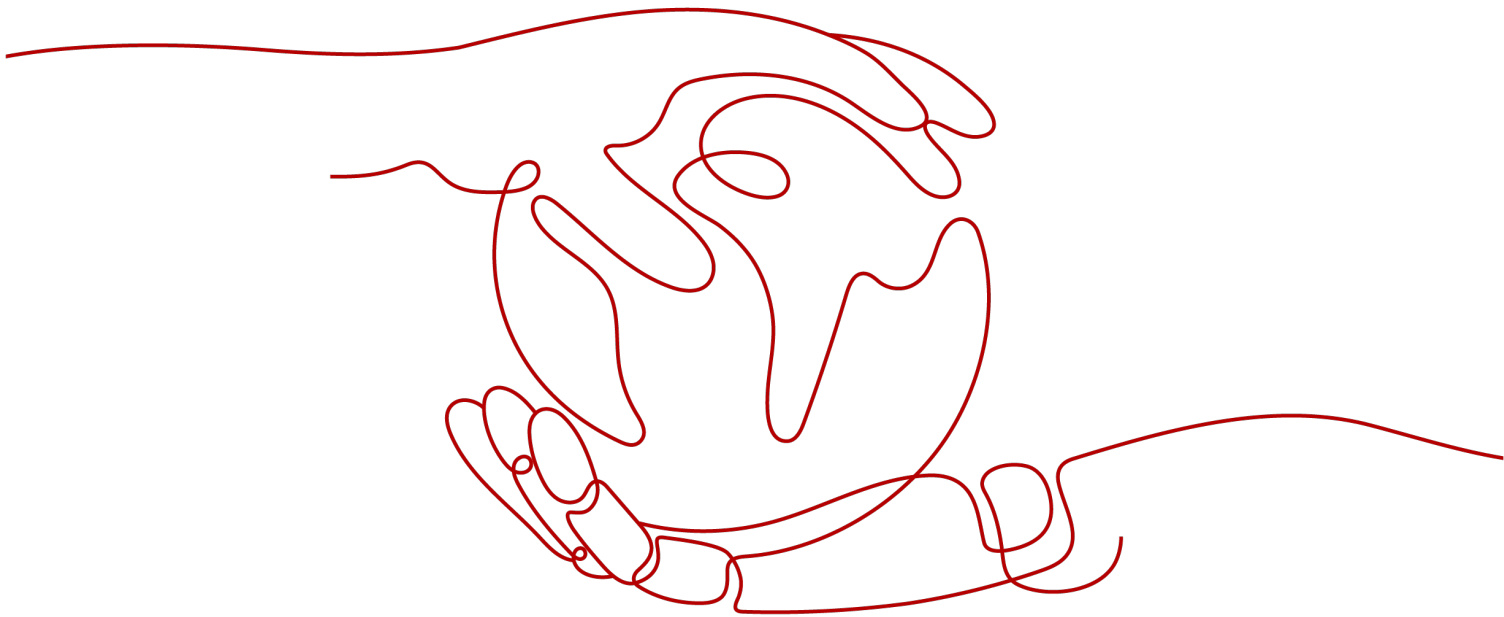


解决方案实践

华为云政务场景数据使能解决方案实践

文档版本 1.2
发布日期 2024-04-24



版权所有 © 华为技术有限公司 2024。保留一切权利。

非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。

商标声明



HUAWEI和其他华为商标均为华为技术有限公司的商标。

本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

注意

您购买的产品、服务或特性等应受华为公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为公司对本文档内容不做任何明示或暗示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

安全声明

漏洞处理流程

华为公司对产品漏洞管理的规定以“漏洞处理流程”为准，该流程的详细内容请参见如下网址：

<https://www.huawei.com/cn/psirt/vul-response-process>

如企业客户须获取漏洞信息，请参见如下网址：

<https://securitybulletin.huawei.com/enterprise/cn/security-advisory>

目录

1 方案概述	1
2 资源和成本规划	9
3 实施步骤	13
3.1 环境确认	13
3.2 系统配置	16
3.3 基础组件部署	25
3.3.1 Nginx	25
3.3.2 全文检索	29
3.3.3 统一支撑平台	30
3.3.4 一体化数据管理平台	34
3.3.5 共享门户	39
3.3.6 开放门户	44
3.3.7 数据交换平台管理端	47
3.3.8 数据交换节点	51
3.4 场景实践	56
3.4.1 概述	56
3.4.2 库表采集	56
3.4.3 数据质检	58
3.4.4 构建人口库模型、分析来源并接入	60
3.4.5 落标与清洗	62
3.4.6 融合与发布	64
3.4.7 挂载资源目录	66
4 附录	68
5 修订记录	70

1 方案概述

应用场景

某市一体化政务数据中台为一体化政务服务平台提供数据支撑，提升一体化政务服务平台服务能力，能够为政务服务智能审批、一证通办等改革创新、“跨区域、跨部门、跨层级”业务协同提供技术和数据支撑，打通数据壁垒，实现“数据多跑路、群众少跑腿”的政务服务目标，将“便民、利民”的目标落到实处。

总体架构设计

- 总体架构

本项目总体架构设计如下图所示：

图 1-1 总体架构



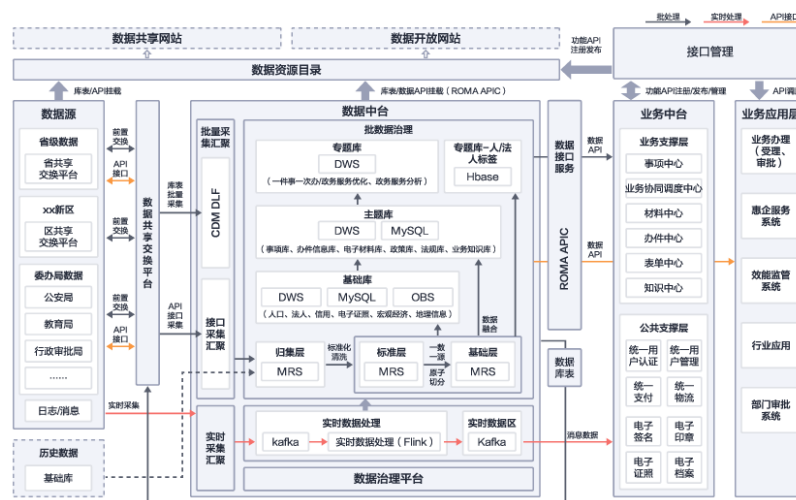
- 基础设施层：云基础设施，为上层应用提供基础支撑。
- 数据资源池层：建设数据资源池，建设内容包括基础库建设、主题库建设以及专题库建设。基础库建设包括了人口库、法人库、电子证照库、信用信息库、自然资源与地理信息库、宏观经济库。主题库建设包括了事项库、电子材料库、办件信息库、政策法规库、业务知识库、工建项目库。专题库建设包括了一件事一次办专题库、政务服务分析专题库、政务服务优化专题库、标签库。

- c. 数据平台层：平台层主要包括了数据支撑建设和接口管理平台。数据支撑建设包括了数据资源目录平台、数据交换平台、数据治理平台、数据共享平台、数据开放平台、数据服务体系、数据平台对接管理。接口管理平台提供接口目录管理功能。

● 数据架构

本建设方案聚焦数据资源层建设方案，为业务层提供数据支撑，并与省级共享交换平台、市级各委办局进行集成对接。某市数据中台对数据实现从归集到沉淀至数据中台仓库的全流程管控。通过在部门终端布置部门交换节点采集部门业务系统数据，运用共享交换能力，形成部门资源库；数据中台通过数据采集汇聚中心端，采集数据形成中台的数据仓库，以业务需求出发，建设主题库、专题数据库；通过对数据的标签化处理，抽离出数据的特性值，建设数据的多场景应用。

图 1-2 数据架构

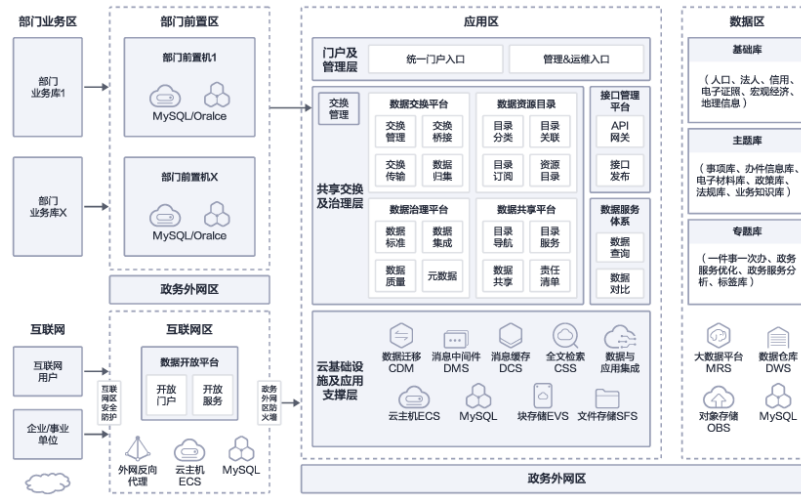


● 部署架构

- a. 本架构基于某一体化政务服务平台项目实践输出，作为最佳实践案例。
- b. 系统部署在政务外网区域，对政务外网区域进行分区分层结构规划，共包括外部门前置区（互联网区）、应用区、数据存储区。

其中，互联网区中部署反向代理服务，所有的web请求通过代理进入业务服务器。应用区包含了数据中台所有子系统的安装部署区域，区域通过防火墙和外网映射区隔离，保证应用安全。数据存储区包含了所有的数据存储服务，通过防火墙进行隔离，进一步加强数据的安全保障。

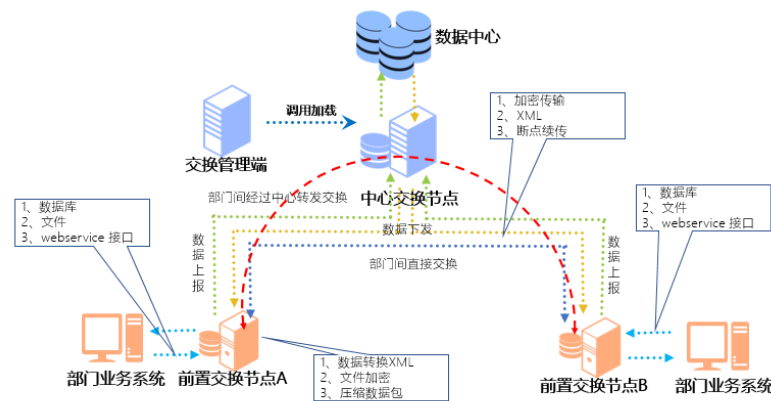
图 1-3 部署架构



● 交换体系架构

交换体系架构主要由交换管理端、交换节点组成。交换节点又可以分成分中心节点和前置节点。数据交换管理端能够管理与监控各交换节点，而交换节点能够根据业务人员的配置，从数据源中抽取数据，并推送到目标数据源中，完成指定业务的交换。通过交换体系可以实现部门到中心、中心到部门和部门到部门的数据交换。

图 1-4 交换体系架构



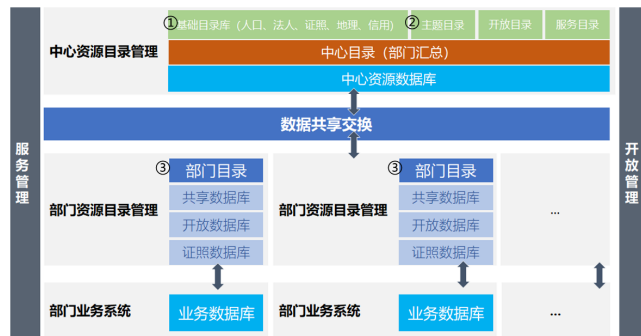
● 目录体系架构

目录体系架构主要包含：部门目录、基础目录、主题目录。

- a. 部门目录：各部门维护注册各部门的信息资源，并将信息资源数据存到部门前置库、汇总各部门的目录。
- b. 基础目录：根据基础库模型，将多个部门数据融合后形成基础目录。
- c. 主题目录：根据主题库模型，将多个部门数据、基础库数据融合后形成主题目录。

所有资源的共享都通过目录驱动，最终呈现为数据共享门户上统一的服务目录。目录系统与数据共享平台对接，实现目录检索、页面联机查询、数据交换、API调用等多种方式的共享。

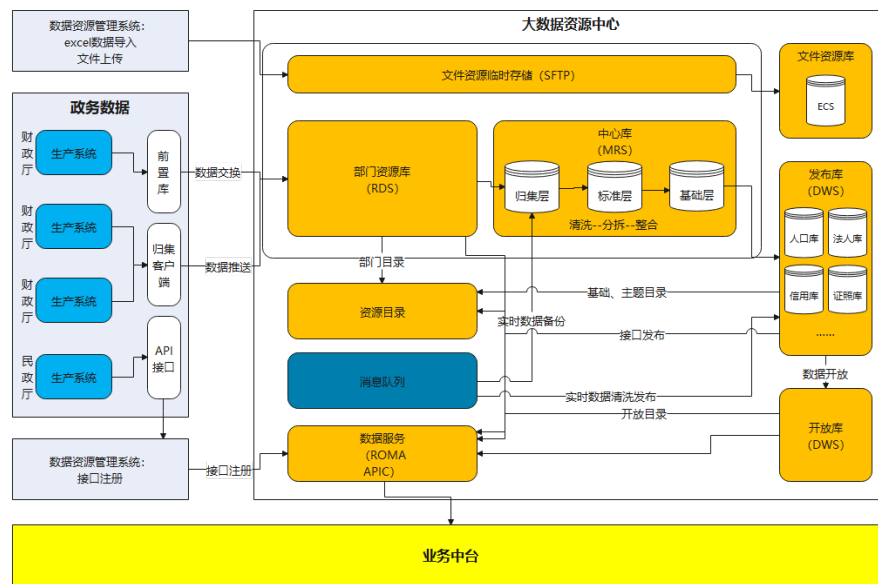
图 1-5 目录体系架构



● 数据流向设计

大数据资源中心依托数据共享交换平台构建，各类数据通过数据共享交换平台数据归集及数据交换系统进行数据归集和交换后进入部门资源库，通过数据共享交换平台数据集成和数据治理系统将数据接入中心库，在中心库进行清洗、融合等体系化治理，并发布到各类基础库，主题库、专题库。数据共享交换平台具体数据共享交换平台构架图如下：

图 1-6 数据流向设计



● 标准规范体系

标准规范建设是数据信息资源整合共享开放实现统一服务、业务协同、安全可靠运行的前提和基础。早期的信息系统建设，普遍存在缺乏标准或

标准不统一的现象，最终导致管理混乱、互联互通不畅、信息共享程度低、信息资源开发利用滞后、安全存在隐患等后果，严重影响了信息化建设的进程。

标准规范体系建设对推进项目建设，保证信息的交流与共享，具有重要意义。整理汇总国家、省标准规范，结合实际情况，制定数据中台数据标准、服务标准、安全标准、管理标准等，确保数据中台与中省平台、与部门自建业务系统无障碍对接，为平台对接、扩展、升级迭代的标准化规范化提供保障。

表 1-1 标准规范体系

分类	子分类	子分类规范	规范名称
01 数据标准	01-01 数据元	数据元规范	数据元规范 第1部分：总则
			数据元规范 第2部分：公共数据元
			数据元规范 第3部分：综合人口数据元
			数据元规范 第4部分：综合法人数据元
			数据元规范 第5部分：社会信用数据元
			数据元规范 第6部分：电子证照数据元
			数据元规范 第7部分：自然资源和空间地理数据元
			数据元规范 第8部分：宏观经济数据元
	01-02 数据库	数据模型设计规范	数据模型设计规范
		基础库数据模型	基础库数据模型 第1部分：综合人口模型设计
			基础库数据模型 第2部分：综合法人模型设计
			基础库数据模型 第3部分：社会信用模型设计
			基础库数据模型 第4部分：电子证照模型设计
			基础库数据模型 第5部分：自然资源和空间地理技术规范
			基础库数据模型 第6部分：宏观经济技术规范
		基础库实施规范	基础库实施规范 第1部分：综合人口实施规范
			基础库实施规范 第2部分：综合法人实施规范
			基础库实施规范 第3部分：社会信用实施规范
			基础库实施规范 第4部分：电子证照实施规范
			基础库实施规范 第5部分：自然资源和空间地理库实施规范
基础库数据模型 第6部分：宏观经济实施规范			

分类	子分类	子分类规范	规范名称
	01-03 资源目录	目录编制规范	编制规范：信息资源目录及数据资源目录编制规范
	01-04 数据共享	共享交换规范	共享交换规范 第1部分：数据归集规范
			共享交换规范 第2部分：部门接入规范
			共享交换规范 第3部分：数据交换规范
			共享交换规范 第4部分：服务接口规范
			数据共享交换平台管理办法（试行）
共享评价指标体系	数据共享评价指标体系		
02 服务标准	02-01 基础库服务	基础库数据接口规范	基础库数据接口规范 第1部分：综合人口接口规范
			基础库数据接口规范 第2部分：综合法人接口规范
			基础库数据接口规范 第3部分：社会信用接口规范
			基础库数据接口规范 第4部分：电子证照接口规范
			基础库数据接口规范 第5部分：自然资源和空间地理接口规范
			基础库数据接口规范 第6部分：宏观经济接口规范
	02-02 主题库服务	主题库数据接口规范	主题库数据接口规范
02-03 专题库服务	专题库数据接口规范	专题库数据接口规范	
02-04 省/市/平台统计接口	省/市/平台统计接口规范	省/市/平台统计数据接口规范	
03 安全标准	数据安全	数据安全防护规范	数据安全防护规范
		数据分级分类管理规范	数据分级分类管理规范
04 管理标准	数据管理	数据全生命周期管理规范	数据全生命周期管理规范

优势说明

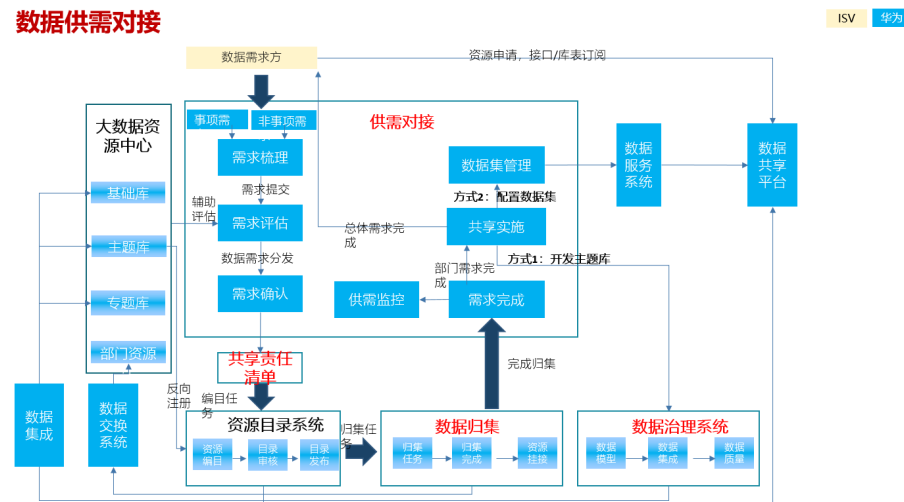
- 政务行业资产沉淀
通过项目实践沉淀基础库/主题库数据模型资产、数据API等集成资产；提升交付效率，降低交付成本；
- 一站式数据治理与管控平台
提供贯穿政务数据全流程一站式治理、开发、运营平台，全方位的数据探索稽查和分析，实现委办局多源数据清洗、标准化和主题联接，生成标准、清洁的数据；
- 数据实时共享和跨库分析
数据实时接入，数据共享时效从小时级缩短至秒级，满足政务联办等实时业务场景诉求；支持跨库数据查询分析，数据免搬迁，数据查询分析效率从小时级缩短到分钟级；

特色应用场景：

1. 原始老系统建设中，数据中台和业务中台的供需对接通过线下完成，效率低，不好跟进，不好闭环，不好回顾，现在通过数据供需对接模块将业务中台建设过程中对数据中台数据的业务数据需求从线下转到线上。

