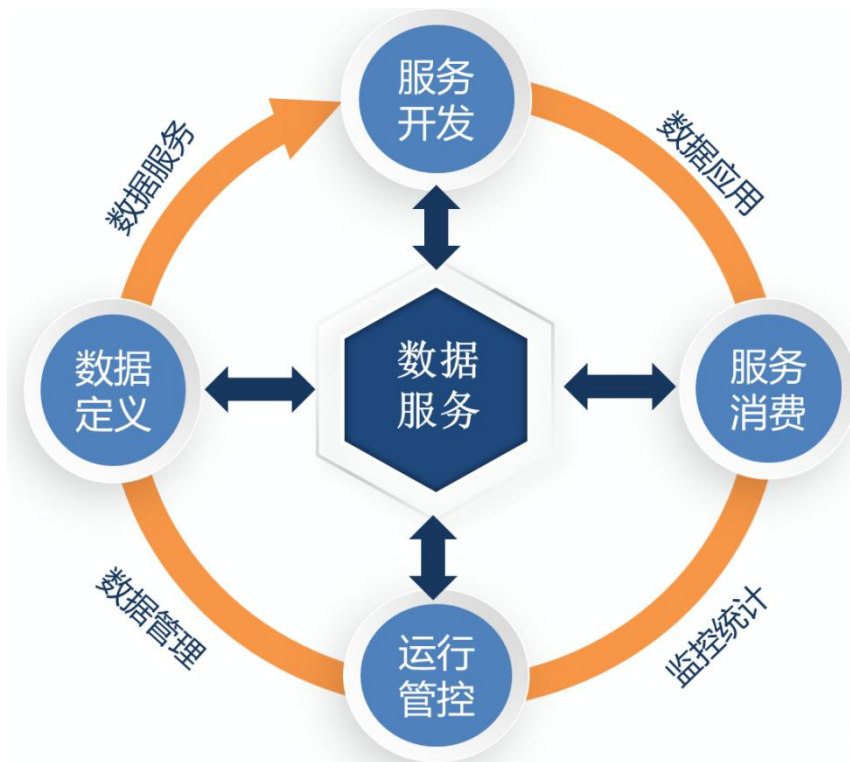
# 1 产品定位

随着企业业务和技术的快速发展，数据资源在企业数字化转型的过程中，成为了最核心资产、最重要的生产资料，成为了企业重塑业务、自我转型的决定性因子，如何有效地管理数据、高效地提供数据服务是目前企业所面临的关键挑战，在此背景下，企业需要一个源源不断的输出数据服务，数据洞察的能力的源泉，为数字化转型提供有力保障；如何改善同质化业务模式，实现以客户为中心的数据驱动精准化、差异化经营模式成为保险企业努力探索的方向。

数据服务共享平台是大数据时代自助化的企业级数据服务共享平台，定位于企业数据资源“纵向贯通”、“横向互联”的共享通道，主要从安全管控、服务治理、服务消费、服务开发效率等系列问题入手，对数据资源进行系统化沉淀、体系化管理，提供对多源、多类型数据进行统一的服务化管控，帮助企业更加高效、可靠的使用数据，充分发挥企业数据资产价值。

# 2 产品概述

数据服务共享平台（Data Service Platform，本文档中后续使用简称 DSP）基于大数据架构提供统一的数据服务能力，是企业数据资源对外开放的共享通道，平台提供实时接口服务、批量作业服务、文件传输服务，从数据定义、服务开发、服务消费、运行管控四方面着手，实现数据资源的闭环管理。



- **数据服务**

数据服务的本质是业务活动的抽象，业务活动需求由业务人员提出，数据服务开发人员依据需求进行开发和数据服务的共享发布。

- **数据定义**

数据定义是数据服务应用的前提，后续数据服务的检索、申请、调度和管理都基于此开展。数据定义基于元数据实现，按照分类、主题等多个角度对数据和服务进行管理、识别、定位及共享发布。通过数据定义，一方面可以描述企业数据和服务的分布情况，另一方面，也可作为一种语义层抽象，用于驱动整个企业数据的加工处理过程。系统具备关系数据库、大数据等多类型技术元数据的自动采集，同时具备业务元数据的维护能力，提供数据资源注册、维护和搜索等功能。

- **服务开发**

平台支持批量作业服务、实时接口服务、文件传输服务三种数据服务形式，

并且提供在线、离线两种服务开发方式。过程中可以根据具体业务场景需求，制定服务发布和安全策略。

- **服务消费**

服务消费是在数据服务化后，面向应用的一系列操作动作。数据服务平台为消费者提供数据预览、申请、审批等功能，具备单个、批量申请的能力，具备向消费方推送或由消费方拉取的服务方式；提供数据流量、权限和优先级的服务控制；支持按照计划、事件触发等多种方式进行服务调度。

- **运行管控**

实现对数据服务全生命周期管理、调度、监控等；通过权限控制验证是否有权限访问数据；对提取的关键数据进行脱敏处理，保障数据安全性；实现对数据流向、批量任务、接口、文件传输等服务的监控。

### 3 客户价值

数据服务共享平台让数据在数据平台和业务系统之间形成了良性的闭环，实现应用与数据解藕，通过集中管理方式实现各个系统之间数据的共享和使用，使企业的数据在各个业务系统之间进行流转，充分发挥数据资产的价值；平台提供资源目录服务、多类型数据服务化的开发、数据资源调度服务、数据接入及访问服务。为对数据资源有存取需求、调度需求的业务应用提供统一服务，帮助客户逐步实现数据服务规范化和标准化。数据服务共享平台不仅仅是技术工具，更是重要的数据运营平台，通过数据服务平台，降低业务对技术依赖，充分发挥业务创新潜能。

数据服务共享平台带给客户的价值主要体现在如下几个方面：

- **统一的数据资源目录：**为企业可交换资源提供统一的注册管理方案，实现技术元数据的自动获取，支持业务元数据的定义，提供资源搜索及血缘分析。实现信息资源的统一发布和集中访问。

- **统一的数据共享通道：**建立公司统一的纵向数据共享通道，提供跨系统、跨单位、跨区域的数据集成、交换、分发、共享机制和平台能力，支持结构化、半结构化、非结构化数据以及大文件的多级交换。
- **统一的数据交换标准：**从服务集成、数据集成、文件传输和数据管理等多个方面，定义统一的标准和规范，并在实施过程中，积累更多符合自身发展的、可复制的最佳实践，不断沉淀、改进和评估，提高系统间集成效率，降低各系统运维成本。
- **统一调度管理：**提供灵活的、多角度的模型作业调度机制，通过简单、灵活、易用的配置管理模式，减轻运维管理工作量，实现运维自动化。帮助运维管理人员更加方便、全面地了解数据交换的运行情况，有助于迅速定位和排除故障，为及时主动的业务优化提供方向。
- **统一的数据运行监控和安全保障体系：**提供数据服务发布、访问授权和运行监控的统一管理。从保障应用系统和数据安全出发，从技术和管理两方面提供事前预防、事中控制和事后追溯能力。

## 4 关键特性

### 4.1 支持结构化数据、大数据、文件数据的共享发布

数据服务共享平台基于大数据架构，支持多种类型数据的发布共享，主要包含以下数据类型：

- **结构化数据：**支持将 Oracle、MySQL、SQLServer、PostgreSQL、MongoDB、Gauss100、DM 数据库数据发布为 RESTful 服务，或者通过批量作业服务方式实现数据共享，供第三方应用使用。
- **大数据：**支持 HBase、Hive 数据发布为 RESTful 服务或者通过批量作业服务进行数据共享。
- **文件数据：**支持文件资源服务注册发布，并以文件传输和批量作业方式传输文件资源。



## 4.2 支持单表、结果集等形式的实时接口服务发布

数据服务共享平台基于微服务架构的数据服务发布组件，支持单表、结果集等形式的实时接口服务发布，将数据库资源由底层编码的方式转变 RESTful 服务的方式提供给消费者。旨在于灵活地适配于各种主流关系型数据库、便捷地切换数据源、消除了底层复杂编码过程。

- **技术实现：**数据服务发布组件基于 SpringBoot 框架实现的一套微服务架构引擎，通过 MyBatis 持久层进行数据访问，提供了实时服务快速发布及服务访问的能力，支持分布式部署，线性扩展。
- **扩展能力：**封装内部技术实现细节，基于数据发布引擎扩展能力将复杂的多表发布为实时服务。
- **接口能力提供：**提供实时服务发布接口，支持单表及结果集等形式的服务发布。

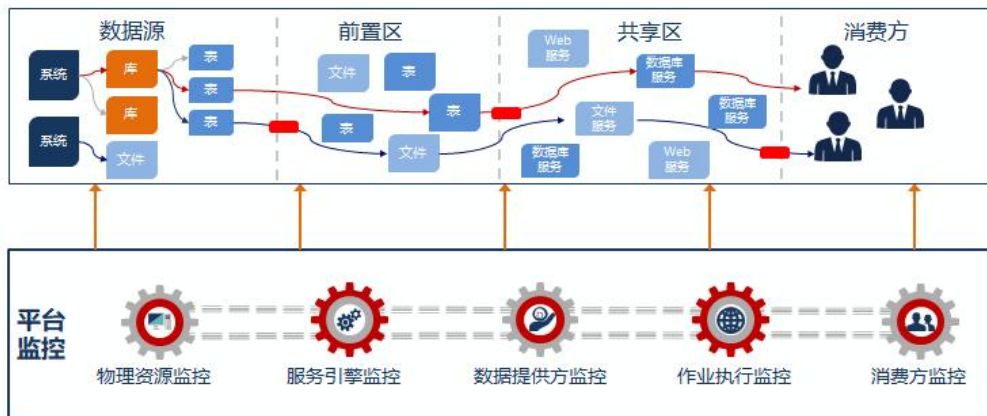
## 4.3 支持在线、离线、模板等多种开发形式

数据服务共享平台支持多种作业的开发形式，实现数据共享发布。

- **在线作业：**支持 HBase 到文件同步、全文对比同步、触发器同步、时间戳同步、数据库到文件同步、全量同步、文件到数据库同步、MySQLtoHBASE 数据脱敏、CanalClientToHBase 同步、DBToHbase、MongoDBToHBaseSingleTable、MongoDBToHBase、DBToDBDesensitization 共十三种类型的在线作业，在数据服务共享平台中经过页面的简单配置即可完成作业开发实现数据共享同步。
- **离线作业：**对于开发逻辑比较复杂，在线作业不能支持的复杂作业开发，可以使用数据服务共享平台中的 DSP Studio 开发工具进行开发，开发完成后将作业部署在数据服务共享平台中，实现数据共享发布。
- **模板开发：**数据服务共享平台提供模板开发标准，用户可以根据项目实际需求开发，将常用的作业模型定制为通用模板（模板开发需要一定的开发工作量），能简化使用难度，缩短单个作业开发时间，提高工作效率。

