

一.Insight大数据智能分析平台概述

二.Insight大数据智能分析平台的架构

三.Insight大数据智能分析平台的使用

四.Insight大数据智能分析平台的特点

- Insight是一个以流程为中心，面向解决方案（Solution），可扩展的商业智能平台。Insight平台会帮助使用者洞悉商业等领域发展的问题和趋势，让使用者可以做出最正确的决策。
- Insight是一个基于java的商业智能(Business Intelligence)平台。它的出现，使得一系列面向商业智能的独立功能可以集成在一起，构成了一个可以提供完整商业智能解决方案的具有强大功能的平台。

- Insight 平台是一个综合平台，您可以通过该平台对数据进行访问、集成、操作、可视化分析等。无论数据存储在线文件、关系数据库、Hadoop 集群、NoSQL 数据库、分析数据库、社交媒体流、操作型存储中还是存储在云中，Insight 平台都会帮助您发现、分析并可视化数据，即使没有编码经验，使用Insight平台也可以找到适合您的解决方案。
- Insight平台可以分为两部分：业务分析 (Business Analysis) 平台和数据集成 (Data Integration) 工具集

- Insight 业务分析 (Business Analysis) 平台主要包含三个部分：Insight用户控制台，设计工具以及大数据。其中最为重要的以及我们经常使用的是 Insight用户控制台。
- Insight用户控制台是一个基于 Web 的设计环境，您可以在其中分析数据、创建交互式报表以及构建集成的仪表盘，为了便于您的组织和互联网上的其他人共享解决方案，Insight用户控制台的开发者把她设计成了一个开放的平台，拥有一些基本权限的用户就可以实现和其他人的资源共享。

- Insight Data Integration 提供提取、转换和加载 (Extract, Transform, and Load , ETL) 功能，以便使用统一且一致的格式来捕获、清理和存储数据，该格式与最终用户和物联网技术相关。
- Insight Data Integration 的常见用途包括：
 - 不同数据库和应用程序之间的数据迁移
 - 将巨大的数据集加载到数据库中，充分利用云、集群和大规模并行处理环境
 - 使用从非常简单到非常复杂的转换步骤进行数据清理
 - 数据集成，包括能够利用实时 ETL 作为 Insight Reporting 的数据源

- Insight 提供了一个支持整个大数据分析流程的完整的大数据分析解决方案。从大数据聚合到交互式可视化、分析和预测，Insight 允许您收集埋藏在大数据存储中的有意义的模式。比如对于企业来说，通过分析大数据集，您可以确定新的收入来源、开发忠实且有利可图的客户关系，以及更高效、更经济地运营您的组织。

- Insight平台的优势：
- 集成化：Insight平台将众多不同的BI功能集成到一个统一的框架中来，使之可以相互协作。以往的BI产品，往往只专注于BI的某一特定领域，如Report主要关注报表的生成，Weka主要进行数据的挖掘等等。然而一个完整的BI应用往往需要这些功能可以快速可靠地集成在一起。

- Insight通过引入“Action”的概念，提供了一个让多种BI产品协作的机制。
“Action”是Insight平台提供的最基本的操作单元，它类似于一种编程语言的基本语句。所有完成具体功能的BI产品作为“插件”集成到Insight平台中，每种插件为Insight平台提供一种或几种“Action”，每个Action有自己的输入和输出，多个Action连接起来就构成了Action序列，完成一个较复杂的功能。Insight平台负责在各个Action之间传递参数，这样多种不同的BI产品便能够协同工作了。

- 面向解决方案（Solution）：Solution本身实质上是一系列Action序列的集合，这些序列如何显示，如何被调用，功能如何实现完全由Insight平台来管理，这使得Solution的开发者，也就是Insight的使用者，可以将开发工作集中于具体的BI业务逻辑的开发上。
- Insight 通过专注于提供性能的功能，提高了甚至是最大的大数据存储的思维分析速度。

- 即时访问：Insight 提供了可视化工具，使您可以轻松定义对于交互式分析非常重要的数据集。这些数据集和相关的分析可以很容易地与其他人共享，并且当出现新的业务问题时，可以为交互式分析定义新的数据视图。
- 高性能平台：Insight 构建在一个现代化、轻量级、高性能的平台上。该平台充分利用 64 位多核处理器和大内存空间，有效利用当代硬件的力量。
- 超大规模的内存缓存：Insight 利用外部数据网格技术将大量数据加载到内存中，以便立即进行思维速度分析，这是独一无二的。
- 联合数据集成：可以从多个来源（包括大数据和传统数据存储）提取数据，将这些数据集集成在一起，然后直接流入报表，而无需企业数据仓库或数据集市。

一.Insight大数据智能分析平台概述

二.Insight大数据智能分析平台的功能架构

三.Insight大数据智能分析平台的使用

四.Insight大数据智能分析平台的特点

- Insight用户控制台，是一个通过Web实现用户控制数据源完成数据分析和管理等操作的平台。
- 通过以下步骤第一次登录Insight User Console (IUC) 非常容易。
- 1.启动任何 Web 浏览器并输入 Insight 服务器的 URL。您的 IT 管理员可以告诉您正确的 URL。该页面加载一个带有 登录 部分的介绍屏幕。
- 2.输入您的用户名和密码并单击 登录 即可。

Insight

大数据智能分析平台

用户控制台

用户名:

密码:

登录

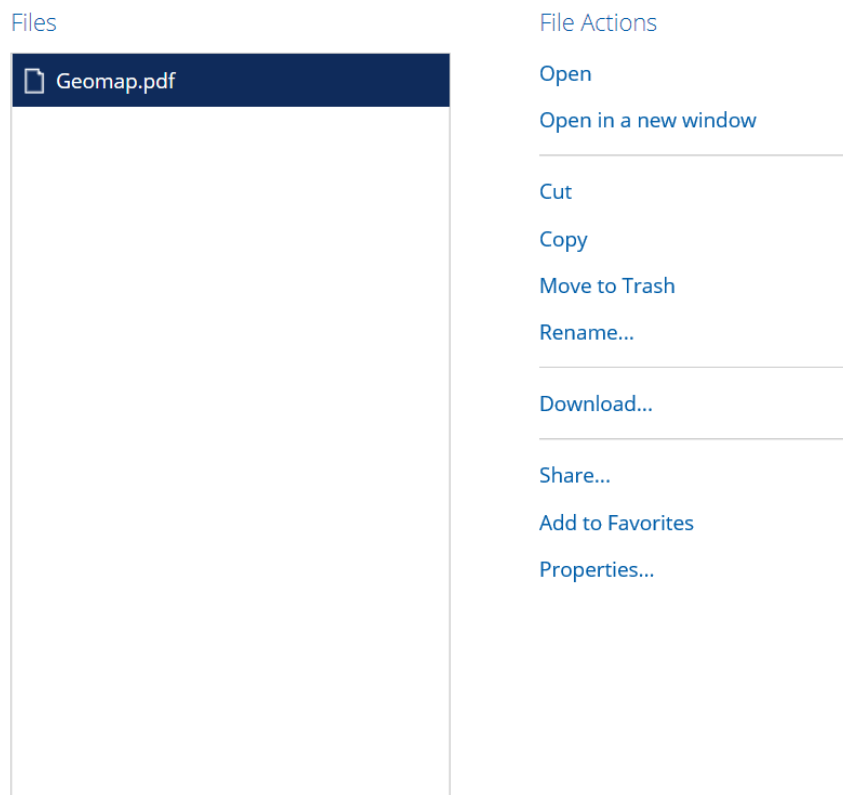
Insight User Console界面使用透视图来扩展在任何给定时间处理不同任务的能力。