图 1-7 特色应用场景 1

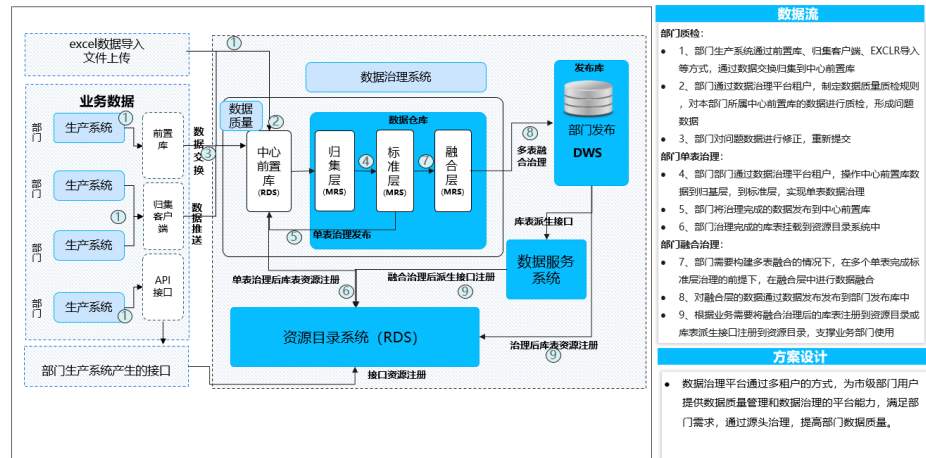
数据供需对接



2. 原先x云服务商xx委办局数据只能由大数据治理平台的团队实现数据治理、部门数据治理需要自建平台。现在通过源头治理的模式，赋能委办局使委办局使用数据治理平台治理自己的数据。

图 1-8 特色应用场景 2

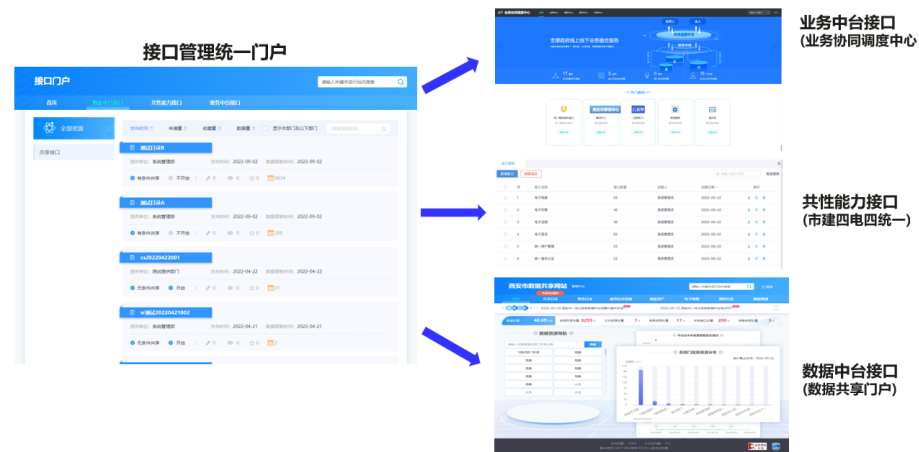
源头治理：标准落工具，委办局自助清洗标准化



3. 接口管理统一门户，为业务中台接口、共性能力接口和数据中台接口提供入口

图 1-9 特色应用场景 3

接口管理统一门户，为三类接口提供统一门户入口



方案成熟度：

通用政务数据中台一般包含以下建设内容：

1. 共享交换平台
2. 共享交换实施
3. 数据治理平台
4. 数据治理实施
5. 基础库
6. 主题库

本项目建设包含以上所有内容。

2 资源和成本规划

资源规划

表 2-1 资源规划

分类	服务名称	配置	量纲	备注
数据资源层	MRS	MRS 管控节点 (32U128G, 4*1.6TB数据盘)	3	2管理+3控制节点, 需要 zookeeper, hdfs, hive, hbase, spark, yarn, ranger, hetu
		MRS 计算节点 (32U128G, 4*1.6TB数据盘)	7	批处理计算, 26TB
		MRS 计算节点 (32U128G, 2*0.8TB数据盘)	3	人法标签库, Hbase
	DWS	DWS 计算密集型 (32U128G 2.98T有效容量) (SSD)	3	基础/主题/专题发布库
	ROMA	企业版, 支持30系统, 200联接	1	生成数据 API, 接口数量 450个, 未估算并发性能, 使用 APIC 功能
基础组件	ECS	通用型 8C16G	7	内网反向代理、网关反向代理、xx容器平台、SFTP、livy、presto、nfs、kafka、flink
	LTS	日志 (给 ROMA 做归档)	1	
	ELB	默认小规格	2	
	ECS	通用型 16C32G	9	
	Redis主备版 16G	主备版 16G	1	系统缓存服务使用

分类	服务名称	配置	量纲	备注
	RDSMySQL 关系型数据库服务	32C128G 主备双实例	1	MHA 集群计算节点系统业务库
	mangoDB	16C64G 三副本	1	地理信息库
	RDSMySQL 关系型数据库服务	32C128G 主备双实例	1	MHA 集群计算节点部门资源库
	RDSMySQL 关系型数据库服务	32C128G 主备双实例	1	
	RDSMySQL 关系型数据库服务	32C128G 主备双实例	1	
	RDSMySQL 关系型数据库服务	32C128G 主备双实例	1	
	RDSMySQL 关系型数据库服务	32C128G 主备双实例	1	
	云数据库 RDS(磁盘)	23000G	1	MHA 集群计算节点系统业务库存储, MHA 集群计算节点部门资源库存储
数据盘	云硬盘存储 EVS	高性能 SSD 云硬盘	1	71TB
应用支撑	ECS	通用型 8C16G	1	统一支撑、统一认证、统一全文检索、网关服务、统一调度、报表服务 (smartBI)
	ECS	通用型 16C32G	2	
			1	

分类	服务名称	配置	量纲	备注
业务系统	ECS	通用型 8C16G	4 3	一体化数据管理平台（edc）、共享门户（sharedportal）、开放门户（extranet）、定时任务执行引擎（edctask）、数据接口调度引擎（edcservice）、数据质检服务引擎（dataqualitytask）、级联系统（cascade）、数据集成开发平台（dpxmanager）、交换服务引擎（dpxnode）、治理服务引擎（dpxnode）、实时任务服务引擎（cdcnode）、标签画像系统（eianalyse）、标签画像服务引擎
	ECS	通用型 16C32G	2	
异机备份服务器	ECS	通用型 8C16G	1	备份服务器
工作台	-	通用型 8C32G	1	工作台（windows sever2012系统）
	ECS	通用型 4C8G	9	
备用服务器	ECS	通用型 8C16G	5	应急备用
测试系统	-	通用型 32C128G	1	测试系统
	ECS	通用型 8C16G	4	
数据开放库	RDSMySQL 关系型数据库服务	8C32G 主备双实例	1	数据开放库
备份数据库	RDSMySQL 关系型数据库服务	8C32G 主备双实例	2	备份数据库
委办局前置库	RDSMySQL 关系型数据库服务	8C16G 主备双实例	3 6	各委办局前置库
	RDSMySQL 关系型数据库服务	16C64G 主备双实例	1	办件上报与省办件回流专库

分类	服务名称	配置	量纲	备注
	云数据库 RDS (磁盘)	32T	1	数据开放库、备份数据库、各委办局前置库、办件上报与省办件回流专库，数据库存储

成本规划

本方案基于 华为云Stack Online 底座部署，属于基地云模式。资源开通向基地发起申请，基地开通后给出。无单独的成本规划。

在项目交付期间，基地云整体未完成验收转维前，由于业务中台、数据中台建设需求，需要提前开通各中台所需云资源。此时根据各中台建设厂商提出资源需求表给基地云交付团队，并抄送客户（大数据公司），基地云交付团队规划、发放相应资源，并将开通资源记录在案。

在基地云完成验证转维后，各中台建设厂商所需资源提交给客户，客户走政府内部工单流程，到大数据中心；大数据中心审批后，将需求单发大数据公司，大数据公司客户将开通资源信息发给基地云经理；基地云经理录入 BSS 系统，由公司完成 BSS 系统资源审批，审批后再由基地云运维工程师开通相关资源服务。

3 实施步骤

- 3.1 环境确认
- 3.2 系统配置
- 3.3 基础组件部署
- 3.4 场景实践

3.1 环境确认

服务器确认

在环境准备完成之后，开始部署前，根据部署清单里面的信息，确认是否满足要求。防止因为不满足要求而导致的部署失败和后期可能出现的使用不正常。

根据部署清单确认 CPU、内存、硬盘是否一致。

查看 CPU

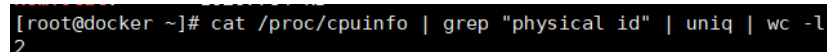
查看 CPU 个数：

输入命令：

```
cat /proc/cpuinfo | grep "physical id" | uniq | wc -l
```

输出：2 即表示2个

图 3-1 输出 1



```
[root@docker ~]# cat /proc/cpuinfo | grep "physical id" | uniq | wc -l
2
```

查看CPU核数：

输入命令：

```
cat /proc/cpuinfo | grep "cpu cores" | uniq
```

输出：cpu cores: 4 2*4=8 一共是8核

图 3-2 输出 2

```
[root@docker ~]# cat /proc/cpuinfo | grep "cpu cores" | uniq  
cpu cores : 4
```

查看 CPU 型号:

输入命令:

```
cat /proc/cpuinfo | grep 'model name' | uniq
```

输出: model name : Intel(R) Xeon(R) CPU E7-4820 v2 @ 2.00GHz

或者直接输入 top 再按 1 可以查看 CPU 核数

图 3-3 输出 3

```
Tasks: 223 total, 1 running, 222 sleeping, 0 stopped, 0 zombie  
%Cpu0 : 1.7 us, 1.3 sy, 0.0 ni, 96.6 id, 0.3 wa, 0.0 hi, 0.0 si, 0.0 st  
%Cpu1 : 3.1 us, 1.0 sy, 0.0 ni, 95.6 id, 0.0 wa, 0.0 hi, 0.3 si, 0.0 st  
%Cpu2 : 5.4 us, 0.3 sy, 0.0 ni, 93.9 id, 0.0 wa, 0.0 hi, 0.3 si, 0.0 st  
%Cpu3 : 5.7 us, 0.7 sy, 0.0 ni, 93.3 id, 0.0 wa, 0.0 hi, 0.3 si, 0.0 st  
%Cpu4 : 1.4 us, 0.7 sy, 0.0 ni, 98.0 id, 0.0 wa, 0.0 hi, 0.0 si, 0.0 st  
%Cpu5 : 2.0 us, 1.4 sy, 0.0 ni, 96.3 id, 0.0 wa, 0.0 hi, 0.3 si, 0.0 st  
%Cpu6 : 2.7 us, 0.7 sy, 0.0 ni, 96.0 id, 0.0 wa, 0.0 hi, 0.7 si, 0.0 st  
%Cpu7 : 2.0 us, 0.7 sy, 0.0 ni, 97.0 id, 0.0 wa, 0.0 hi, 0.3 si, 0.0 st
```

查看内存

输入命令:

```
cat /proc/meminfo | grep MemTotal
```

输出: MemTotal:16267784KB 即16G内存

图 3-4 查看内存

```
[root@docker ~]# cat /proc/meminfo | grep MemTotal  
MemTotal: 16267784 kB
```

查看硬盘

输入命令:

```
fdisk -l
```

图 3-5 查看硬盘

```
[root@docker ~]# fdisk -l

磁盘 /dev/sda: 536.9 GB, 536870912000 字节, 1048576000 个扇区
Units = 扇区 of 1 * 512 = 512 bytes
扇区大小(逻辑/物理): 512 字节 / 512 字节
I/O 大小(最小/最佳): 512 字节 / 512 字节
磁盘标签类型: dos
磁盘标识符: 0x00012348

   设备 Boot      Start         End      Blocks   Id  System
/dev/sda1  *          2048         2099199     1048576   83  Linux
/dev/sda2                2099200     1048575999     523238400   8e  Linux LVM

磁盘 /dev/sdb: 214.7 GB, 214748364800 字节, 419430400 个扇区
Units = 扇区 of 1 * 512 = 512 bytes
扇区大小(逻辑/物理): 512 字节 / 512 字节
I/O 大小(最小/最佳): 512 字节 / 512 字节

磁盘 /dev/mapper/cl-root: 268.4 GB, 268435456000 字节, 524288000 个扇区
Units = 扇区 of 1 * 512 = 512 bytes
扇区大小(逻辑/物理): 512 字节 / 512 字节
I/O 大小(最小/最佳): 512 字节 / 512 字节

磁盘 /dev/mapper/cl-swap: 8455 MB, 8455716864 字节, 16515072 个扇区
Units = 扇区 of 1 * 512 = 512 bytes
扇区大小(逻辑/物理): 512 字节 / 512 字节
I/O 大小(最小/最佳): 512 字节 / 512 字节

磁盘 /dev/mapper/cl-home: 473.6 GB, 473645973504 字节, 925089792 个扇区
Units = 扇区 of 1 * 512 = 512 bytes
扇区大小(逻辑/物理): 512 字节 / 512 字节
I/O 大小(最小/最佳): 512 字节 / 512 字节
```

可查看硬盘分区情况、磁盘大小等情况。

查看系统版本

需要使用最小化安装，最低要 openEuler20.03(LTS-SP1) 版本

输入命令：

```
cat /etc/os-release
```

输出：openEuler20.03(LTS-SP1)

持续集成配置

使用了持续集成，流程如下：

1. 开发或者实施将部署包通过 SVN 送到现场的 SVN（仓库安装后自带SVN服务）
2. 界面中配置持续集成服务（界面向导）
3. 镜像自动构建完成
4. 开发或实施将更新的 jar 包推送到 svn 即可

从上述方式来看，使用了持续集成后，开发和实施不再需要和 linux 进行交互，仅需要使用 svn 和页面向导，即可将部署包生成为镜像。

流水线管理-构建镜像

1. 使用平台管理员账号登录 xx容器平台，在“后台管理 > 系统管理 > 平台配置”中有如下配置，默认配置找实施人员获取：

Jenkins_IP: Jenkins服务器的主机ip





Jenkins_Port: jenkins服务器的端口号，默认*****

- Jenkins_Username: jenkins的登录账号, 默认*****
- Jenkins_Password: jenkins的登录密码, 默认*****
- Svn_Service_IP: SVN服务器的ip, 默认与jenkins装在一台主机上
- Svn_Service_Rest_Port: SVN接口的端口, 默认*****
- Jenkins_Build_Cron: jenkins构建的计划任务参数, 默认H/5 * * * *
2. “持续集成 > 流水线管理 > 导入文件 > pipeline.yaml” 或者单击“新增流水线”手动创建
 3. 单击详细信息, 查看源码地址
SVN 先 check out

 **注意**

操作 SVN 的时候 URL 地址如果 IP 地址做了映射用映射之后的 IP 不然 SVN 会卡死

再把部署包放在该文件夹下面, 提交上去。数据库链接修改一下 (jdbc.properties)

4. 进入 WEB-INF\classes\Docker, 修改 dockerfile 文件, 根据仓库版本选择相应的版本号, 修改完之后上传修改文件
5. 单击构建项目, 开始构建项目
构建项目 : 单击开始构建项目
查看控制台 : 可以查看构建情况
详细信息 : 查看流水线的详细信息
修改 : 修改流水线
可能会因为 IP 的问题导致构建失败 (做了映射的可能会出现这个问题)
解决方案:
后台管理-系统配置-平台配置, 新增 local_ecloud_uri

3.2 系统配置

本节介绍 xx容器平台中系统的配置, 该章节是配置一个系统的开始。

通常情况下, 在项目上通常都会有两套系统, 分别是测试系统、正式系统, 一般的流程是在测试系统上验证开发侧的代码, 与业主方确认无误后在更新至正式系统。

系统创建

在“部署编排 > 系统分类”中单击新增系统按钮, 在界面中可以输入系统名称与选择划归到该系统下的主机, 还可以选择配置是否在主题页中现实监控信息。在实际使用中可能会发现选择不到主机, 此时需确认

1. 是否正确安装 Agent, 主机是否已经自动注册至 Server 端
2. 在多租户模式下, 租户管理员是否将主机规划到登录用户下 (此时联系平台管理员, 详见“多租户”章节)

3. 是否该主机已经被其他系统纳管，所以无法新建系统时选择主机

主机配置

依次单击“资源管理 > 主机管理”菜单即可打开主机配置管理页面，该页面主要分为两个部分，系统分类视图和主机概览列表。

在服务器正确安装了 Agent 之后会自动显示在这里。

应用配置

在 xx 容器平台中依次单击“部署编排 > 应用分类”即可打开应用管理界面，选择左侧系统分类视图中对应的系统新增应用即可为相应的系统添加应用。

输入应用名称、选择监控方法。

服务配置

在 xx 容器平台中，服务是一个具体的功能的封装。例如统一认证要使其正常的运转，需要一个 Tomcat 容器作为 WEB 页面服务的提供，但是他还需要 PGSQL 和 KONG 才能够使的整个统一认证正常的运转，所以在 xx 容器平台 中编排时，就会为这两种类型的程序分别编排两种服务以用来管理与监控。

在 xx 容器平台 中，服务可以分为两大类，一类是容器类服务，该类服务是容器组成，相应的监控配置在新建容器时配置；一类是非容器类服务，该类型服务是一个虚拟机或实体机应用，例如是一个数据库服务，相应的监控配置在新建服务配置时设定。

容器配置

1. 进入 pass 平台“部署编排 > 容器管理 > 新增容器”；在界面中包含两个配置模块

基本设置：主要配置容器运行时的网络、镜像版本、所属系统和服、监控方法配置等

高级设置：主要配置容器运行时文件数据库逻辑卷挂载，环境变量，运行时 CPU 与内存等资源限制规则

首先在“新增容器”页面中选择该容器属于哪一个系统，再选择主机名称，此处是一个级联操作，如果发现主机下拉框中无法选择到想要的主机，需要确认该主机是否上线（可以在“主机管理”中查看到）或是否被归纳为所选的系统，如何归纳主机请参见“主机配置”章节，通过以上两步即明确了开篇中的容器调度的配置。

接下来配置容器名称，xx 容器平台 在处理时并不会在意该名称，但是为了方便后续问题的排查与描述，请命名为一个有意义的名称（不单单是排查问题，对于平台级项目多个部门协作的，只要说一个容器名称别人就可以找到该容器，大大的提高了排错效率），随后选择容器类型，目前支持的选项支持有 Tomcat、Nginx、RabbitMQ、Redis、Weblogic 几种类型，选择类型有助于 xx 容器平台 自动为创建者生成容器挂载与监控等配置信息。然后选择需要运行的容器镜像，容器镜像名称包含两个部分组成，容器镜像的名称和容器镜像的版本号，选择完成后即明确了容器内运行的程序，容器镜像是由 Jenkins 自动构建编译而成，在项目现场一般是通过 Dockerfile 构建出来。

2. 所属服务：新增系统-新增应用-新增服务

- 输入系统名称，选择系统所在主机 IP 地址
- 输入应用名称，选择关联系统，监控方法选择 HTTP
- 输入服务名称，选择服务类型，业务平台一般选择 tomcat
- 3. 选择主机名称 输入容器名称
业务文件夹名称：SVN 里面业务系统对应\WEB-INF\classes\Docke里面的 Dockerfile
选择建好对应的镜像名称
配置相应的配置文件
如果2个业务系统在一个主机下，端口不能冲突需要调整
- 4. 选择单击“启动”
单击启动按钮，xx容器平台 将会在宿主机上运行刚刚配置的容器。

MRS 配置

hive连接字符串：

```
jdbc:hive2://XX.XX.XX.XX:24002,XX.XX.XX.XX:24002,XX.XX.XX.XX:24002/  
default;serviceDiscoveryMode=zooKeeper;zooKeeperNamespace=hiveserver2;sasl.qop=auth-  
conf;auth=KERBEROS;principal=hive/hadoop.hadoophw.com@HADOOPHW.COM
```

