

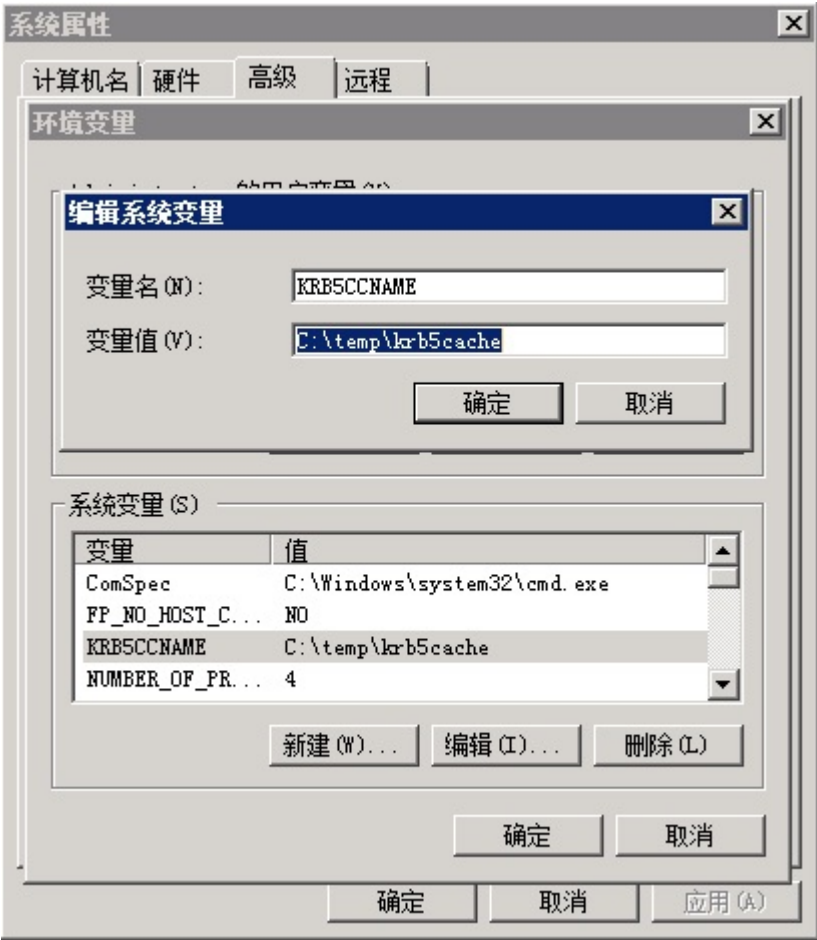
Tableau对接FusionInsight

适用场景

Tableau 10.0.0 <-> FusionInsight HD V100R002C30
Tableau 10.0.0 <-> FusionInsight HD V100R002C50
Tableau 10.0.0 <-> FusionInsight HD V100R002C60U10
Tableau 10.1.4 <-> FusionInsight HD V100R002C60U20
Tableau 10.3.2 <-> FusionInsight HD V100R002C70SPC200
Tableau 10.5.0 <-> FusionInsight HD V100R002C80SPC100