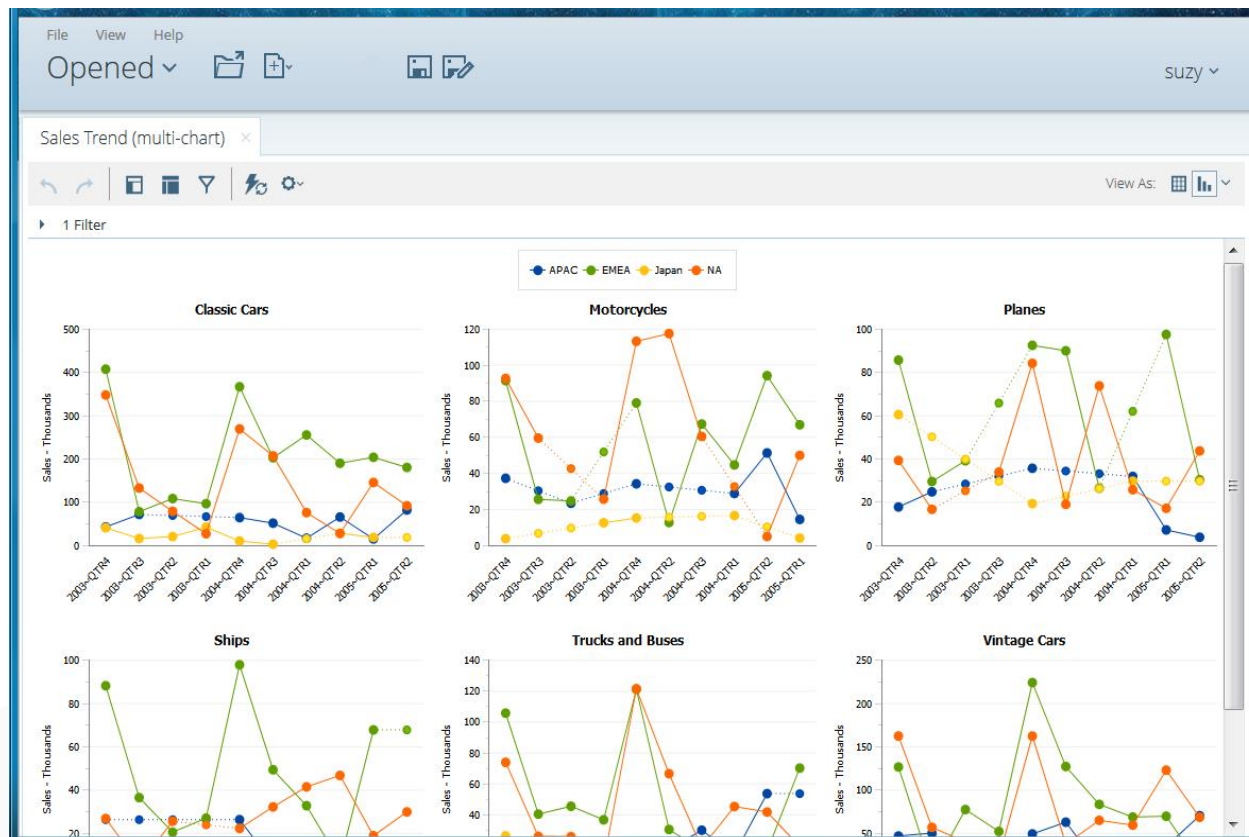
浏览页面帮助您组织文件以方便您查找和使用。

The screenshot displays the file management interface of the Insight User Console. At the top, a dark blue navigation bar contains the text '文件 视图 工具' (File View Tools) on the left, '浏览文件' (Browse Files) with a dropdown arrow in the center, and 'admin' with a dropdown arrow on the right. Below this, the interface is divided into three main sections: 'Folders', 'Files', and 'Folder Actions'. The 'Folders' section on the left shows a tree view with 'Home' expanded to reveal sub-folders: 'admin' (highlighted), 'pat', 'suzy', 'tiffany', and 'zktr'. Below these are 'Public' and 'Trash'. The 'Files' section in the center shows a single file named 'Geomap.pdf'. The 'Folder Actions' section on the right lists several options: 'New Folder...', 'Move to Trash', 'Rename...' (with an input field), 'Paste' (with an input field), 'Upload...', 'Download...' (with an input field), and 'Properties...'. A refresh icon is located between the 'Folders' and 'Files' sections.

默认是文件夹操作，选中文件后，可以对文件进行编辑，下载，收藏，查看属性等操作，如图所示。



- 通过Insight User Console您可以创建出各式各样精美的Report。并且您可以根据您的需要选择不同类型的Report。



- Analyzer Reports : 是一个直观的分析可视化工具，可过滤和深入到 Insight Analysis数据源中包含的业务信息。
- Interactive Reports : 是一个基于 Web 的设计界面，可用于创建简单和按需操作的报表，而不依赖于 IT 或报表开发人员。
- Dashboard : 允许您在经过少量培训或者不经过任何培训的情况下创建仪表盘。仪表盘是在一个屏幕内汇集的几个不同的报表。

- 管理透视图：在 Insight User Console 中，分配了具有管理安全性权限的角色的已登录用户可以使用 管理 来执行系统配置和维护任务。如果您看到 管理 在 主页 页面上的左侧下拉菜单中，您可以单击它以显示特定于 Insight Server 的管理的菜单项。如果您没有管理权限，则 管理 不会出现在控制台中。



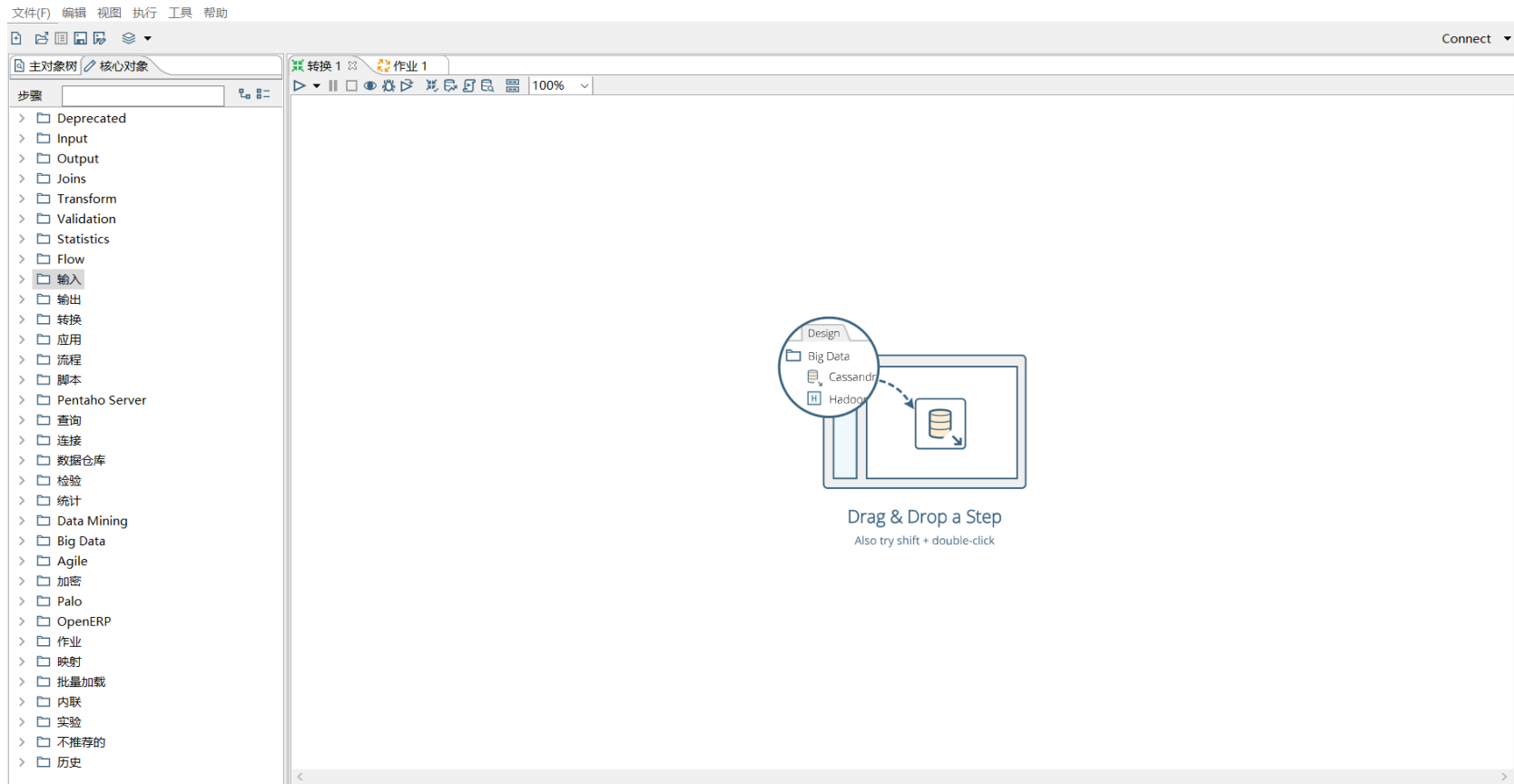
The screenshot displays the 'Users & Roles' management interface. At the top, there is a navigation bar with '文件', '视图', and '工具' menus, and a '管理' dropdown menu. The user 'admin' is logged in. On the left, a sidebar lists 'Users & Roles', 'Authentication', 'Mail Server', and 'Settings'. The main content area is titled 'Users / Roles' and has three tabs: 'Manage Users', 'Manage Roles', and 'System Roles'. The 'Manage Users' tab is active, showing a list of users: 'admin', 'pat', 'suzy', 'tiffany', and 'zktr'. The 'admin' user is selected. To the right of the list is a 'Password:' field with a masked password and an 'Edit...' button. Below this is a 'Roles' section with two columns: 'Available:' containing 'Business Analyst', 'Power User', and 'Report Author'; and 'Selected:' containing 'Administrator'. Navigation arrows are present between the columns.