1. 参照图中所标查找

图 3-6 查找 1



图 3-7 查找 2

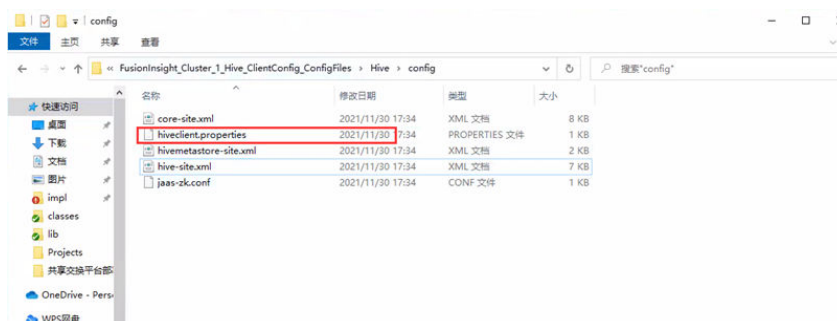


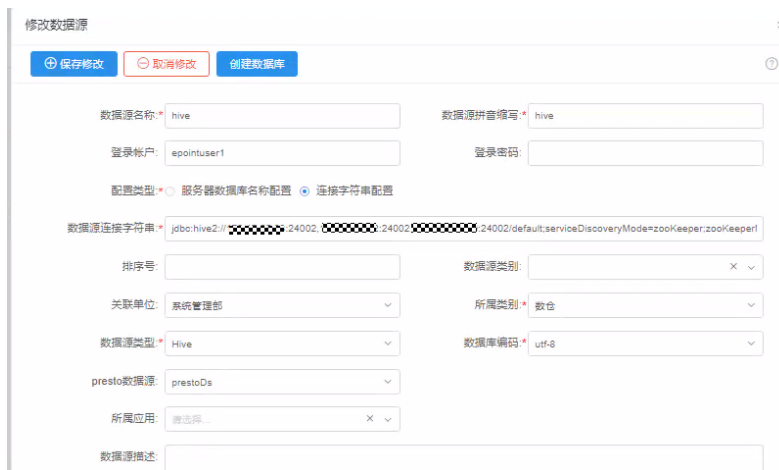
图 3-8 查找 3



2. 按顺序替换相关参数即可

jdbc:hive2://ip地址/
default;serviceDiscoveryMode=xxx;zooKeeperNamespace=xxx;sasl.qop=xxxx;au
th=xxx;principal=XXX

图 3-9 替换参数



Hetu配置:

XXXXframe 添加 hetu 所需参数

hetuUser: mrs 的认证用户, 如 XXXXuser。

hetuPassword: 认证用户的密码

hetuTenant: hive 的模式名称, 默认为default。

添加 hetu 的认证文件至 classes 路径下, 调整 user.keytab 路径

ROMA 配置

如何访问 roma 平台

- 步骤1** 本地电脑接入vpn, 连上跳板机(华为提供)。
- 步骤2** 跳板机 hosts 文件中配置域名映射(华为提供, 如截图中第二个圈所示)(部署测试系统时, 容器里的 hosts 文件要挂载出来, 并且也要配置域名映射)。
- 步骤3** 跳板机上访问平台地址, 输入用户名密码即可登录(华为提供, 如截图中第一个圈所示)。

图 3-10 ROMA 配置 1

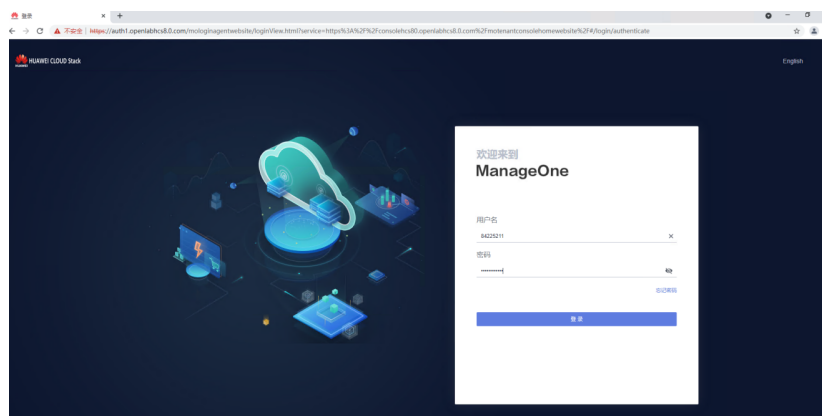
用户名	密码
84225211	XXXXXXXXXX

访问地址：<https://consolehcs80.openlabhcs8.0.com>
视hcs8.0上的公网访问地址均没有开通，需要的话需说明理由再申请。

注意事项：

- 1、请在本地hosts文件中配置域名映射。
即在 C:\Windows\System32\drivers\etc\hosts 中添加如下域名映射：
172.16.144.41 consolehcs80.openlabhcs8.0.com
172.16.144.41 auth1.openlabhcs8.0.com
172.16.144.41 docs.openlabhcs8.0.com
172.16.136.71 ocl.hcs8.0.2.com
172.16.144.41 console-staticsa-fb-1.openlabhcs8.0.com
172.16.141.15 obs.sa-fb-1.openlabhcs8.0.com
172.16.141.15 obs.openlabhcs8.0.com
- 2、请使用roma.openlab实例（SSL），请勿新建或删除实例；
- 3、需新建应用（不能与其他用户创建的应用重名），**新建MQS Topic**，需在Topic描述里添加工号+使用截止时间（例如：8419***+20211031）
- 4、MQS内网连接地址（VPC内互通）：
193.168.200.221.9093.193.168.200.134.9093.193.168.200.250.9093
公网连接地址（与实验室环境互通）：
172.16.140.226.9096.172.16.140.226.9097.172.16.140.226.9095
RESTFUL接口地址（与实验室环境互通）：
<https://172.16.140.226.9292>
- 5、APIC公网（与实验室环境互通）访问地址
<https://172.16.140.226.443>
<https://172.16.140.226.80>
- 6、Roma各组件页面各功能进行操作的接口的连接地址如下：
<https://172.16.144.30.443>
- 7、注意不能创建cce集群、ecs虚拟机，以及其他高阶服务的实例都不能创建,如有需求请联系相关负责人来处理。

图 3-11 ROMA 配置 2




步骤4 登录成功之后，单击左上角  服务列表，选择“应用与数据集成平台 ROMA Connect”，就进入了 roma 平台。

图 3-12 ROMA 配置 3

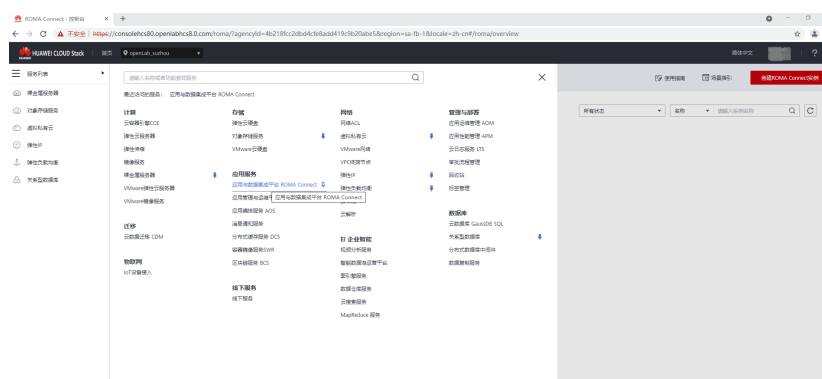
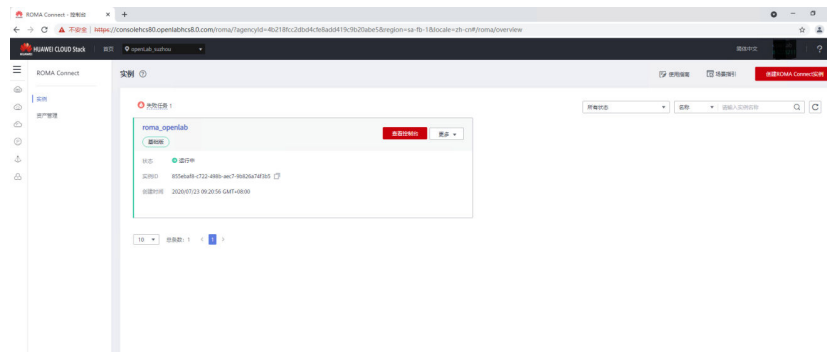




图 3-13 ROMA 配置 4



1. 系统参数配置：
管理平台 edc 配置以下系统参数

表 3-1 系统参数配置

系统参数名称	值示例	系统参数描述	获取方式
AMType	ROMA	【必配】网关类型，配置成 ROMA 才会走 ROMA 网关实现	无
extranetAmVersion	ROMA	同上（开放网站的）	无
roma_token_address	https://auth1.openlabhcs8.0.com/v3/auth/tokens	【必填】华为提供，请求该地址获取 token，token 在返回的头部参数 X-Subject-Token 里	图 3-14 图示 1 
roma_gateway_address	https://172.16.144.30:443	【必填】华为提供，注册应用、接口时的请求地址，拼接对应的 uri 请求接口	图 3-15 图示 2 

系统参数名称	值示例	系统参数描述	获取方式
roma_callad dress_http	http:// 172.16.140.226 :80	【必填】华为提供，接口发布到 roma 网关后的 http 代理地址（APIC 公网访问地址）	图 3-16 图示 3 
roma_callad dress_https	https:// 172.16.140.226 :443	【必填】华为提供，接口发布到 roma 网关后的 https 代理地址（APIC 公网访问地址）	
roma_usern ame	84225211	【必填】华为提供，登录用户名	图 3-17 图示 4 
roma_passw ord	CQMYG@#14d ss	【必填】华为提供，登录用户密码	
roma_doma inname	openlab	【必填】租户名称，获取方式：Roma 平台页面右上角“个人设置 > 基本信息 > 租户”	图 3-18 图示 5 
roma_projec t_id	dc9295fd39814 0338fe435554e 9a7200	【必填】项目标识，获取方式：Roma 平台右上角“个人设置 > 资源集列表 > 列表上的 id”	图 3-19 图示 6 
roma_instan ce_id	855ebaf8- c722-498b- aec7-9b826a74 f3b5	【必填】实例标识，获取方式：Roma 平台右上角“应用与数据集成平台 ROMA Connect > 实例上的实例 ID”	图 3-20 图示 7 

2. 脚本更新

ROMA 台里的应用返回的 appsecret 长度有点长，所以需要修改下字段长度。

检查下管理平台edc业务库“应用系统表（operation_clientapp）”里的字段“AppSecret”长度，如果长度小于500，执行下面语句修改字段长度：

```
alter table operation_clientapp modify column appsecret varchar(500);
```

图 3-21 脚本更新

行	abc DeveloperGuid	abc AppSecret	abc IP	abc Status	abc AD
1	[NULL]	f3d5ac0f-557d-473f-a9c7-b3f849e8612a	10.23.11.132	1	79f6f
2	[NULL]	eab784be-9bbc-4535-a55e-45377470eff3	127.0.0.1	1	
3	[NULL]	w%gg=#-hWd4MQZ+B2nOR.UblOO#.+ +38/N4Cl1kp=h4+.o2a2@r2b\$5KH2riGo% N\$c#0udoit@%b# +ny-Y18#Z9yU-.Q7LoQ6p.95+Y8w\$+26n2qPsbUIW9+ x+#0v5q	127.0.0.1	1	
4	[NULL]	d59b2d15-Udee-42fe-8747-6c54392b356d	127.0.0.1	1	
5	[NULL]	13196bce-2958-4b5c-8d92-1ff1009428a7	10.23.11.132	1	
6	[NULL]				

3. 访问网关代理接口地址

以上步骤配置完毕后，就可以在管理平台 edc 里创建应用、注册接口资源、接口授权了。

接口授权给应用系统后，就可以访问发布后的接口地址了。

以 postman 为例，请求地址是华为提供的 APIC 公网地址（“系统参数配置”步骤里的roma_calladdress_http、roma_calladdress_https 参数，根据源地址是否以 https 开头选择对应地址）拼上接口资源的 uri。

此外，还需额外传入 3 个头部参数：

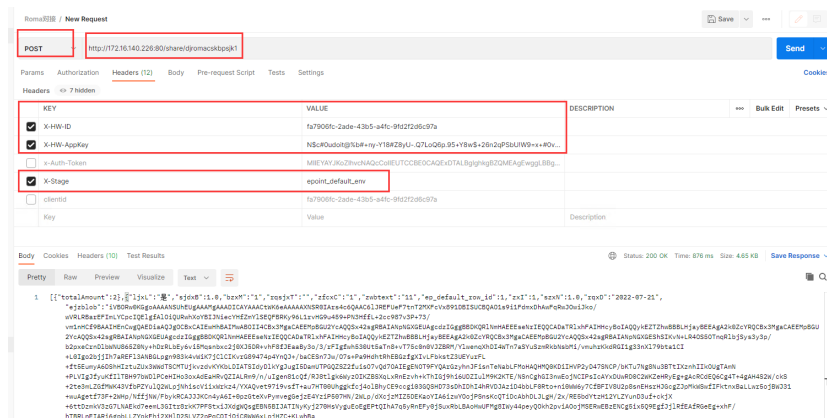
X-HW-ID: 应用系统的 appkey

X-HW-AppKey: 应用系统的 appsecret

X-Stage: 环境名称，系统里默认创建的环境名称为“XXXX_default_env”，可通过系统参数“roma_default_envname”修改，但是不建议。

后续操作正常传参调接口就行。文档参考：<https://support.huaweicloud.com/usermanual-roma-roma/apic-ug-190215022.html>

图 3-22 访问网关代理接口地址



4. 其它说明

对接步骤

- 参考华为提供的文档，编写 roma 工具类 **RomaAmUtil.java**
<https://support.huaweicloud.com/wtsnew-roma/index.html>

创建 roma 网关实现类 **RomaAMService.java**，实现管理平台网关业务接口，底层调用工具类 **RomaAmUtil.java** 里的方法。

管理平台和门户里接口访问地址展示、接口测试等地方特殊处理，适配 roma。目前只有管理平台4.1、4.1.1版本对接了 roma 网关，其它版本或者其它系统需要对接 roma 网关，可以参考上述步骤（RomaAmUtil.java 应该可以通用，RomaAMService.java 需要根据自身业务调整）。

- 未实现点

roma 暂无从授权应用维度统计接口调用情况的 API（已向华为说明，可能后面会提供出来），所以管理平台部分接口统计相关的报表、中屏暂时没有数据展示。

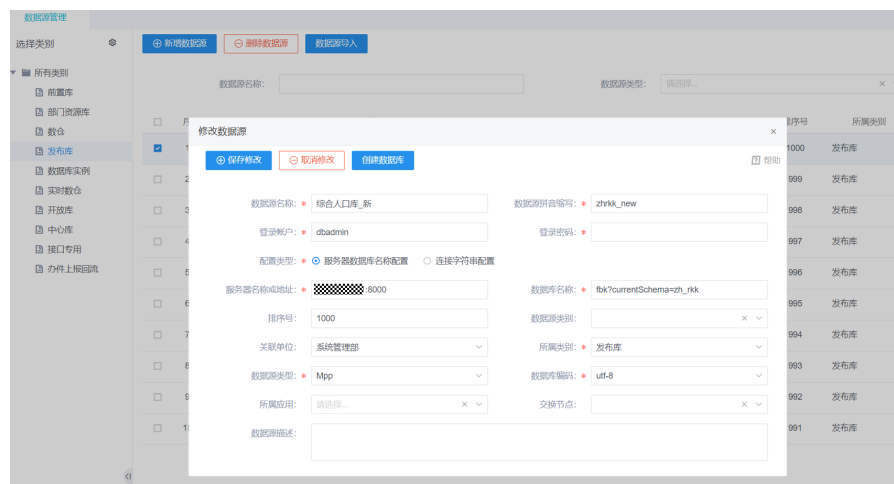
----结束

DWS 配置

DWS 连接只需要在数据源管理配置 jdbc 连接就可以。

获取 DWS 的 jdbc 链接地址可参照指导文档“[获取集群连接地址](#)”。

图 3-23 DWS 配置



DCS 配置

Redis 需配置 redisSetting、redisKeyPrefix、RedisSubscriberScope 三个参数。

详细参数说明参考下图。

图 3-24 DCS 配置



3.3 基础组件部署

3.3.1 Nginx

- **组件说明**
FAQ: 此章节为项目上需要使用 nginx 负载均衡或者使用 https 时才会使用到
Nginx 是一个高性能的 HTTP和反向代理服务器。
- **容器创建**
所属服务: 新增服务的时候需要注意一定要选择 Nginx 类型。
主机名称: 按照规划的服务器选择用于部署 Nginx 的服务器即可。
容器名称: 按照实际情况取名。
镜像名称: docker.epoint.com.cn/base/epoint_alpine_nginx
镜像版本: 选择v1.1版本。
配置文件: nginx.conf
编辑配置文件: 把需要接入 nginx 的业务平台地址写入配置文件

图 3-25 编辑配置文件 1



第一个是后缀名第二个是 IP: 端口

图 3-26 编辑配置文件 2



client_max_body_size 限制上传附件大小，默认为 64M。改为 0 则无限制。

图 3-27 64M

```
55
56 server {
57     listen      80;
58     server_name localhost;
59     client_max_body_size 64M;
60     proxy_connect_timeout 75s;
61     proxy_read_timeout 60s;
62     proxy_send_timeout 60s;
63     proxy_set_header Host $http_host;
64     proxy_set_header X-Real-IP $remote_addr;
65     proxy_set_header X-Forwarded-For $proxy_add_x_forwarded_for;
66
67     location / {
68         root    html;
```

如果出现跳转到原地址，需要修改2个文件
SSOClient.properties 文件里面的原地址修改为 nginx 的地址

图 3-28 修改 1

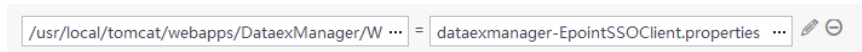
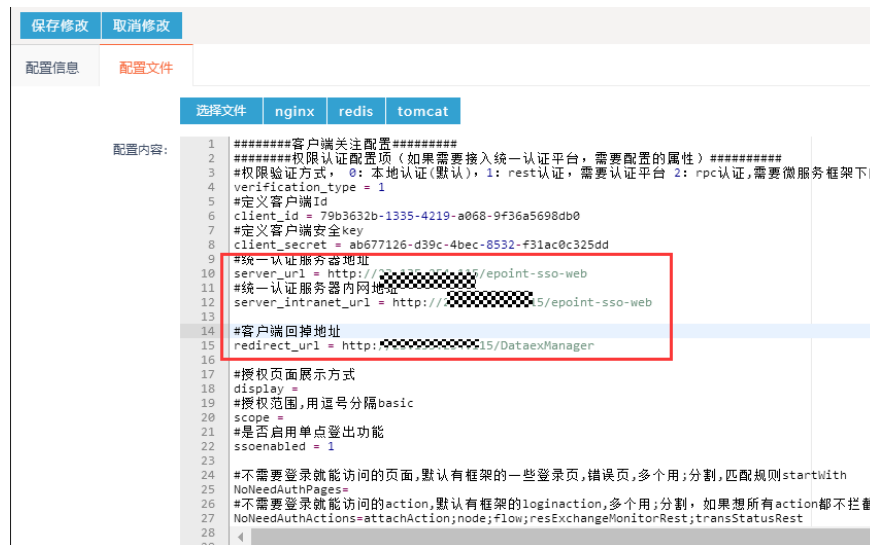


图 3-29 修改 2



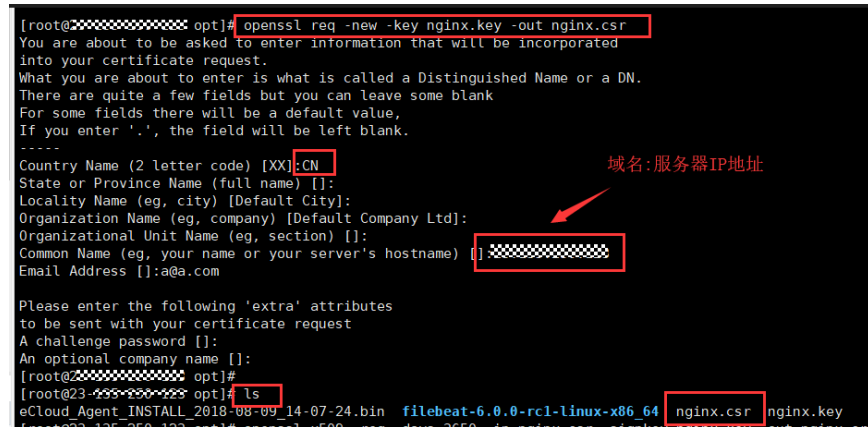
- 部署验证
使用原地址访问和使用 nginx 地址加后缀名访问，访问一致，表示接入 nginx 成功。
- Nginx SSL 证书申请
 - a. 进入 nginx 服务器，生成 nginx.key 文件服务器密钥文件
openssl genrsa -out nginx.key 2048

图 3-30 密钥文件

```
[root@23-250-123-123 ~]# openssl genrsa -out nginx.key 2048
Generating RSA private key, 2048 bit long modulus
.....+++
e is 65537 (0x10001)
[root@23-250-123-123 ~]# ls
eCloud_Agent_INSTALL_2018-08-09_14-07-24.bin  filebeat-6.0.0-rc1-linux-x86_64  nginx.key
```