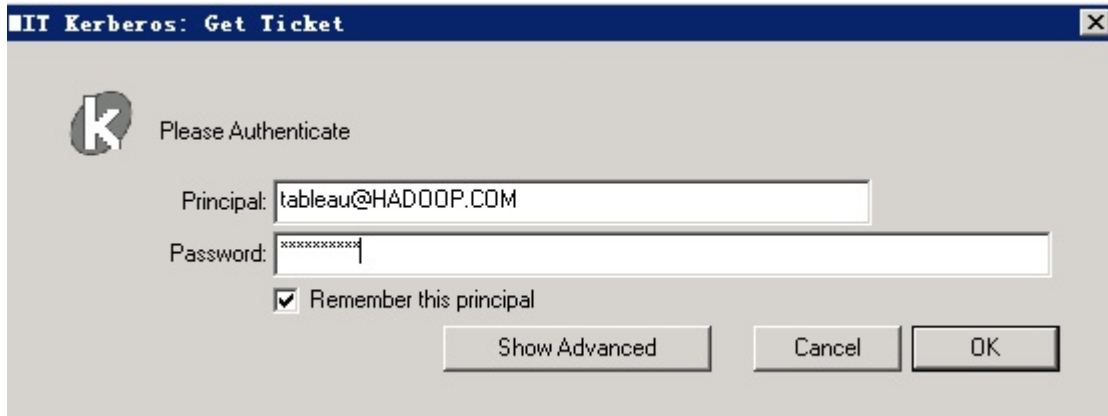
配置Windows的kerberos认证

- 下载并安装MIT Kerberos
下载网址: <http://web.mit.edu/kerberos/dist/#kfw-4.0>
版本与操作系统位数保持一致, 本文版本kfw-4.1-amd64.msi。
- 确认客户端机器的时间与FusionInsight HD集群的时间一致, 时间差要小于5分钟
- 设置Kerberos的配置文件
在FusionInsight Manager创建一个角色与“人机”用户, 具体请参见《FusionInsight HD 管理员指南》的创建用户章节。角色需要根据业务需要授予Hive的访问权限, 并将用户加入角色。例如, 创建用户“tableau”并下载对应的keytab文件userkeytab以及krb5.conf文件, 把krb5.conf文件重命名为krb5.ini, 并放到 **C:\ProgramData\MIT\Kerberos5** 目录中。
- 设置Kerberos票据的缓存文件
 - 创建存放票据的目录, 例如“C:\temp”。
 - 设置Windows的系统环境变量, 变量名为“KRB5CCNAME”, 变量值为“C:\temp\krb5cache”。



- 重启机器。

- 在Windows上进行认证
 - 使用上述创建的用户名密码登录，用户名的格式为：用户名@Kerberos域名。
 - 打开MIT Kerberos，单击“get Ticket”，在弹出的“MIT Kerberos: Get Ticket”窗口中，“Principal”输入用户名，“Password”输入密码，单击“OK”。



配置Hive数据源

Tableau中配置Hive数据源，对接Hive的ODBC接口。

- 下载并安装ODBC驱动：[下载地址](#)

根据操作系统类型选择对应的ODBC版本，下载并安装。
- 配置ODBC驱动
 - 创建DSN(Data Source Name): 选择 开始 -> **Simba Spark ODBC Driver** -> **ODBC Administrator**。
 - 选择 **User DSN** -> **Add** -> **Cloudera ODBC Driver for Apache Hive** -> **Finish**

按实际配置相应的变量，

- Host(s): Hive Service主节点
- Port: Hive Service端口21066
- Mechanism: Kerberos
- Host FQDN: [hadoop.hadoop.com](#)
- Service Name: hive
- Realm: 留空

如下图

Cloudera ODBC Driver for Apache Hive DSN Setup

Data Source Name:

Description:

Hive Server Type:

Service Discovery Mode:

Host(s):

Port:

Database:

ZooKeeper Namespace:

Authentication

Mechanism:

Realm:

Host FQDN:

Service Name:

☐ Delegate Kerberos Credentials

User Name:

Password:

☐ Save Password (Encrypted)

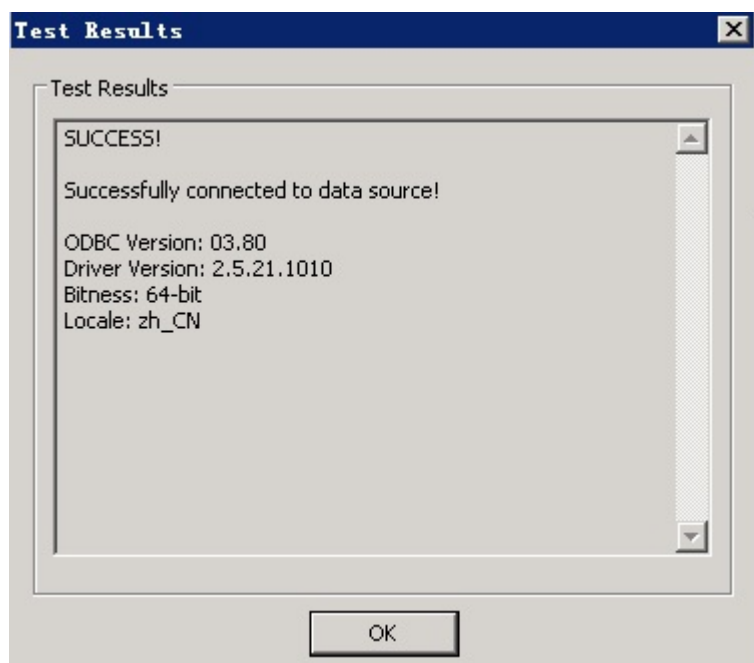
Delegation UID:

Thrift Transport:

v2.5.21.1010 (64 bit)