## 4.4 提供多视角、全链路数据资源使用状况管控

数据服务共享平台从系统管理、服务开发、服务消费等多个视角对数据服务使用状况进行全链路管控。对数据源、前置区、共享区、消费方四个区域的资源进行全链路管控（物理资源监控、服务引擎监控、数据提供方监控、作业执行监控、消费方监控），如下图所示。



PRIMETON · 普元

## 4.5 元数据驱动实现数据资源目录

数据服务共享平台基于元数据驱动实现数据资源目录管理，支持元数据自动采集、数据分类管理、数据服务开发与注册、数据服务浏览与审批。

- **元数据自动采集：**基于元数据驱动支持在数据源、前置区、共享区、消费方四个区域配置数据库资源（含关系型数据库 Oracle、MySQL、SQLServer、PostgreSQL、MongoDB、Gauss100、DM 与大数据 HBase、Hive），实现数据库资源自动采集、预览，为后续数据服务化与数据共享打下基础。
- **数据分类管理：**支持按照分类、主题、应用等多个层次对数据进行分类管理、识别、定位和共享。
- **数据服务开发与注册：**支持通过服务发布组件实现数据服务的快速开发，开发完成的 RESTful 服务自动注册在数据服务共享平台中。
- **数据服务浏览与审批：**已将开发完成的数据服务消费方可以平台中浏览查看，如果消费方需要使用数据服务，则需要在平台中发起申请，审批通过后根据

平台中提供的服务信息使用数据。

## 4.6 支持大文件传输

数据服务共享平台基于大文件传输架构实现文件传输服务，支持断点续传、脱敏、加密等能力，大文件传输用于源到前置区和前置区到共享区的文件交换。通过文件传输配置能够将源中的文件资源传输到前置区，也能够将前置区的资源传输到共享区。

## 4.7 统一的作业调度中心

数据服务共享平台提供统一的作业调度中心，实现对批量作业服务、文件传输服务的调度管理。

### 全面性调度

- 全面的、专业的调度管理，有效提高了数据管理水平。
- 提供日历方案、频度方案、作业、消息配置等按需提供任务调度，提高开发效率，减少开发成本。

### 灵活的调度机制

- 支持事件机制和定时机制的调度方式。
- 支持灵活的、多角度的作业调度管理，包括事件、文件到达和计划调度以及手工触发。

### 灵活配置，可维护性强

- 定制简单化，仅需简单的鼠标点击拖拉即可完成任务调度。
- 调度活动痕迹化，为管理风险做到可控化。
- 灵活的消息机制，把各种消息自动分发给用户（例：将作业的执行状态通知给管理员等）。
- 组件化，耦合度低，接口简单，易开发。
- 三层架构，结构清晰，具有良好的稳定性和容错性。

- 采用了先进的 **JavaEE** 技术架构，平台具有很强的跨平台性和良好的可移植性。
- 可配置的代理负载加权算法，可以在执行作业时支持多机集群，根据负载情况合理分担作业执行的服务器，从而能承载大规模数据的高负荷运行。

## 4.8 数据服务安全管理

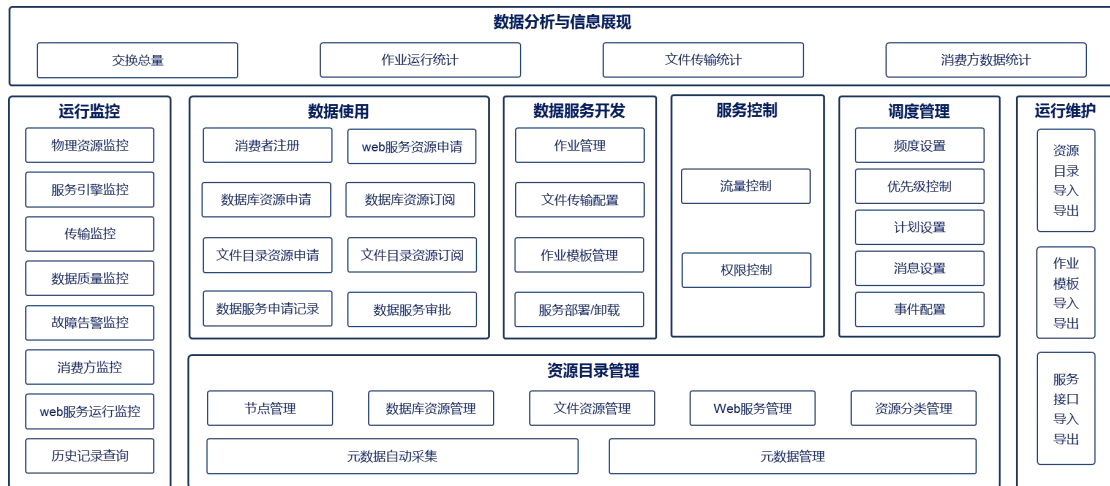
数据服务共享平台从功能权限、客户端权限、服务权限等几个方面实现数据服务安全管理。

- **功能权限：**平台支持按照角色划分系统功能权限，支持到最底层菜单功能的权限划分，例如数据管理员、数据开发人员、数据消费方等用户权限的划分都可以满足。
- **客户端权限：**消费方系统或者客户端访问依次通过平台中 IP 白名单、访问频度、服务状态等多重控制策略进行控制，保证服务访问安全。
- **服务权限：**消费方系统获取服务地址、参数等信息，必须通过完整的数据服务申请、审批流程才能获取服务信息进行访问，并且支持在审批过程中对数据行列级权限进行控制。

# 5 核心架构

## 5.1 应用架构

数据服务共享平台包括资源目录管理、数据使用、数据服务开发、服务控制、调度管理、运行监控、运行维护和数据分析与信息展现，应用架构如下图所示：



- 1) **资源目录管理**：是可交换数据元数据的结构化展现，需支持数据库、大数据、文件、Web 服务等多类型数据资源技术元数据的自动采集和业务元数据的维护能力；支持面向消费者业务视图（比如按主题的划分） 的创建；提供资源注册、维护和搜索等基本功能；
- 2) **数据使用**：主要面向消费方，是消费者如何通过平台申请数据资源及数据管理方审批处理的过程；消费方可通过注册功能自行在平台注册，支持消费方浏览数据资源后提交资源申请（拉）和订阅（推）的方式；消费方可查看所有申请的状态及历史记录，可查看申请的汇总情况；支持单个或批量申请；
- 3) **数据服务开发**：主要面向开发人员，提供在线和离线两种开发方式。对于如全量开发同步、数据库到文件同步等的作业，开发人员可在 Web 页面中进行简单配置完成在线作业开发；对于复杂模型，可通过离线方式开发。数据服务开发提供作业目录管理、作业流的可视化配置等；
- 4) **服务控制**：平台提供用户对数据服务过程中的权限控制，包含 IP 白名单、服务状态、调用关系管理等；
- 5) **调度管理**：作业运行的指挥中心，可通过调度管理配置任务的调度策略，配置任务运行的优先级及触发方式等；
- 6) **运行监控**：对整个平台运行过程中的状态进行监控，包括物理资源、服务引擎、传输监控、故障告警、消费方等，同时提供对日志和历史记录的查看功能。
- 7) **运维维护**：为方便平台使用的功能，支持资源目录、作业模板、服务接

口的导入导出；

- 8) **数据分析与展现：**针对平台作业、数据交换总量、文件传输、消费方等进行统计分析。

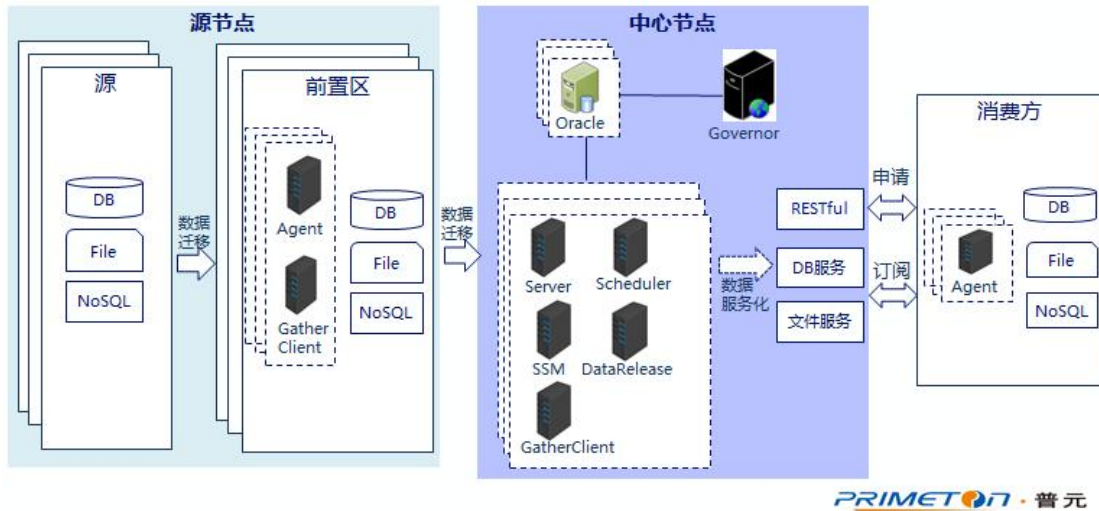
## 5.2 数据架构

数据架构反映平台中，各种类型数据的分布及流转情况。在数据服务共享平台中，数据可分为业务数据、元数据（含资源目录）、质量数据、作业模板及运行监控数据、统计分析数据等。