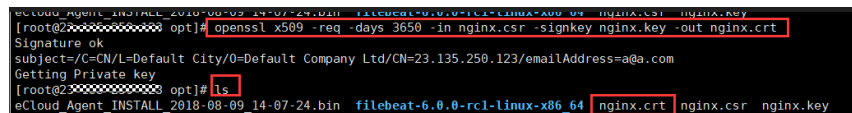
- b. 生成 nginx.csr 文件及证书申请文件
openssl req -new -key nginx.key -out nginx.csr

图 3-31 证书申请文件



- c. 将证书请求文件生成证书文件
openssl x509 -req -days 3650 -in nginx.csr -signkey nginx.key -out nginx.crt

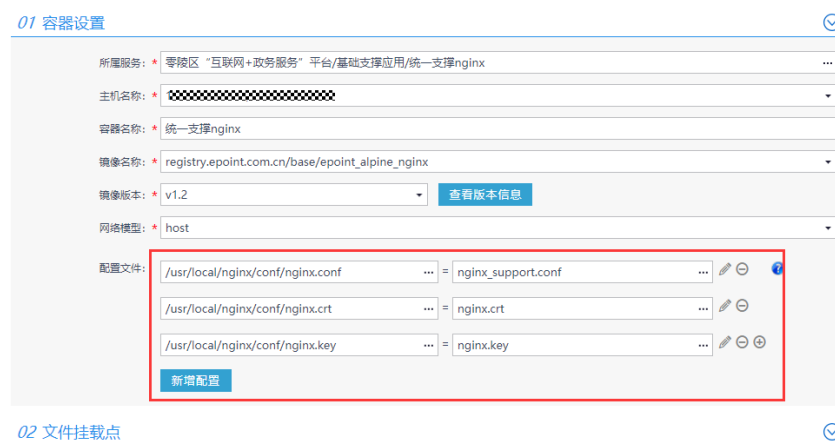
图 3-32 证书文件



- Nginx SSL 部署验证

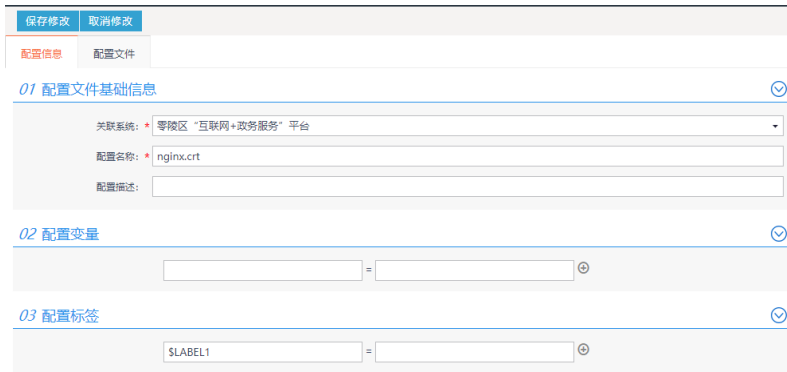
配置文件:

图 3-33 Nginx SSL 部署验证 1



- a. nginx.crt

图 3-34 Nginx SSL 部署验证 2



b. nginx.key

图 3-35 Nginx SSL 部署验证 3

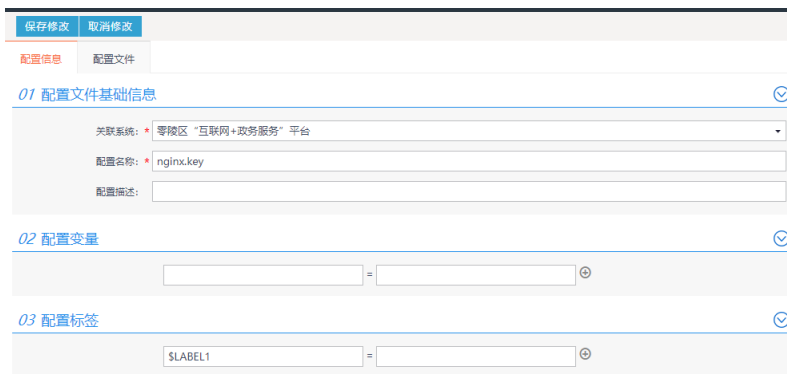
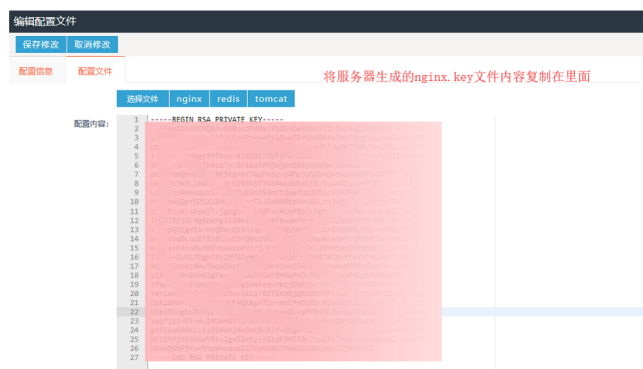
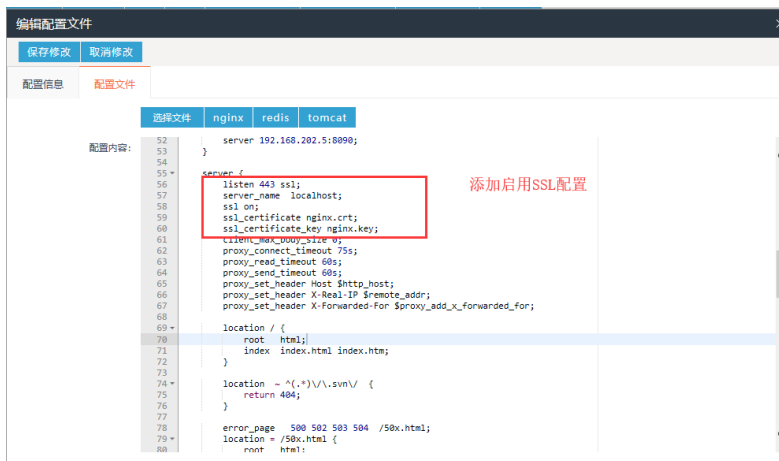


图 3-36 Nginx SSL 部署验证 4



c. nginx.conf

图 3-37 Nginx SSL 部署验证 5

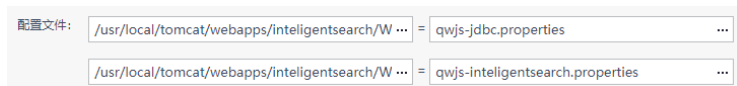


- **访问验证**
浏览器打开 nginx 默认页面，看到 Welcome to nginx! 验证成功。

3.3.2 全文检索

- **组件说明**
提供快速检索功能
- **容器创建**
所属服务：新增服务的时候需要注意一定要选择 Tomcat 类型。
主机名称：按照规划的服务器选择用于部署全文检索的服务器即可。
容器名称：按照实际情况取名。
镜像名称：docker.epoint.com.cn/library/intelligentsearch
镜像版本：选择 v1.0-snapshot 版本。
配置文件：

图 3-38 配置文件



- a. **数据库配置文件 (jdbc.properties) :**
/usr/local/tomcat/webapps/intelligentsearch/WEB-INF/classes/jdbc.properties
url: 填写数据库 jdbc 连接;
Username: 填写数据库用户名;
password: 填写数据库密码。
或者在环境变量里面新增

图 3-39 数据库配置文件



- b. 修改 intelligentsearch.properties，配置智能检索系统参数
/usr/local/tomcat/webapps/intelligentsearch/WEB-INF/classes/intelligentsearch.properties
zkIpConfig: 前面安装的 zookeeper 的 ip 和端口（端口默认为 2181）
replicNums: 索引备份数，如果 solr 部署了三台，那么这个数字可以改为 2，否则不用修改
redisSetting: redis 连接配置，这个配置非常重要，一定要配置
maxPageNumber: 索引分页查询可以查询的最大页数，建议不要配置太大
NULLREPLAY: 机器人默认不知道的回答内容，以封号隔开

图 3-40 数据库配置文件

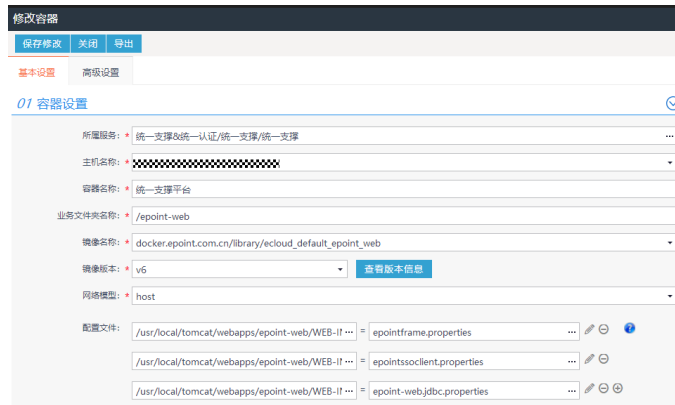


- 部署验证
访问地址：http://localhost:端口/intelligentsearch 用户名：admin 密码：111111
- 访问验证
登录相应页面，正常显示。业务平台部署成功。

3.3.3 统一支撑平台

- 组件说明
统一支撑平台用于统一的应用管理、组织架构管理、API 服务管理、附件管理等。
- 数据库还原
统一认证和统一支撑使用同一个数据库，两个平台仅需要还原一次数据库。
- 容器创建

图 3-41 容器创建 1



所属服务：新增服务的时候需要注意服务类型一定要选择 Tomcat 类型。

图 3-42 容器创建 2

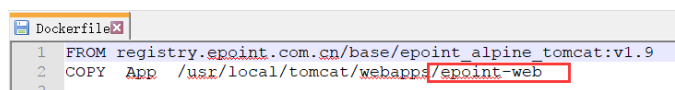


主机名称：按照规划的服务器选择用于部署统一支撑的服务器即可。

容器名称：按照实际情况取名。

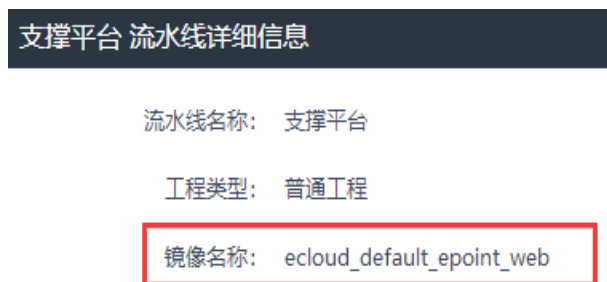
业务文件夹名称：需要与 DockerFile 文件中的项目名称一致。

图 3-43 容器创建 3



镜像名称：使用流水线构建完成之后的镜像地址。

图 3-44 容器创建 4



镜像版本：按照实际情况选择。一般更新之后构建容器选择最新的（最高的）版本。

- **配置修改**

所有的配置文件都通过配置文件覆盖的方式进行配置，一定要在原来的配置文件的基础上进行修改，不要复制其他的项目中同名的配置文件，避免特殊配置项遗漏。

- a. 数据库配置文件（jdbc.properties）：

`/usr/local/tomcat/webapps/-web/WEB-INF/classes/jdbc.properties`

url：填写数据库 jdbc 连接；

Username：填写数据库用户名；

password：填写数据库密码。

- b. 框架配置文件（frame.properties）：按部署包里面的原样拿过来

如果是集群部署：

redisSetting：配置 3.4.6.1.1 中部署的 redis 数据库地址，格式：redis://host:port/dbIndex，与其他配置文件中的 redis 地址不要重复

enableMutiClient：1

- c. 框架配置文件（SSOClient.properties）：按部署包里面的原样拿过来

NoNeedAuthAction：加上 am；addACLPluginsForApp 以分号隔开

如果是集群部署：

clustered_enabeld：1

- **部署验证**

启动容器后访问 `http://IP:PORT/-web`，出现登录页面，输入 `admin/11111`，可以登录成功的话，部署成功。

- **网关配置**

打开统一应用与服务-统一服务管理-服务参数模块进行配置。

- a. 勾选启用服务网关；

- b. 网关注册地址、调用地址、授权地址填写 KONG 部署完成的地址；

- c. 日志监听地址填写支撑平台地址（自己），如果加了外网映射，需要使用映射之前的地址。

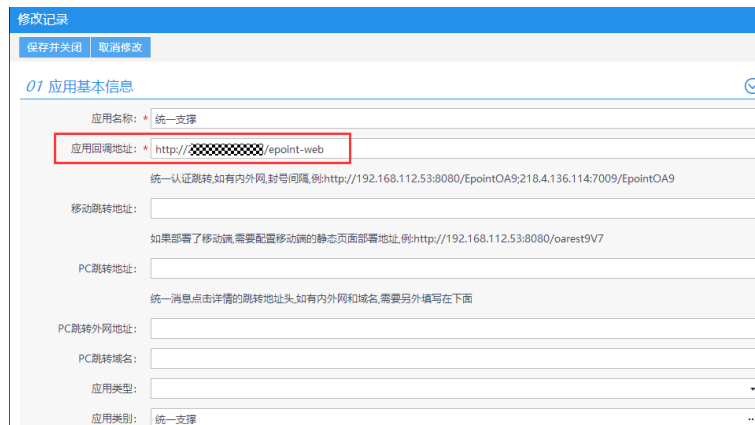
- d. redis 地址填写统一认证 frame.properties 文件中的配置一致。

- e. 保存参数，完成配置。

- **接入统一认证**

- a. 注册应用：打开支撑平台的统一应用与服务-统一应用管理-应用注册，单击新增应用，填写应用名称、应用回调地址（接入了 nginx 的 IP 地址填写 nginx 的地址）、PC 跳转地址、移动跳转地址。地址都使用支撑平台的浏览器访问地址。单击保存并关闭。接下来查看应用详情，可以看到应用 key 和应用密钥，注册完成。

图 3-45 注册应用 1

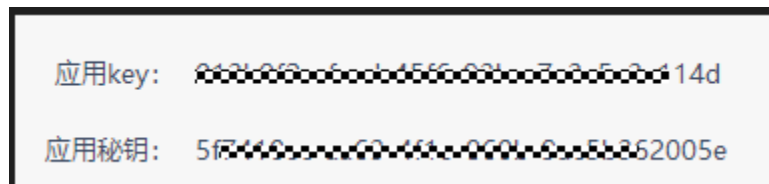


The screenshot shows a web interface for registering an application. At the top, there are buttons for '修改记录' (Modify Record), '保存并关闭' (Save and Close), and '取消修改' (Cancel Modify). Below this is a section titled '01 应用基本信息' (01 Application Basic Information). The form contains several input fields: '应用名称' (Application Name) with the value '统一支撑'; '应用回调地址' (Application Callback Address) with the value 'http://[redacted]/epoint-web', which is highlighted with a red box; '统一认证跳转,如有内外网,封号间隔,例:http://192.168.112.53:8080/EpintOA9;218.4.136.114:7009/EpintOA9'; '移动跳转地址'; 'PC跳转地址'; 'PC跳转外网地址'; 'PC跳转域名'; '应用类型' (Application Type) with a dropdown menu; and '应用类别' (Application Category) with the value '统一支撑'.

图 3-46 注册应用 2

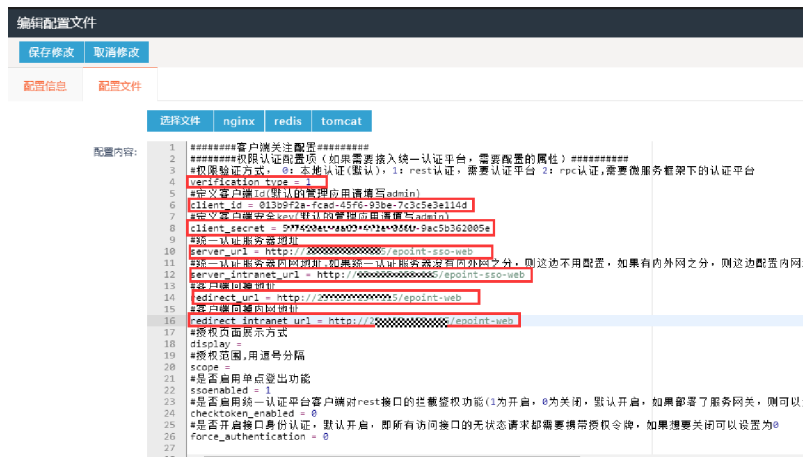


图 3-47 注册应用 3



- b. 修改支撑平台配置 (SSOClient.properties)
/usr/local/tomcat/webapps/-web/WEB-INF/classes/SSOClient.properties
verification_type: 1;
client_id: 上个步骤的应用 key;
client_secret: 上个步骤的应用密钥;
server_url: 统一认证的浏览器访问地址;
server_intranet_url: 非必填, 如果统一认证配置了映射, 内外网地址不一样的情况, 需要在此处配置统一认证的内网地址;
redirect_url: 支撑平台的浏览器访问地址, 与注册应用时填写的一致。

图 3-48 修改支撑平台配置

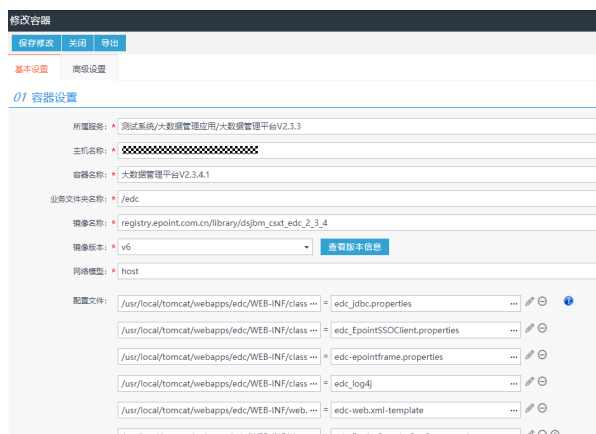


- c. 重启容器，访问支撑平台会跳转到统一认证，输入账号密码可以登录到首页，配置完成。
- 访问验证
登录相应页面，正常显示。业务平台部署成功。

3.3.4 一体化数据管理平台

- 组件说明
一体化数据管理平台包括目录管理模块、资源管理模块、数据供需模块、绩效考核模块、运行监控模块、数据治理模块。
首先需要在平台发布下载最新的发布包，由项目开发打成部署包，然后在 PAAS 平台中构建镜像。
- 数据库还原
一体化数据管理平台与开放共享门户三个部署包使用同一个数据库。
 - a. 在 MySQL 实例上面创建数据库，数据库名称为 edc;
 - b. 从数据资源目录平台部署包中复制出数据库备份文件，还原到上一步创建的数据库中。
- 容器创建

图 3-49 容器创建 1



所属服务：新增服务的时候需要注意一定要选择 Tomcat 类型。

图 3-50 容器创建 2

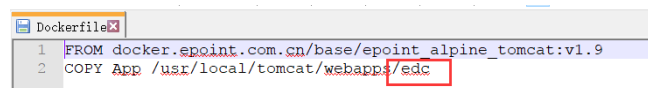


主机名称：按照规划的服务器选择用于部署资源目录管理平台的服务器即可。

容器名称：按照实际情况取名。

业务文件夹名称：需要与 DockerFile 文件中的项目名称一致。

图 3-51 容器创建 3



镜像名称：使用流水线构建完成之后的镜像地址。

图 3-52 容器创建 4



镜像版本：按照实际情况选择。一般更新之后构建容器选择最新的（最高的）版本。

- **配置修改**

所有的配置文件都通过配置文件覆盖的方式进行配置，一定要在原来的配置文件的基础上进行修改，不要复制其他的项目中同名的配置文件，避免特殊配置项遗漏。

数据库配置文件（jdbc.properties）：

/usr/local/tomcat/webapps/edc/WEB-INF/classes/jdbc.properties

url：填写数据库 jdbc 连接；

Username：填写数据库用户名；

password：填写数据库密码。

- **部署验证**

启动容器后访问 `http://IP:PORT/edc`，
出现机器码 URL 申请上传服务器 `/opt/apache-tomcat-7.0.85/webapps/edc/
WEB-INF/lib`
出现登录页面，输入 `admin/11111`，可以登录成功的话，部署成功。

图 3-53 部署验证



- 接入统一支撑

- a. 注册应用：打开支撑平台的统一应用与服务-统一应用管理-应用注册，单击新增应用，填写应用名称、应用回掉地址、PC 跳转地址、移动跳转地址。地址都使用资源目录平台的浏览器访问地址。单击保存并关闭。接下来查看应用详情，可以看到应用 key 和应用密钥，注册完成。

