# 元数据采集

[元数据采集 1](#_Toc453515608)

[1 视图创建 3](#_Toc453515609)

[1.1 简介 3](#_Toc453515610)

[1.2 系统视图 3](#_Toc453515611)

[1.3 用户视图 4](#_Toc453515612)

[2 DB 数据记录采集配置 4](#_Toc453515613)

[2.1 简介 4](#_Toc453515614)

[2.2 步骤 5](#_Toc453515615)

[2.2.1 创建视图 5](#_Toc453515616)

[2.2.2 创建元数据 5](#_Toc453515617)

[2.2.3 创建数据源 6](#_Toc453515618)

[2.2.4 创建数据源和元数据的映射关系 8](#_Toc453515619)

[2.2.5 任务配置 9](#_Toc453515620)

[2.2.6 元数据审批 11](#_Toc453515621)

[3 DB数据字典采集适配器 11](#_Toc453515622)

[3.1 简介 11](#_Toc453515623)

[3.2 步骤 11](#_Toc453515624)

[3.2.1 创建视图 11](#_Toc453515625)

[3.2.2 创建元数据 11](#_Toc453515626)

[3.2.3 创建数据源 11](#_Toc453515627)

[3.2.4 任务配置 13](#_Toc453515628)

[3.2.5 元数据审批 13](#_Toc453515629)

[4 JDBC驱动采集适配器 13](#_Toc453515630)

[4.1 简介 13](#_Toc453515631)

[4.2 步骤 14](#_Toc453515632)

[4.2.1 创建视图 14](#_Toc453515633)

[4.2.2 创建元数据 14](#_Toc453515634)

[4.2.3 创建数据源 15](#_Toc453515635)

[4.2.4 任务配置 16](#_Toc453515636)

[4.2.5 元数据审批 16](#_Toc453515637)

[4.2.6 采集结果展示 16](#_Toc453515638)

[5 Kettle采集适配器 17](#_Toc453515639)

[5.1 简介 17](#_Toc453515640)

[5.2 步骤 17](#_Toc453515641)

[5.2.1 创建视图 17](#_Toc453515642)

[5.2.2 创建元数据 18](#_Toc453515643)

[5.2.3 创建数据源 18](#_Toc453515644)

[5.2.4 采集两个数据字典 19](#_Toc453515645)

[5.2.5 生成kettle配置文件 19](#_Toc453515646)

[5.2.6 手动采集 20](#_Toc453515647)

[5.2.7元数据审批 20](#_Toc453515648)

[6 数据地图 20](#_Toc453515649)

[6.1 简介 20](#_Toc453515650)

[6.2 手工创建数据地图 21](#_Toc453515651)

[6.2.1 添加系统层次元数据 21](#_Toc453515652)

[6.2.2新增系统 21](#_Toc453515653)

[6.2.3 维护关系 22](#_Toc453515654)

[6.2.4 数据地图显示效果 22](#_Toc453515655)

[6.3 采集创建数据地图 22](#_Toc453515656)

[6.3.1新建悬挂点 22](#_Toc453515657)

[6.3.2新建模板映射 23](#_Toc453515658)

[6.3.3 创建数据源 24](#_Toc453515659)

[6.3.4 手动采集 24](#_Toc453515660)

[6.4 数据地图设置保存 25](#_Toc453515661)

[6.4.1 选中图中需要调整位置的节点拖动 25](#_Toc453515662)

[6.4.2 保存位置信息 25](#_Toc453515663)

[6.5 数据地图的下钻 26](#_Toc453515664)

[7 元数据分析 27](#_Toc453515665)

[7.1影响分析 27](#_Toc453515666)

[7.2血统分析 27](#_Toc453515667)

[7.3全链分析 28](#_Toc453515668)

[7.4 表关联度分析 28](#_Toc453515669)

[7.5 属性值差异分析 28](#_Toc453515670)

[8 元数据介绍 29](#_Toc453515671)

[8.1 元数据概念 29](#_Toc453515672)

[8.2 适配器 29](#_Toc453515673)

[8.3 元模型 29](#_Toc453515674)

[8.4 视图 30](#_Toc453515675)

[8.5 MedaCube简介 30](#_Toc453515676)

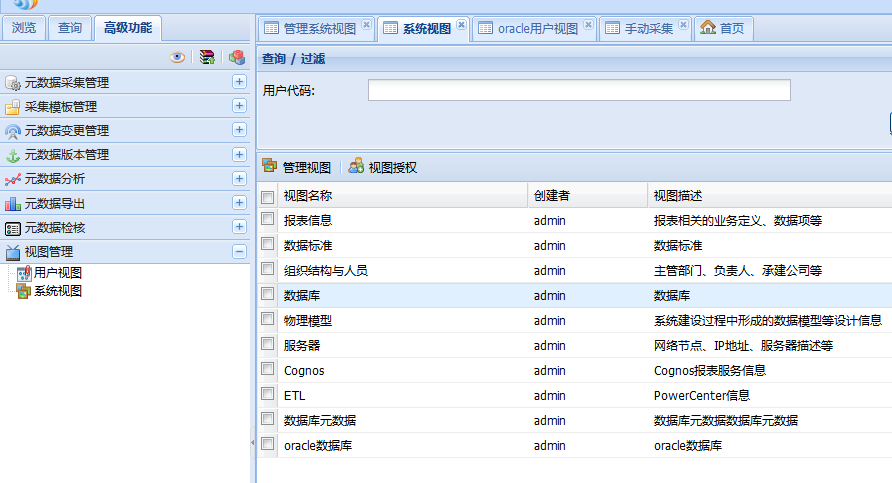
## 1 视图创建

### 1.1 简介

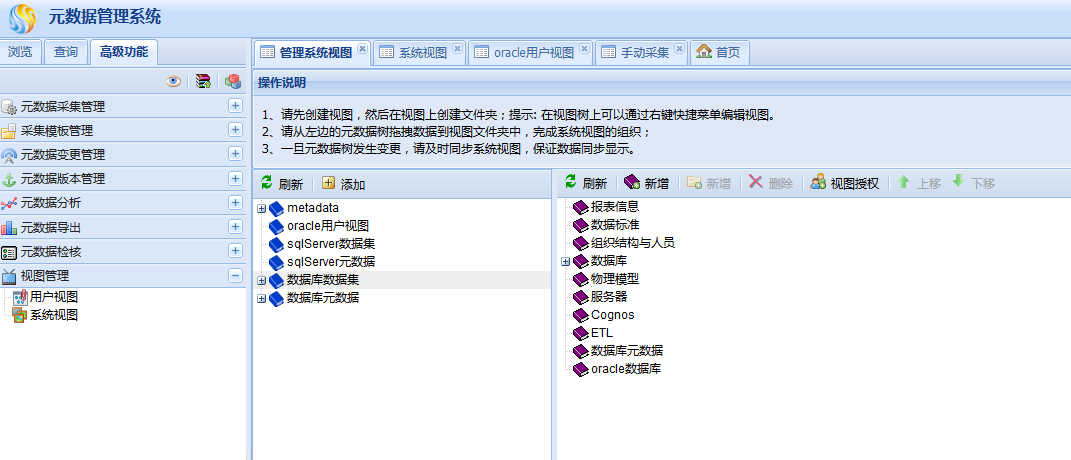
浦元元数据管理系统是将数据源挂载在视图上的，因此必须创建系统视图或者用户视图。（在浏览tab页面可以展示系统视图和用户视图）