- 1) **业务数据：**指承载业务含义（大多来源于企业内部各业务系统及少量外部数据），由各总部及各分子公司业务部门（数据提供方）提供，转移至前置区（数据准备）后，经过相应的处理和扩展，提供给消费方。具备数据提供、数据获取、数据准备、数据处理及服务化开发、数据共享发布和数据消费的完善数据交换过程。业务数据一般存储于 ODS 区，也可直接提供点对点的方式，业务数据在平台中不存储；
- 2) **元数据（资源目录）：**业务数据交换的整个过程，都是以元数据（资源目录）为驱动。元数据可分为技术元数据和业务元数据两种，在数据交换过程中，通过手动和自动的方式进行管理；元数据存储于平台数据库中；
- 3) **质量数据：**针对业务数据所开展质量评估后产生的数据，一般包括质量检核规则和检查结果数据；质量数据存储于平台数据库中，也可将结果数据生成文档后发送给业务系统主责部门；
- 4) **模型及运行监控数据：**平台运行过程中，各类作业的模式（模板）及配置信息，作业运行过程中的告警、监控信息和日志信息等，作业运行过程中的调度策略及调度信息；模型及运行监控数据存储于平台数据库中；
- 5) **统计分析数据：**针对作业、数据交换量、消费方等进行的统计分析；部分数据直接取自模型及运行监控数据，部分数据直接存储于平台数据库。

## 5.3 部署架构

数据服务共享平台部署机构如下所示：



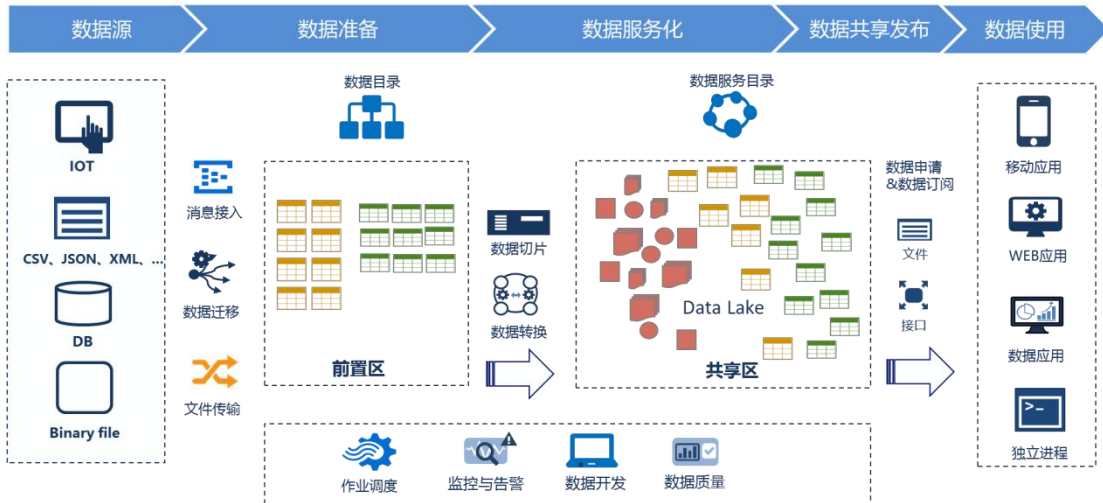
数据服务共享平台适应于单级或多级的部署方式：

- 1) **单级部署：**当企业没有上级单位、平级单位、下级单位数据共享需求时可以采用单级部署方式，单级部署时依据生产环境实际情况，部署 DSP 产品所有组件于物理服务器（建议服务器数量最少为二台及以上）中，其中 SSM 与 DataRelease 组件必须关联部署在同一台服务器中，当数据交换的作业量较大时，可以对 Server 组件进行集群部署，当数据服务（RESTful）的访问量大时，可以对 SSM 与 DataRelease 组件进行集群部署。
- 2) **多级部署：**当企业对上级单位、平级单位、下级单位有数据共享需求时可以采用多级部署方式，多级部署时将 Agent 和 GatherClient 部署于有数据共享需求的上级单位、平级单位、下级单位的物理服务器中，其他产品组件部署于企业内部服务器中，其中 SSM 与 DataRelease 组件必须关联部署在同一台服务器中，当数据交换的作业量较大时，可以对 Server 组件进行集群部署，当数据服务（RESTful）的访问量大时，可以对 SSM 与 DataRelease 组件进行集群部署。

## 5.4 技术架构

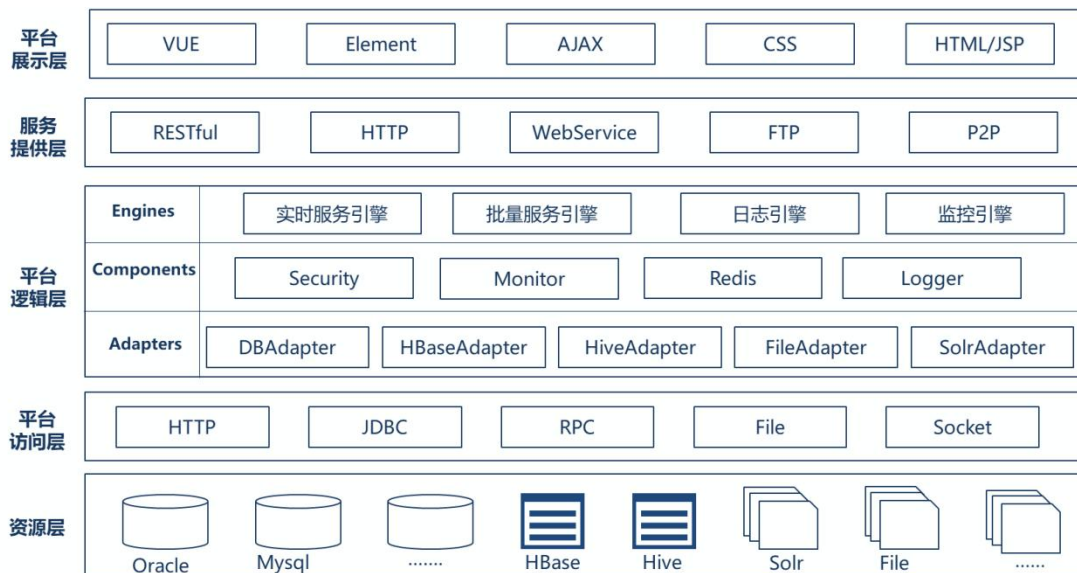
### 5.4.1 技术视图

数据服务共享平台整体技术框架及各过程所采用技术如下图所示：



数据服务共享平台通过多种数据（结构化数据、非结构化数据、文件数据）、协议（Http、JDBC、Socket）接入方式将已有的业务数据接入 DSP 平台，在平台中通过数据转换、加工计算等逻辑操作，发布为数据服务，通过数据共享区对外提供多种类型的数据服务。

### 5.4.2 技术实现



- 1) **资源层:** 支持对主流关系型数据库（Oracle、MySQL、SQLServer、PostgreSQL、MongoDB、Gauss100、DM），大数据（HBase、Hive）、文件等多种类型数据资源技术元数据的自动采集和业务元数据的维护。
- 2) **平台访问层:** 基于 JDBC、HTTP、RPC 等通信协议实现与底层数据和文件

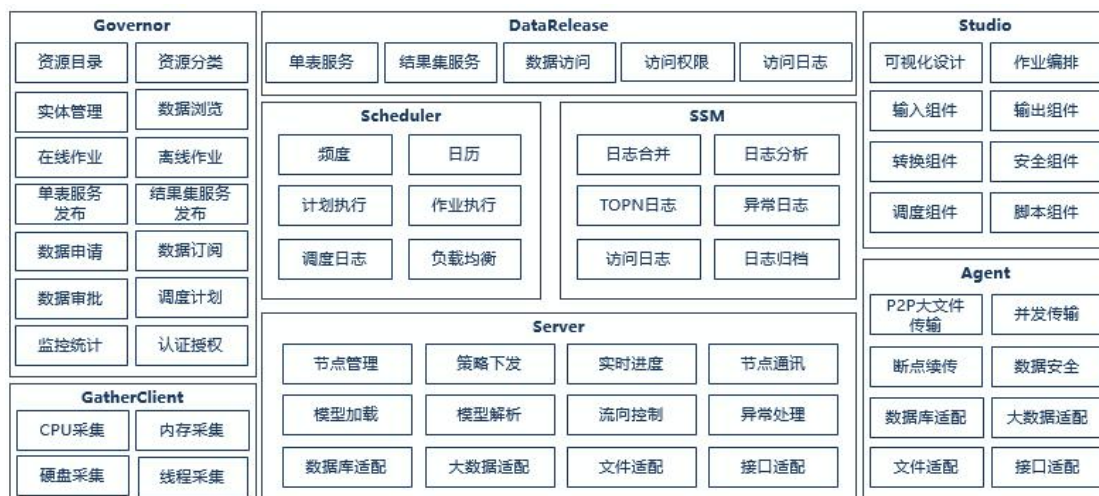


间的访问交互，基于 Socket 通信实现平台各组件间的数据通信，确保平台中数据间交互与通讯高效、安全、可靠。

- 3) **平台逻辑层：**平台逻辑层依托于各类适配器、组件、引擎保证平台的安全稳定运行，适配器有数据库（DBAdapter）、HBase（HBaseAdapter）、Hive（HiveAdapter）、文件（FileAdapter）等类型；组件包含安全、监控、Redis 缓存、日志等组件；引擎包含实时服务引擎、批量服务引擎、监控引擎、日志引擎等。
- 4) **服务提供层：**服务提供层采用轻量级的 Web Service 架构（RESTful），其实现和操作比 SOAP 和 XML-RPC 更为简洁，可以完全通过 HTTP 协议实现对外部应用提供统一的 RESTful Web 服务。
- 5) **平台展示层：**平台展示层使用 VUE+ElementUI 框架，结合 CSS、AJAX 等业界前沿技术实现丰富的界面展示效果，提升使用者的交互体验。