Advanced Options不需要进行配置默认的参数即可连接成功。

- 点击中的Test进行测试连接，如果出现下图，则表示ODBC连接Hive成功。



- Tableau使用数据源
 - Tableau启动后会进入连接选择界面，点击“更多服务器”，再点击“其他数据库（ODBC）”；
 - DSN选择hive_odbc（上一步中设置ODBC的名称），点击连接，如下图所示，点击“连接”，然后登陆。

其他数据库 (ODBC)

hive_odbc (ODBC)

连接方式(C)

☒ DSN:

hive_odbc

☐ 驱动程序:

Cloudera ODBC Driver for Apache Hive

连接

连接属性(A)

服务器(V):

端口(R):

数据库(D):

HIVE

用户名(U):

密码(P):

字符串附加部分:

登录

- 查询百万级数据表数据

连接

hive_odbc (ODBC)

数据库

HIVE

架构

default

表

web_sales (tpcd_text_2.web_sales)

workers_info(workers_info)

新自定义 SQL

store_sales (tpcds_text_2.store_sales)+ (HIVE)

web_sales

连接

实时

数据提取

筛选器

0

添加

排序字段

数据源顺序

显示别名

显示隐藏字段

1,000

行

#	#	#	#	#	#	#	#	#
web_sales	web_sales	web_sales	web_sales	web_sales	web_sales	web_sales	web_sales	web_sales
ws_sold_date_sk	ws_sold_time_sk	ws_ship_date_sk	ws_item_sk	ws_bill_custom...	ws_bill_demo_sk	ws_bill_demo_sk	ws_bill_addr_sk	ws_ship_custom...
2,451,383	73,313	2,451,482	13,801	91,074	596,485	1,096	56,907	
2,451,383	73,313	2,451,411	4,700	91,074	596,485	1,096	56,907	
2,451,383	73,313	2,451,413	19,418	91,074	596,485	1,096	56,907	
2,451,383	73,313	2,451,393	19,048	91,074	596,485	1,096	56,907	
2,451,383	73,313	2,451,502	12,595	91,074	596,485	1,096	56,907	
2,451,383	73,313	2,451,421	5,605	91,074	596,485	1,096	56,907	
2,451,383	73,313	2,451,457	10,594	91,074	596,485	1,096	56,907	
2,451,383	73,313	2,451,430	10,172	91,074	596,485	1,096	56,907	
2,451,383	73,313	2,451,458	20,860	91,074	596,485	1,096	56,907	
2,452,625	34,964	2,452,702	25,521	130,296	436,090	2,684	25,278	

数据源

工作表 1

数据源

数据源

数据源

- 查询多表数据

连接

hive_odbc (ODBC)

添加

数据库

HIVE

架构

tpcd_text_2

表

输入表名称

精确 包含 开头为

catalog_retalog_returns)

catalog_salalog_sales)

customer (t...2.customer)

customer_ad...er_address)

customer_de...ographics)

date_dis (t...2.date_dim)

household...ographics)

income_band...ncome_band)

inventory (...inventory)

item (tpcd_text_2.item)

promotion (...promotion)

reason (tpc...t_2.reason)

thin...s...thin...s...)

新自定义 SQL

转到工作表

store_sales (tpcds_text_2.store_sales)+ (HIVE)

连接 实时 数据提取

筛选器 0 添加

store_sales item

排序字段 数据源顺序

显示别名 显示隐藏字段 1,000 行

非	非	非	非	非	非	非	非	非	非
item	item	item	item	item	item	item	item	item	item
i_item_sk	i_item_id	i_rec_start_date	i_rec_end_date	i_item_desc	i_current_price	i_wholesale_cost	i_brand_id	i_brand	
7,649	AAAAAAAAACNFAAAA	1999-10-28	2001-10-26	So-called steps would not see very origina...	6.9300	29.3200	5003002	expertischolar #2	
19,874	AAAAAAAAACNFAAAA	1997-10-27	2000-10-26	Fresh islands could develop normal, gene...	6.7800	4.5400	5002001	importescholar #1	
18,371	AAAAAAAAACNFAAAA	1999-10-28	2001-10-26	Financial, good	4.1400	6.4800	4003002	expertiedu pack #2	
9,401	AAAAAAAAALECAAAA	1999-10-28	2001-10-26	Rough properties find meanwhile different ...	22.0700	1.2000	7002002	importbrand #2	
10,772	AAAAAAAAACNFAAAA	1997-10-27	2000-10-26	Grounds will maintain nerely white faces: ...	5.4700	3.3300	7012003	importnameless #3	
4,370	AAAAAAAAACNFAAAA	1997-10-27	2000-10-26	Arms will pay scottish, wrong user...	1.8700	1.1700	7009001	maizbrand #1	
19,637	AAAAAAAAALEMFAAAA	1999-10-28	2001-10-26	Members avoid personal authorities. Very in...	4.3000	1.4600	9013005	edu packedu pack #2	
24,617	AAAAAAAAICAGAAAA	1999-10-28	2001-10-26	Only new women place quite of course func...	0.7400	3.5800	10006001	amalgamalg #2	
25,616	AAAAAAAAABEGAAAA	1997-10-27	2000-10-26	Drivers will not hear simultaneously with ...	4.8600	1.6500	7016007	corpnameless #7	
2,630	AAAAAAAAACNFAAAA	1997-10-27	2000-10-26	Other things get now Quite eastern system...	0.2700	0.0900	8004001	edu packnameless #1	