图 3-54 接入统一支撑 1

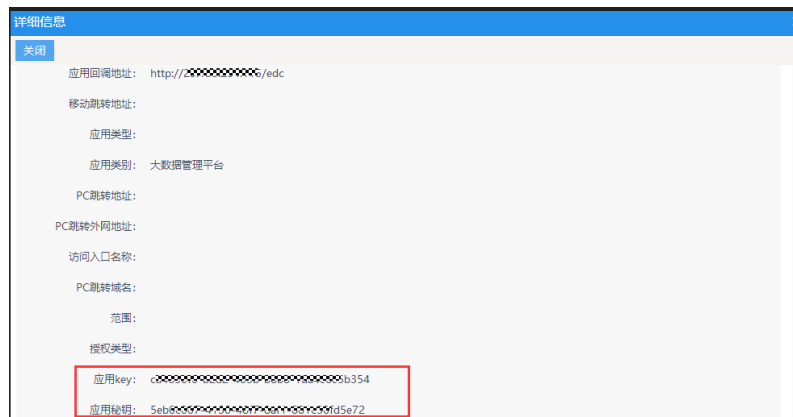
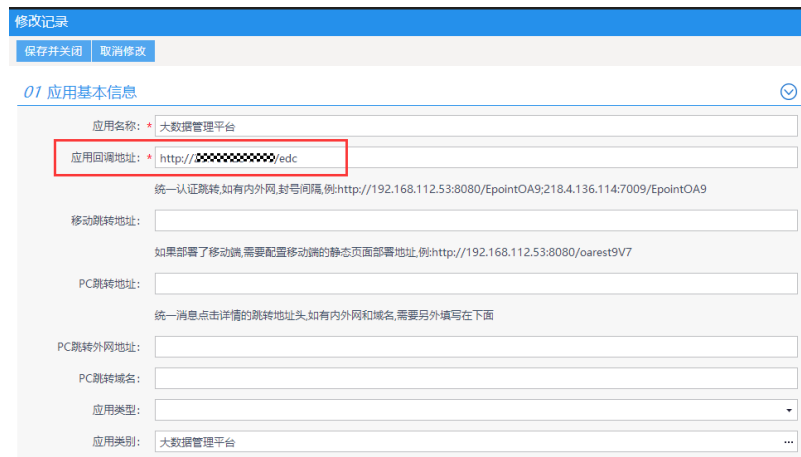


图 3-55 接入统一支撑 2



b. 订阅用户，订阅角色：打开支撑平台的统一应用与服务-统一应用管理-应用管理。

订阅用户：用于设置需要同步哪些用户到应用系统，公开同步所有；

订阅角色：用于设置需要同步哪些角色到应用系统，需要勾选 adminGroup 和 ouAdminGroup

图 3-56 接入统一支撑 3

应用名称	应用类别	设置数据同步	订阅用户	已订阅用户	订阅角色	子应用管理	应用管理	在应用中	订阅中	操作管理
1 本地应用门户	应用门户	是	否	否	否	否	否	否	否	5
2 应用门户	应用门户	是	否	否	否	否	否	否	否	5
3 数据应用管理	数据应用管理	是	否	否	否	否	否	否	否	5
4 大数据管理平台本地版	大数据管理平台	是	否	否	否	否	否	否	否	5
5 统一支撑	统一支撑	是	否	否	否	否	否	否	否	5
6 大数据管理平台	大数据管理平台	是	否	否	否	否	否	否	否	5
7 开发-数据	开发-数据	是	否	否	否	否	否	否	否	5
8 区五+移动应用系统	区五+移动应用系统	是	否	否	否	否	否	否	否	5
9 区五+应用系统	区五+应用系统	是	否	否	否	否	否	否	否	5
10 大数据管理平台应用系统	大数据管理平台应用系统	是	否	否	否	否	否	否	否	5

图 3-57 接入统一支撑 4

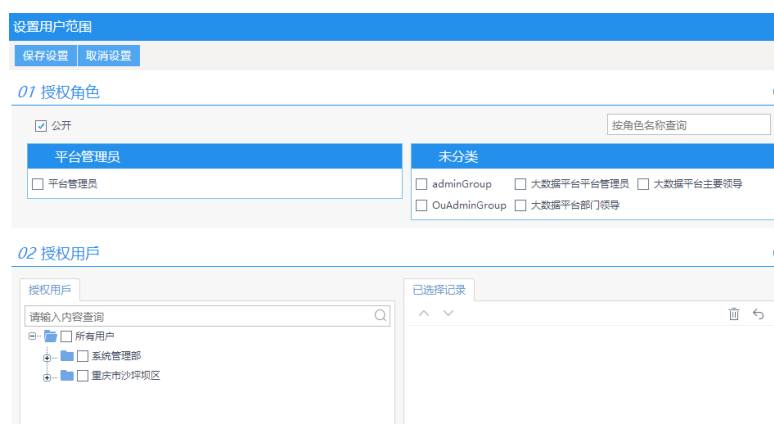


图 3-58 接入统一支撑 5



c. 修改框架配置文件 (frame.properties)

/usr/local/tomcat/webapps/edc/WEB-INF/classes/frame.properties

SOAAddress: 支撑平台地址 +/rest/organization, 如果配置了外网映射, 需要配置内网地址;

SOA_Appkey_secret: 第一步的应用 key 和应用密钥, 使用分号隔开;

IsOAuthEnabled: 1

图 3-59 接入统一支撑 6

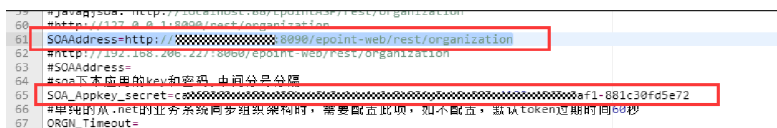


图 3-60 接入统一支撑 7



如果是集群部署:

redisSetting: 配置已经部署的redis数据库地址, 格式: redis://host:port/dbIndex, 与其他配置文件中的 redis 地址不要重复

enableEditorTempUpload: true

- 接入统一认证

修改支撑平台配置 (SSOClient.properties)

/usr/local/tomcat/webapps/edc/WEB-INF/classes/SSOClient.properties

verification_type: 1;

client_id: 上个步骤的应用 key;

client_secret: 上个步骤的应用密钥;

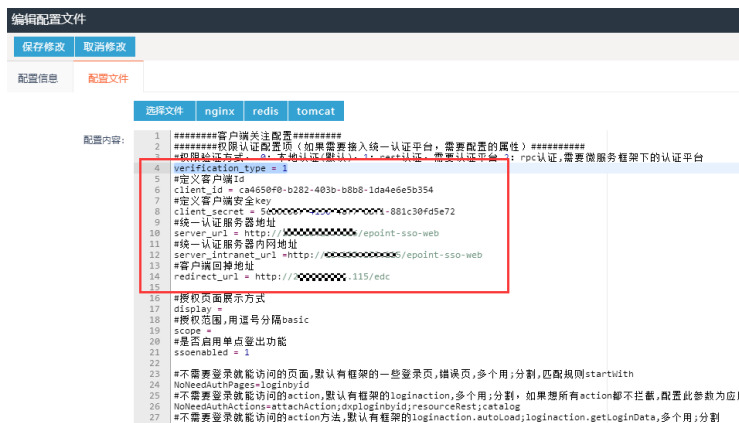
server_url: 统一认证的浏览器访问地址;

server_intranet_url: 非必填, 如果统一认证配置了映射, 内外网地址不一样的情况, 需要在此处配置统一认证的内网地址;

redirect_url: 资源目录平台的浏览器访问地址, 与注册应用时填写的一致。

重启容器, 访问支撑平台会跳转到统一认证, 输入账号密码可以登录到首页, 配置完成。

图 3-61 接入统一认证



如果是集群部署：

clustered_enablend: 1

- 资源目录配置

资源目录的配置项都在后台管理-系统配置-系统参数模块中，下面介绍每个系统参数的作用和配置的值。

- a. API_ADDRESS

说明：用于服务管理中的应用系统注册和API服务注册，如果发现上述功能无法正常使用时，请优先检查改配置是否正确。

配置：配置支撑平台地址。如果支撑平台配置了外网映射，此处需要配置内网地址。

- b. APIPrefix

说明：资源目录平台如果配置了外网映射，发布的API服务会使用外网地址，如果KONG服务器无法访问外网地址，需要配置该参数。

配置：配置资源目录平台的内网访问地址。

- c. DataEXUrl

说明：用于中心资源管理-交换配置和资源订阅下发，如果发现上述功能无法正常使用时，请优先检查改配置是否正确。

配置：配置交换平台地址。如果交换平台配置了外网映射，此处需要配置内网地址。

- d. TOKEN_ADDRESS

说明：用于门户测试API接口，如果发现上述功能无法正常使用时，请优先检查改配置是否正确。

配置：配置统一认证平台地址。如果统一认证平台配置了外网映射，此处需要配置内网地址。

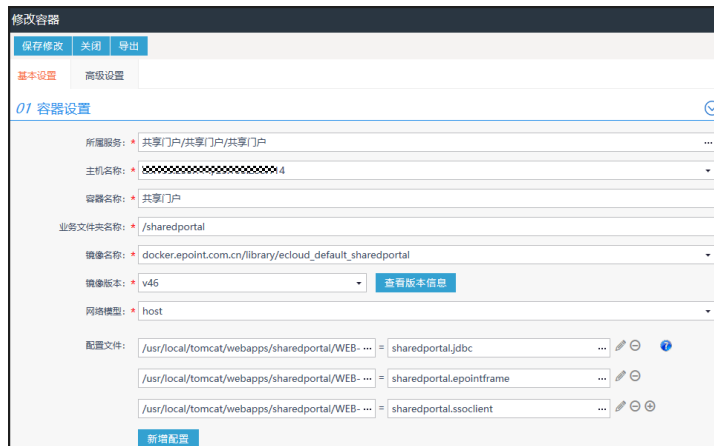
- 访问验证

登录相应页面，正常显示。业务平台部署成功。

3.3.5 共享门户

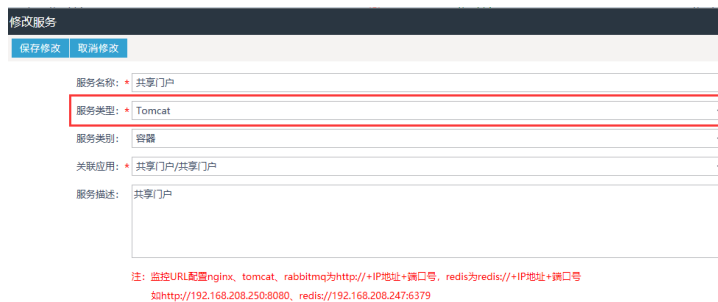
- 容器创建

图 3-62 容器创建 1



所属服务：新增服务的时候需要注意一定要选择 Tomcat 类型。

图 3-63 容器创建 2

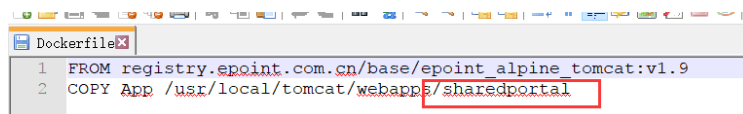


主机名称：按照规划的服务器选择用于共享门户的服务器即可。

容器名称：按照实际情况取名。

业务文件夹名称：需要与 DockerFile 文件中的项目名称一致。

图 3-64 容器创建 3



镜像名称：使用流水线构建完成之后的镜像地址。

图 3-65 容器创建 4



镜像版本：按照实际情况选择。一般更新之后构建容器选择最新的（最高的）版本。

- **配置修改**

所有的配置文件都通过配置文件覆盖的方式进行配置，一定要在原来的配置文件的基础上进行修改，不要复制其他的项目中同名的配置文件，避免特殊配置项遗漏。

数据库配置文件（jdbc.properties）：

```
/usr/local/tomcat/webapps/sharedportal/WEB-INF/classes/jdbc.properties
```

url：填写数据库 jdbc 连接；

Username：填写数据库用户名；

password：填写数据库密码。

- **部署验证**

启动容器后访问 <http://IP:PORT/sharedportal/pages/default/index>，

出现机器码 URL 申请上传服务器 /opt/apache-tomcat-7.0.85/webapps/extranet/WEB-INF/lib

出现下面的页面，输入 admin/11111，可以登录成功的话，部署成功[在获取机器码需要上传授权到服务器无法页面上传 如果上传请给出地址。

图 3-66 部署验证



- **接入统一支撑**

- a. 注册应用：打开支撑平台的统一应用与服务“统一应用管理 > 应用注册”，单击新增应用，填写应用名称、应用回掉地址、PC 跳转地址、移动跳转地址。地址都使用资源目录平台的浏览器访问地址。单击保存并关闭。接下来查看应用详情，可以看到应用 key 和应用密钥，注册完成。

图 3-67 支撑 1

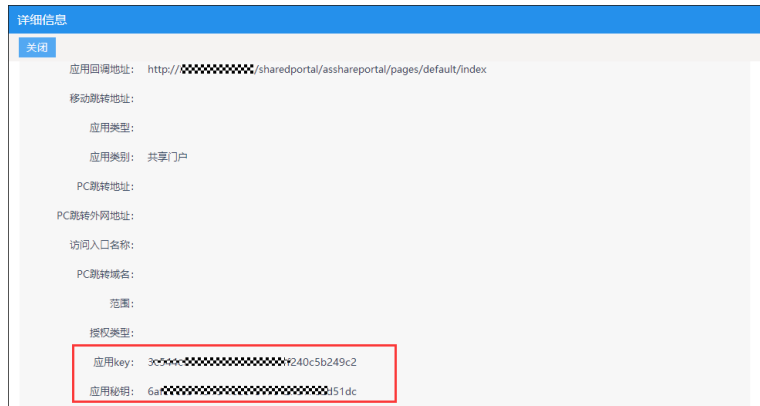
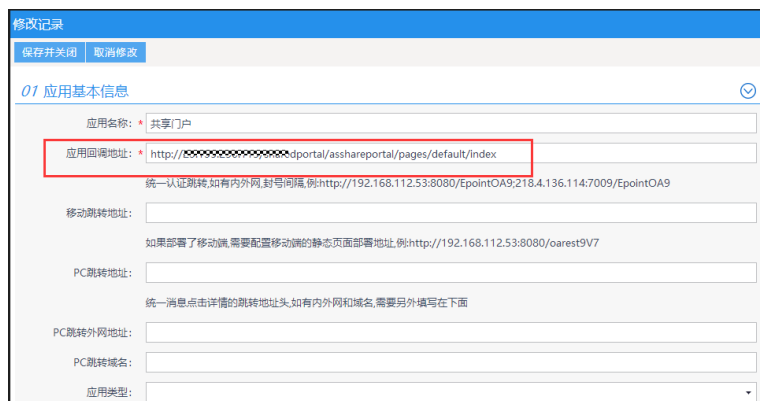


图 3-68 支撑 2



- b. 订阅用户，订阅角色：打开支撑平台的统一应用与服务-统一应用管理-应用管理。

订阅用户：用于设置需要同步哪些用户到应用系统，公开同步所有；

订阅角色：用于设置需要同步哪些角色到应用系统，需要勾选 adminGroup 和 ouAdminGroup

图 3-69 支撑 3

应用名称	应用图标	应用类型	应用地址	应用密钥	应用key	应用secret	应用名称	应用图标	应用地址	应用密钥	应用key	应用secret
1	应用图标	应用类型	应用地址	应用密钥	应用key	应用secret	应用名称	应用图标	应用地址	应用密钥	应用key	应用secret
2	应用图标	应用类型	应用地址	应用密钥	应用key	应用secret	应用名称	应用图标	应用地址	应用密钥	应用key	应用secret
3	应用图标	应用类型	应用地址	应用密钥	应用key	应用secret	应用名称	应用图标	应用地址	应用密钥	应用key	应用secret
4	应用图标	应用类型	应用地址	应用密钥	应用key	应用secret	应用名称	应用图标	应用地址	应用密钥	应用key	应用secret
5	应用图标	应用类型	应用地址	应用密钥	应用key	应用secret	应用名称	应用图标	应用地址	应用密钥	应用key	应用secret
6	应用图标	应用类型	应用地址	应用密钥	应用key	应用secret	应用名称	应用图标	应用地址	应用密钥	应用key	应用secret
7	应用图标	应用类型	应用地址	应用密钥	应用key	应用secret	应用名称	应用图标	应用地址	应用密钥	应用key	应用secret
8	应用图标	应用类型	应用地址	应用密钥	应用key	应用secret	应用名称	应用图标	应用地址	应用密钥	应用key	应用secret
9	应用图标	应用类型	应用地址	应用密钥	应用key	应用secret	应用名称	应用图标	应用地址	应用密钥	应用key	应用secret
10	应用图标	应用类型	应用地址	应用密钥	应用key	应用secret	应用名称	应用图标	应用地址	应用密钥	应用key	应用secret

图 3-70 支撑 4



图 3-71 支撑 5



- c. 修改框架配置文件 (frame.properties)
/usr/local/tomcat/webapps/sharedportal/WEB-INF/classes/frame.properties
SOAAddress: 支撑平台地址 +/rest/organization, 如果配置了外网映射, 需要配置内网地址;
SOA_Appkey_secret: 第一步的应用key和应用密钥, 使用分号隔开;
IsOAuthEnabled: 1

图 3-72 支撑 6

```
60 #http://192.168.209.50:8090/epoint-web/rest/organization http://127.0.0.1:8090/rest/organization ht  
61 SOAAddress=http://192.168.209.50:8090/epoint-web/rest/organization  
62 #http://725.155.234.116:8090/epoint-web/rest/organization  
63 #soa下本应用的key和密钥,中间分号分隔 53fech93-608b-4115-bd65-51d641958fa5;67aa55e8-2938-4753-948d-4  
64 SOA_Appkey_secret=380002-0507-f240c3b249c2,6af00ceb-191c-46be-8170-2781782d51dc  
65 45c744ca-37a3-4002-0507-f240c3b249c2,6af00ceb-191c-46be-8170-2781782d51dc
```

图 3-73 支撑 7

```
144  
145 #是否开启了oauth认证:1:开启  
146 IsOAuthEnabled=1  
147
```

- 如果是集群部署:
redisSetting: 配置DCS 配置中部署的 redis 数据库地址, 格式: redis://host:port/dbIndex, 与其他配置文件中的 redis 地址不要重复
- 接入统一认证
修改支撑平台配置 (SSOClient.properties)
/usr/local/tomcat/webapps/sharedportal/WEB-INF/classes/SSOClient.properties

verification_type: 1;
client_id: 上个步骤的应用key;
client_secret: 上个步骤的应用密钥;
server_url: 统一认证的浏览器访问地址;
server_intranet_url: 非必填, 如果统一认证配置了映射, 内外网地址不一样的情况, 需要在此处配置统一认证的内网地址;
redirect_url: 共享门户的浏览器访问地址, 与注册应用时填写的一致。
配置完成之后重启容器, 使用支撑平台新建的其他的用户也可以登录的话, 配置就完成了。

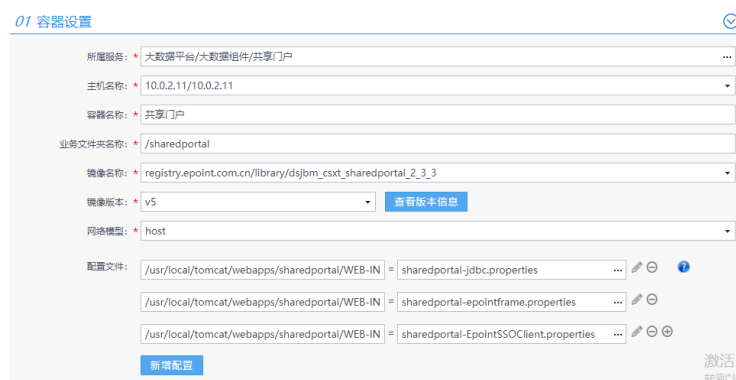
图 3-74 支撑 8



3.3.6 开放门户

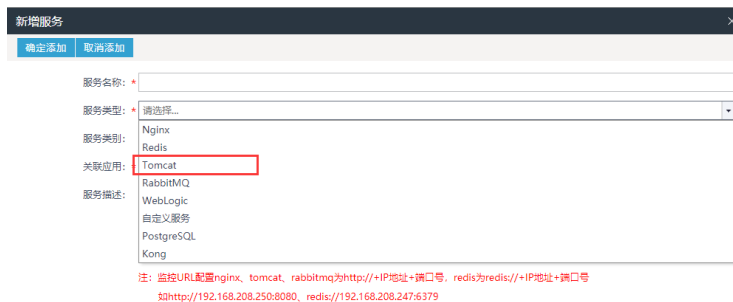
- 访问验证
登录相应页面, 正常显示。业务平台部署成功。
- 容器创建
下面介绍容器创建页面配置项, 未提到的使用默认配置:

图 3-75 创建 1



所属服务: 新增服务的时候需要注意一定要选择 Tomcat 类型。

图 3-76 创建 2



主机名称：按照规划的服务器选择用于部署开放门户的服务器即可。

容器名称：按照实际情况取名。

业务文件夹名称：需要与 DockerFile 文件中的项目名称一致。下面以 extranet 为例，下面所有提到的 extranet 项目按照实际情况进行修改。

镜像名称：使用流水线构建完成之后的镜像地址。

镜像版本：按照实际情况选择。

● **配置修改**

所有的配置文件都通过配置文件覆盖的方式进行配置，一定要在原来的配置文件的基础上进行修改，不要复制其他的项目中同名的配置文件，避免特殊配置项遗漏。

1、数据库配置文件（jdbc.properties）：

/usr/local/tomcat/webapps/extranet/WEB-INF/classes/jdbc.properties

url：填写数据库 jdbc 连接；

Username：填写数据库用户名；

password：填写数据库密码。

● **部署验证**

启动容器后访问 <http://IP:PORT/extranet/openportal/pages/default/index>；出现下面的页面，部署成功。

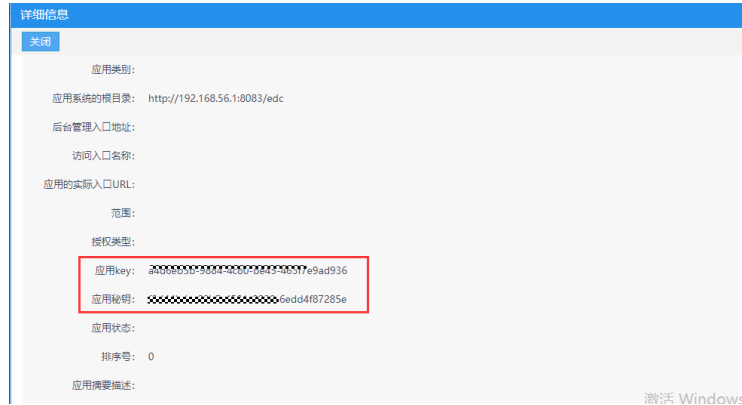
图 3-77 部署验证



● **接入统一支撑**

- a. 注册应用：打开支撑平台的统一应用与服务“统一应用管理 > 应用注册”，单击“新增应用”，填写应用名称、应用回掉地址、PC 跳转地址、移动跳转地址。地址都使用共享门户的浏览器访问地址。单击保存并关闭。接下来查看应用详情，可以看到应用 key 和应用密钥，注册完成。

图 3-78 支撑 1



- b. 订阅用户，订阅角色：打开支撑平台的统一应用与服务-统一应用管理-应用管理。
订阅用户：用于设置需要同步哪些用户到应用系统，公开同步所有；
订阅角色：用于设置需要同步哪些角色到应用系统，需要勾选 adminGroup 和 ouAdminGroup。

图 3-79 支撑 2



- c. 修改框架配置文件（frame.properties）
`/usr/local/tomcat/webapps/extranet/WEB-INF/classes/frame.properties`
SOAAddress：支撑平台地址 +/rest/organization，如果配置了外网映射，需要配置内网地址；
SOA_Appkey_secret：第一步的应用 key 和应用密钥，使用分号隔开；
IsOAuthEnabled：1
- 接入统一认证
修改支撑平台配置（SSOClient.properties）
`/usr/local/tomcat/webapps/extranet/WEB-INF/classes/SSOClient.properties`
verification_type：0；请注意此处与其他平台不一致。
client_id：上个步骤的应用 key；
client_secret：上个步骤的应用密钥；
server_url：统一认证的浏览器访问地址；
server_intranet_url：非必填，如果统一认证配置了映射，内外网地址不一样的情况，需要在此处配置统一认证的内网地址；

redirect_url: 共享门户的浏览器访问地址, 与注册应用时填写的一致。

- 访问验证
登录相应页面, 正常显示。业务平台部署成功。

3.3.7 数据交换平台管理端

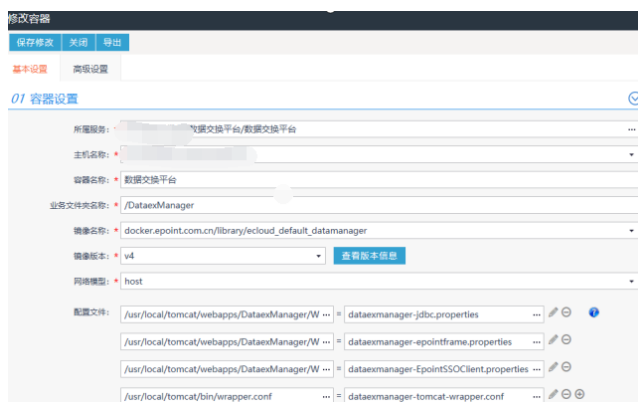
- 组件说明
交换管理端用于对交换节点的统一管控, 包括数据源的新增修改、作业的批量生成、作业的启动停止、作业状态的监控、作业执行情况的监控等。

图 3-80 组件说明



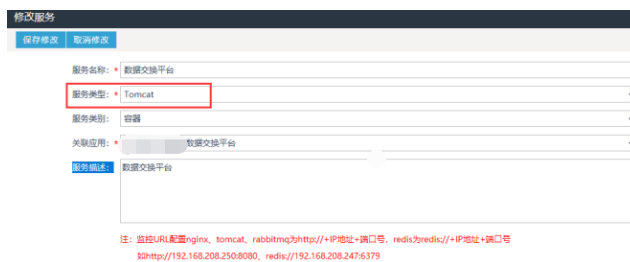
- 数据库还原
 - a. 在 MySQL 实例上面创建数据库, 数据库名称为 `dxp_epointdpx` ;
 - b. 从数据交换管理平台部署包中复制出数据库备份文件, 还原到上一步创建的数据库中。
- 容器创建

图 3-81 容器创建 1



所属服务: 新增服务的时候需要注意一定要选择 Tomcat 类型。

图 3-82 容器创建 2

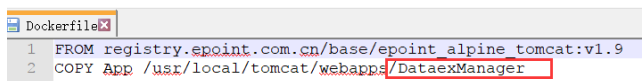


主机名称：按照规划的服务器选择用于部署资源目录管理平台的服务器即可。

容器名称：按照实际情况取名。

业务文件夹名称：需要与 Dockerfile 文件中的项目名称一致。

图 3-83 容器创建 3



镜像名称：使用流水线构建完成之后的镜像地址。

图 3-84 容器创建 4



镜像版本：按照实际情况选择。一般更新之后构建容器选择最新的（最高的）版本。

● **配置修改**

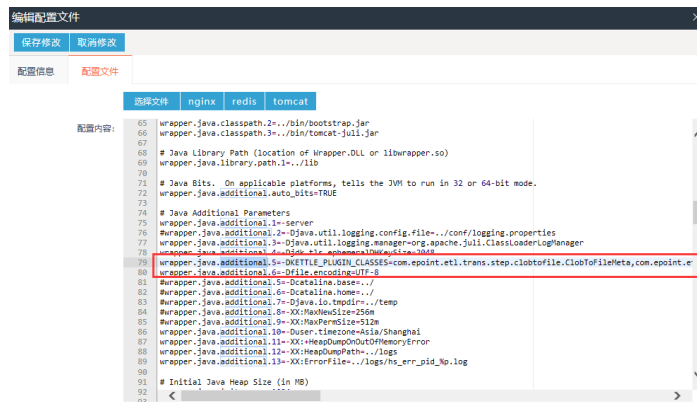
所有的配置文件都通过配置文件覆盖的方式进行配置，一定要在原来的配置文件的基础上进行修改，不要复制其他的项目中同名的配置文件，避免特殊配置项遗漏。

- a. 数据库配置文件（jdbc.properties）：
/usr/local/tomcat/webapps/DataexManager/WEB-INF/classes/jdbc.properties
url：填写数据库jdbc连接；
Username：填写数据库用户名；
password：填写数据库密码。wrapper.conf文件：
- b. wrapper.conf文件
/usr/local/tomcat/bin/wrapper.conf

注意

从部署的DataexManager中下载该文件后，修改以下地方，加入如下参数：

图 3-85 编辑配置文件



```
wrapper.java.additional.5=-  
DKETTLE_PLUGIN_CLASSES=com..etl.trans.step.clobtofile.ClobToFileMeta,com..etl.trans.step.fileto  
clob.FileToClobMeta,org.pentaho.di.trans.steps.xmloutput.XMLOutputMeta,com..etl.trans.step.slic  
eupload.SliceUpload,org.pentaho.di.trans.steps.getxmldata.GetXMLDataMeta,com..etl.job.unzip.U  
nZip,com..etl.trans.step.transstatuslog.TransStatusLogMeta  
wrapper.java.additional.6=-Dfile.encoding=UTF-8
```

- **部署验证**

启动容器后访问 <http://IP:PORT/DataexManager>，出现登录页面，输入 admin/11111，可以登录成功的话，部署成功。

- **接入统一支撑**

- a. **注册应用**：打开支撑平台的统一应用与服务“统一应用管理 > 应用注册”，单击新增应用，填写应用名称、应用回掉地址、PC 跳转地址、移动跳转地址。地址都使用资源目录平台的浏览器访问地址。单击保存并关闭。接下来查看应用详情，可以看到应用 key 和应用密钥，注册完成。

图 3-86 详细信息

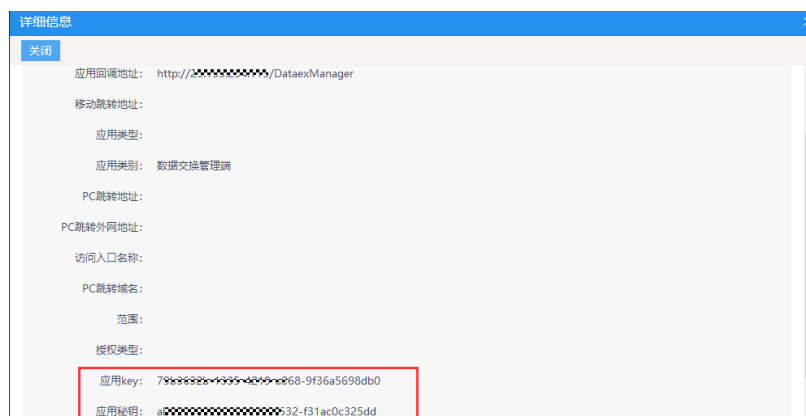
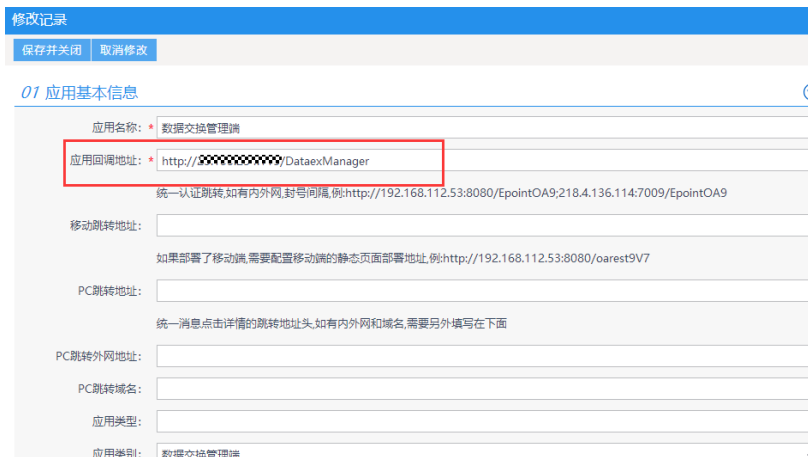


图 3-87 修改记录



- b. 订阅用户，订阅角色：打开支撑平台的统一应用与服务-统一应用管理-应用管理。

订阅用户：用于设置需要同步哪些用户到应用系统，公开同步所有；

订阅角色：用于设置需要同步哪些角色到应用系统，需要勾选 adminGroup 和 ouAdminGroup

图 3-88 勾选

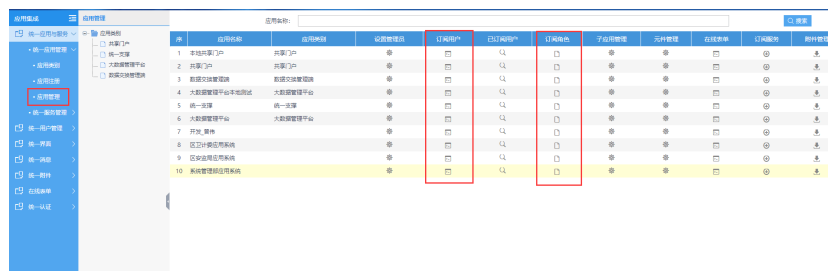


图 3-89 设置用户范围

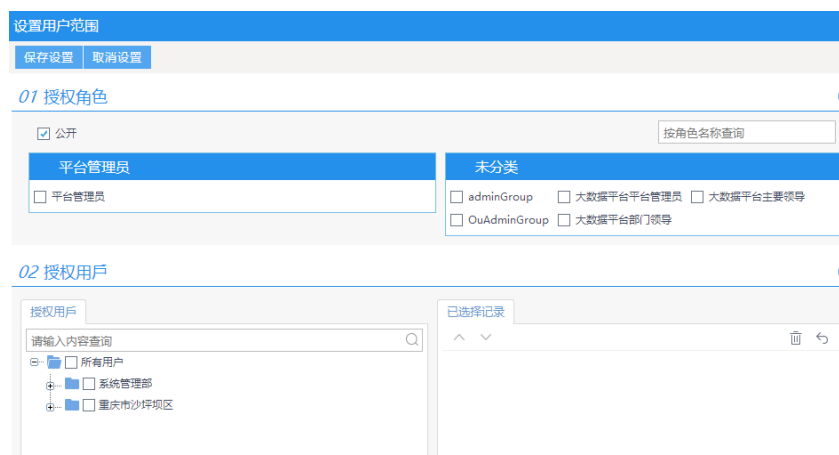


图 3-90 角色分发



c. 修改框架配置文件 (frame.properties)

/usr/local/tomcat/webapps/edc/WEB-INF/classes/frame.properties

SOAAddress: 支撑平台地址+/rest/organization, 如果配置了外网映射, 需要配置内网地址;

SOA_Appkey_secret: 第一步的应用 key 和应用密钥, 使用分号隔开;

IsOAuthEnabled: 1

图 3-91 修改 1



图 3-92 修改 2



• 访问验证

登录相应页面, 正常显示。业务平台部署成功。

3.3.8 数据交换节点

• 组件说明

交换节点用于执行各自的交换作业。

• 数据库还原

不同部门可能使用不同的交换节点, 每个交换节点需要使用单独的数据库。

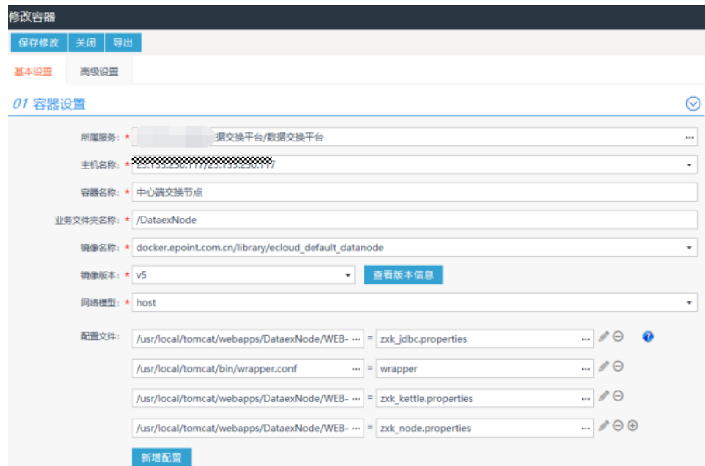
a. 在 MySQL 实例上面创建数据库, 数据库名称为 dxp_node;

中心节点: epointdxp_node 部门前置节点: epointdxp_node_qzk

b. 从数据交换节点部署包中复制出数据库备份文件, 还原到上一步创建的数据库中。

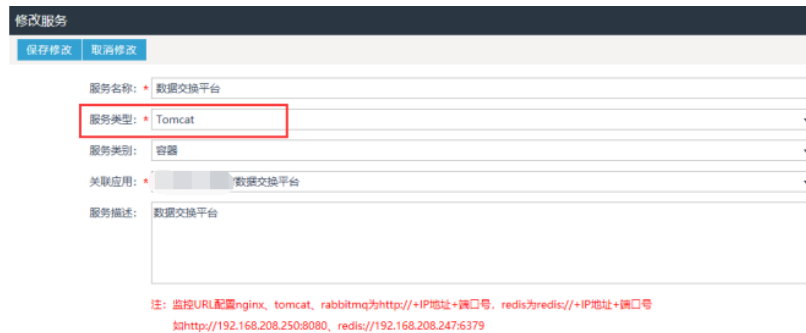
• 容器创建

图 3-93 修改容器



所属服务：新增服务的时候需要注意一定要选择 Tomcat 类型。

图 3-94 修改服务

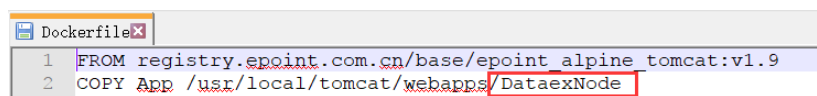


主机名称：按照规划的服务器选择用于部署资源目录管理平台的服务器即可。

容器名称：按照实际情况取名。

业务文件夹名称：需要与 Dockerfile 文件中的项目名称一致。

图 3-95 修改 1



镜像名称：使用流水线构建完成之后的镜像地址。

图 3-96 修改 2



镜像版本：按照实际情况选择。一般更新之后构建容器选择最新的（最高的）版本。

- **配置修改**

所有的配置文件都通过配置文件覆盖的方式进行配置，一定要在原来的配置文件的基础上进行修改，不要复制其他的项目中同名的配置文件，避免特殊配置项遗漏。

- a. 数据库配置文件（jdbc.properties）：

/usr/local/tomcat/webapps/DataexNode/WEB-INF/classes/jdbc.properties

url：填写数据库jdbc连接；

Username：填写数据库用户名；

password：填写数据库密码。

- b. 交换配置文件（kettle.properties）：

/usr/local/tomcat/webapps/DataexNode/WEB-INF/classes/kettle.properties

LOGURL：配置交换管理端的地址，如果配置了外网映射，请填写内网地址；

- c. 节点配置文件（node.properties）：

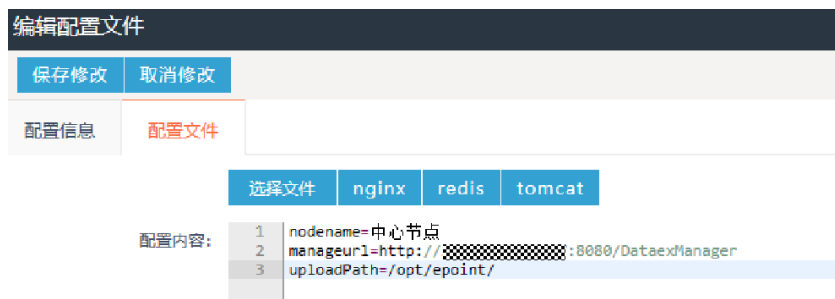
/usr/local/tomcat/webapps/DataexNode/WEB-INF/classes/node.properties

nodename：配置节点名称不能与其他节点重复；

manageurl：配置交换管理端的地址，如果配置了外网映射，请填写内网地址；

uploadPath：文件接受路径，配置本地文件夹路径，文件夹最后需要带分隔符例如：C:\Users\Administrator\Desktop\uploadpath\ 或者 /opt//

图 3-97 编辑配置文件



- **部署验证**

访问该地址：<http://IP:端口号/DataexNode/rest/upload/checkMD5>（注意替换IP和端口号），如果页面返回空白，说明未放注册码，需到日志中查看注册码信息，并申请后将注册码放到 DataexNode\WEB-INF\lib 文件夹下（通过更新镜像包的方式上传机器码）。返回以下页面说明部署成功。