## 6 产品组成

数据服务共享平台（DSP）产品套件主要包括：Governor(管控平台)、Server（服务引擎）、SSM（服务状态监控）、Studio（可视化开发工具）、Agent（传输代理）、Scheduler（调度引擎）、DataRelease（数据发布引擎）、GatherClient（资源采集客户端），产品功能架构图如下所示。



- **Governor（管控平台）**

数据服务共享统一管控平台，提供资源目录、数据开发、数据使用、服务引擎管理、调度计划、统计监控、服务权限配置等系列功能，对数据资源和服务资源统一注册管理。

- **Server（服务引擎）**

批量作业服务及文件传输服务引擎，负责批量作业模型解析、批量作业执行、文件传输服务；提供多协议、多数据源适配支持；为服务运行提供高性能、高可靠的运行环境。

- **SSM（服务状态监控）**

提供日志解析及监控能力，对事前预警、事中告警、事后统计分析等功能提供后台支撑。

- **Studio（可视化开发工具）**

基于 Eclipse 技术开发的可视化批量作业设计器，提供可视化批量作业模型定义与调试、可视化的模型性能监控、元数据管理以及数据处理模型部署等功能。

- **Agent（传输代理）**

数据抽取及文件传输代理，支持数据库、大数据等数据源的数据抽取，支持任意节点及任意大小文件一对一、一对多的传输，并支持压缩、加密等多种文件传输策略。

- **Scheduler（调度引擎）**

作业调度引擎，支持作业与作业流的串行及并行调度，提供日历、频度、事件等多种规则的调度，为作业和作业流运行提供多样化的调度方式。

- **DataRelease**（数据发布引擎）

基于 SpringBoot 架构提供数据服务发布及数据访问能力，支持单表、结果集等形式的实时接口服务发布。

- **GatherClient**（资源采集客户端）

实时采集各引擎所在物理机的 CPU、内存、磁盘、网络等指标，通过 Governor 进行展示。

## 7 应用功能概述

数据服务共享平台（DSP）的典型用户分为系统管理员、数据管理员、开发人员、消费方四类，不同用户在数据服务共享平台中所承担的角色和功能有所不同，具体如下：

- **系统管理员**：数据服务共享平台的管理人员，负责完成供数平台的安装部署工作，包括环境检查、产品安装与验证、数据接入、权限配置等，负责对系统运行环境进行维护，对系统运行过程监控及故障告警信息的及时有效处理，以保证系统运行的不间断性和工作性能的良好性；同时负责平台用户角色管理和功能授权、系统功能配置管理，包括系统日志、系统相关参数等。
- **数据管理员**：数据管理员是企业数据资源的管理人员，负责数据目录、服务目录的注册和维护，负责数据服务的审批和数据服务的使用控制（谁用数据、数据权限是什么）等；负责数据整体共享交换过程中数据资源的监控和统计分析。
- **开发人员**：开发人员负责数据资源服务化前的转换、脱敏、核检，负责实时和批量服务的开发工作。
- **数据消费方**：消费方是数据资源的使用者，消费方在数据生产及服务过程中占据中心地位，因为消费是所有数据活动的起点和终点。消费方通过资源目

录或检索的方式查找相应的数据服务；如果数据服务不存在，可通过数据目录查看是否存在所需的数据，若存在，则可向数据管理员提交数据服务申请。

## 7.1 数据资源目录管理

数据资源目录是以树型结构管理和展现的包括数据库、大数据、Web 服务和文件等资源在内的注册信息。数据资源目录管理包括节点管理、数据库资源管理、大数据管理、接口服务管理、业务分类管理等。

数据管理员登录数据服务共享平台，对资源目录进行配置管理。

### 7.1.1 节点管理

节点是各类资源的挂接点，在本平台中，对应于公司总部及各分、子公司（或组织机构）。其主要功能包括：

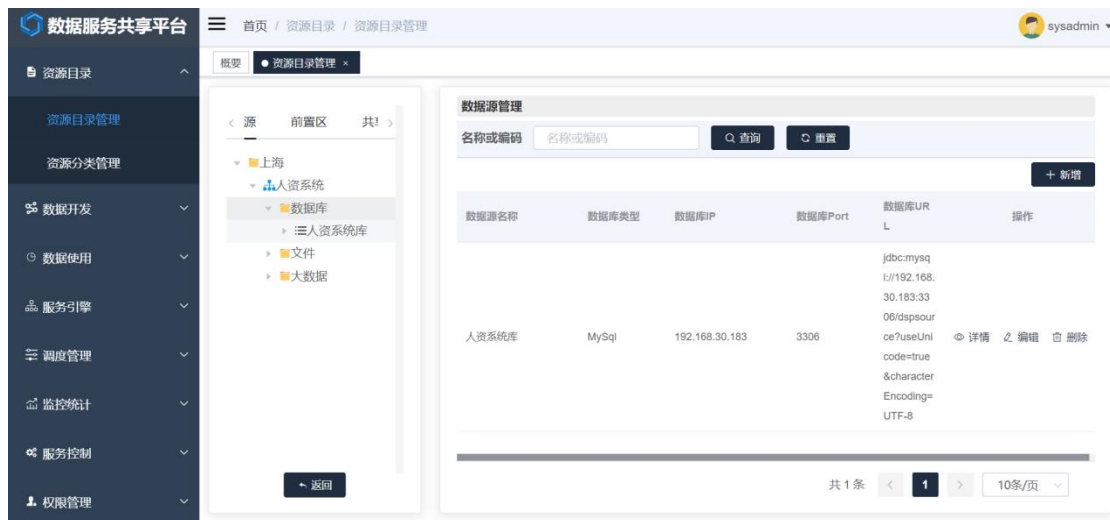
- 节点管理需要能够进行节点的新增、修改和删除；
- 能够实现对节点的检索；
- 自动创建分区（源、前置区、共享区、消费方）
- 自动创建资源类型（如数据库、大数据、Web 服务、文件等），资源类型可通过数据字典可自行定义；
- 支持资源管理功能，包括资源同步、校验等。



## 7.1.2 数据库资源管理

数据管理员和开发人员可以通过平台注册、修改、删除源区、前置区、共享区、消费方数据库资源。数据库资源管理支持 Oracle、SQLServer、MySQL 关系型数据库连接；平台支持在资源目录中数据库资源节点可以展开，分层预览显示数据库表和字段并且可以添加业务描述。其主要功能包括：

- 支持数据库资源管理功能：包括数据库资源的注册、修改、删除及连接测试等
- 支持多种数据库类型：Oracle、MySQL、SQLServer 等常用关系型数据库；
- 支持数据库表、字段信息和数据的自动采集与预览展示；
- 支持为数据库添加业务信息；
- 数据库资源与数据服务开发功能的交互；
- 支持数据库资源的检索。



## 7.1.3 文件资源管理

数据管理员和开发人员可以通过数据服务共享平台注册、修改、删除源区、前置区、共享区、消费方的文件资源，文件资源可以配置 Agent，用以文件传输。其主要功能包括：

- 文件资源管理，包括文件资源的注册、修改、删除等；
- 文件资源注册页面自动关联信息和用户的输入校验；
- 文件资源与数据服务开发功能的交互；
- 文件传输 Agent，目录授权和目录关联到 Agent；
- 支持文件资源的检索。



## 7.1.4 接口服务管理

数据服务共享平台通过数据服务发布组件发布的数据服务在此进行管理，其主要功能包括：

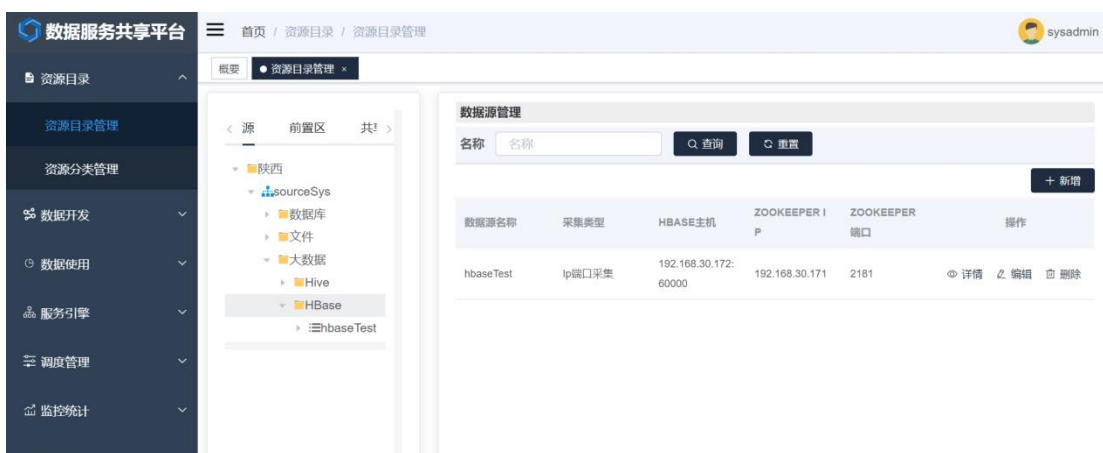
- Web 服务资源管理，包括 Web 服务资源修改，删除功能；
- Web 服务资源注册页面的自动关联信息；
- Web 服务资源与数据服务开发功能的交互，数据服务开发后自动注册在此处；
- 支持 Web 服务资源的检索。



## 7.1.5 大数据资源管理

数据管理员可以通过平台注册、修改、删除大数据资源。大数据服务管理支持 HBase、Hive 等大数据连接。其主要功能包括：

- 大数据资源管理，包括大数据资源的注册、修改、删除及连接测试等；
- 支持 HBase、Hive 的大数据连接；
- 支持 HBase、Hive 元数据的自动采集与展示；
- 大数据库与数据服务开发功能的交互；
- 支持数据库资源的检索。