### 1.2 系统视图

高级功能菜单栏->视图管理->系统视图。



点击视图管理



左侧显示的是元数据，可以将元数据拖放到相应的系统视图下，实现元数据的挂载。

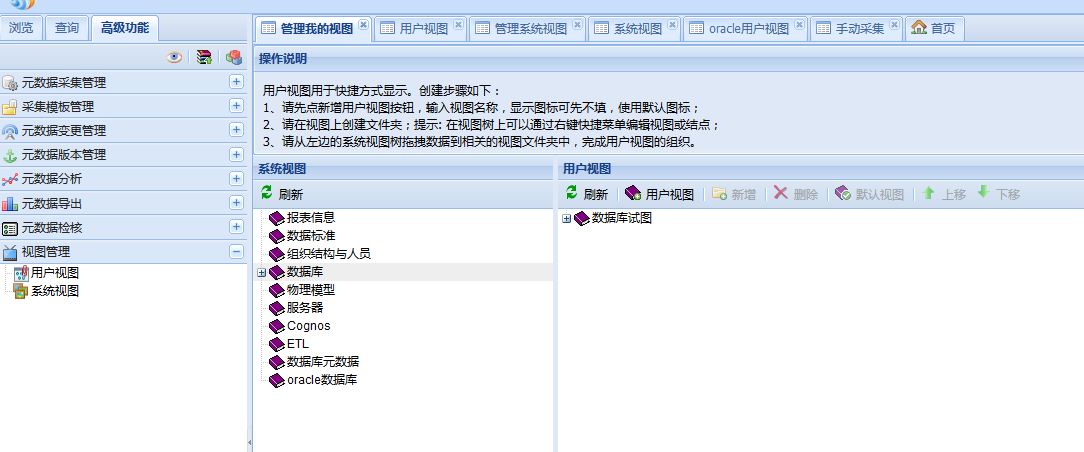
右侧展示的是系统视图。点击右侧的新增添加系统视图即可。

### 1.3 用户视图

高级功能菜单栏->视图管理->用户视图。



点击视图管理 创建新的视图



左侧显示的是系统视图（只可读）

右侧显示的是当前用户创建的用户视图。点击用户视图即可创建用户视图。

## 2 DB 数据记录采集配置

### 2.1 简介

为解决通过 DB 数据记录采集适配器采集元数据，通过 DB 数据采集配置将 DB 记录中记录与元模型进行映射。

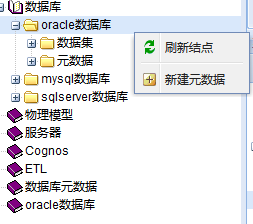
### 2.2 步骤

#### 2.2.1 创建视图

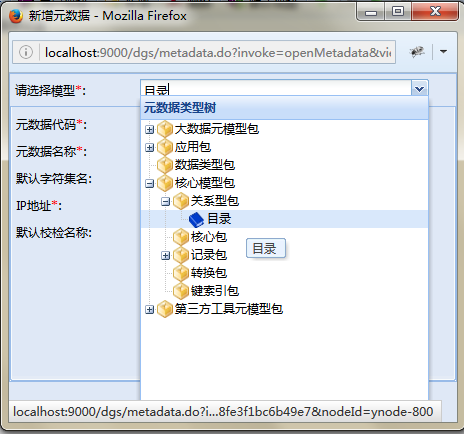
参照 [1视图创建](#_1_视图创建)

#### 2.2.2 创建元数据

在预览tab，找到创建的系统视图或者用户视图。右键点击创建元数据



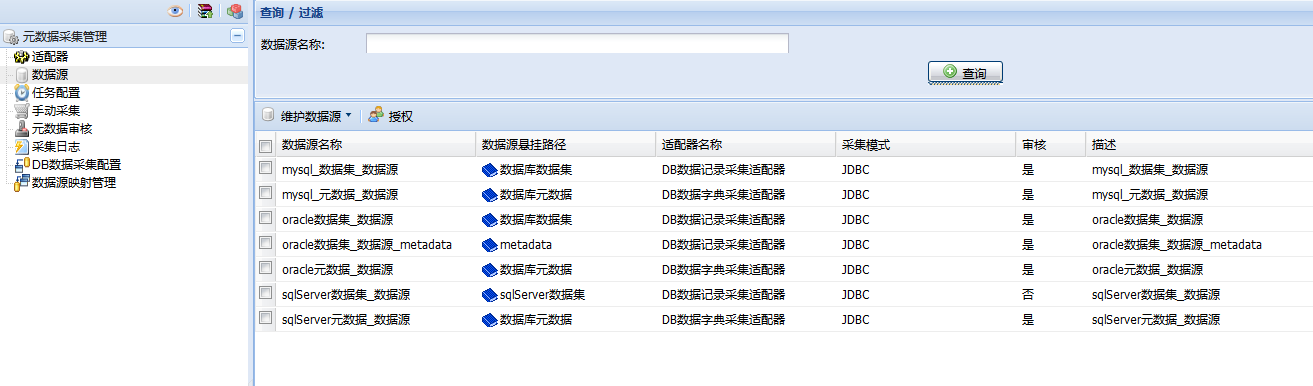
在弹出来的界面，选择核心模型包->关系型包->目录（数据库元数据类型）



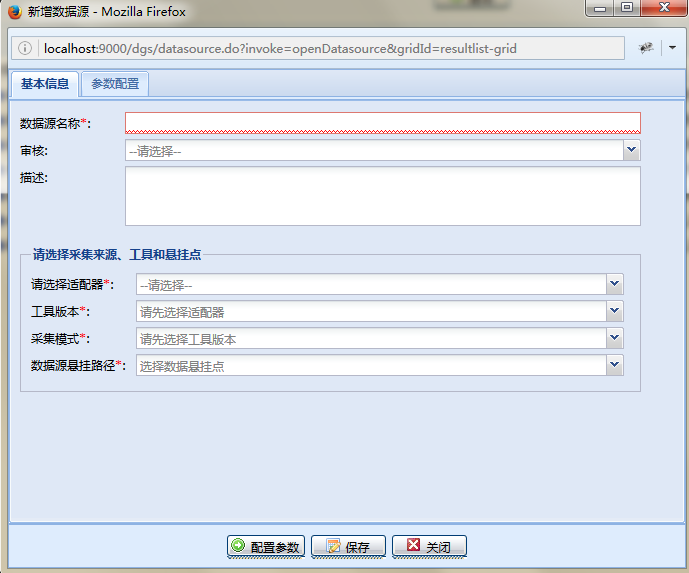
点击目录 然后输入元数据信息来创建元数据。创建完毕后，可以刷新视图，显示挂载的元数据。

#### 2.2.3 创建数据源

高级功能->元数据采集管理->数据源



点击维护数据源->增加

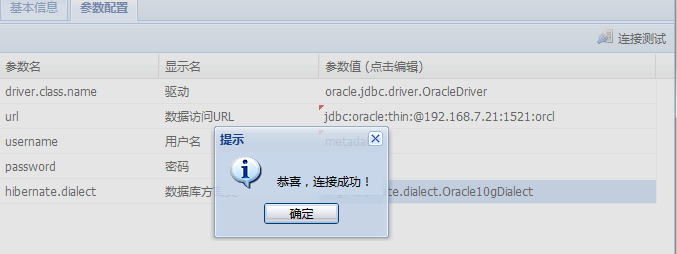


输入必要的数据源信息

请选择适配器 ：选择 DB数据记录采集适配器

数据源悬挂路径 ： 选择创建的元数据路径

在配置参数页面 输入数据库的详细信息

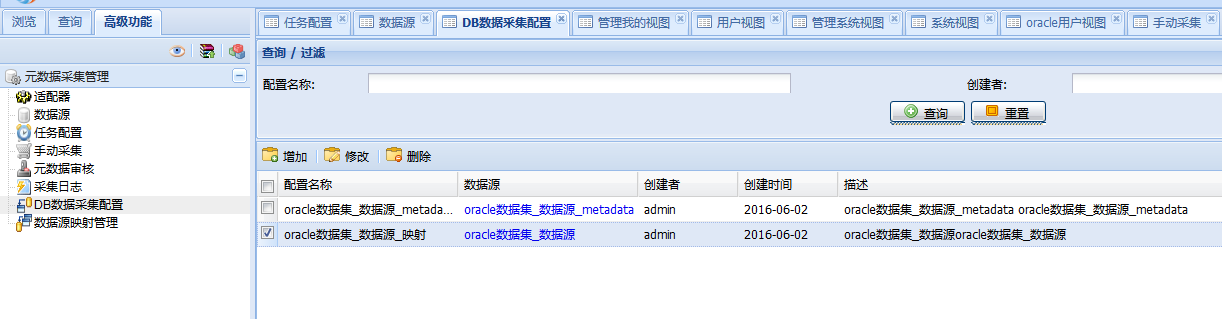


点击保存后，然后对该数据源授权。

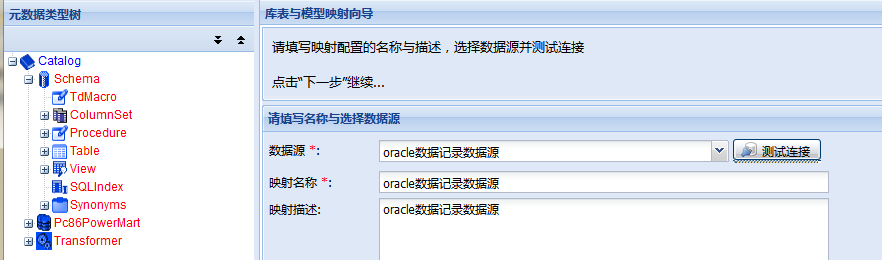
#### 2.2.4 创建数据源和元数据的映射关系

企业元数据资产存在同一元数据流经不同数据层的情况以及元数据在设计和实现不同的生命周期中有不同的  
存在形式情况，为建立不同数据层、不同生命周期的同一元数据的关系，引入元数据映射管理概念，其目标是通过  
配置目录之间的映射关系，在有映射关系的目录下如果元数据的代码相同则认为是同一元数据。