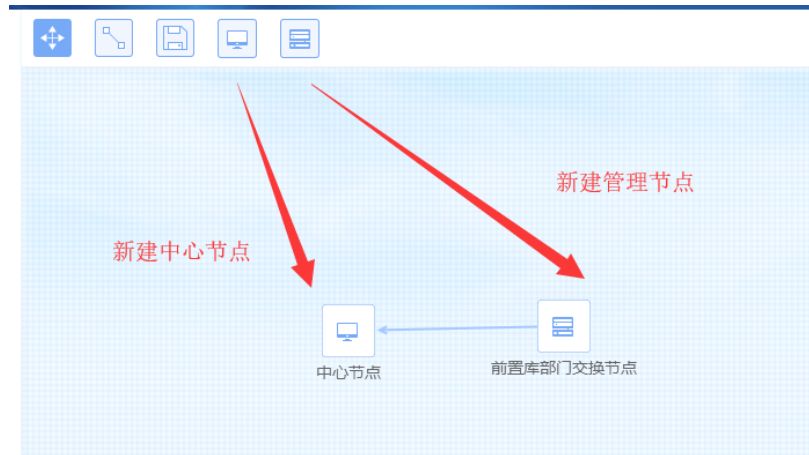
图 3-98 地址



- **节点注册**

- a. 打开交换管理端，登录，首页拖动图标新建节点。

图 3-99 新建节点



- b. 填写节点基本信息，保存并关闭：
- 节点名称：需要与节点的 node.properties 配置文件中一致；
 - 节点 IP：交换节点部署的容器 IP；
 - 节点分类：按照业务自行划分；
 - 端口号：交换节点部署的 tomcat 端口号；
 - 主机名：新建的数据库的 IP 地址；
 - 数据库名：新建的数据库的名称；
 - 端口号：新建的数据库的端口号；
 - 用户名：新建的数据库的用户名；
 - 密码：新建的数据库的密码；

图 3-100 编辑节点

- c. 添加数据源，右键新增的节点，选择数据源管理，单击新增数据源，填写资源库的连接信息。其中连接名称一定要与数据库名称一致。
- 至此交换平台部署完毕，需要注意的是：Kettle.properties文件中的*_DB 配置项与交换节点 MySQL 数据库名称与数据源管理中新增的数据源名称这三者必须保持一致。

图 3-101 数据源管理 1

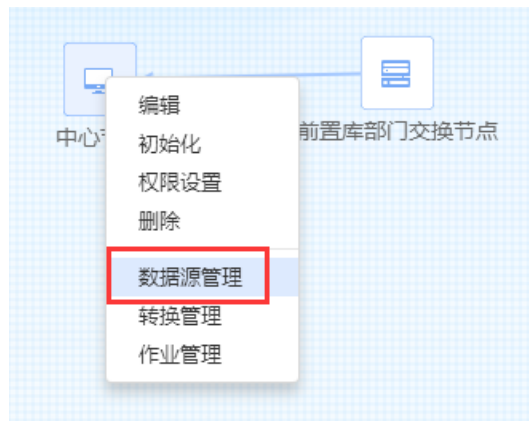


图 3-102 数据源管理 2



- **接入统一认证**

修改支撑平台配置 (SSOClient.properties)

```
/usr/local/tomcat/webapps/DataexManager/WEB-INF/classes/SSOClient.properties
```

verification_type: 1;

client_id: 上个步骤的应用 key;

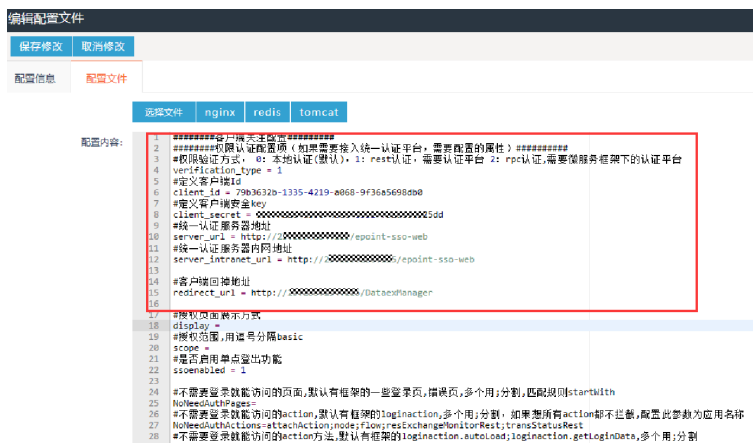
client_secret: 上个步骤的应用密钥;

server_url: 统一认证的浏览器访问地址;

server_intranet_url: 非必填, 如果统一认证配置了映射, 内外网地址不一样的情况, 需要在此处配置统一认证的内网地址;

redirect_url: 交换管理端的浏览器访问地址, 与注册应用时填写的一致。

图 3-103 编辑配置文件



- 访问验证

登录相应页面，正常显示。业务平台部署成功。

3.4 场景实践

3.4.1 概述

场景介绍：以人口库建设为场景，介绍数据从部门前置库进入到最后从资源目录挂载出来的整个流程的实现。

前置条件：部门已经建设好前置库，并且在前置库中推了数据。

整体流程：

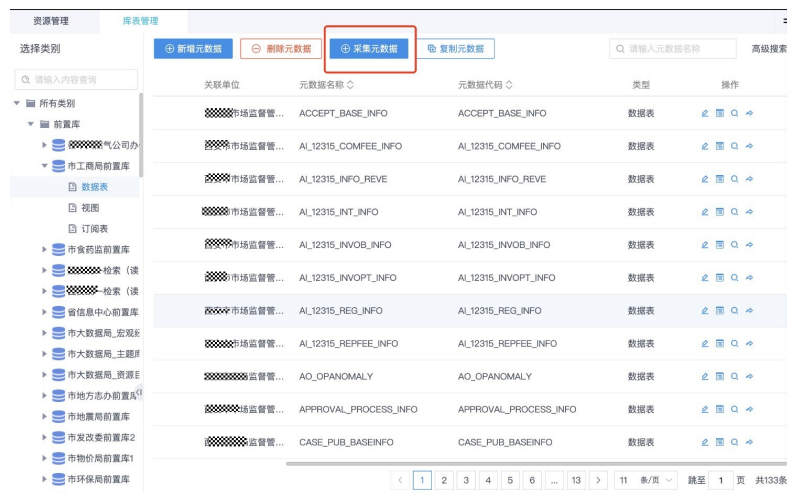
1. 部门前置库数据经过库表采集环节进入资源库（中心前置库）。
2. 资源库数据经过数据质检、分析来源并接入到数据治理平台对应的模型中（对应人口库模型构建环节）。
3. 数据治理平台的数据经过落标与清洗、融合与发布将数据发布到DWS发布库上。
4. 发布库的数据经过挂载资源目录环节将资源挂载到共享门户上。

3.4.2 库表采集

左侧导航栏依次选择单击资源管理、库表管理、采集元数据，配置所需要的数据源表，并依照数据标准规范，合理命名元数据名称，以及代码等项，将采集到的数据资源进行复制至部门资源库，找到部门资源库对应库表物化库表资源，最后通过数据桥接同步对应库表数据。

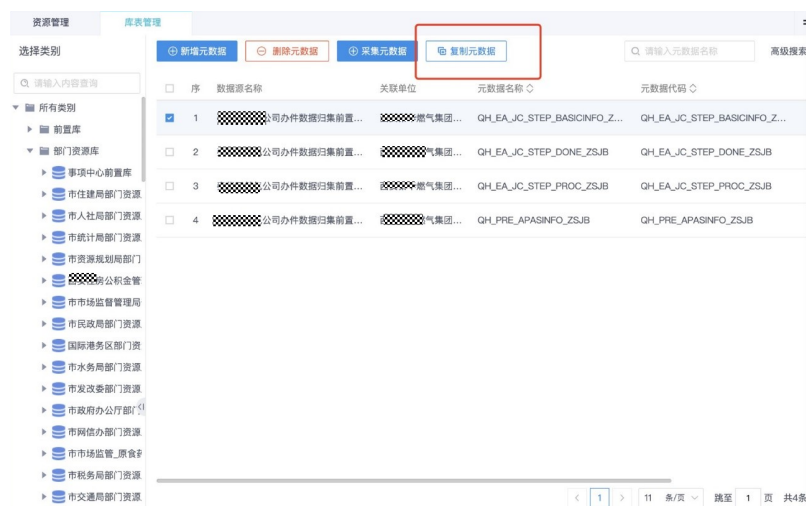
采集元数据：

图 3-104 采集



复制元数据:

图 3-105 复制



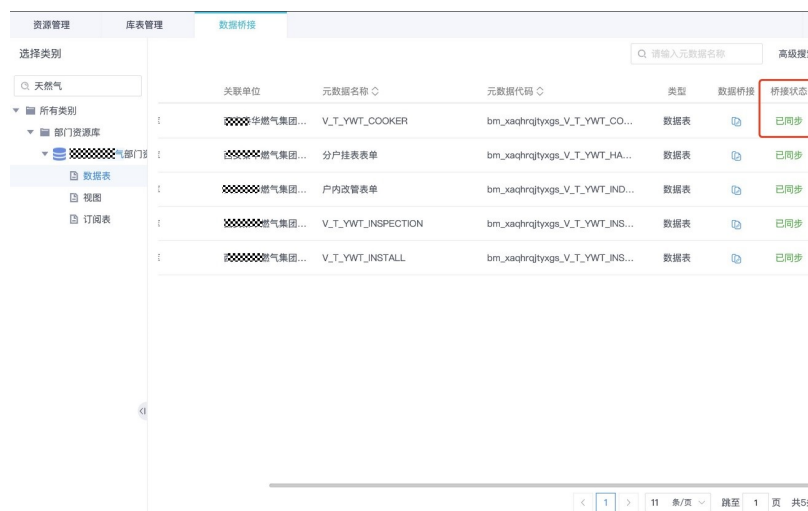
库表物化:

图 3-106 物化



数据桥接：

图 3-107 桥接

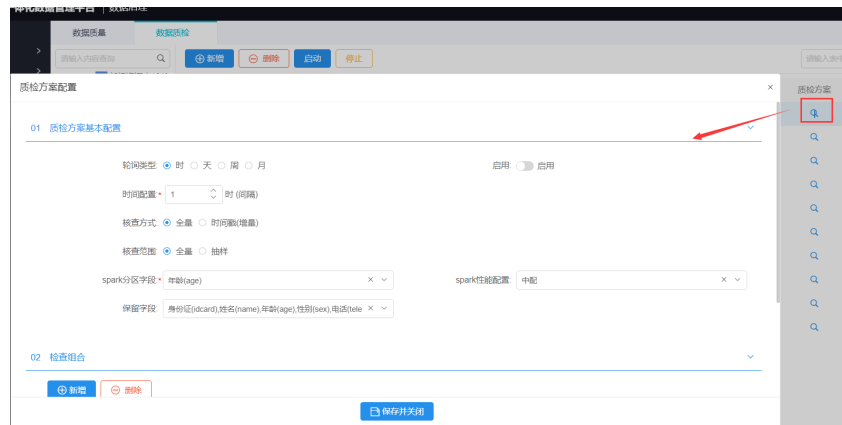


3.4.3 数据质检

数据质检主要是将库中错误数据检测出来，与治理流程的清洗定义中的过滤规则不同的是，独立质检是将不规范的数据检测出来，并不改变质检表的数据，治理流程中的过滤规则是将不符合规范的数据过滤掉，改变质检表的数据。检测出错误数据后，可以查看错误数据原因、导出错误数据到 excel 中、导出质检报告供部门整改用，数据质检完对错误数据可以发起工单给到具体部门的部门进行错误数据整改。

如果刚开始就要对人口基本信息 t_rk_baseinfo 表进行检测它的数据是否规范，是否有错误数据时，就可以将这张表先质检一下，配置质检方案、质检规则

图 3-108 数据质检 1



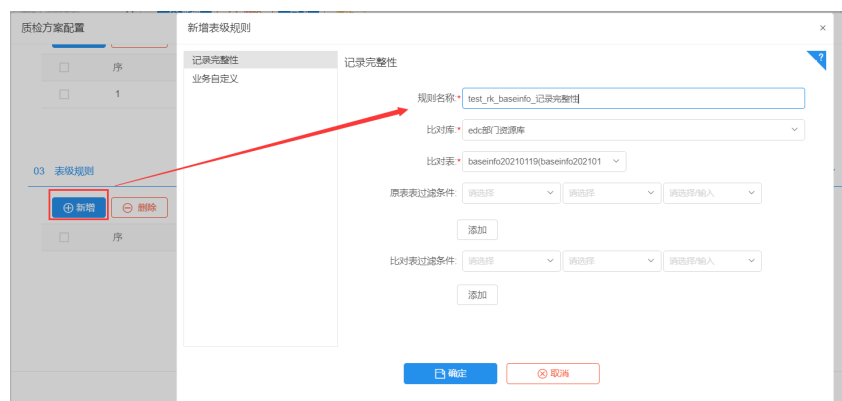
检查组合是新增记录级的质检规则，比如可以对身份证字段 idcard 配一个身份证的规范检查，将不符合身份证号码标准的数据检查出来。

图 3-109 数据质检 2



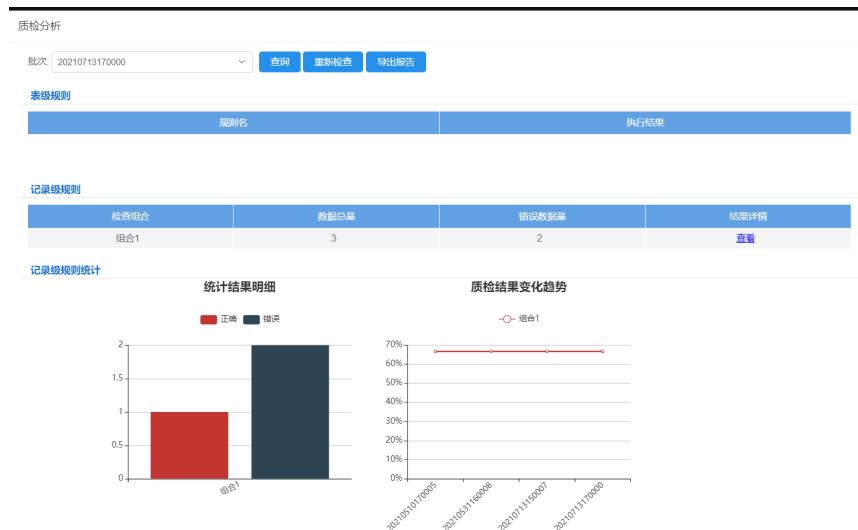
表级规则是以表为维度来检查的，比如可以配置对照表来检查检测表的数量是否一致，无需比较其中的具体数据

图 3-110 数据质检 3



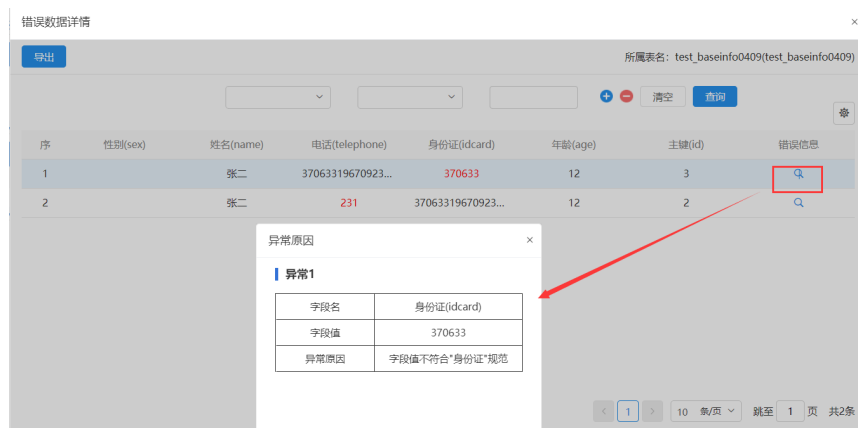
配置好质检方案之后在列表页面去启动任务，执行状态为成功说明该质检任务跑成功了，在质检分析页面可以查看是否有错误数据，具体的数据数据，以及统计质检结果的变化趋势。

图 3-111 数据质检 4



单击查看详情可以查看具体的错误数据以及错误信息，单击导出可以将错误数据导出到 excel 中

图 3-112 数据质检 5



单击质检分析页面的“导出报告”按钮可以将质检的报告导出来

3.4.4 构建人口库模型、分析来源并接入

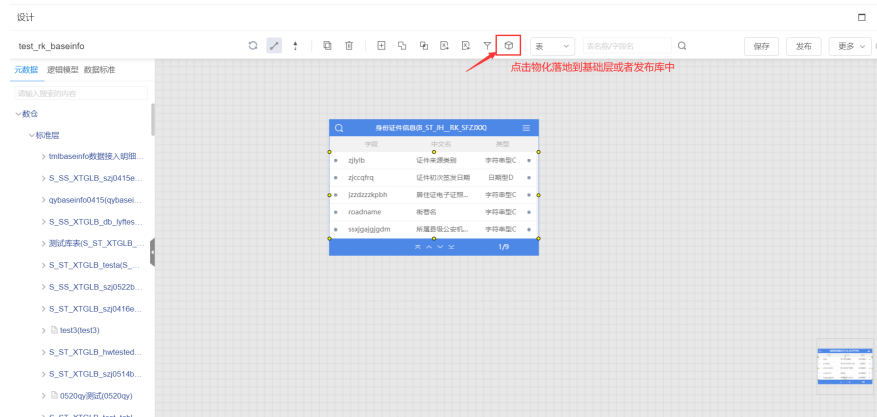
构建人口库模型

在标准的人口库中应包含人口基本信息表 T_RK_BaseInfo、人口婚姻 T_RK_Marriage 等表信息，先结合实际需求将人口库所需表维护至模型管理中（下面4模型设计）。

融合、发布所需目标表可于【数据标准】-【数据模型】-【模型管理】-【逻辑模型】处设计，管理员从左侧树中拖出需要的表、字段或直接手动添加，拼接为最终的可发

布表结构，通过物化功能将其物化至基础层、发布库中。例如可将人口信息表和人口婚姻表拖拽到画板中，通过表合并将两个表字段合并形成一张新表，通过物化落地到实际的数仓标准层中。

图 3-113 构建人口库模型



分析来源并接入

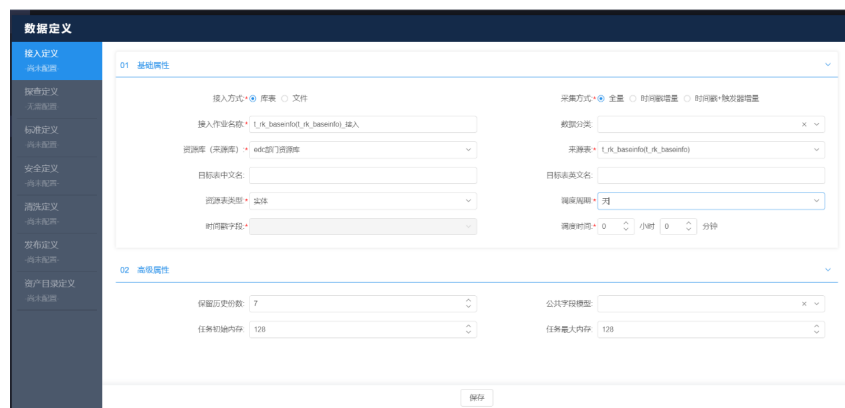
分析人口库中所需数据，得出数据来源，例：户籍人口信息 rk_baseinfo、民政局 婚姻状况 rk_marriage、人社局 社保信息 Insurance、教育局 教育信息 Education。将这些部门资源库配置入系统，并配置接入作业

首先要根据需求将不同部门的人口信息实体表 t_rk_baseinfo 从部门资源库采集到数仓贴源层中。

选择部门资源库（来源库）、来源表，根据选择的来源表判断表的类型是实体还是明细，比如如果设定好调度周期和调度时间，保存成功后单击【推送并启用】，等待作业跑成功，可以先关闭页面，任务跑结束后会发送一个是否跑成功的系统提醒。在数仓的贴源层会生成一个贴源层表，如果是实体表格式为 O_ST_{@dept}_{@table}，明细表格式为 O_SS_{@dept}_{@table}（{@dept}与{@table} 分别代表部门简称及原表名）。

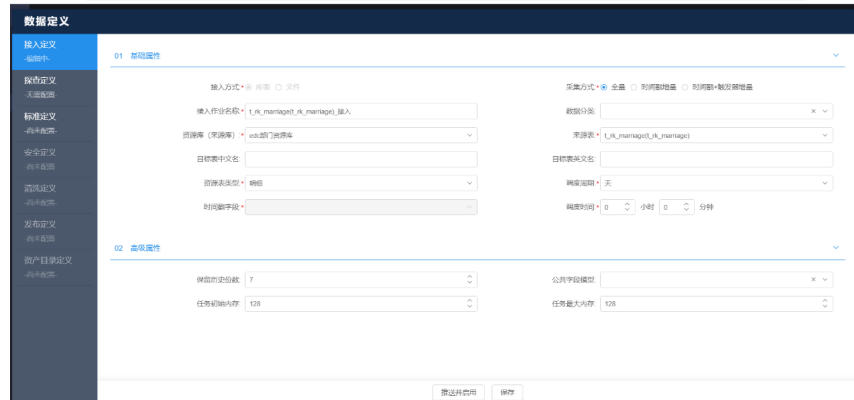
如人口信息实体表t_rk_baseinfo接入成功后贴源层生成O_ST_XTGLB_t_rk_baseinfo表。