## 7.1.6 资源分类管理

数据管理员根据业务主题和业务实体（分类及名称可根据需求定义），实现

对数据库、大数据、Web 服务、文件资源的分类管理。其主要功能包括：

- 业务主题管理，包括业务主题的新增，修改，删除功能；
- 业务实体管理，包括业务实体的新增，修改，删除功能；
- 业务实体资源分配和资源查看，其中资源分配要求可实现数据库、文件、接口和大数据的关联；
- 支持分类资源的检索；
- 业务实体已分配的资源，支持快速的服务发布。



## 7.2 数据使用

### 7.2.1 消费方注册

消费者可通过系统管理员进行帐号的创建和权限分配。主要功能包括：

- 基本信息维护；
- 密码初始化、修改；
- 消费者权限分配。



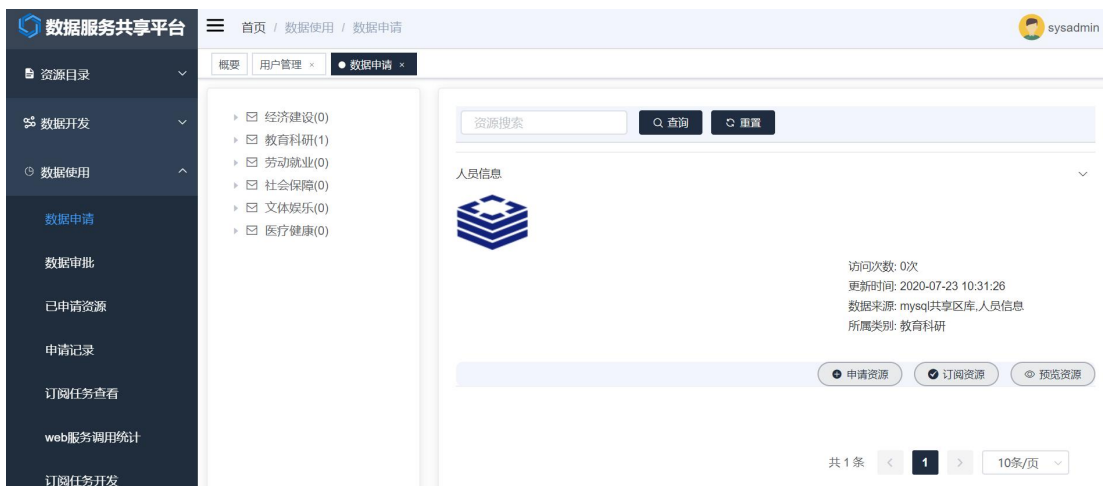


## 7.2.2 数据服务使用

### 7.2.2.1 数据资源申请

消费方通过资源申请方式申请主题分类资源树中的数据实体资源，相对平台而言“拉”的方式，消费方申请资源后并且审批通过，向消费方开放 Web 服务及数据信息，由消费方主动获取资源，主要功能包括：

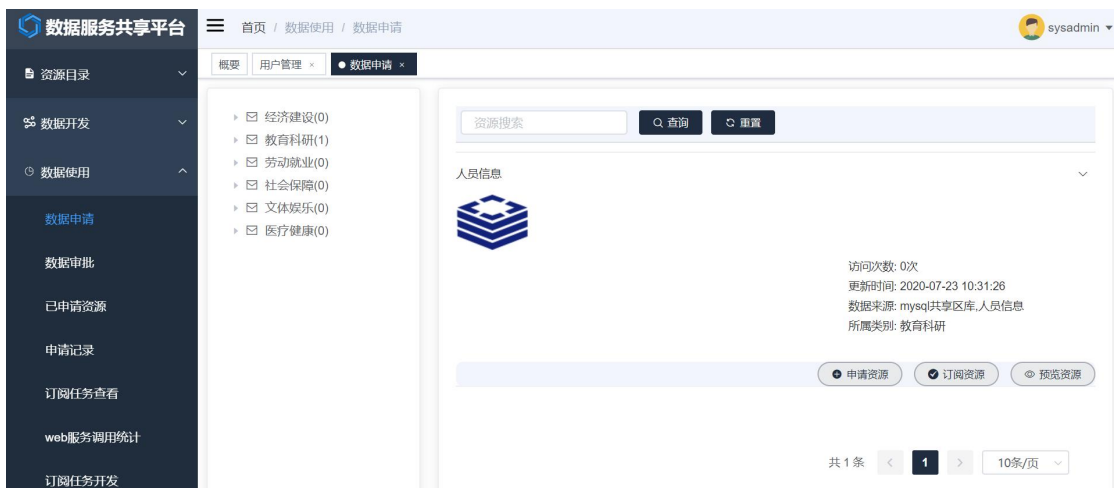
- 数据资源分类展现，数据资源搜索；
- 消费方申请 Web 服务或 DB 类型资源；
- 业务实体汇总；
- 数据库资源预览；
- 支持以 web 服务和表格下载的方式提供数据；
- 支持单表、多表、结果集数据服务发布（发布 RESTful 服务）；
- 资源申请信息维护；
- 已申请记录列表展示。



### 7.2.2.2 数据资源订阅

消费方通过资源订阅在主题分类资源树中的数据实体资源中订阅资源，相对数据服务共享平台而言“推”的方式，消费方订阅资源，向平台提供数据库、文件目录地址，订阅任务审批通过后，由平台向消费方推送资源，主要功能包括：

- 数据资源分类展现，数据资源搜索；
- 业务实体汇总；
- 支持订阅文件和数据库两类资源；
- 支持数据库资源转文件；
- 资源订阅申请信息维护；
- 订阅记录列表展示；
- 数据资源预览。



### 7.2.3 数据服务审批

消费者提交数据资源申请（包括申请和订阅两种方式）后，由数据管理员对申请记录进行审批。申请方式（拉）的申请记录审批通过，消费者可以直接查看到相关资源（包括 Web 服务、数据信息）的地址及获取方式信息；订阅方式的申请记录审批通过后需要数据开发人员根据需要开发相应的数据服务（同时需要消费者注册其对应的数据库地址、文件目录，以向其推送）开发完成后，提供给消费者，开发人员查看申请记录列表。其主要功能包括：

- 申请记录列表查询；
- 申请信息查看；

- 数据服务审批，审批信息维护；
- 服务审批流程，审批过程中对数据行列级权限进行控制；
- 开发人员查看审批结果；
- 订阅方式订阅记录服务开发-在线开发；
- 订阅方式申请记录服务开发-离线开发。

记录名称	实体名称	消费方系统	申请人	申请方式	需求描述	申请时间	审批状态	审批意见	审批时间
singletable 的申请	singletable	consumer.conSys	sysadmin	资源申请		2020-08-19 14:21:25	待审批		
mysql的申请	mysql	consumer.conSys	consumer	资源申请		2020-07-02 17:01:16	审批通过		2020-07-02 17:02:02
hive的订阅申请	hive	consumer.conSys	consumer	资源订阅		2020-07-03 16:10:11	审批通过		2020-07-03 16:10:38
hbase的申请	hbase	consumer.conSys	consumer	资源申请		2020-07-13 09:43:44	审批通过		2020-07-24 11:41:44

## 7.3 数据服务开发

数据服务共享平台（DSP）数据服务开发可分为在线开发和离线开发两种。离线开发是只通过 Studio 进行数据模型开发和 Web 服务的开发，离线开发的模型只需通过平台部署即可，此处不再赘述。本节内容主要是面向在线开发。

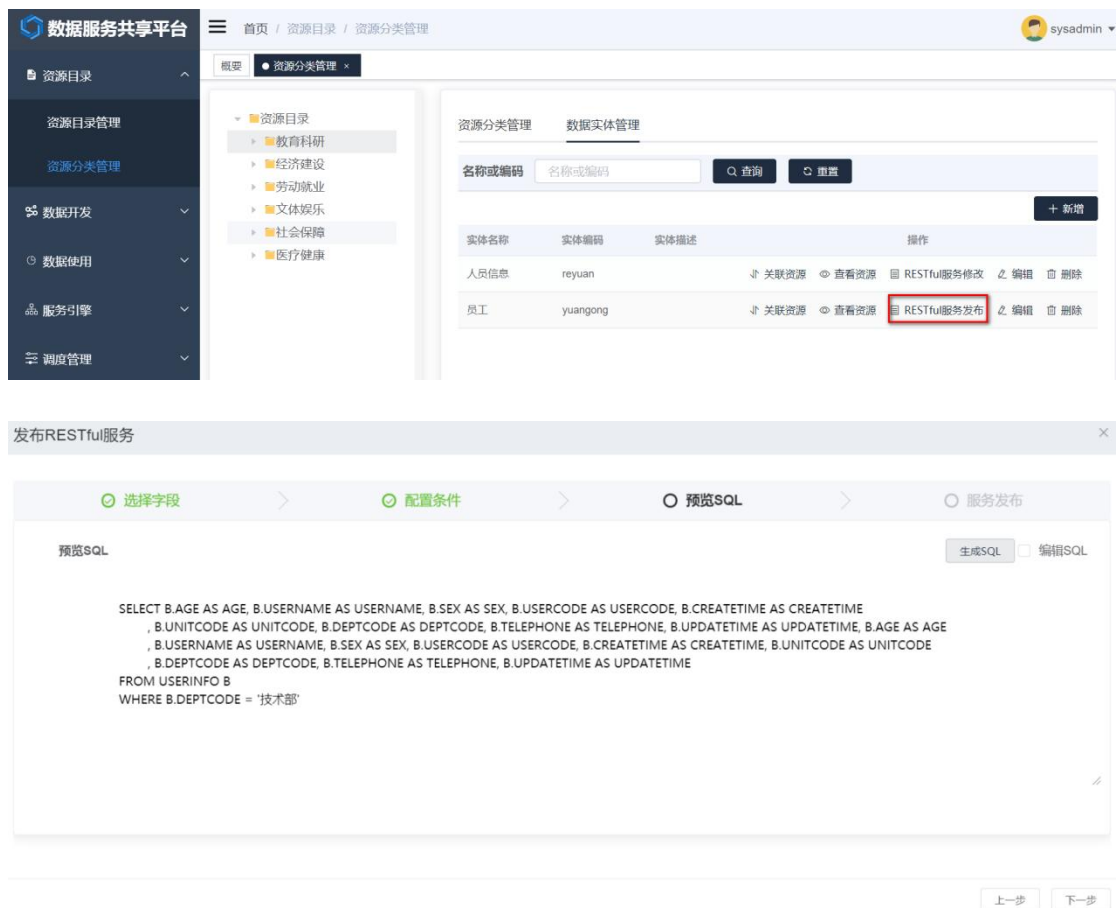
### 7.3.1 实时数据服务

#### 1) 单表服务

通过平台数据资源目录，可以关联单个表资源发布为数据服务，通过界面化配置选择需要发布的表字段、配置查询条件、自动生成 SQL（支持对 SQL 语句进行编辑）几个步骤完成服务发布，并支持服务的自动注册发布。