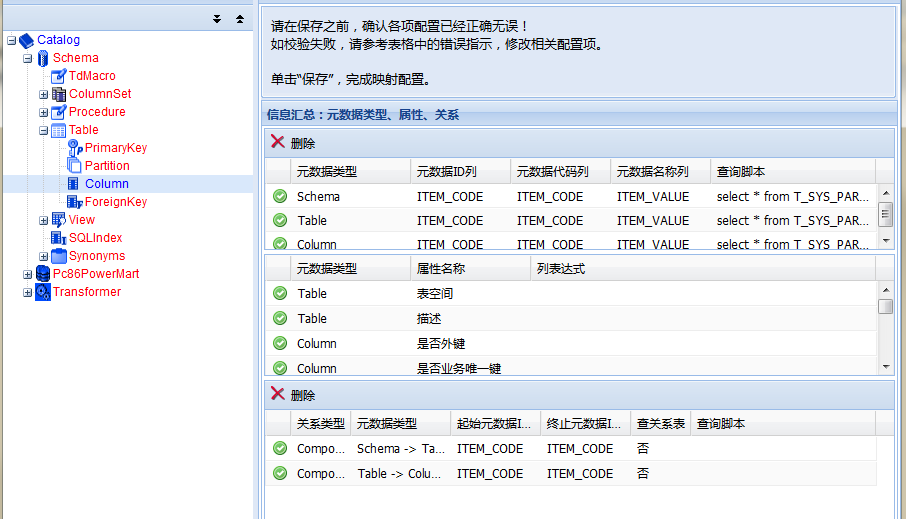
高级功能->元数据采集管理->DB数据采集管理



点击增加 添加映射。选择数据源 左侧会自动显示该数据源的catalog。



选择页面左边元模型树要配置的节点，然后在页面右方输入查询脚本，属性映射，组合关系和依赖关系，  
点击下一步并完成保存



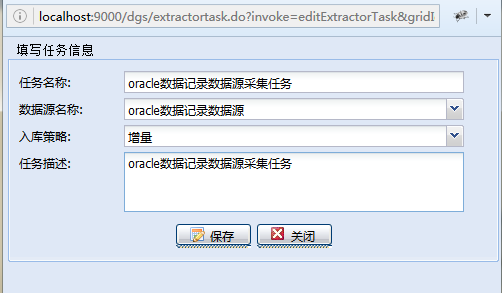
#### 2.2.5 任务配置

高级功能->元数据采集管理->任务配置



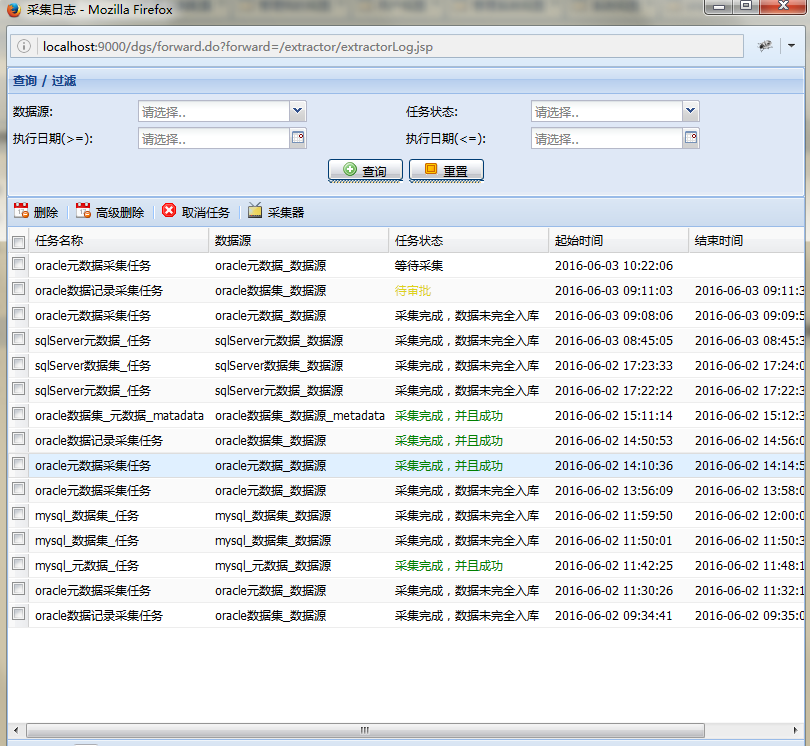
维护采集任务->增加

添加新的采集任务 选择创建的数据源



点击保存 保存采集任务

选择创建的采集任务 可以配置定时采集，也可以手动点击 立即运行按钮 来实现采集任务

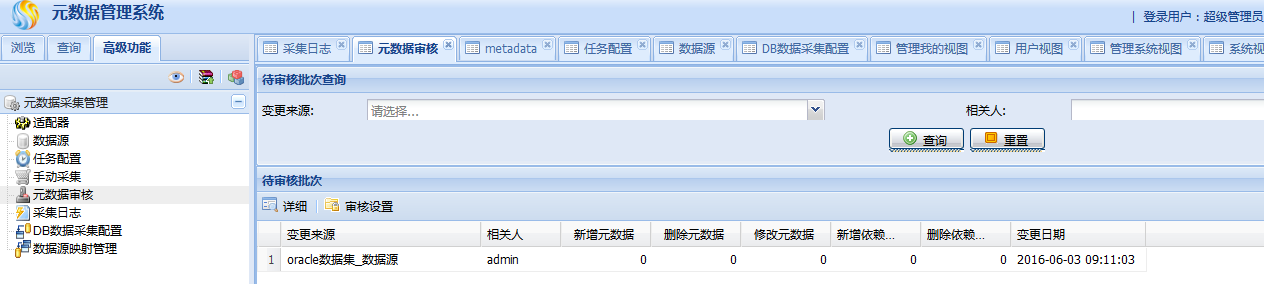


双击 采集任务 或者高级功能->元数据采集管理->采集日志

可以显示采集日志。（采集过程中可能会出现bug而中断采集任务）

#### 2.2.6 元数据审批

创建数据源的时候，如果选择 审核：是，那么采集成功后，数据不会立即进库，而是等待审核。



双击采集任务或者点击详情 来审核采集数据。

审核完毕后 在视图目录下就 可以显示采集的元数据。

## 3 DB数据字典采集适配器

### 3.1 简介

采集数据结构及相关对象。采集数据库的存储过程、表、视图。

### 3.2 步骤

#### 3.2.1 创建视图

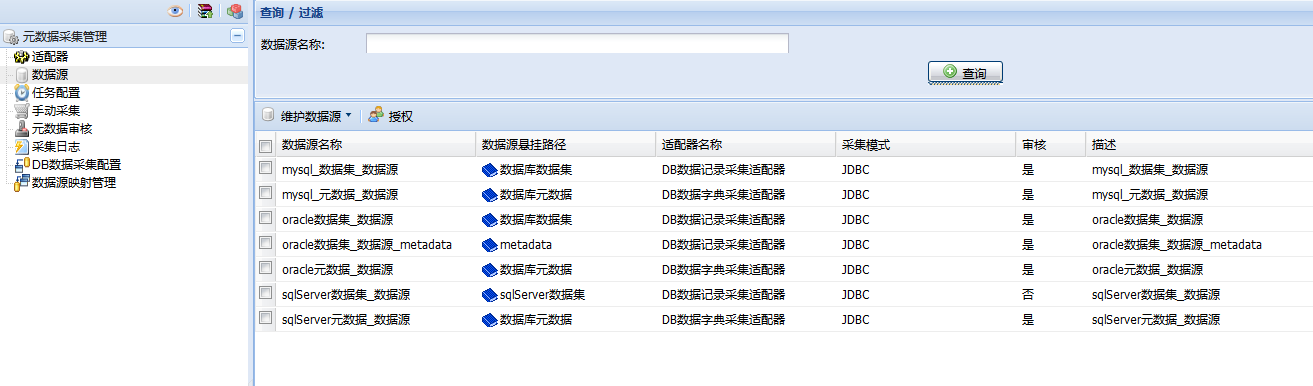
参照 [1视图创建](#_1_视图创建)