- Schedules
- 所有活动的计划报表都会显示在计划列表中，您可以通过单击 Home 下拉菜单，然后单击 IUC 页面左上角的 Schedules 来查看。
- 计划列表显示哪些报表已计划运行、计划的重复模式、何时是最后一次运行、何时设置为再次运行以及计划的当前状态。您可以使用工具栏右侧计划列表上方的控件来编辑和维护每个计划。

The screenshot displays the 'Schedules' section of the Insight User Console. It features a table titled 'My Schedules' with the following data:

Schedule Name	Repeats	Source File	Output Location	Last Run	Next Run	Status
Inventory List (report)	The first Sunday of every month at 17:00:00	/public/Steel Wheels/Inventory List (report)	/home/suzy	-	2015 Feb 1 17:00:00	Normal
Leading Product Lines (pivot table)	Every Friday at 23:00:00	/public/Steel Wheels/Leading Product Lines (pivot table)	/home/suzy	-	2015 Jan 16 23:00:00	Normal
European Sales (geo map)	Every Sunday at 23:00:00	/public/Steel Wheels/European Sales (geo map)	/home/suzy	-	2015 Jan 18 23:00:00	Normal

- Insight Data Integration是一个用来数据整合的工具集



- Insight Data Integration 透视图允许您创建两个基本的文件类型：转换和作业。转换用于描述 ETL 的数据流，例如从某个来源读取，转换数据并将其加载到某个目标位置中。作业用于协调 ETL 活动，例如定义有关应按照什么顺序运行转换的流和依赖关系，或者通过检查“我的源文件是否可用？(Is my source file available?)”或“我的数据库中是否存在表？(Does a table exist in my database?)”等条件来为执行做准备

功能：INSIGHT DATA INTEGRATION能够清洗数据，把复杂的，无规律性的数据通过一定的流程转换成整洁，干净的数据源。即实现完整的ETL过程。

INSIGHT DATA INTEGRATION的部分步骤如下：

输入

- Access 输入
- CSV文件输入
- Cube 文件输入
- Excel输入
- GZIP CSV Input
- Google Analytics
- HL7 Input
- LDAP 输入
- LDIF 输入
- Mondrian 输入
- OLAP 输入
- RSS 输入
- SAS 输入

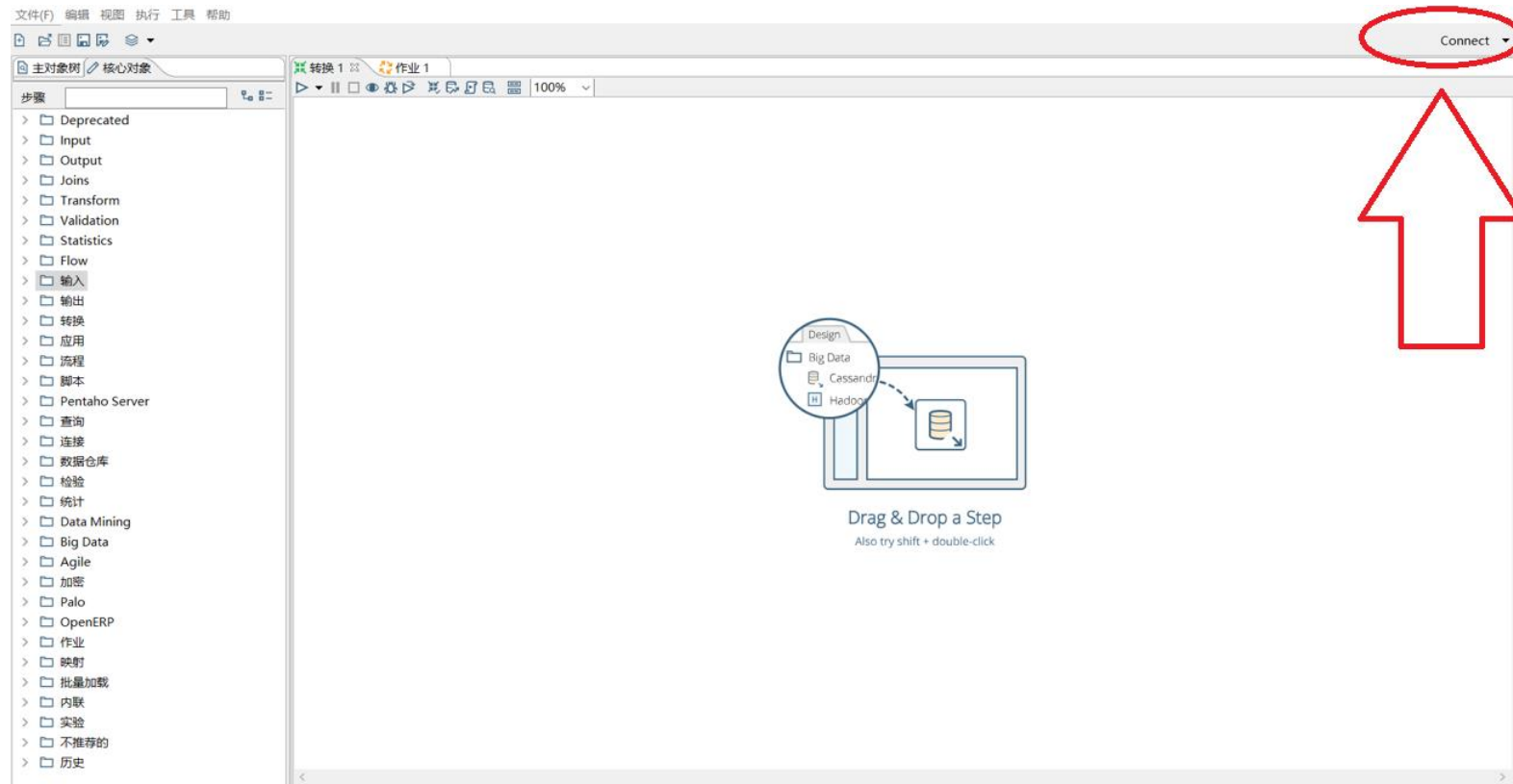
转换

- Closure Generator
- Concat Fields
- Get ID from slave server
- 值映射
- 列拆分为多行
- 列转行
- 剪切字符串
- 去除重复记录
- 唯一行 (哈希值)
- 增加常量
- 增加序列
- 增加校验列
- 字段选择

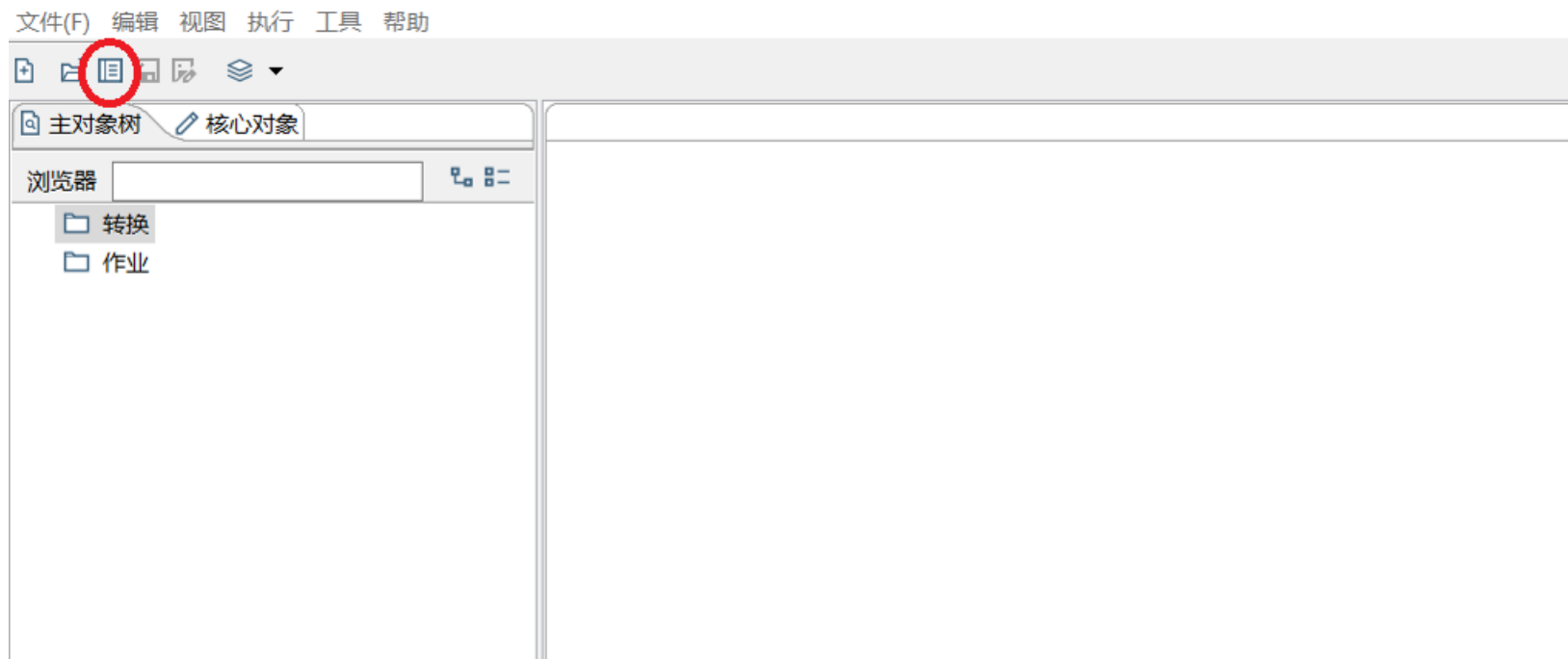
输出

- Access 输出
- Cube输出
- Excel输出
- LDAP 输出
- Microsoft Excel 输出
- Pentaho 报表输出
- RSS 输出
- S3 File Output
- SQL 文件输出
- Salesforce Delete
- Salesforce Insert
- Salesforce Update
- Salesforce Upsert