数据源 工作表 1

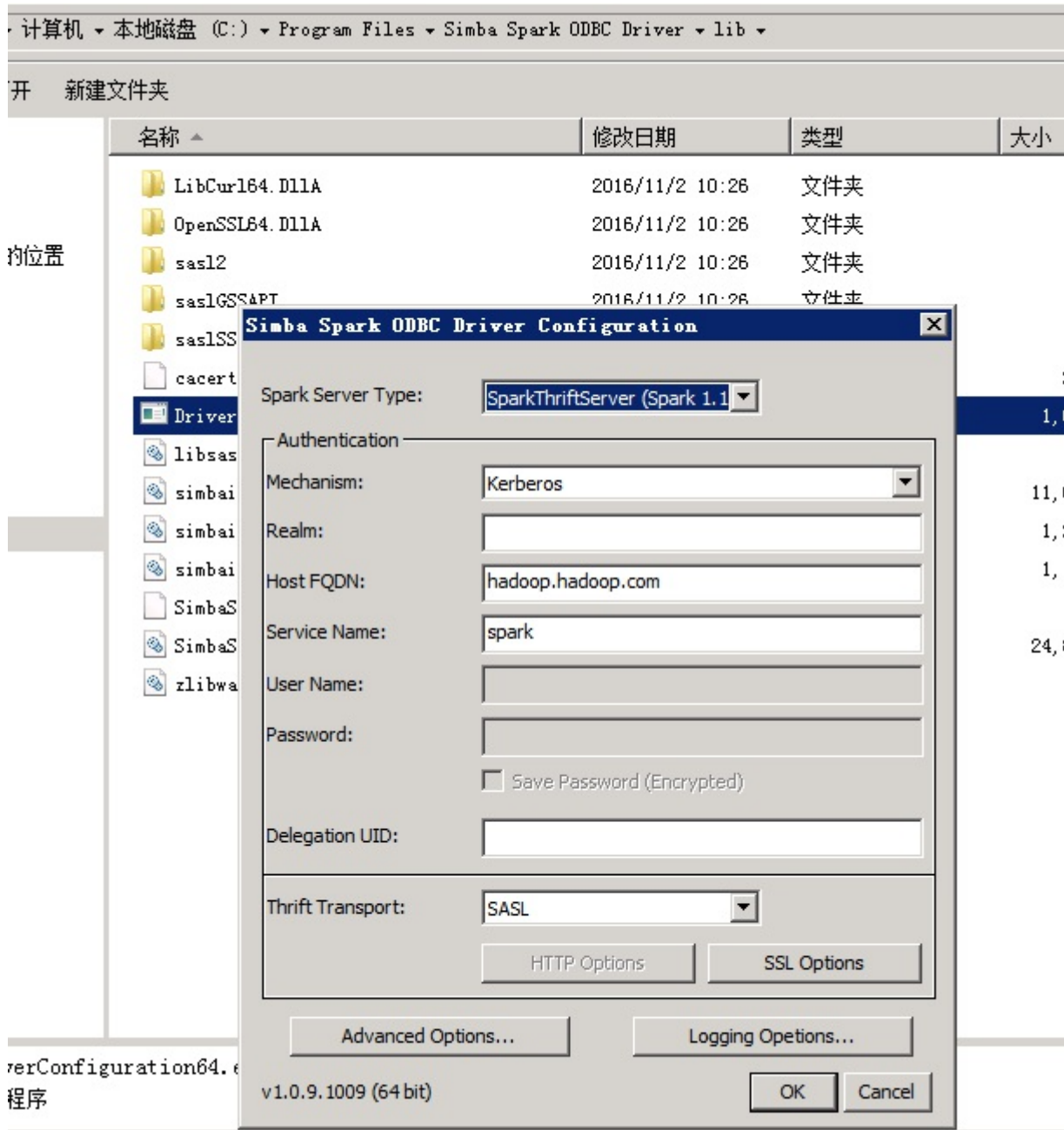
配置Spark数据源

- 下载并安装spark的ODBC驱动 ODBC驱动下载地址：<http://www.tableau.com/support/drivers>

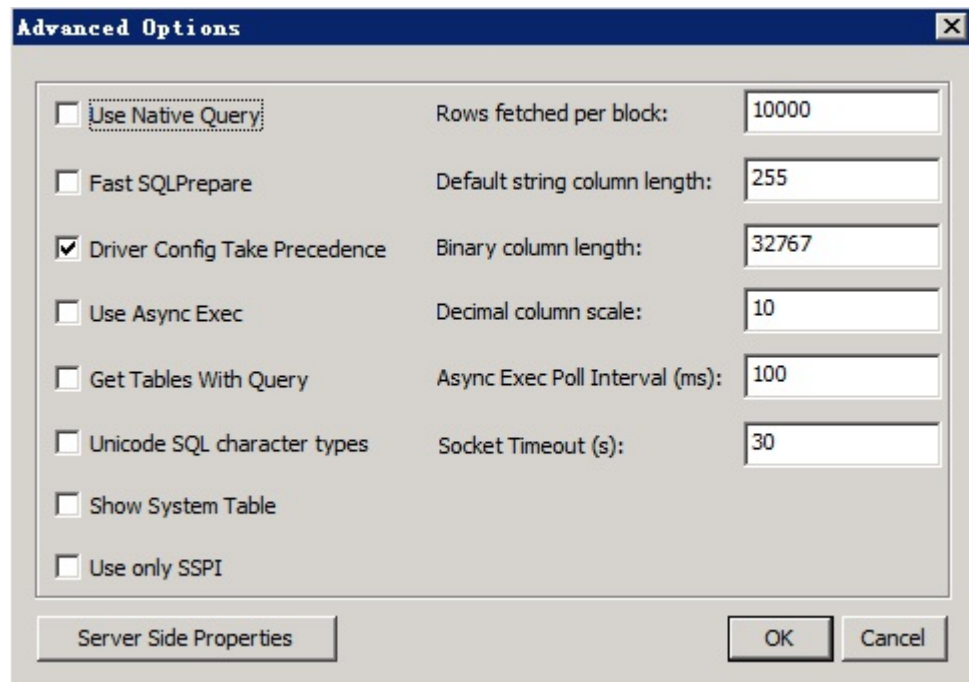
Spark SQL (64-bit)	8.2.8 and later, 8.3.3 and later, 9.0 and later	Use the Spark SQL connector to connect to a Spark cluster on Azure HDInsight, Databricks, or Apache Spark.
--------------------	---	--

- 创建DSN（Data Source Name）
- 打开 C:\Program Files\Simba Spark ODBC Driver\lib\DriverConfiguration64.exe
- 按实际配置相应的变量
 - Mechanism：Kerberos
 - Host FQDN：hadoop.hadoop.com
 - Service Name：spark
 - Realm：留空

如下图：



- 点击“Advanced Options”，勾选如下选项：



- 点击OK，保存配置。
- Tableau使用Spark数据源
 - Tableau启动后会进入连接选择界面，点击“更多服务器”，再点击“Spark SQL”，作如下配置：

Spark SQL

172.21.3.103

服务器(V): 172.21.3.103 端口(R): 23040

输入服务器登录信息:

类型: SparkThriftServer (Spark 1.1 及更高版本)

身份验证: 无身份验证

用户名(U):

密码:

域:

主机 FQDN:

服务名称:

HTTP 路径:

初始 SQL(I)...