#### 3.2.2 创建元数据

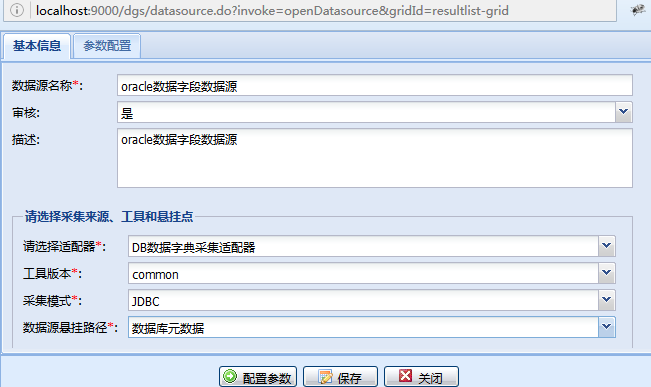
参照 [2.2.2 创建元数据](#_2.2.2_创建元数据)

#### 3.2.3 创建数据源

高级功能->元数据采集管理->数据源



点击维护数据源->增加



输入必要的数据源信息

请选择适配器 ：选择 DB数据字段采集适配器

数据源悬挂路径 ： 选择创建的元数据路径

点击配置参数页面 填写数据库信息 并且在schemas中 选择要采集的schema



#### 3.2.4 任务配置

参照 [2.2.4 任务配置](#_2.3.4_任务配置)

#### 3.2.5 元数据审批

参照 [2.2.5 元数据审批](#_2.3.5_元数据审批)

## 4 JDBC驱动采集适配器

### 4.1 简介

采集数据结构及相关对象。包括存储过程、表、视图。

存储过程: 存储过程的名称，

存储过程的参数列表

表 ： 表名称

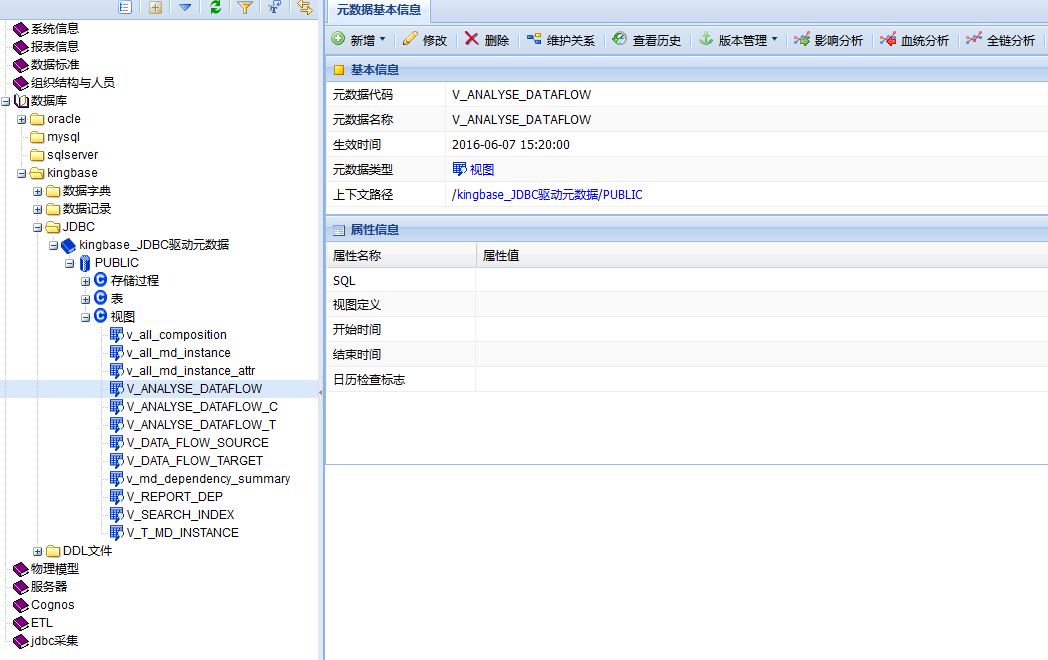
主键信息

外键信息

字段信息

视图 ： 视图名称

视图基本信息（没有显示有效的数据）



### 4.2 步骤

#### 4.2.1 创建视图

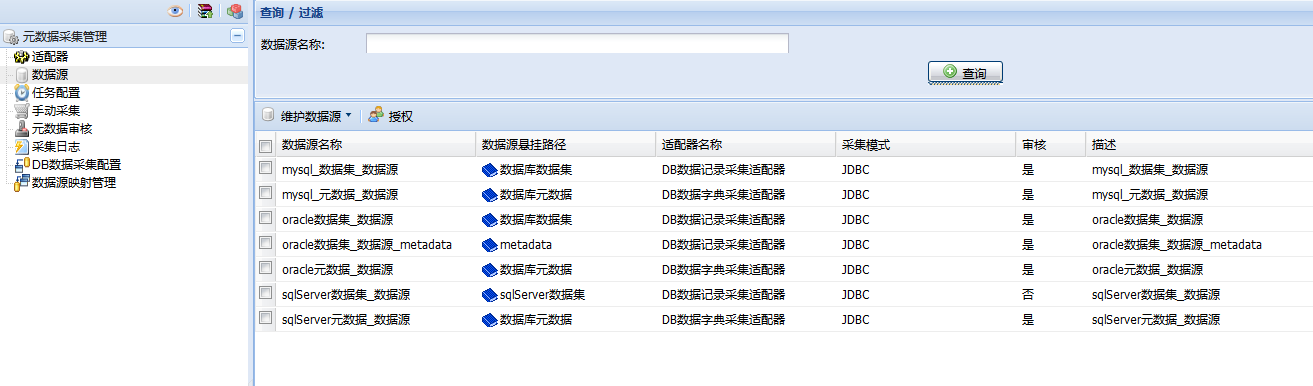
参照 [1视图创建](#_1_视图创建)

#### 4.2.2 创建元数据

参照 [2.2.2 创建元数据](#_2.2.2_创建元数据)