- 通过connect，INSIGHT DATA INTEGRATION可以连接到repository进行转换文件和作业文件的存储和提取，以使得协同作业，共享资源变得简单易行。




- 通过Explore repository图标可点击进入探索资源库界面，在这个界面中可以直接操作存储在IUC中的文件



文件 视图 工具


浏览文件 ▾

Folders 

- Home
 - admin
 - dashboard
 - job
 - report
 - transformation
 - pat
 - suzy
 - tiffany
 - zktr
- Public
- Trash

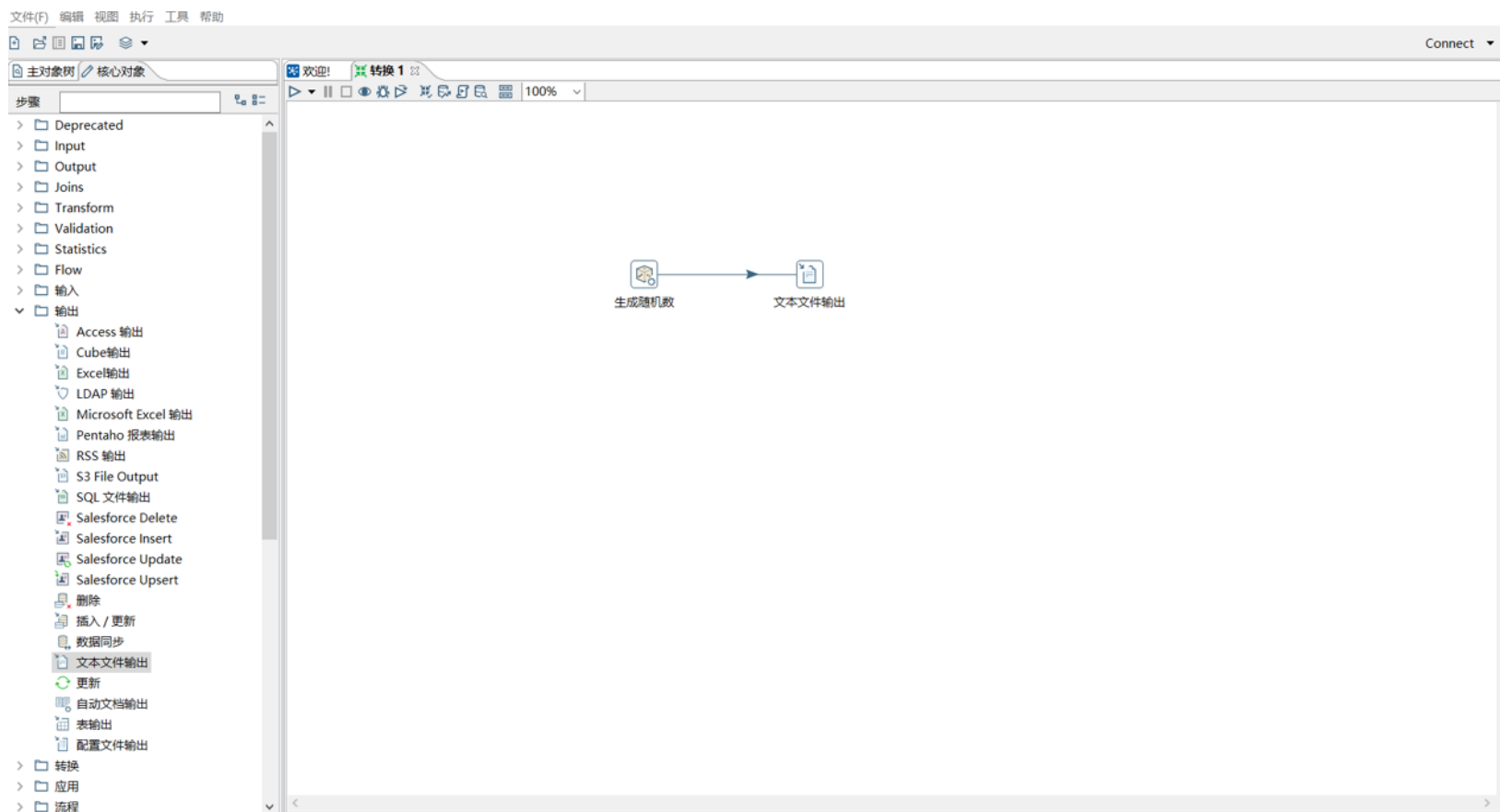
探索资源库 [admin]

浏览 连接 Hadoop Clusters 安全 子服务器 Partition

目录: 

- home
 - admin
 - dashboard
 - job
 - report
 - transformation
 - pat
 - suzy
 - tiffany
 - zktr
- public
- Trash

- 目标：生成10个随机数，并输出到一个文本文件中。可建立如下转换。



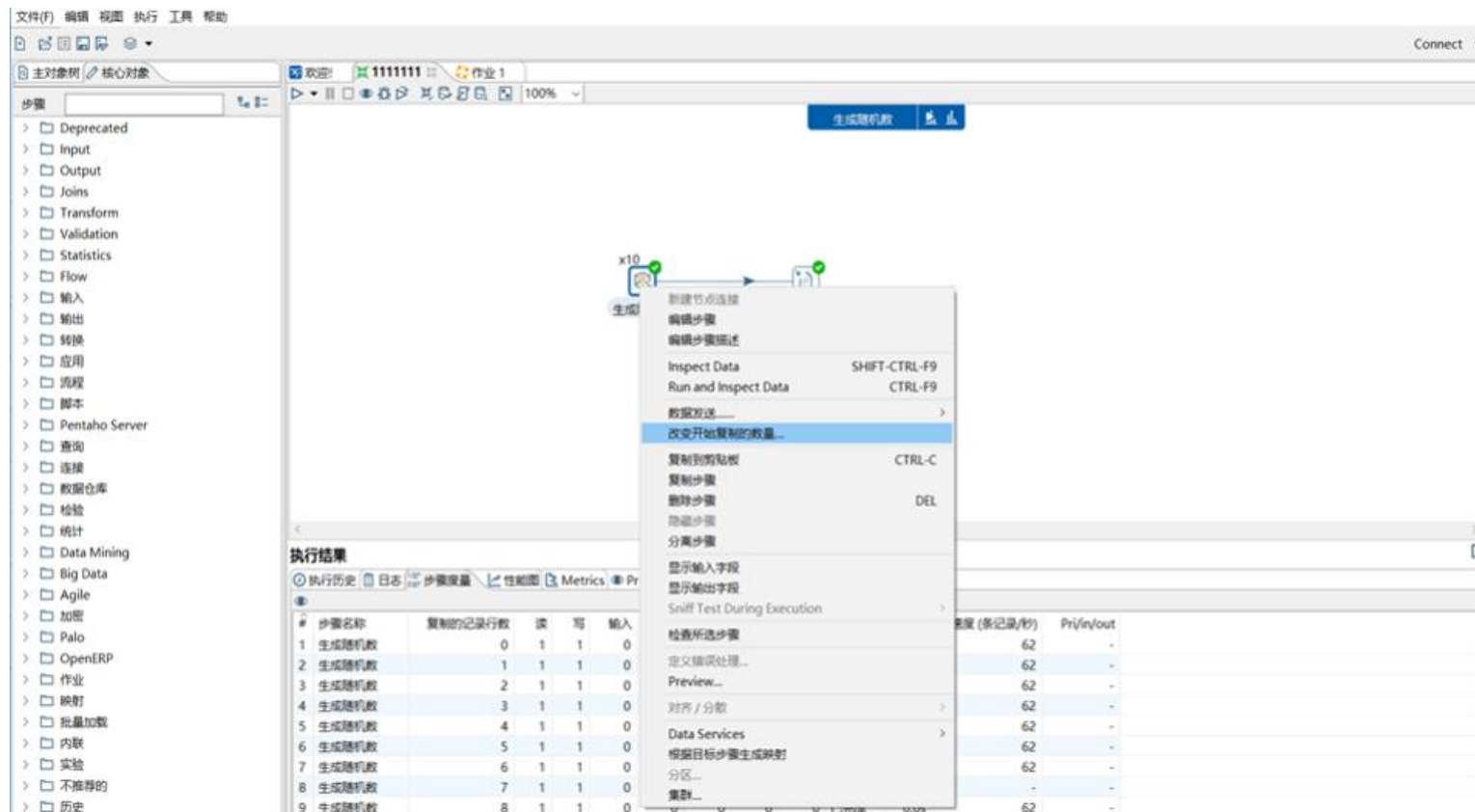
生成随机值

步骤名称

字段

#	名称	类型
1	随机数	随机数字

- 点击右键，选择改变开始复制的数量，选择想要生成的随机数的个数



文本文件输出

步骤名称 文本文件输出

文件 内容 字段

文件名称 C:\Users\dell\Desktop\输出文件 浏览(B)...