#### 2) 结果集服务

支持通过平台数据资源目录，可以关联多张表进行数据组装形成新的结果集，并支持对结果集的自定义，并支持按定义的结果集进行服务的自动注册发布。



## 7.3.2 批量数据服务

### 7.3.2.1 结构化

#### 1) 在线作业服务开发

对于在线开发，支持通过页面的方式，根据已经导入到平台中的作业模板配置映射关系，而不用通过 **studio** 进行模型开发即可完成批量数据处理工作。其整体过程为：选择模板→选择区→选择库→批量表配置→变量配置→服务器配，平台在线开发基础模板包含以下几类：

- 全量同步；
- HBase 到文件同步
- 全文对比同步、
- 触发器同步

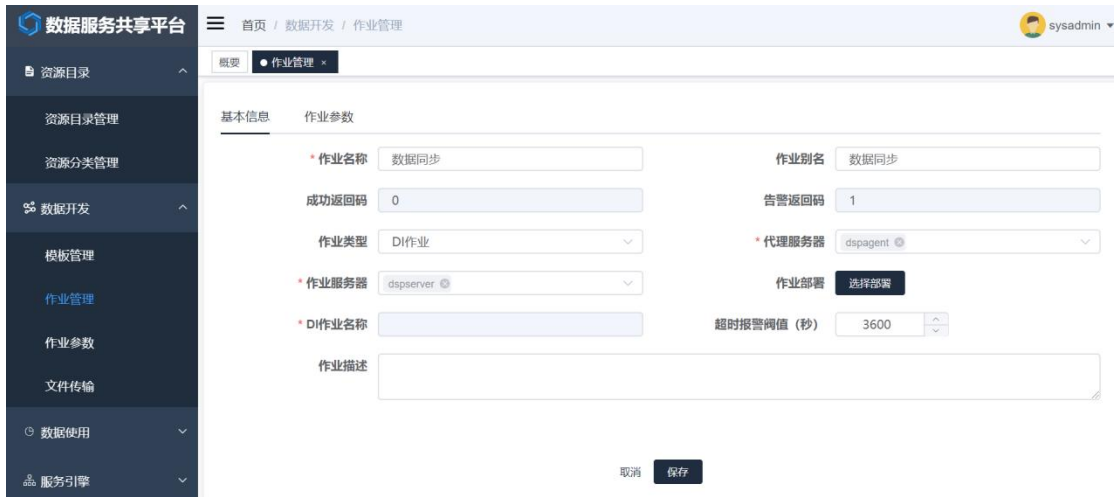
- 时间戳同步
- 数据库到文件同步
- 全量同步
- 文件到数据库同步
- MySQLtoHBASE 数据脱敏
- CanalClientToHBase 同步
- DBToHbase
- MongoDBToHBaseSingleTable
- MongoDBToHBase
- DBToDBDesensitization。



## 2) 离线作业服务开发

对于复杂的作业服务，提供独立的开发工具(DSP Studio)，使开发期和运行期分开，支持多种协议，可通过使用图形化界面进行快速开发，满足用户个性化的需求。功能上与平台具备从设计、开发、组装、调试、维护、部署、管理和发布一体的集成开发环境。同时，具有向导、视图和编辑器等工具供开发人员在开发过程中使用。

对于离线作业开发，支持离线数据模型部署在数据服务共享平台。



### 7.3.2.2 非结构化

非结构化数据服务支持文件的传输配置。通过文件传输配置能够将相应的文件资源传输给消费方。其主要功能包括：

- 新建文件传输配置。
- 编辑文件传输配置。
- 删除文件传输配置。

文件传输配置支持一对一、一对多传输；支持文件传输策略的配置。



新增传输策略
✕

---

基本信息
控制参数

\* 策略名称

\* 发送区

\* 发送文件目录

发送代理

发送目录路径

服务类型

文件类型

发送成功后操作

发送失败后操作

重名处理

\* 接收区

\* 接收文件目录

接收代理

接收目录路径

扫描频率(分钟)

表达式类型

### 7.3.3 订阅任务开发

数据开发人员登录数据服务共享平台，对数据管理员已经审批通过的消费方订阅任务进行开发，订阅任务开发分为离线开发和在线开发，且离线开发和在线开发功能与章节 7.3.2.1 [结构化](#) 相同。

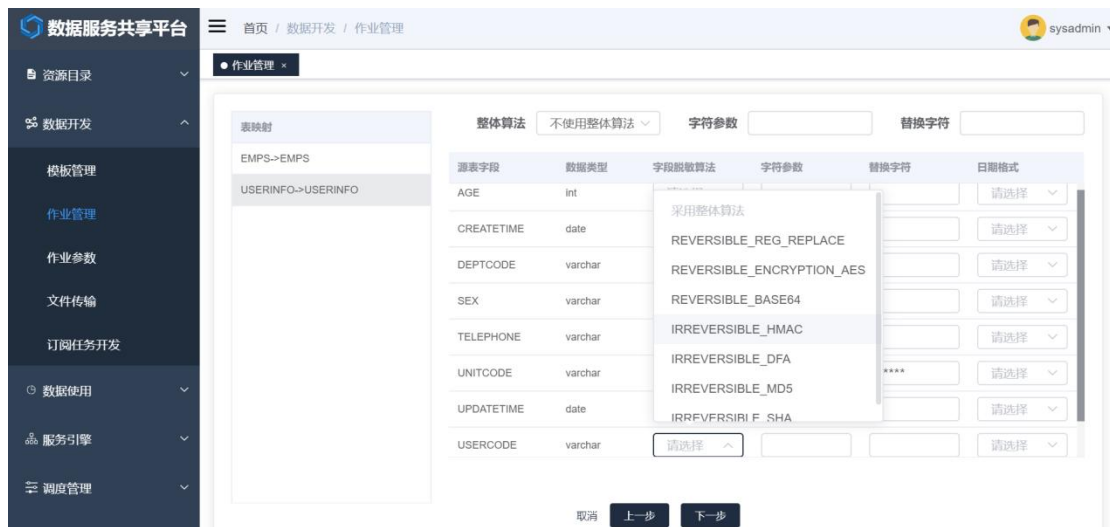
记录名称	实体名称	申请人	需求描述	申请时间	审批状态	审批意见	审批时间	开发完成时间
hive的订阅申请	hive	consumer		2020-07-03 16:10:11	审批通过		2020-07-03 16:10:38	
hive2的订阅申请	hive2	consumer		2020-07-03 16:26:57	数据已发布		2020-07-03 16:27:11	2020-07-03 16:37:45
hbase的订阅申请	hbase	consumer		2020-07-13 09:44:48	数据已发布		2020-07-13 09:45:27	2020-08-07 15:31:30

### 7.3.4 脱敏算法配置

数据服务共享平台数据脱敏模板内置多种数据脱敏算法，可实现替换、加密等处理。其主要功能包括：



- 目标数据源选择;
- 算法配置
  - DFA 算法
  - AES 算法
  - BASE64 算法
  - MD5 算法
- 替换处理。



### 7.3.5 作业模板开发

作业模板实际上是参数化的数据模型。在线作业开发中用到的作业类型都是通过对模板的开发产生的，模板的开发是通过 studio 开发 dbr 或 dsr 模板文件，开发的过程与 ETL 模型开发相同，不同的是模板开发对数据源和表都有新的标记定义，如：数据源填写时用%{source}进行标记，表填写时用[sourcetablename]进行标记。在开发场景中使用到该模板时可对数据源和表的变量进行映射。

在作业中需要进行数据转换、数据质量检验和数据脱敏的时候需要使用到业务调度 dsr 来进行模型组合完成，dsr 能够在模型之间传递变量和结果集。这个是作业流目前所做不到的，需要共享平台提供底层工程化的支撑来完成有前后条件的这种模型的执行。

当前基础模板包含以下几类：

- 全量同步；
- HBase 到文件同步
- 全文对比同步、
- 触发器同步
- 时间戳同步
- 数据库到文件同步
- 全量同步
- 文件到数据库同步
- MySQLtoHBASE 数据脱敏
- CanalClientToHBase 同步
- DBToHbase
- MongoDBToHBaseSingleTable
- MongoDBToHBase
- DBToDBDesensitization。。

除上述模板之外，可依据项目实际需求进行开发。采用模板化的在线开发，能简化使用难度，不需要专业开发人员进行模型的开发和设计。

模板名称	模板别名	创建时间	操作
HBase到文件同步	HBase到文件同步	2017-04-24 00:00:00	
全文对比同步	全文对比同步	2017-04-24 00:00:00	
触发器同步	触发器同步	2017-04-24 00:00:00	
时间戳同步	时间戳同步	2017-04-24 00:00:00	
数据库到文件同步	数据库到文件同步	2017-04-24 00:00:00	
全量同步	全量同步	2017-04-17 00:00:00	
文件到数据库同步	文件到数据库同步	2017-04-17 00:00:00	
MySQLtoHBASE数据脱敏	数据库到HBase数据脱敏	2017-04-17 00:00:00	