图 3-114 分析来源并接入 1



人口婚姻表 t_rk_marriage 资源表类型选择明细，接入成功后贴源层生成 O_SS_XTGLB_t_rk_marriage 表

图 3-115 分析来源并接入 2



采集方式说明

全量模式接入：推荐 1000w 以下数据量使用，平均 100ws 数据只需 60s 即可接入至数仓中

时间戳接入：只支持数据是增量插入的，存在删除、修改数据的情况不建议使用，数据会不准确

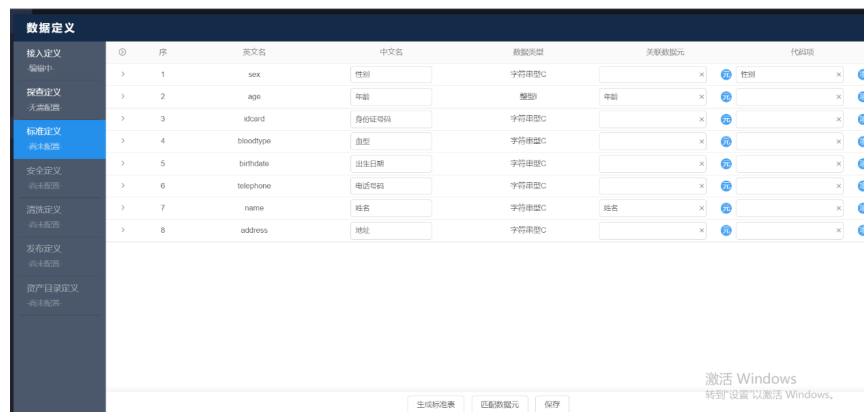
时间戳+触发器接入：推荐 1000w 数据量以上使用，但贴源层表数据不能直接使用，需数据清洗至标准层，hive 中数据处理存在临时表，处理数据时间比前两种方式长

3.4.5 落标与清洗

由于部门资源库表由各部门维护，可能会存在相同数据结构不一致等状况（例：国标姓名长度为 30，部门可能存在长度过长数据），故需通过标准定义为部门资源库表配置数据元并基于数据元为其生成标准表。

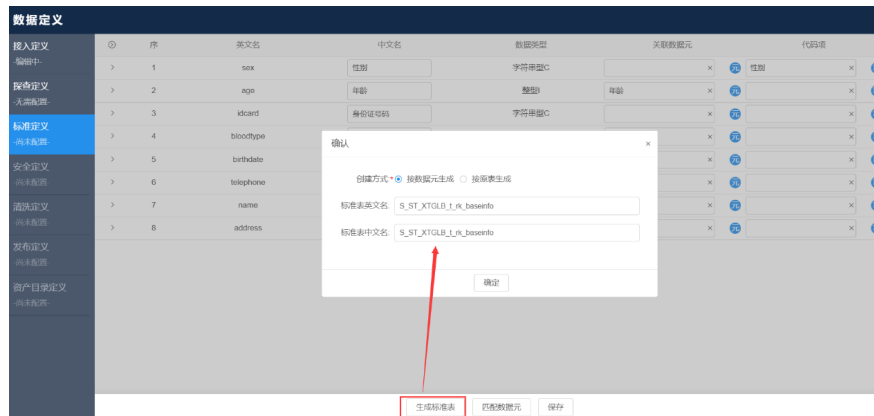
不同部门的业务库，字段规则并不统一，要形成统一的库，就必须指定统一标准，此时便可借助数据定义，规范字段的中文名、关联标准的数据元，如果字段是代码项的也可以关联代码项，帮助管理员建立有统一标准的标准表。

图 3-116 落标与清洗 1



默认按数据元生成标准表，默认的标准表名格式为 S_ST_{@dept}_{@table}，{@dept}与{@table}，分别代表部门简称及原表名。比如人口信息表生成的标准表为 S_ST_XTGLB_t_rk_baseinfo。

图 3-117 落标与清洗 2



构建完标准表后可通过清洗定义将部门资源库表数据交换到标准表，在这个过程中，可以处理掉一些较为明显的问题数据，例如人口基本信息中家庭住址存在特殊字符需替换掉，婚姻信息中注册时间不符合标准时间格式 yyyy-MM-dd hh:mm:ss 需做时间格式转换，人口基本信息中身份证号码 idcard 不符合身份证号码规范需过滤掉这些错误数据等等。

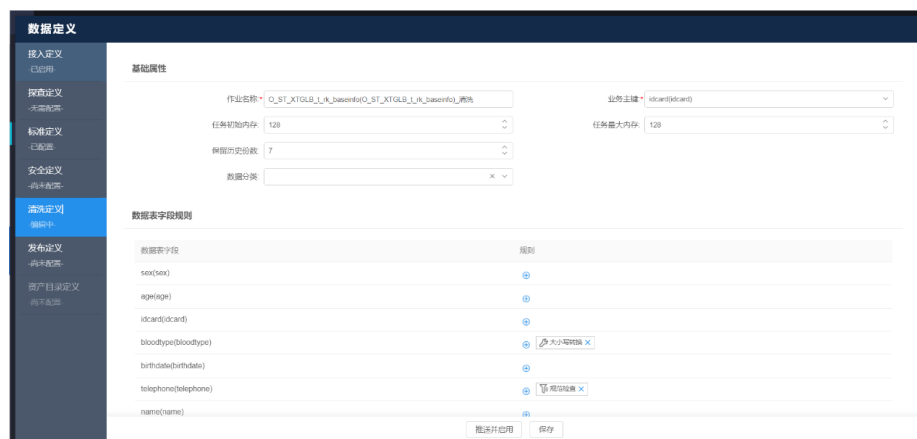
注意

清洗定义中清洗规则为已知异常数据及处理方式，直接对数据进行处理，来源与目标数据量一致。

清洗定义中过滤规则将错误数据过滤掉，目标表数量可能小于来源表数量，对于这些质检产生的异常数据，通过质检工单形式告知部门，进行修正，中心无法直接处理这些数据，故需告知部门整改后在下个批次识别为增量数据进行同步。

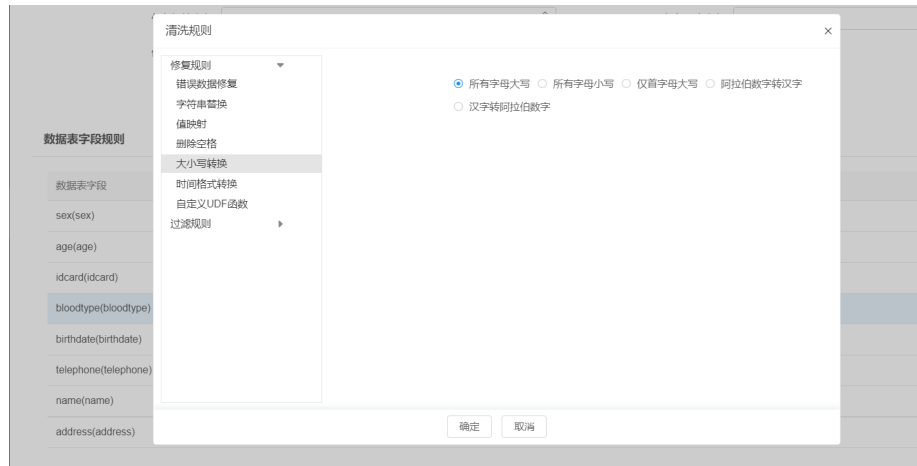
有了人口库标准表后，对于原业务库中不符合标准规则的数据进行处理即数据清洗，标准表生成成功后，通过配置修复规则对贴源层数据简单修复、值处理等操作和通过过滤规则将不符合规则的错误数据过滤掉从而形成数仓中的标准层数据。

图 3-118 落标与清洗 3



比如对人口信息基本表的标准表 S_ST_XTGLB_t_rk_baseinfo 的电话号码 telephone 字段配置电话规范检查，即可将不规范的电话号码的数据过滤掉，对血型字段 bloodtype 配置修复规则-大小写转换，将血型字段全部转为大写，对地址字段 address 配置去除空格规则，将该字段的数据去除空格。

图 3-119 落标与清洗 4



3.4.6 融合与发布

融合不是一个必须的过程，对于例举的人口基本信息、概况信息等常用宽表，可用到融合，而如果仅公安临时需要人社、婚姻及教育局部分数据字段，可直接通过发布作业进行拼接。

应用开始构建的人口库模型，将表物化到基础层、发布库。通过融合作业，讲前序产生的户籍、教育等标准表进行拼接，数据汇聚入人口概况宽表。再通过发布作业将概况宽表发布至发布库。

数据融合非治理流程中的必须步骤，主要用于将多张业务表合并为一张业务宽表，在人口库形成场景中，公民的基本信息、教育信息、车辆信息、劳务信息等可能存在于多张业务表中，如有新的业务需求于一张表中查询某公民的全部信息，则需要将多张业务表的关键信息融合至一张业务宽表，此时便需要使用数据融合功能。

现在融合有两种类型，一种是通过建模平台配置，另一种是通过自定义 sql 来融合。

- 建模方式融合

如下图，是通过建模来配置融合作业，拖入人口基本信息表和人口婚姻表，将两表进行表关联，数据流向本地库中，选择本地库中存在的表进行字段一一映射，本地库存在的的表是指基础层的表，一般通过数据模型-逻辑模型创建。

图 3-120 建模方式融合 1

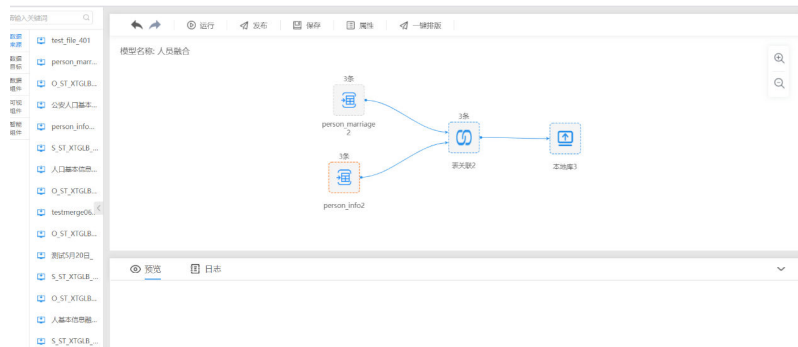
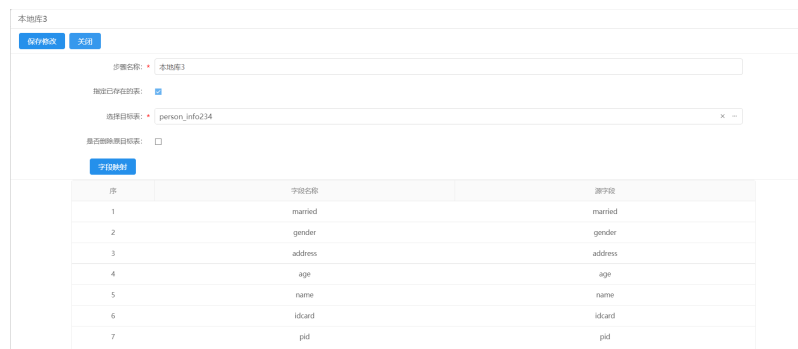


图 3-121 建模方式融合 2



图 3-122 建模方式融合 3



- 自定义 sql 融合

选择来源表和目标表，目标表是基础层的表，要确保来源表的表结构表名称和目标表一一对应，填写融合的 sql 语句，保存完之后在列表页启动作业。

交换任务成功运行后，系统将根据融合配置将于数仓基础层用张业务表合并为一张宽表。

这边的发布和发布定义的不同之处是这边的发布主要用于融合之后的表的发布，可由基础层表或经过融合的基础层表发布至发布库供部门使用。

图 3-123 自定义 sql 融合 1

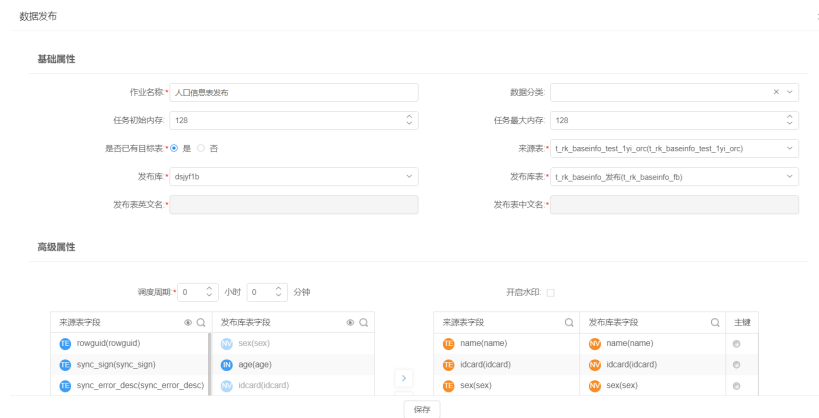
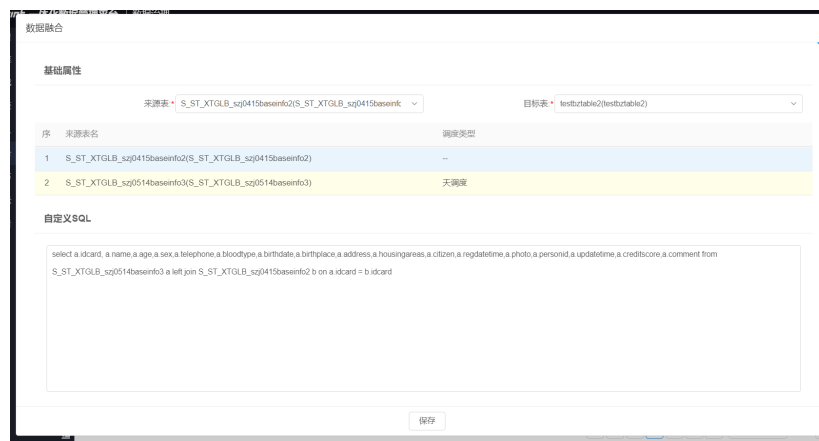


图 3-124 自定义 sql 融合 2



3.4.7 挂载资源目录

左侧导航栏依次选择单击资源管理、资源注册、按库表注册，在跳转页面选择想要挂载的数据，单击页面右侧的库表注册按键，再次跳转至图三后，填写数据对应的基本信息，以及共享属性后提交审核，并完成挂载

资源注册：

图 3-125 资源注册

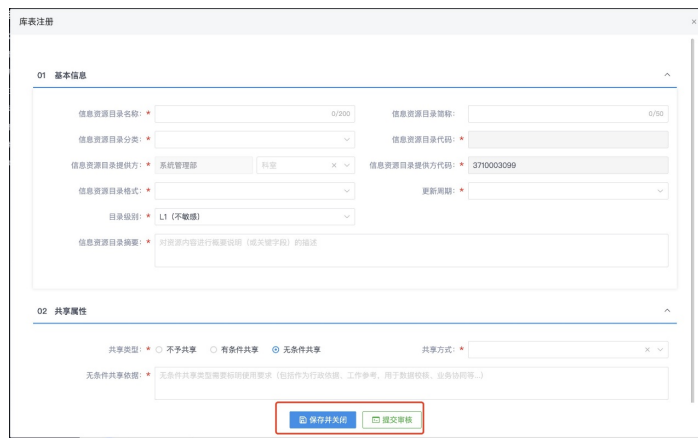


库表注册：

图 3-126 库表注册 1



图 3-127 库表注册 2



4 附录

背景信息：

自党的十八大以来，党中央、国务院多次研究部署并加快推进“互联网+政务服务”，陆续出台审批服务便民化、“互联网+政务服务”、优化营商环境等一系列政策文件；十九大以后，更是明确提出打造数字政府、培育数字经济、构建数字社会，尤其是在《中华人民共和国国民经济和社会发展第十四个五年规划和 2035年远景目标纲要》中，着重提出提高数字政府建设水平，加强公共数据开放共享，推动政务信息化共建共用，提高数字化政务服务效能，推动政府治理体系和治理能力现代化，建设人民满意的服务型政府。

作为数字政府建设的一部分，围绕推进政务服务一体化建设的相关问题，国务院办公厅和相关部门出台了一系列指导全国一体化政务服务平台建设的重要文件。2018年8月，国务院出台《关于加快推进全国一体化在线政务服务平台建设的指导意见》，重点围绕十六类三十三项任务，提出加快建设全国一体化在线政务服务平台，全面实现“一网通办”。2020年9月，《国务院办公厅关于加快推进政务服务“跨省通办”的指导意见》提出依托全国一体化政务服务平台和各级政务服务机构，着力打通业务链条和数据共享堵点，推动更多政务服务事项“跨省通办”，为建设人民满意的服务型政府提供有力保障。2021年11月，针对政务服务平台移动端建设管理分散、标准规范不统一、数据共享不充分、技术支撑和安全保障体系不完备等突出问题，出台《全国一体化政务服务平台移动端建设指南》，从政务服务平台移动端总体架构、支撑能力、服务功能、服务方式等方面加强和规范全国一体化政务服务平台移动端建设。2021年12月，中央网络安全和信息化委员会印发《“十四五”国家信息化规划》，表明大力推进“一网通办”，优化全国一体化政务服务平台服务水平，促进政务服务线上线下融合，完善统一身份认证、电子证照等共性支撑体系，有力支撑政务服务标准化、规范化、便利化，实现全国范围内“一次认证、全网通办”。2022年3月，针对政务服务标准不统一、线上线下服务不协同、数据共享不充分、区域和城乡政务服务发展不平衡等问题，国务院出台《关于加快推进政务服务标准化规范化便利化的指导意见》，明确提出加强跨层级、跨地域、跨系统、跨部门、跨业务协同管理和服务，充分发挥全国一体化政务服务平台“一网通办”支撑作用，进一步推进政务服务运行标准化、服务供给规范化、企业和群众办事便利化，有效服务生产要素自由流动和畅通国民经济循环，更好满足人民日益增长的美好生活需要，为推动高质量发展、创造高品质生活、推进国家治理体系和治理能力现代化提供有力支撑。2022年06月23日，国务院发布《关于加强数字政府建设的指导意见》，进一步明确提出充分发挥全国一体化政务服务平台“一网通办”枢纽作用，推动政务服务线上线下标准统一、全面融合、服务同质，构建全时在线、渠道多元、全国通办的一体化政务服务体系，推行政务服务事项集成化办理，推广“免申即享”、“民生直达”等服务方式，打造掌上办事服务新模式，提高主动服务、精准服务、协同服务、智慧服务能力。

为贯彻落实党中央、国务院关于建设一体化政务服务平台，提高数字化政务服务效能的文件精神，XX委省政府高度重视“数字政府”改革建设，于2021年10月印发了《XX数字政府建设“十四五”规划》，提出全面提升一体化政务服务平台功能，构建线上线下融合的政务服务体系，推动政务服务平台、政务服务大厅（便民中心、站点）、移动端、自助端、便民服务热线等线上线下服务渠道向基层、向偏远地区延伸，不断满足群众多层次多样化服务需求。XX积极响应国家和省级要求，围绕政府治理和公共服务的改革需要，加快构建全市一体化网上政务服务体系。XXX人民政府分别于2019年1月和2021年3月印发《XXX全面推进“一网通办”加快数字化转型行动方案》和《政务信息系统打通与数据共享工作实施方案》，提出全面推进一体化在线政务服务平台建设，紧紧围绕政府治理和公共服务的改革需要，不断完善数据赋能和数字惠民服务，加快构建全市一体化网上政务服务体系，推进跨层级、跨区域、跨系统的协同管理和服务，转变政府职能，推进政务服务标准化、规范化、便利化。

通过本项目建设加强标准规范体系建设、建强数据中台能力，规范平台安全运行和运营保障能力等内容，积极落实和响应国家、省、市关于数字政府、一体化政务服务平台建设工作要求和整体目标，建成纵向贯通、横向协同、全面覆盖的一体化政务服务体系，提升XXX的政务服务品质，数据共享交换取得突破。

5 修订记录

表 5-1 修订记录

发布日期	修订记录
2024-04-24	规范词、敏感词专项处理，章节优化
2023-12-04	更新发布
2023-09-19	第一次正式发布。