结果输送至命令行或脚本

输出传递到servlet

创建父目录

启动时不创建文件

从字段中获取文件名?

文件名字段

扩展名 txt

文件名里包含步骤数?

文件名里包含数据分区号?

文件名里包含日期?

文件名里包含时间?

指定日期时间格式

日期时间格式

显示文件名...

结果中添加文件名

? Help 确定(O) 取消(C)

- 转换会按照顺序执行，如果正在执行的步骤没有问题，会在右上角出现一个绿色的对勾，当所有的步骤都有绿色的对勾出现时，表示运行通过。

文件(F) 编辑 视图 执行 工具 帮助

Connect

主对象树 核心对象

步骤

- Deprecated
- Input
- Output
- Joins
- Transform
- Validation
- Statistics
- Flow
- 输入
- 输出
- 转换
- 应用
- 流程
- 脚本
- Pentaho Server
- 查询
- 连接
- 数据仓库
- 检验
- 统计
- Data Mining
- Big Data
- Agile
- 加密
- Palo
- OpenERP
- 作业
- 映射
- 批量加载
- 内联
- 实验
- 不推荐的
- 历史

生成随机数 x10 → 文本文件输出

执行结果

执行历史 日志 步骤度量 性能图 Metrics Preview data

#	步骤名称	复制的记录行数	读	写	输入	输出	更新	拒绝	错误	激活	时间	速度 (条记录/秒)	Pri/in/out
1	生成随机数	0	1	1	0	0	0	0	0	已完成	0.0s	-	-
2	生成随机数	1	1	1	0	0	0	0	0	已完成	0.0s	-	-
3	生成随机数	2	1	1	0	0	0	0	0	已完成	0.0s	-	-
4	生成随机数	3	1	1	0	0	0	0	0	已完成	0.0s	-	-
5	生成随机数	4	1	1	0	0	0	0	0	已完成	0.0s	-	-
6	生成随机数	5	1	1	0	0	0	0	0	已完成	0.0s	-	-
7	生成随机数	6	1	1	0	0	0	0	0	已完成	0.0s	-	-
8	生成随机数	7	1	1	0	0	0	0	0	已完成	0.0s	-	-
9	生成随机数	8	1	1	0	0	0	0	0	已完成	0.0s	-	-
10	生成随机数	9	1	1	0	0	0	0	0	已完成	0.0s	-	-

输出文件.txt - 记事本

文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)

随机数

0.1

0.6

0.1

0.7

0.6

0.4

0.8

0

0.9

0.9

一.Insight大数据智能分析平台概述

二.Insight大数据智能分析平台的架构

三.Insight大数据智能分析平台的使用

四.Insight大数据智能分析平台的特点

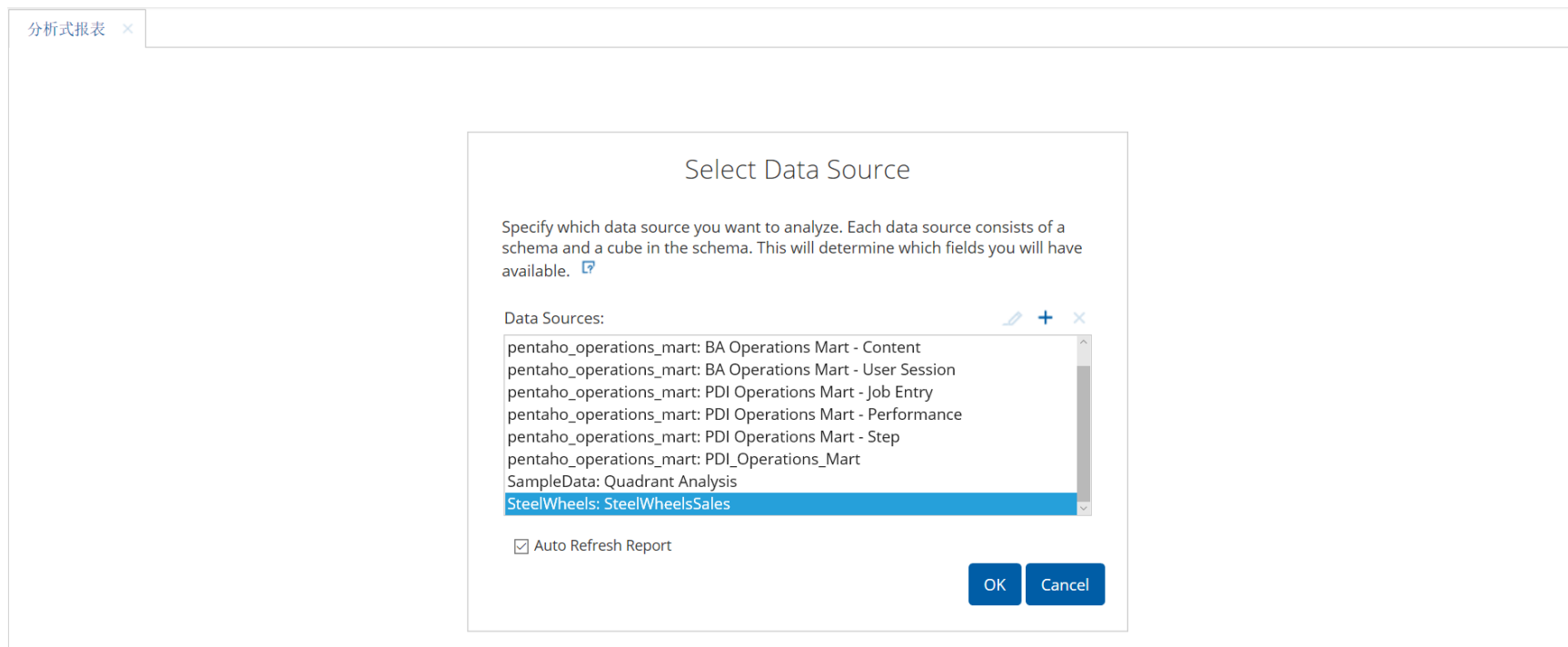
- 大数据时代下的数据分析师是企业顺利发展的重要的引路人，一个优秀分析师的成熟分析可以让一个企业做出最正确的决策。IUC的出现将是数据分析师们的福音。数据分析师的一个基本工作创建报表的流程如下：



- 在主页透视图中点击创建后，可以选择分析式报表、交互式报表、仪表盘和数据源进行创建。此处我们以分析式报表为例。



- 选择要做成分析式报表的数据源，点击OK



- 分析式报表的制作界面

分析式报表

Available fields (14) for: SteelWheelsSales

Find: View ▼

- Customers
 - Customer (6)
- Markets
 - Territory
 - Country
 - State Province
 - City
- Measures
 - Quantity
 - Sales
- Order Status
 - Type
- Product
 - Line
 - Vendor

Layout

Rows ☰
Drop Level Here *

Columns ☰
Drop Level Here *

Measures ☰
Drop Measure Here *

Properties

Report Options...

No Filters

View As: ☰ 📊 ▼

Table

Drag an available field to the required layout zones.
(Marked with a red asterisk in the layout panel.)

Need help getting started? [Help](#)

- 按照需求，拖动左边一系列的选项，可以构造出想要得到的报表
- 右上角表示可选择的报表的不同表现形式

分析式报表

Available fields (14) for: SteelWheelsSales

Find: View

- Customer (6)
- Markets
 - Territory
 - Country
 - State Province
 - City
- Measures
 - Quantity
 - Sales
- Order Status
 - Type
- Product
 - Line

Layout

Rows

Drop Level Here *

Columns

Drop Level Here *

Measures

Drop Measure Here

Properties

Report Options...