## 7.4 服务控制

### 7.4.1 系统调用关系管理

消费者访问数据服务共享平台中 web 服务共享资源，数据管理员给消费者分配 web 服务的访问权限。其主要功能包括：

- 支持 web 服务调用关系管理；
- 支持 Server 端调用关系控制处理；
- 支持 Server 端调用关系控制告警信息处理；
- 支持消费者与 Web 服务调用关系控制关系配置。



### 7.4.2 服务状态管理

消费者访问数据服务共享平台中 Web 服务共享资源，数据管理员对服务状态进行管理，控制服务能否被消费方访问调用。

- 支持对已发布的 web 服务进行启动、暂停、恢复、注销；
- 支持按照操作名称、操作编码、接口名称、接口编码查询。



### 7.4.3 系统 IP 地址管理

消费者访问数据服务共享平台中 web 服务资源，数据管理员给消费者分配 web 服务的 IP 白名单访问权限，对于不在 IP 此名单中的用户，不能访问数据服务资源。其主要功能包括：

- web 服务 IP 白名单控制管理；
- Server 端 IP 白名单控制处理；
- Server 端 IP 白名单控制告警信息处理；
- 消费者与 Web 服务 IP 白名单控制关系配置。



## 7.5 调度管理

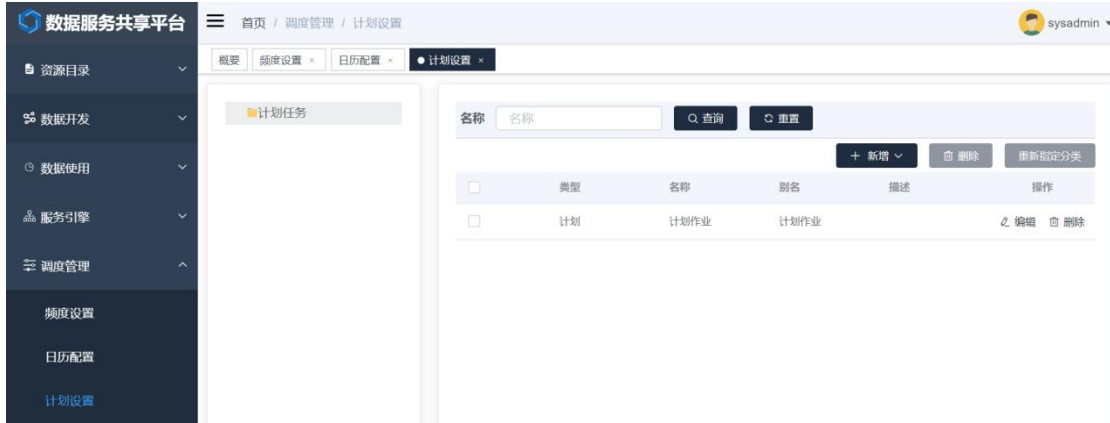
### 7.5.1 作业策略管理

支持自定义作业执行的频度、日历信息，可以为不同作业定义不同调度频率和日历。



### 7.5.2 作业计划场景

数据服务共享平台为开发人员提供按照计划自动执行作业的能力，支持在计划中配置自定义的频度、日历、作业、作业执行优先级等内容。



### 7.5.3 作业图形编排

数据服务共享平台为开发人员提供图形编排调度作业的能力，包括作业的设计优先级，依赖关系等。



## 7.5.4 事件触发

数据服务共享平台为开发人员提供基本的调度管理触发功能，支持通过事件触发作业的执行，通过文件到达的方式触发作业、作业流的运行，其主要功能包括：

- 支持按照名称进行查询；
- 时间基本信息维护、修改、查看；
- 与已创建的作业、作业流关联。



## 7.6 运行监控

### 7.6.1 交换监控

数据服务共享平台上的作业和文件传输运行状态的监控，其中作业状态展示最近一次作业的运行状态数据，文件传输则显示当前的实时传输进度。

- 前置交换监控查询浏览
  - 从前置区角度查看最近一次作业运行状态
  - 从前置区角度查看文件传输实时进度
- 共享交换监控查询浏览
  - 从共享区角度查看最近一次作业运行状态
  - 从共享区角度查看文件传输实时进度

作业名称	触发类型	状态	开始时间	结束时间	运行时长	增加	修改	删除	操作
数据同步	单作业	成功	2020-07-27 17:02:04	2020-07-27 17:02:05	0天0小时0分1秒	10	0	0	🔍 📄 🗑️ 🔄
数据同步	计划触发	成功	2020-07-22 18:00:15	2020-07-22 18:00:16	0天0小时0分1秒				🔍 📄 🗑️ 🔄
在线作业	单作业	失败	2020-07-23 11:06:45	2020-07-23 11:06:45	0天0小时0分1秒				🔍 📄 🗑️ 🔄
自定义条件过滤	单作业	成功	2020-07-27 17:02:54	2020-07-27 17:02:55	0天0小时0分1秒	0	2	0	🔍 📄 🗑️ 🔄

### 7.6.2 消费者监控

监控已经注册到系统当中的消费者提出的数据申请记录以及实际进行共享区数据的使用记录。使用记录则记录包括了：

1. 通过作业将共享区资源传输到消费者指定数据源中的数据量；
2. 通过文件传输推送到消费者指定文件夹的文件记录。



主要功能包括：

- 根据消费方进行文件、作业数据使用量统计查询

The screenshot shows the 'Data Service Sharing Platform' (数据服务共享平台) interface. The breadcrumb path is '首页 / 监控统计 / 消费方统计'. The user is 'sysadmin'. The page title is '消费方统计'. The main content area is titled '作业 文件服务'. There are search filters for '消费方' (Consumer), '开始时间' (Start Time), and '结束时间' (End Time). Below the filters is a table with the following data:

消费方	业务实体名称	作业名称	新增量	更新量	删除量	错误数	成功数	失败数	操作
sysadmin		hive2xt在线	0	0	0	0	0	0	详情
sysadmin		hbase2xt0708	0	0	0	0	0	1	详情
sysadmin		mysql02hbase	0	0	0	0	0	1	详情

At the bottom right, there is a pagination bar showing '共 3 条' (Total 3 items), page '1', and '10条/页' (10 items per page).

### 7.6.3 计划、作业流监控

对作业计划的执行情况进行监控，对作业流执行情况监控，主要包括：

- 作业计划信息查看
- 作业计划执行历史信息查查
- 作业计划执行日志信息查看
- 作业流详情信息查看
- 作业流执行历史信息查看
- 作业流执行详细日志查看

路径	计划任务名称	计划上次时间	计划下次时间	操作
计划任务\hive\订阅计划703	hive\订阅计划703	2020-07-20 18:46:00	<NO>	🔍 📄 🗑️
计划任务\jhua2	jhua2	<NO>	<NO>	🔍 📄 🗑️
计划任务\jhua444	jhua444	<NO>	<NO>	🔍 📄 🗑️
计划任务\plan1	plan1	<NO>	<NO>	🔍 📄 🗑️
计划任务\全量同步\orc11g	全量同步\orc11g	2020-08-03 18:02:00	<NO>	🔍 📄 🗑️
计划任务\计划0807	计划0807	2020-08-10 10:58:00	<NO>	🔍 📄 🗑️

计划名称	作业/作业流名称	类型	状态	执行时间	日志信息	操作
hive\订阅计划703	hive2的订阅申请的在线作业	作业	🟢	2020-07-20 18:44:15	Perform operation command sent successfully!	📄
hive\订阅计划703	hive2的订阅申请的在线作业	作业	🟢	2020-07-20 18:42:17	Perform operation command sent successfully!	📄
hive\订阅计划703	hive2的订阅申请的在线作业	作业	🟢	2020-07-20 18:40:16	Perform operation command sent successfully!	📄
hive\订阅计划703	hive2的订阅申请的在线作业	作业	🟢	2020-07-20 18:38:16	Perform operation command sent successfully!	📄

## 7.6.4 故障告警监控

数据服务共享平台中作业、文件传输任务和 Web 服务在中执行时，可能存在着运行时的错误和异常，这些异常信息可以在故障告警监控中进行浏览，浏览告警信息可以查阅到故障的服务和具体的作业、传输任务，通过这些信息为排除故障做出调整提供参考依据。其主要功能包括：

- 收集各个服务引擎的故障信息
  - 存储来自服务引擎接口的故障信息
  - 故障信息的定时清理
- 查询故障信息
  - 查询故障信息，其中包含足以定位故障位置的信息

- 异常信息预警
  - 对出现故障的信息，通过消息、短信或邮件的方式向管理人员进行提示

## 7.6.5 元数据采集监控

提供采集日志界面，对元数据采集结果进行监控，主要包括：

- 采集器连接测试
- 采集任务日志查询
- 采集任务详情查看

The screenshot shows the '元数据采集日志' (Metadata Collection Log) page in the '数据服务共享平台'. The page features a search bar for '数据源名称' (Data Source Name) with '查询' (Search) and '重置' (Reset) buttons. Below the search bar are two buttons: '测试采集器连接' (Test Collector Connection) and '取消任务' (Cancel Task). The main content is a table with the following columns: '任务名称' (Task Name), '数据源名称' (Data Source Name), '任务状态' (Task Status), '起始时间' (Start Time), '更新时间' (Update Time), and '操作' (Action). The table lists several tasks, all with a status of '入库完毕' (Imported). The tasks are as follows:

任务名称	数据源名称	任务状态	起始时间	更新时间	操作
DB321	mysql0818	入库完毕	2020-08-18 09:28:27	2020-08-18 09:29:14	详情
DB321	mysql0818	入库完毕	2020-08-18 09:23:59	2020-08-18 09:24:47	详情
DB301	test1	入库完毕	2020-08-13 16:45:38	2020-08-13 16:46:26	详情
DB301	test1	入库完毕	2020-08-13 16:43:52	2020-08-13 16:44:14	详情
DB301	test1	入库完毕	2020-08-13 10:10:46	2020-08-13 10:11:20	详情
DB301	test1	入库完毕	2020-08-13 10:02:05	2020-08-13 10:02:54	详情