#### 4.2.3 创建数据源

高级功能->元数据采集管理->数据源



点击维护数据源->增加



输入必要的数据源信息

请选择适配器 ：选择 JDBC驱动采集适配器

数据源悬挂路径 ： 选择创建的元数据路径

点击配置参数页面 填写数据库信息 并且在schemas中 手动输入schema（不同于DB数据字段采集适配器 点击选择schema）



#### 4.2.4 任务配置

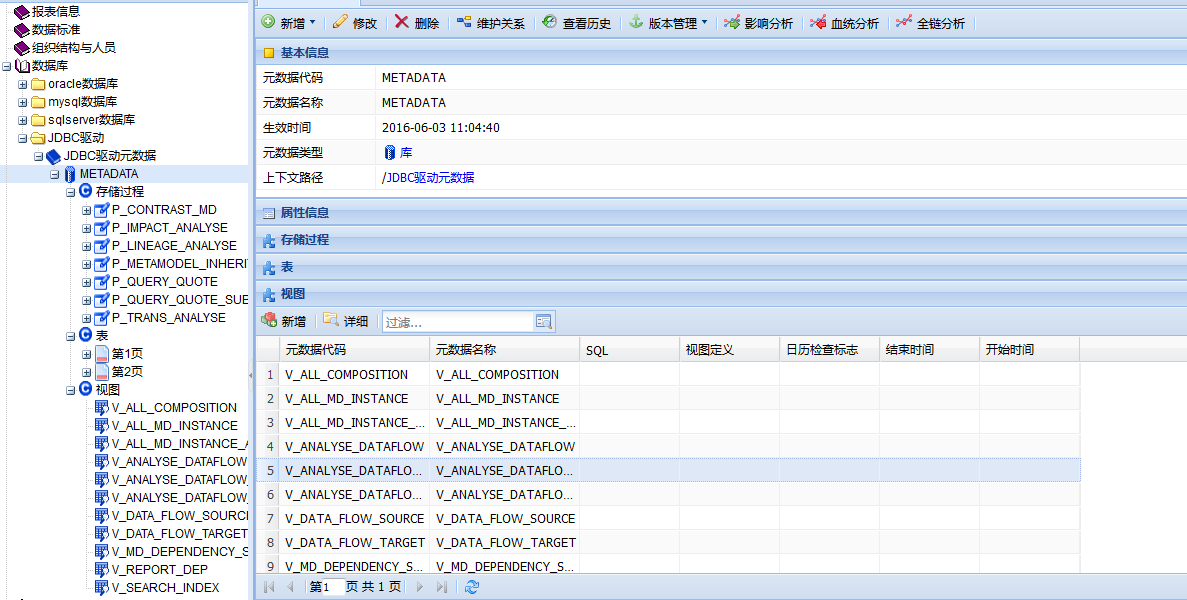
参照 [2.2.4 任务配置](#_2.3.4_任务配置)

#### 4.2.5 元数据审批

参照 [2.2.5 元数据审批](#_2.3.5_元数据审批)

#### 4.2.6 采集结果展示

使用jdbc驱动采集 会将方案下的 存储过程 表 视图都采集出来



## 5 Kettle采集适配器

### 5.1 简介

Kettle采集适配器为 xml 采集，仅支持 XML 文件，对 Kettle 工具导出的知识库文件进行解析，采集元数据包含 DB 数据源、子服务器、资源目录结构以及转换，不对作业进行采集。首先创建元数据的悬挂点， 悬挂点元模型选择“ 第三方工具元模型包/Kettle 包/知识库”（此悬挂点为采集到的文件夹内容在其下面显示）。

(kettle采集的时候，由于金仓不同数据库的默认模式都为public，导致采集结果混乱。)

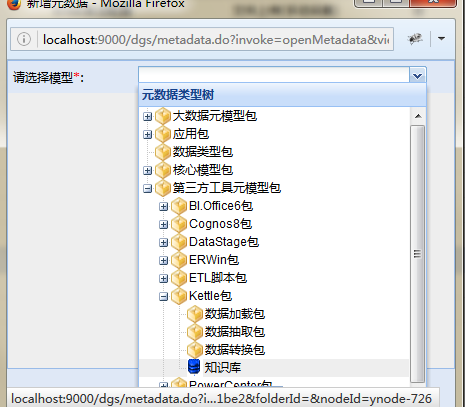
### 5.2 步骤

#### 5.2.1 创建视图

参照 [1视图创建](#_1_视图创建)

#### 5.2.2 创建元数据

在预览tab，找到创建的系统视图或者用户视图。右键点击创建元数据

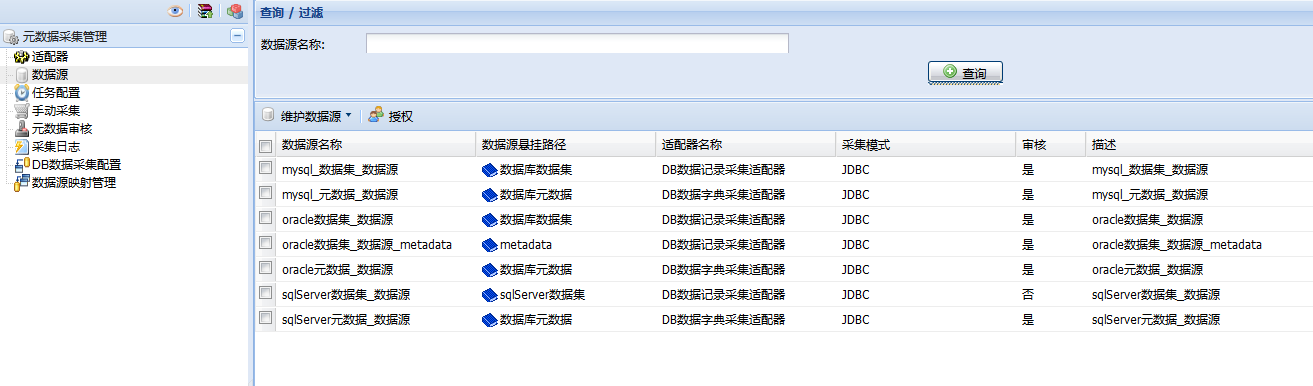
在弹出来的界面，选择第三方工具元模型包->Kettle包->知识库（

点击知识库 然后输入元数据信息来创建元数据。创建完毕后，可以刷新视图，显示挂载的元数据。

#### 5.2.3 创建数据源

创建数据源，适配器选择“ Kettle 采集适配器”，版本选择“ common”，采集方式选择“服务器文件(定时采集)”或者“文件上传(手动采集)”，数据源悬挂路径则选择刚才创建的“ 知识库” —“ Kettle 知识库”

高级功能->元数据采集管理->数据源。



点击维护数据源->增加



输入必要的数据源信息

采集模式 ：文件上传（手动采集）、服务器文件(定时采集)

请选择适配器 ：选择 Kettle采集适配器

#### 5.2.4 采集两个数据字典

采集两个数据字典。参照 [3 DB数据字典采集适配器](#_3_DB数据字典采集适配器)

#### 5.2.5 生成kettle配置文件

使用kettle ETL工具生成 [数据库的](#_5.2.4_采集两个数据字典)关系文件。

#### 5.2.6 手动采集

高级功能->元数据采集管理->手动采集



点击下一步 选择上传xml文件 然后点击上传 最后点击完成 即完成Kettle采集

#### 5.2.7元数据审批

参照 [2.2.5 元数据审批](#_2.3.5_元数据审批)

审批完成后，即可在视图上查看采集结果。

## 6 数据地图

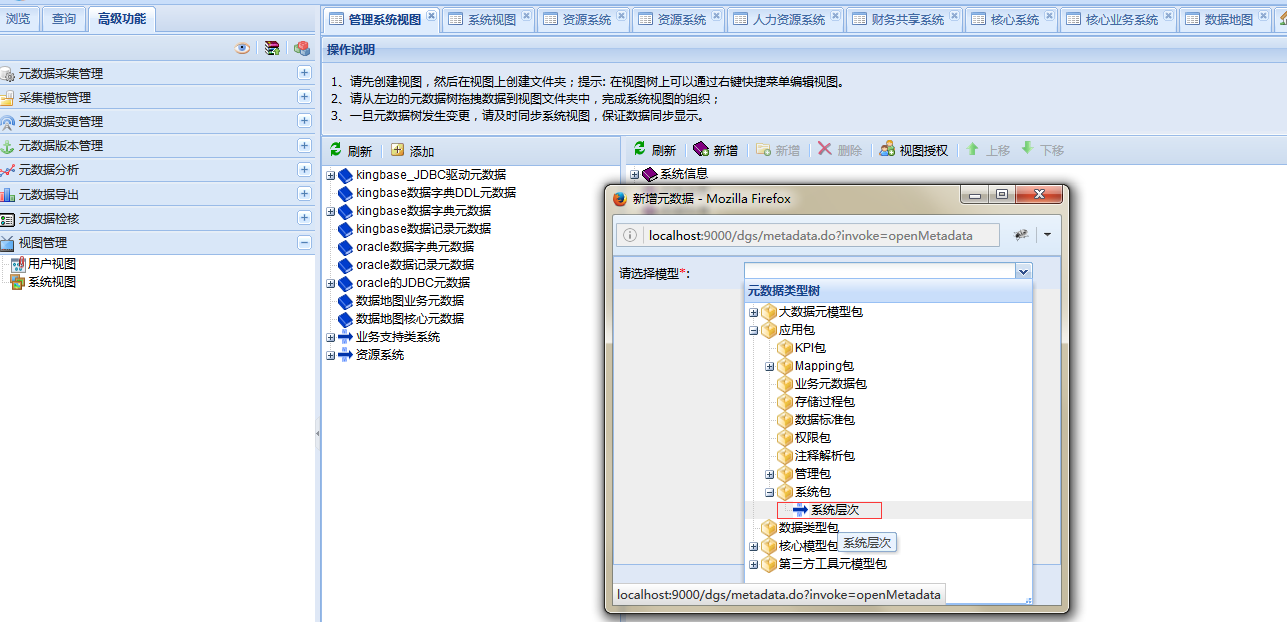
### 6.1 简介

数据地图为您从宏观角度展示系统以及系统之间数据流向关系，并支持从系统到数据库的下钻及相关信息的展示。 通过 WEB 方式编辑元数据或通过模板方式采集元数据都可进行元数据创建；对于创建好的初始数据地图，如果您不满意数据地图的布局，可通过数据地图位置调整及保存来调整数据地图；对于数据地图中单节点的信息可通过双击节点展示数据地图节点详细信息； 双击两个系统节点之间连线上的锚点 可下钻到这两个节点下的数据库之后。