View As:

Table

Drag an available field to the required layout zones.
(Marked with a red asterisk in the layout panel.)

Need help getting started?

• 拖拽之后形成的报表

分析式报表 ×

Available fields (14) for: SteelWheelsSales

Find: View ▾

- Customers (6)
 - Customer
- Markets
 - Territory
 - Country
 - State Province
 - City
- Measures
 - Quantity
 - Sales
- Order Status
 - Type
- Product
 - Line

Layout

Rows

- Years
- Months
- Customer
- City

Drop Level Here

Columns

Drop Level Here

Measures

- Quantity

Drop Measure Here

Properties

Report Options...

View As: ▾

No Filters

Rows: 305 Cols: 1

Years	Months	Customer	City	Quantity
2003	Sep	Anna's Decorations, Ltd	North Sydney	430
		Australian Gift Network, Co	South Brisbane	116
		Collectables For Less Inc.	Brickhaven	341
		Dragon Souvenirs, Ltd.	Singapore	380
		Euro+ Shopping Channel	Madrid	458
		Gift Ideas Corp.	Glendale	76
		Oulu Toy Supplies, Inc.	Oulu	342
	Oct	Signal Collectibles Ltd.	Brisbane	367
		Baane Mini Imports	Stavern	22
		Blauer See Auto, Co.	Frankfurt	251
		Boards & Toys Co.	Glendale	67
		Classic Legends Inc.	NYC	225
		Corporate Gift Ideas Co.	San Francisco	976
		Dragon Souvenirs, Ltd.	Singapore	670
		Euro+ Shopping Channel	Madrid	68
		Frau da Collezione	Milan	216
		FunGiftIdeas.com	New Bedford	98
Heintze Collectables	Århus	358		
Men 'R' US Retailers, Ltd.	Los Angeles	231		

- 点击右上角，改变变现形式

分析式报表

Available fields (14) for:
SteelWheelsSales

Find: View

Customers

Customer (6)

Markets

- Territory
- Country
- State Province
- City

Measures

- Quantity
- Sales

Order Status

Type

Product

Line

Layout

X-Axis

- Years
- Months
- Customer
- City

Drop Level Here

Series

Drop Level Here

Measures

- Quantity

Drop Measure Here

Properties

Data Labels:

None

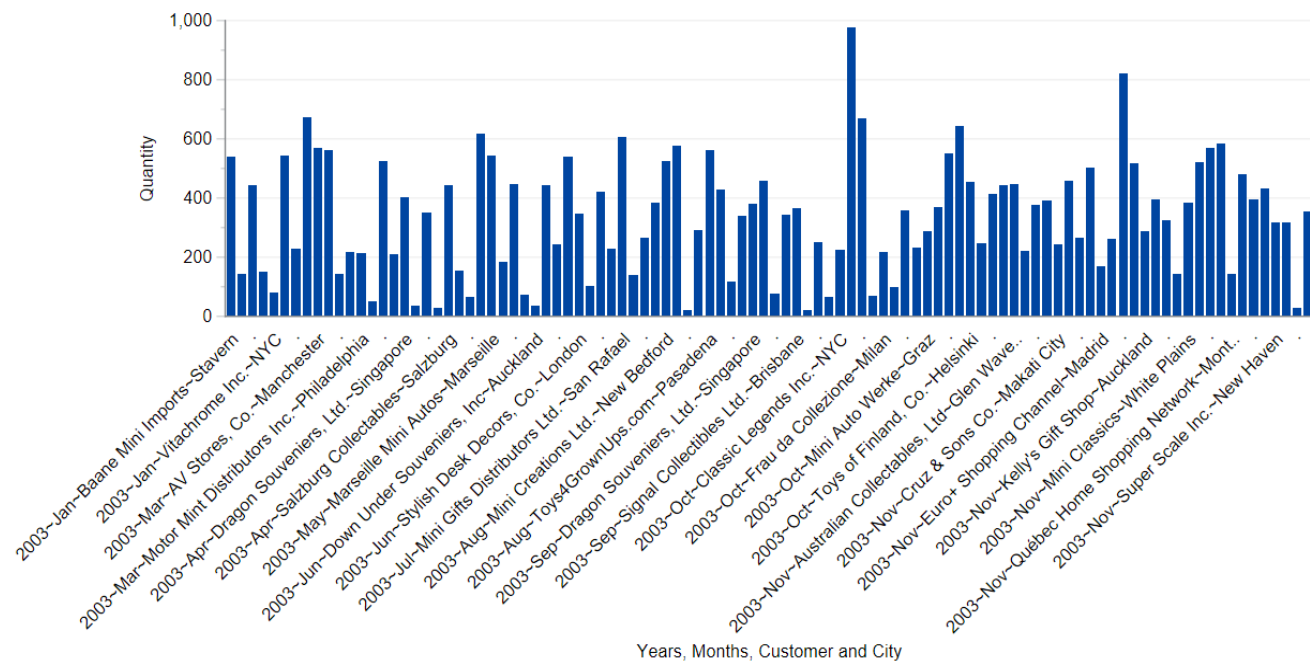
Trend Type

None

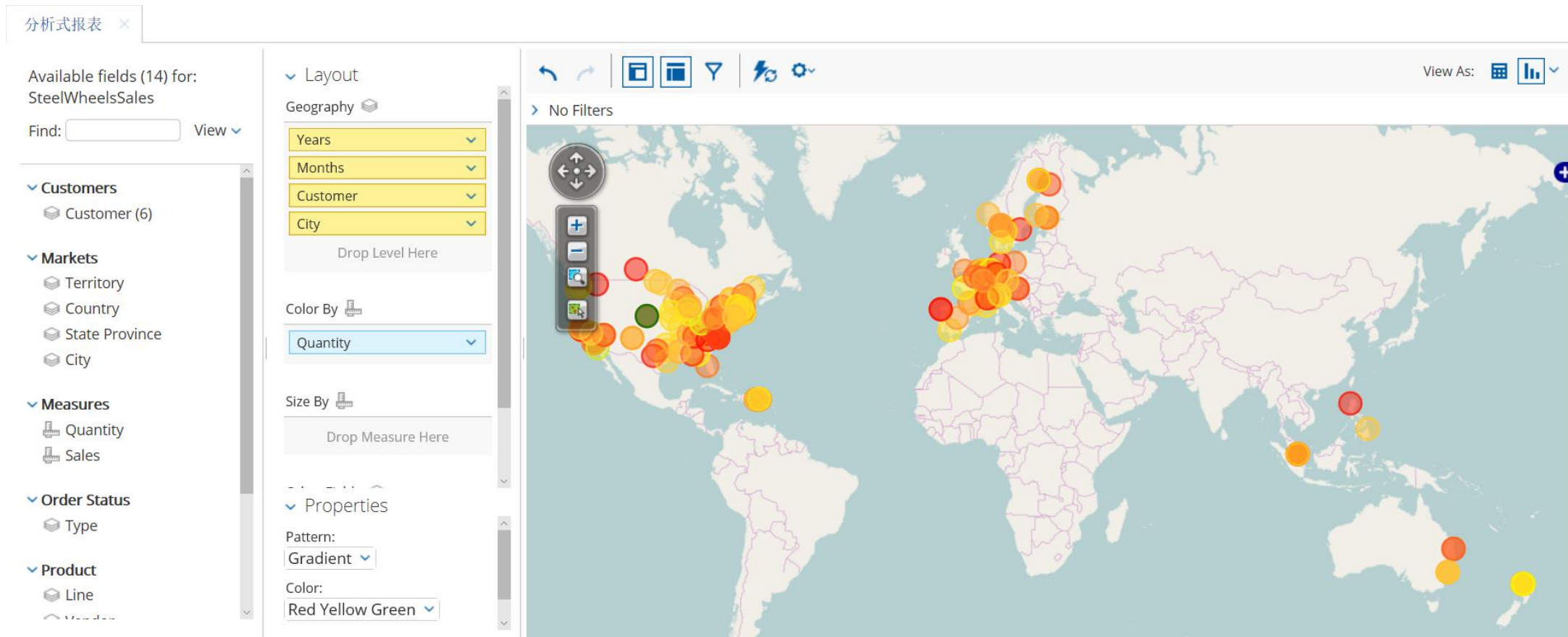


View As:

No Filters



- 数据源中含有不同国家的信息，故可以用地图的形式呈现







- 报表制作完成，可以保存在Insight User Console的文件夹当中

保存

Filename:

Location:

Name	Type	Date Modified
 dashboard	Folder	
 job	Folder	
 report	Folder	
 transformation	Folder	
example1	File	Jun 20, 2017 4:40:11 PM
Geomap.pdf	File	Feb 16, 2017 12:00:02 AM