## 7.6.6 历史记录查询

可以查询共享区或者交换区中作业、文件传输任务的执行历史，历史记录中记录每次作业或者任务执行的最终状态。其主要功能包括：

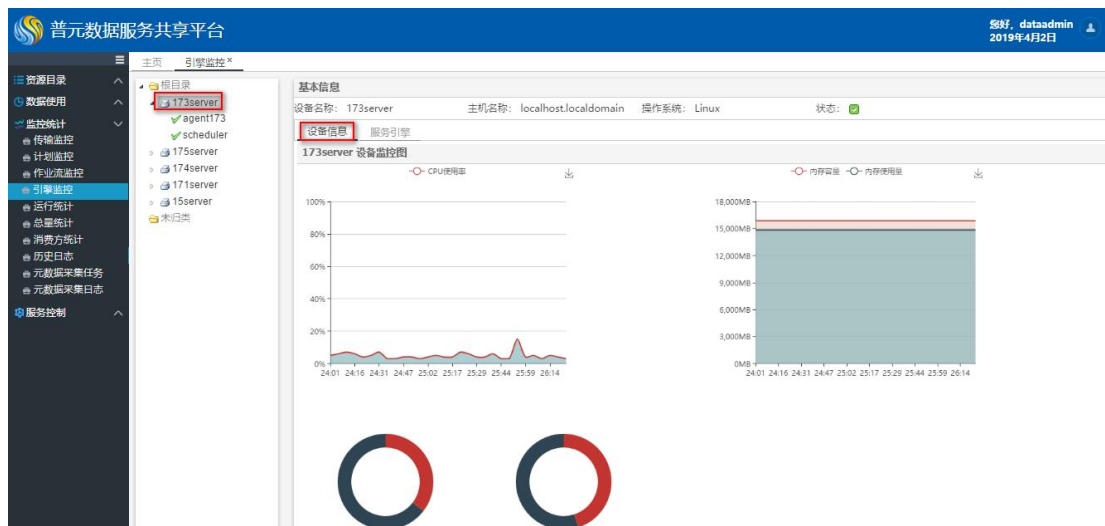
- 文件传输历史记录查询
- 作业历史记录查询



## 7.6.7 物理资源监控

交换平台所有物理设备之间的通信关系，服务引擎所在设备的物理资源使用情况,监控的物理资源包括 CPU、内存、硬盘。其主要功能包括：

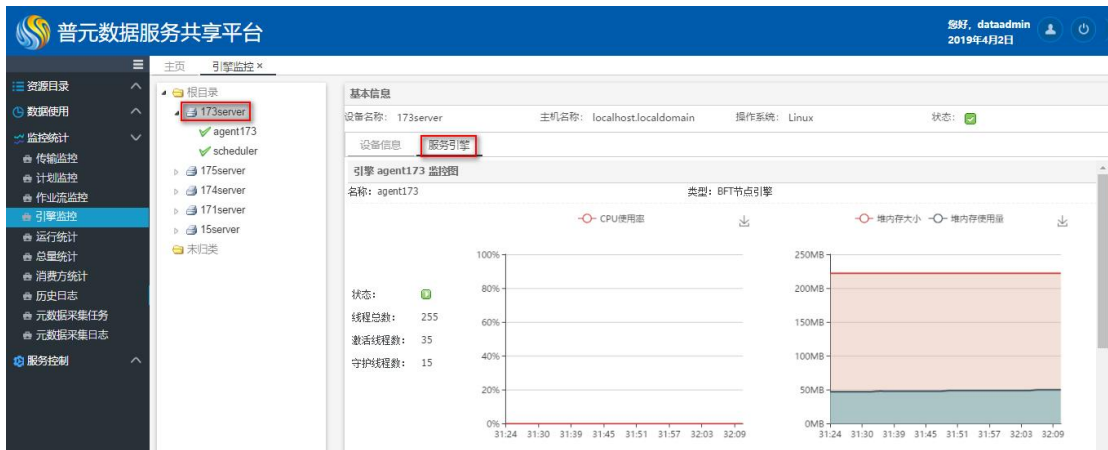
- 收集各个物理设备资源的使用状况
  - 收集 CPU 负载
  - 收集内存负载
  - 收集存储设备信息
  - 物理资源使用状况的刷新
- 展示物理资源使用数据
  - 图表方式展示各个物理设备的运行状况；
- 物理设备信息维护
  - 物理设备信息的注册、注销、更新
  - 物理设备的浏览



## 7.6.8 服务引擎监控

部署在物理设备上的服务引擎运行状态监控，以及各个引擎产生的系统日志。运行状态和日志查看为排查进程异常和服务错误提供依据。引擎状态数据包括了服务进程是否正常、JVM 运行情况等。其主要功能包括：

- 收集服务引擎运行状态数据
  - 收集服务引擎 JVM 内存区域状态
  - 收集服务引擎的进程状态
  - 收集服务引擎 JVM 的 CPU 负载状态
- 展示物理设备上的服务引擎运行状态
  - 服务引擎域部署的物理设备之间关联
  - 提供服务引擎状态接口，定时获取状态信息
  - 图表方式浏览指定设备上的所有服务引擎信息
- 服务引擎系统日志查看
  - web 浏览器中查看指定服务引擎的系统日志
  - 服务引擎系统日志集中收集



## 7.7 统计分析

### 7.7.1 交换总量

对源和前置区或者前置区和共享区、共享区和消费者之间数据交换的数据量进行监控。对作业执行的数据总量（条数）等进行监控统计。 通过从提供者的纬度来对数据指标进行显示。其主要功能包括：

- 节点到节点之间数据交换总量统计
- 库到库之间数据交换总量统计
- 表和表之间数据交换总量统计
- 缓存统计（定期计算统计维度指标）



## 7.7.2 作业统计

数据管理员根据配置好的源和前置区、前置区和共享区之间的作业模型调度进行数据交换，对作业执行情况进行统计。其主要功能包括：

- 基础数据指标计算和统计
- 页面图表展示
- 页面查询列表展示



作业ID	作业名称	新增量	更新量	删除量	错误数	成功数	失败数	操作
101	orc11g211g-全量	4	0	0	0	0	0	详情
142	hive2txt在线	194	0	0	0	0	0	详情

## 7.7.3 文件统计

数据管理员根据配置好的源和前置区，前置区和共享区之间进行文件传输，对文件传输的数据进行监控。其主要功能包括：

- 基础数据指标计算和统计
- 页面图表展示
- 页面查询列表展示

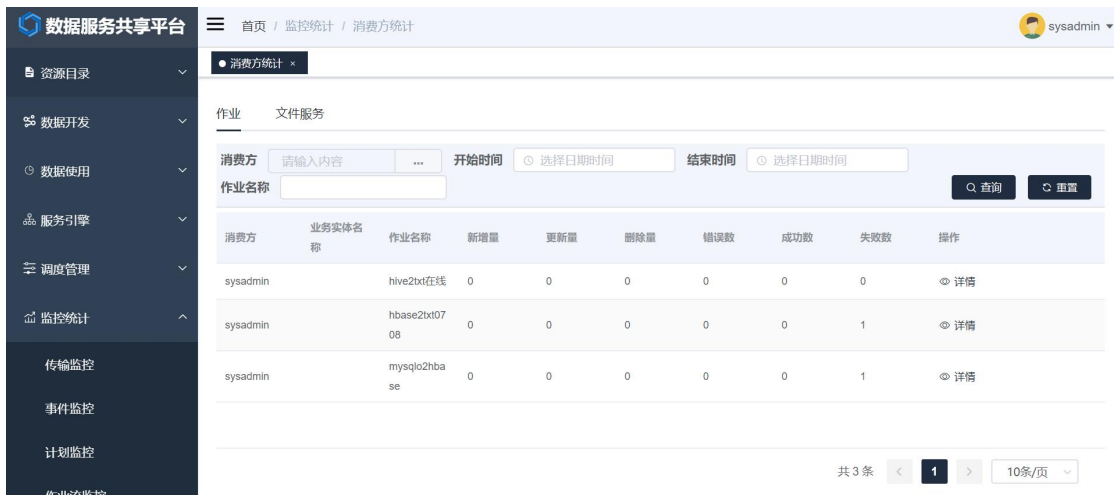
源	目标	服务名称	服务类型	总量	成功文件数	失败文件数	操作
资源	资源		本地传输	257.65 KB	1	0	执行日志
资源	资源		本地传输	0.00 KB	0	0	执行日志
资源	资源		本地传输	0.00 KB	0	0	执行日志
资源	资源		本地传输	1.24 KB	2	0	执行日志
资源	资源		本地传输	68.26 KB	2	0	执行日志

## 7.7.4 消费方数据统计

消费方数据统计主要是消费者和共享区之间进行数据交换,对传输的数据量进行监控和统计,主要针对 DI 作业执行情况、文件传输、Web 服务调用情况进行监控和统计。其主要功能包括:

- DI 作业运行产生数据流转的统计和分析
  - 基础数据指标计算和统计
  - 页面图表展示
  - 页面查询列表展示
- 文件传输产生的数据进行统计和分析
  - 历史数据统计分析
  - 页面图表展示
  - 页面查询列表展示
- Web 服务接口产生数据进行统计和分析
  - web 服务调用过程中基础数据采集
  - 基础历史数据指标计算和统计
  - 页面图表展示
  - 页面查询列表展示





## 7.7.5 Web 服务统计

对服务 Web 服务接口调用情况提供统计查询界面，方便用户对服务的运行情况进行统计监控。其主要功能包括：

- 服务消费方维度 Web 服务统计
- 服务提供者维度 Web 服务统计
- Web 服务调用情况统计
- 查询缓存