### 6.2 手工创建数据地图

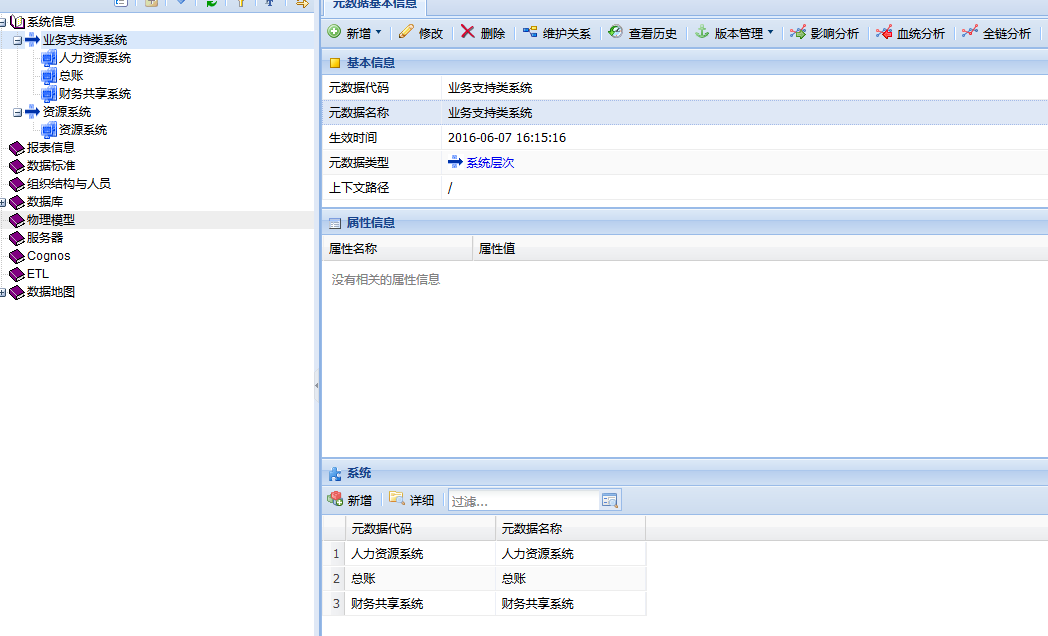
#### 6.2.1 添加系统层次元数据

在高级功能 Tab 中点击：系统视图->管理视图->添加

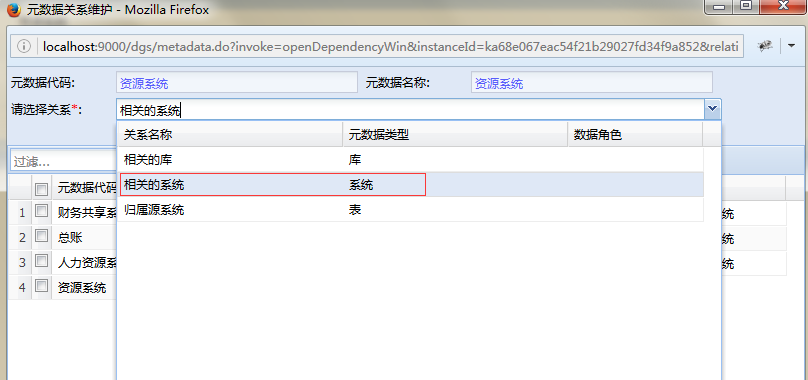


#### 6.2.2新增系统

添加多个系统，以便于维护系统之间的关系。

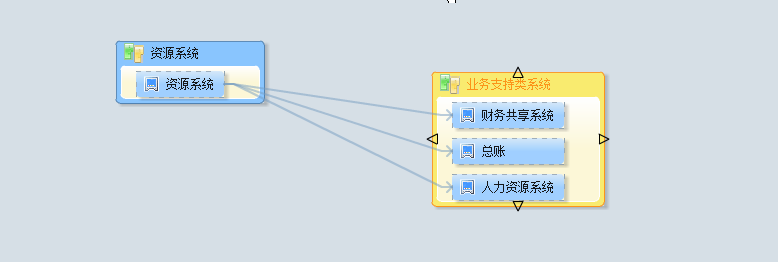


#### 6.2.3 维护关系



选择系统，添加建立关系

#### 6.2.4 数据地图显示效果

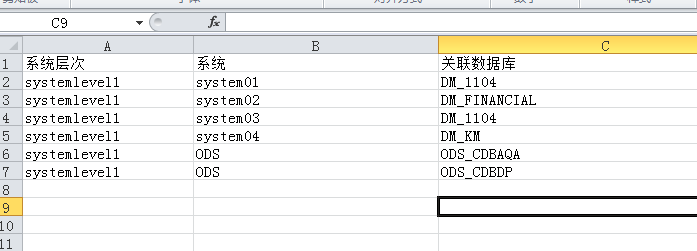


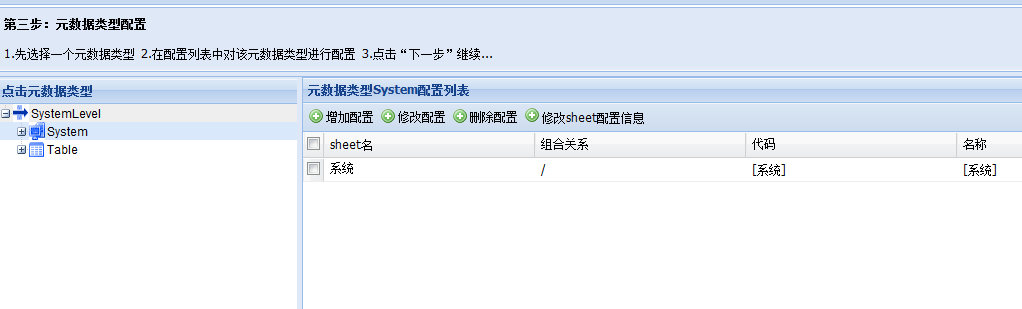
### 6.3 采集创建数据地图

#### 6.3.1新建悬挂点

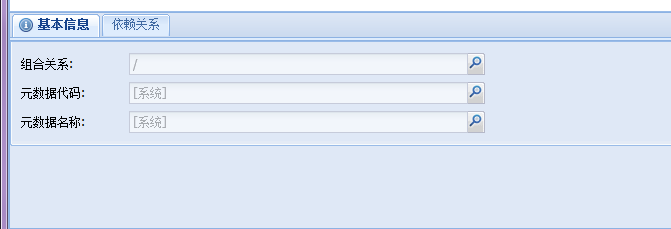
悬挂点的类型为系统层次

#### 6.3.2新建模板映射

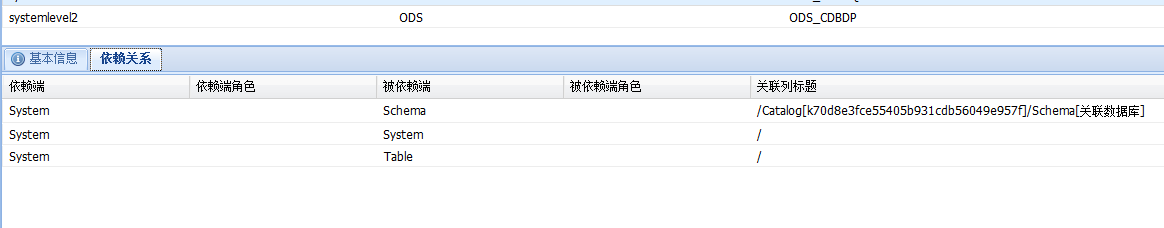




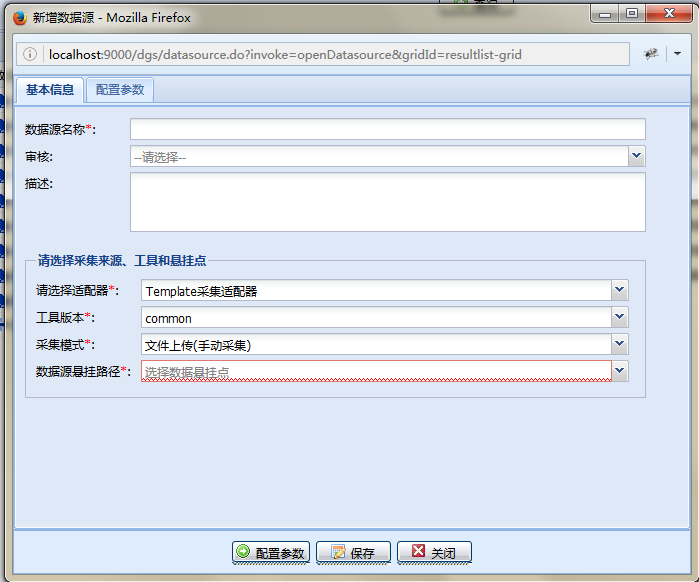
填写基本信息



建立依赖关系



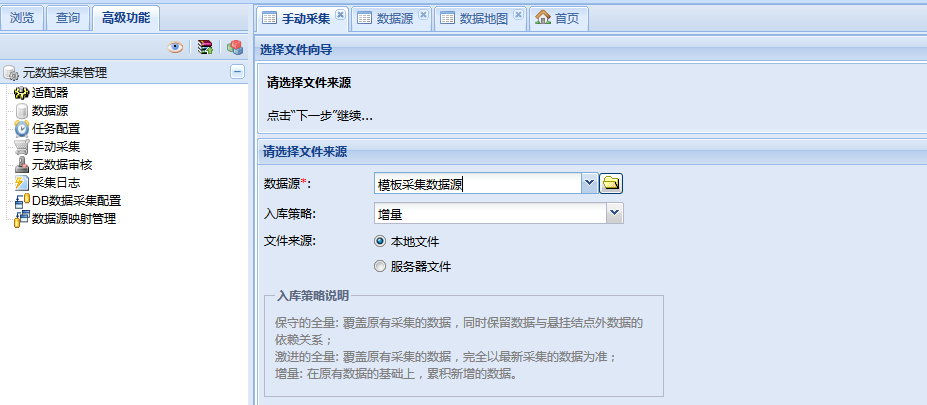
#### 6.3.3 创建数据源



建立模板数据源。选择Template采集适配器。并且将数据源悬挂在创建的系统层次。

#### 6.3.4 手动采集

高级功能->元数据采集管理->手动采集