保存

Cancel

- 数据的清洗是一个繁琐复杂的工作，一个有效的工具可以大大提高ETL工程师进行数据抽取，转换，加载的工作。
- 建立如图所示的一个转换



- 使用步骤“生成记录”：名称定为“生成记录”；限制生成10条记录；定义三个字段，名称、类型和值如下：

生成记录

步骤名称

限制

Never stop generating rows

Interval in ms (delay)

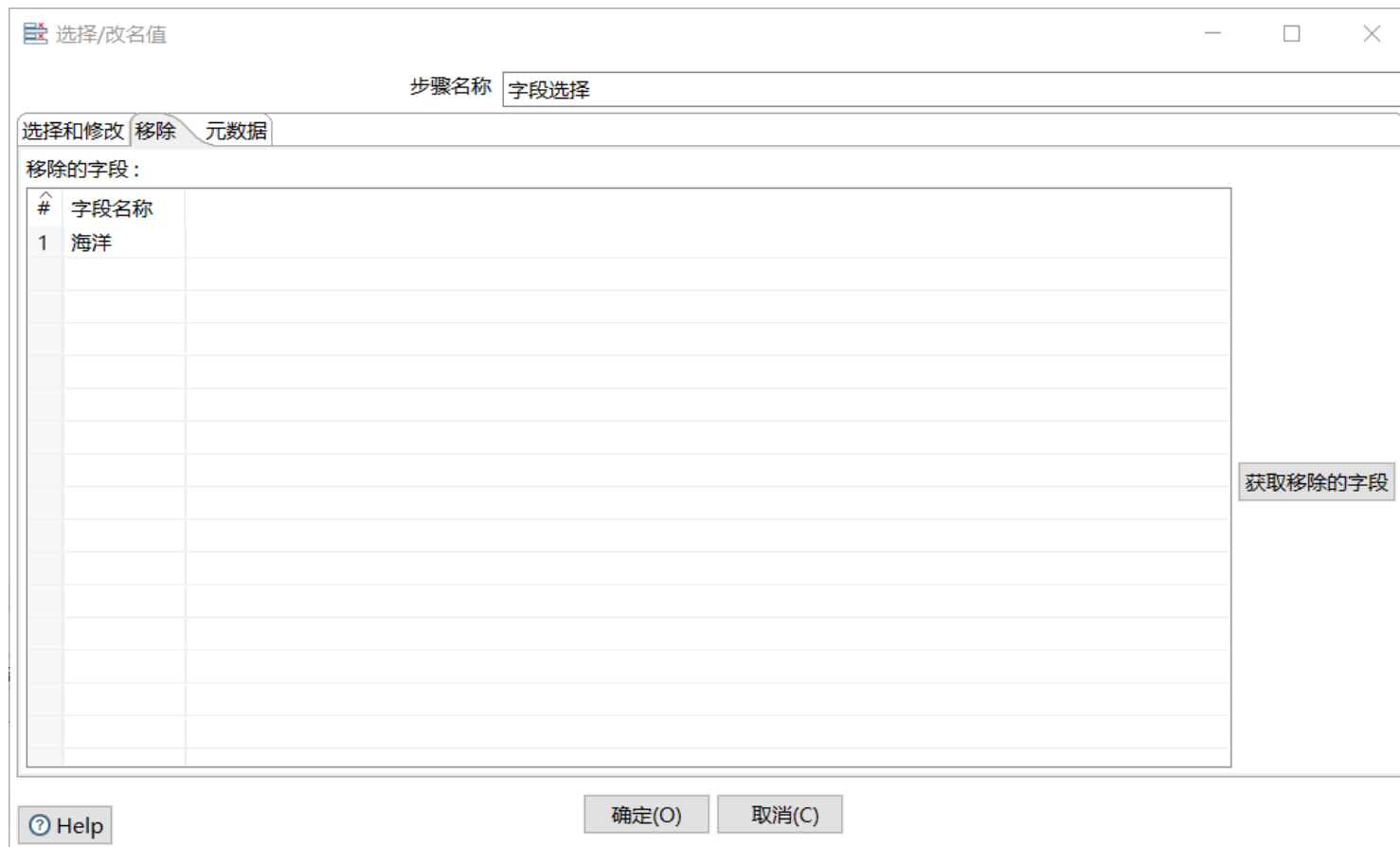
Current row time field name

Previous row time field name

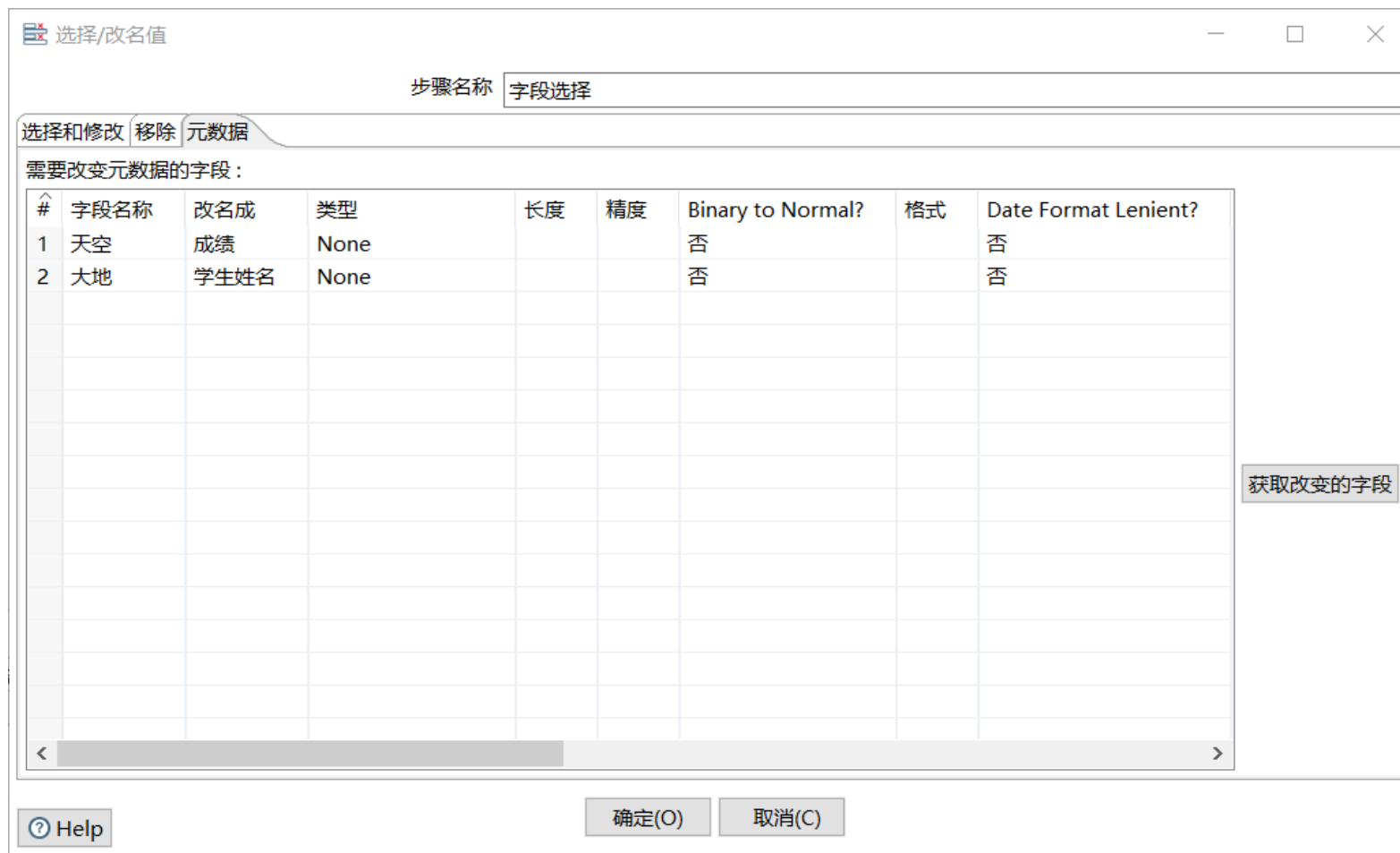
字段:

#	名称	类型	格式	长度	精度	货币类型	小数	分组	值	设为空串?
1	天空	Number							100	否
2	大地	String							张三	否
3	海洋	Boolean							0	否


- 在步骤“字段选择”中移除字段“海洋”



- 修改元数据，把“天空”修改成“成绩”；把“大地”修改成“学生姓名”

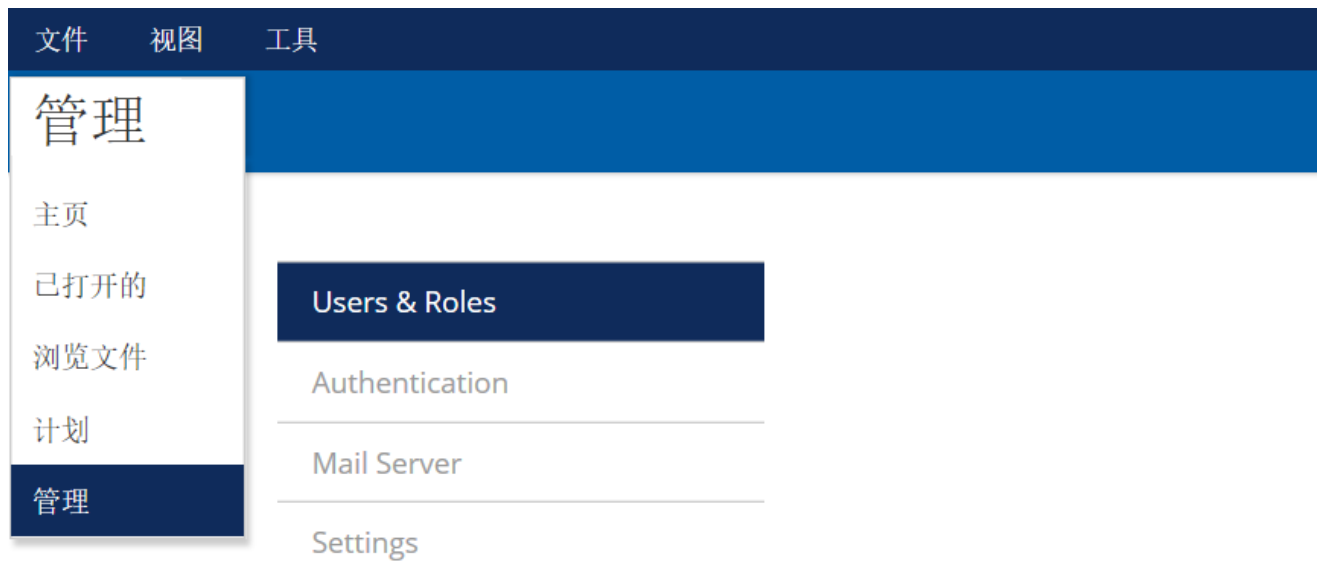


- 运行转化，可以输出到文本文档中，信息如下：



```
学生成绩.txt - 记事本
文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)
成绩;学生姓名
100.0;张三
100.0;张三
100.0;张三
100.0;张三
100.0;张三
100.0;张三
100.0;张三
100.0;张三
100.0;张三
100.0;张三
100.0;张三
```

- 管理员可以在较高权限的层面上对INSIGHT进行管理。



一.Insight大数据智能分析平台概述

二.Insight大数据智能分析平台的架构

三.Insight大数据智能分析平台的使用

四.Insight大数据智能分析平台的特点

- 面向解决方案
- 交互式可视化
- 支持大数据接入
- 可扩展性
- 安全性

- 面向解决方案（Solution）：

Solution本身实质上是一系列Action序列的集合，这些序列如何显示，如何被调用，功能如何实现完全由Insight平台来管理，这使得Solution的开发者，也就是Insight的使用者，可以将开发工作集中于具体的BI业务逻辑的开发上。

- Insight采用非常友好的图形界面实现人机交互。
- 项目拖拽功能

分析式报表

Available fields (14) for: SteelWheelsSales

Find: View

- Customers
 - Customer (6)
- Markets
 - Territory
 - Country
 - State Province
 - City
- Measures
 - Quantity
 - Sales
- Order Status
 - Type
- Product
 - Line

Layout

Rows

- Years
- Months
- Customer
- City

Drop Level Here

Columns

Drop Level Here

Measures

- Quantity

Drop Measure Here

Properties

Report Options...

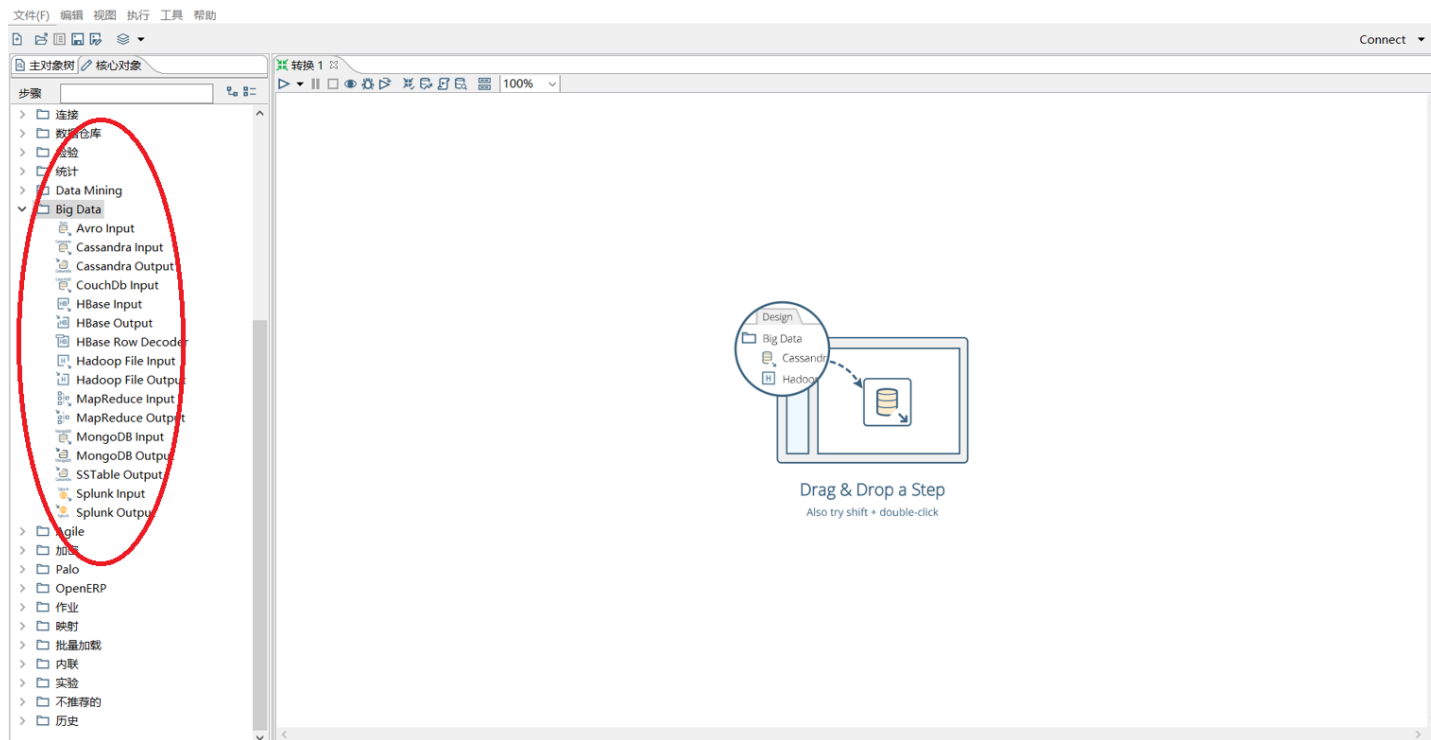
No Filters

View As:

Rows: 305 Cols: 1

Years	Months	Customer	City	Quantity
2003	Sep	Anna's Decorations, Ltd	North Sydney	430
		Australian Gift Network, Co	South Brisbane	116
		Collectables For Less Inc.	Brickhaven	341
		Dragon Souvenirs, Ltd.	Singapore	380
		Euro+ Shopping Channel	Madrid	458
		Gift Ideas Corp.	Glendale	76
		Oulu Toy Supplies, Inc.	Oulu	342
	Signal Collectibles Ltd.	Brisbane	367	
	Oct	Baane Mini Imports	Stavern	22
		Blauer See Auto, Co.	Frankfurt	251
		Boards & Toys Co.	Glendale	67
		Classic Legends Inc.	NYC	225
		Corporate Gift Ideas Co.	San Francisco	976
		Dragon Souvenirs, Ltd.	Singapore	670
Euro+ Shopping Channel		Madrid	68	
Frau da Collezione	Milan	216		
FunGiftIdeas.com	New Bedford	98		
Heintze Collectables	Århus	358		
Men 'R' US Retailers, Ltd.	Los Angeles	231		

- 在Insight Data Integration的步骤中，Big Data包含各种大数据的输入和输出，为数据的处理带来了极大的便利



- 一个系统平台的可扩展性,是指该平台适应变化的能力。
- 在软件范畴上,可扩展性是软件系统本身的属性,或者进一步说是设计的属性,代码的属性。
- 软件环境的变化(可能是业务环境,运行环境)导致软件要进行改动才能满足人们对它的要求,这种适应变化的能力就是可扩展性。

- 用户名密码认证机制
- 管理员管理机制
- 系统定期维护





计算 决定未来

曙光信息产业股份有限公司

THANKS

通讯地址：北京市海淀区东北旺西路8号中关村软件园36号