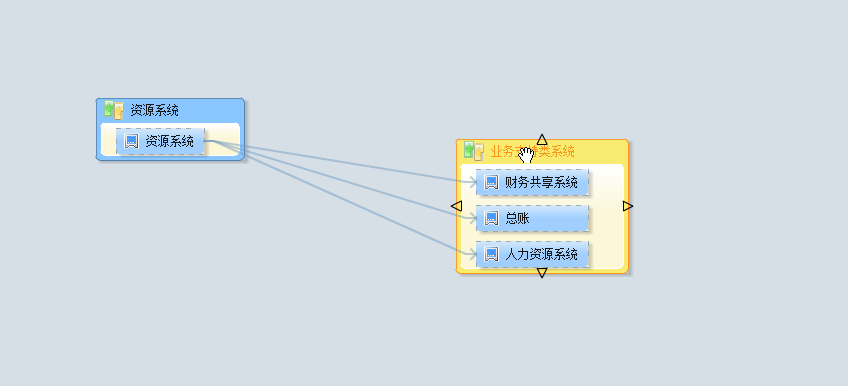
选择创建的模板采集数据源。

然后点击下一步->增加->上传->下一步->完成 即开始采集

采集完成后，点击数据地图。即可展示采集的数据地图。

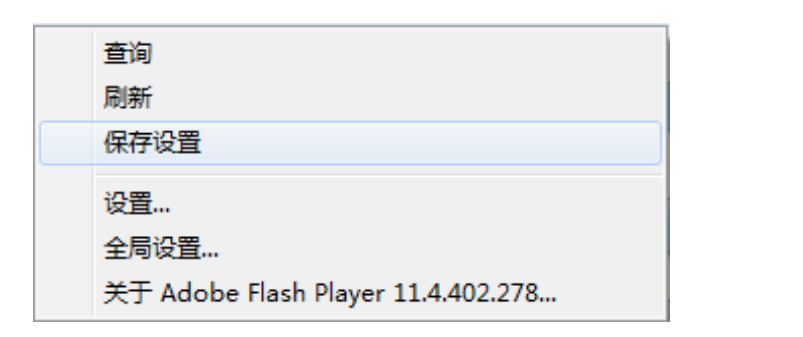
### 6.4 数据地图设置保存

#### 6.4.1 选中图中需要调整位置的节点拖动



#### 6.4.2 保存位置信息

右键面板，弹出右键菜单框。点击保存设置 即可以保存位置信息。



如果右键数据地图空白处，菜单中没有“保存设置”项，则需要申请管理员给自己的角色赋予数据地图权限。

### 6.5 数据地图的下钻

为进一步查看两个系统之间所属数据库之间关系，您可通过双击两个系统节点之间连线上的锚点下钻到这两个节点下的数据库之间关系。

1 鼠标悬到两个系统之间的连线上，系统会自动显示锚点

2双击锚点，产品会展示两个系统所属数据库之间的关系

3为进一步查看数据库节点的详细信息，您可通过双击该节点，产品会打开单独的 Tab 页面展示该节点详细信息

4鼠标悬到两个库之间的连线上，系统会自动显示锚点

5双击锚点会打开单独的 Tab 页面，显示这两个库之间表级间的关系以及他们之间的 ETL 作业

6选择‘表间关系’的源 schema 分类，目标 schema 分类，源关键字，目标关键字，并点击‘查询’按钮，可以对记录进行过滤

7点击‘表间关系’记录上的链接，可以单独打开该元数据的详细信息 Tab

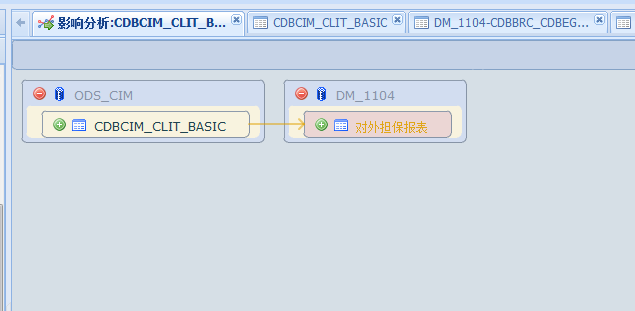
8点击‘表间关系’记录上的‘下钻’图片上的连接，可以下钻到该源表和目标表之间的字段级关系。

9点击‘字段间关系’记录上的链接，可以单独打开该元数据的详细信息 Tab。

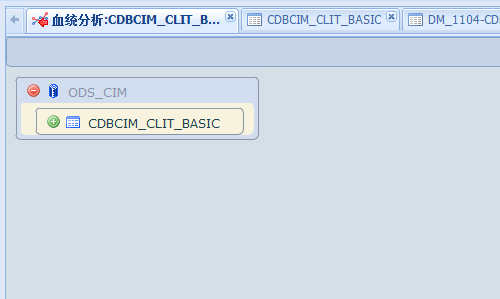
## 7 元数据分析

### 7.1影响分析

影响分析采用图形方式展示了以某个元数据为起始节点，其后与其有关系的所有元数据，反应数据的流向与加工过程，使用影响分析可分析数据流向和数据转换中错误的定位。



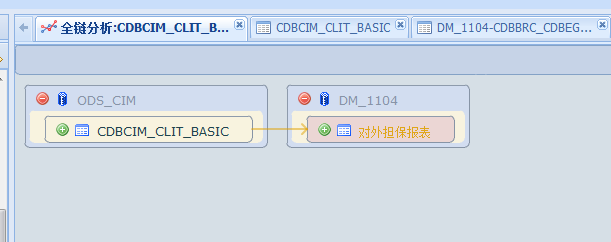
### 7.2血统分析

血统分析采用图形方式展示了以某个元数据为终止节点，其前与其有关系的所有元数据，反应数据的来源与加工过程，使用血统分析可分析数据来源和数据质量问题的定位。

### 7.3全链分析

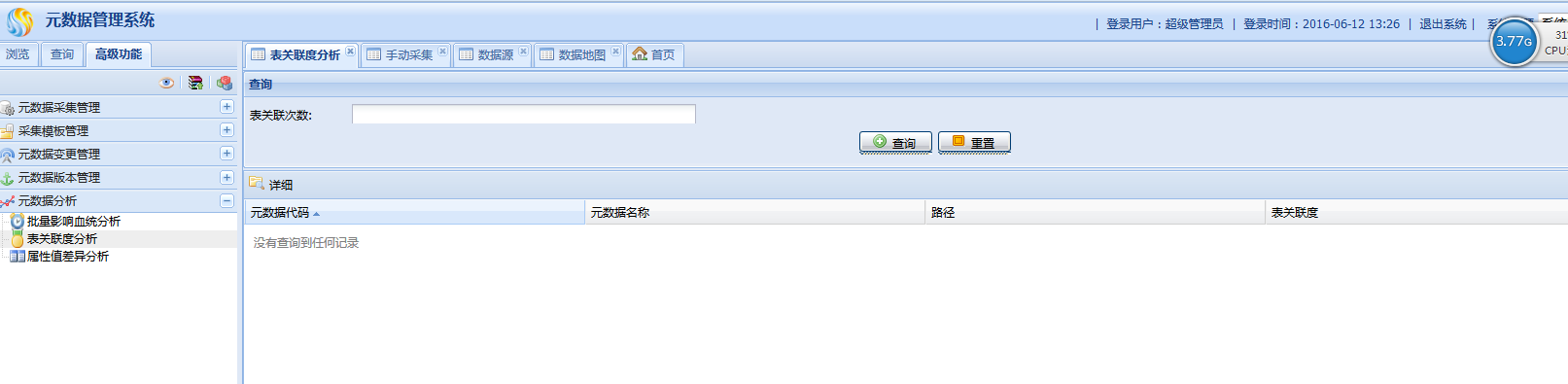
全链分析采用图形方式展示了以某个元数据为目标节点，其前后与其有关系的所有元数据，

反应数据的来源与加工过程，使用全链分析可分析数据来源和数据质量问题的定位。



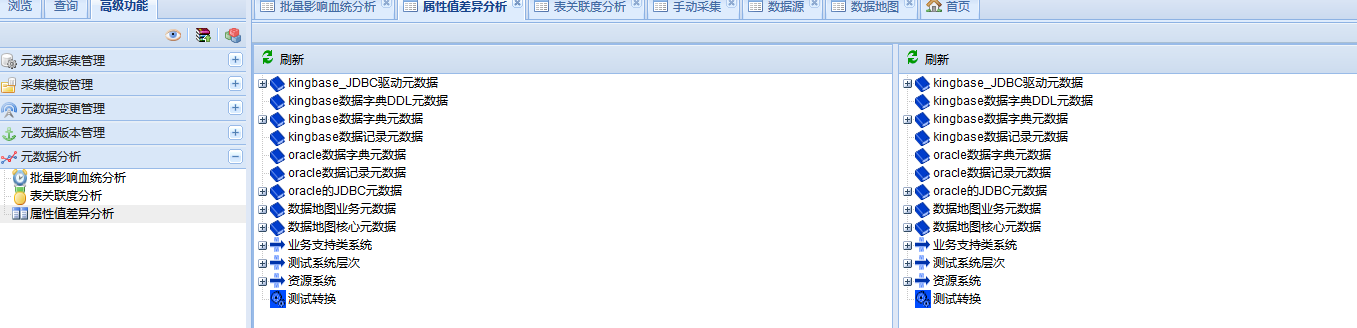
### 7.4 表关联度分析

表关联程度分析描述了数据库表元数据与其他元数据的关系出现次数情况，如：表与 ETL 程序、表与 OLAP、表与指标等关系次数情况，产品通过分析数据库表关系次数的多少，来体现表关联程度的高低。



### 7.5 属性值差异分析

产品提供同类元数据进行元数据属性值差异分析功能，您可通过属性值差异分析功能直观获知两个同类元数据之间的相同和不同之处。



## 8 元数据介绍

### 8.1 元数据概念

关于数据的数据，描述数据及其环境数据。包括业务元数据、技术元数据、管理元数据。

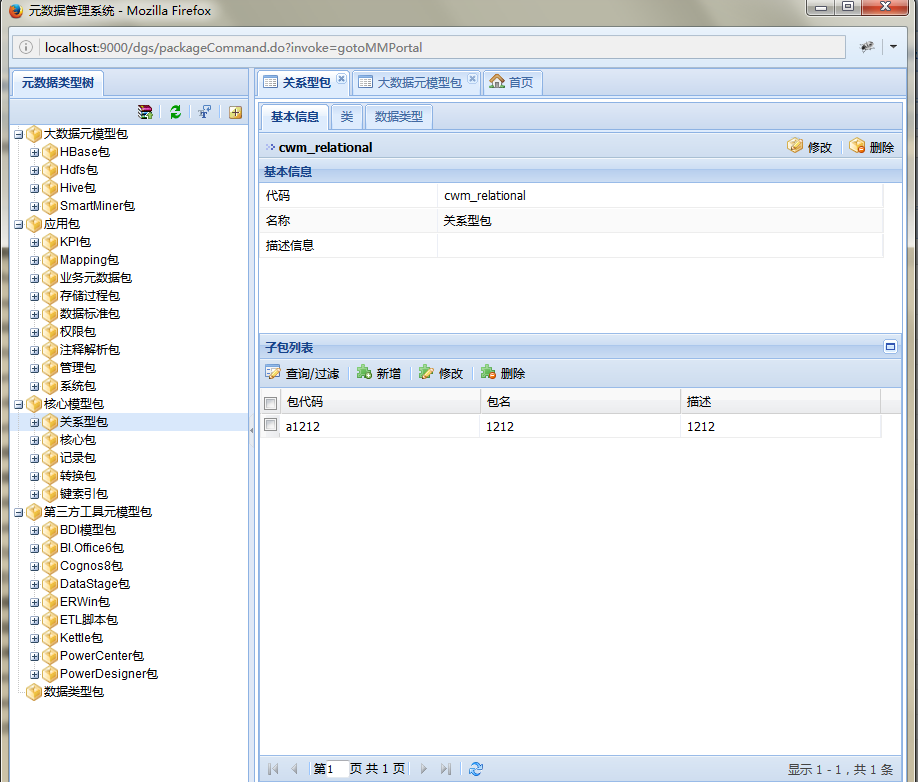
### 8.2 适配器

元数据对采集是通过统一接口来导入数据的，为了能够导入不同数据源或数据载体的元数据，需要针对不同的数据源或数据载体提供数据适配，来屏蔽元数据采集存储端对不同数据源或数据载体的变化。



### 8.3 元模型

用于描述元数据的结构和关系的数据模型。



### 8.4 视图

MetaCube中视图，是一个前台展示接口。用户将元数据挂载到视图中，方便用户查看采集的数据。而实际的数据则是保存在元数据中的。

### 8.5 MedaCube简介

MetaCube 是一款基于 WEB 方式的元数据管理工具，采用这个工具能够整合游离于企业各环节的元数据资产，便于用户浏览及分析元数据。

MetaCube 有助于帮助用户了解和管理信息和加工处理过程的来源，也有助于用户理解信

息与加工过程之间的关系以及它们如何被使用。（尤其在于元数据版本变更）

MetaCube 支持采集多种元数据来源，包括：关系数据库、建模工具、数据集成工具、 BI 工具以及客户化元数据等多种元数据来源， MetaCube 对采集的元数据进行统一存储到 MetaCube 知识库并集中管理，为上层元数据应用提供服务。

MetaCube 采用元数据树方式来统一展示 MetaCube 知识库的元数据，层级结构清晰，便于用户直观了解元数据组织结构。