登录

- 其中服务器为JDBCServer(主)的业务IP。

服务 > Spark 实例

服务状态 实例 服务配置 资源贡献排名

+ 添加实例 更多操作

	角色	主机名	管理IP	业务IP
<input type="checkbox"/>	JDBCServer(备)	172-21-2-102	172.21.3.102	172.21.3.102
<input type="checkbox"/>	JDBCServer(主)	172-21-3-103	172.21.3.103	172.21.3.103
<input type="checkbox"/>	JobHistory	172-21-2-102	172.21.3.102	172.21.3.102
<input type="checkbox"/>	JobHistory	172-21-3-103	172.21.3.103	172.21.3.103
<input type="checkbox"/>	SparkResource	172-21-3-101	172.21.3.101	172.21.3.101
<input type="checkbox"/>	SparkResource	172-21-2-102	172.21.3.102	172.21.3.102
<input type="checkbox"/>	SparkResource	172-21-3-103	172.21.3.103	172.21.3.103

- 端口为FusionInsight中Spark服务配置，导出服务配置文件，其中hive.server2.thrift.port对应值。

```
- <configuration-definition>
  <name>hive.server2.thrift.port</name>
  <value>23040</value>
</configuration-definition>
```

- 点击“登录”，进入tableau页面，选择架构和表，结果如下。

Tableau - 工作簿 1

文件(F) 数据(D) 服务器(S) 窗口(W) 帮助(H)

连接 172.21.3.103 Spark SQL

架构 default

表

输入表名称

精确 包含 开头为

employees_co...ees_contact)

employees_in...loyees_info)

employees_sa...yees_salary)

sample_07 (d...t.sample_07)

sample_08 (d...t.sample_08)

新自定义 SQL

employees_info (default.employees_info) (default)

employees_info

排序字段 数据源顺序

#	Abc	Abc	#	Abc	Abc
employee...	employees_info	employees_info	employees_in...	employees_info	employees_info
id	name	usd_flag	salary	address	entrytime
1	Wang	R	8,000.01	China:Shenzhen	2014
2	Tom	D	12,000.02	America:NewYork	2014
3	Jack	D	24,000.03	China:America:Manh...	2015
4	zhang	R	9,000.05	China:shanghai	2015
5	Jim	R	10,000.05	China:shenzhen	2015

- 用Tableau做实时连接，打开工作簿，对该表进行图形化分析。

Tableau - 工作簿 1

文件(F) 数据(D) 工作表(W) 仪表板(B) 故事(T) 分析(A) 地图(M) 设置格式(O) 服务器(S) 窗口(W) 帮助(H)

数据 分析

employees_info (default...)

维度

Abc address

Abc entrytime

id

Abc name

Abc usd_flag

Abc 度量名称

度量

salary

记录数

度量值

工作表 1

entrytime / name

2014 2015

salary

25K

20K

15K

10K

5K

0K

Tom Wang Jack zhang

12,000 8,000 24,000 9,000

智能显示

对于水平条，尝试

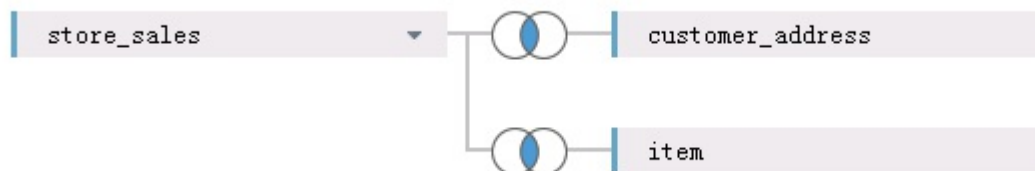
0 个或多个 维度

1 个或多个 度量

4 个标记 1 行 x 4 列 总计(salary): 53,000

- 性能测试

- 查询包含百万条数据的表web_sales



查询结果:

文件(F)

数据(D)

服务器(S)

窗口(W)

帮助(H)

←

→

↺

↻

连接

162.1.93.103

Spark SQL

添加

架构

tpcds_text_2

表

输入表名称

精确

包含

开头为

call_center('all_center')

catalog_page('atalog_page')

catalog_returns('catalog_returns')

catalog_sales('alog_sales')

customer ('t'=2.customer)

customer_address('customer_address')

customer_demographics('customer_demographics')

date_dim ('t'=2.date_dim)

household_income_demographics('household_income_demographics')

income_band('ncome_band')

inventory ('..inventory')

item (tpcds_text_2.item)

promotion ('..promotion')

reason (tpcds_text_2.reason)

ship_mode ('..ship_mode')

新自定义 SQL

store_sales (tpcds_text_2.store_sales)+ (tpcds...

连接

实时

数据提取

筛选器

0

添加

store_sales

customer_address

item

排序字段

数据源顺序

显示别名

显示隐藏字段

1,000

行

#	customer_address_sk	customer_address_id	customer_address_street_number	customer_address_street_name	customer_address_street_type	customer_address_suite_number	customer_address_city	customer_address_county	customer_address_state
69,029	AAAAAAAFIXABAAA	595	Pine	Ln	Suite 440	Centerville	Clinton County	PA	
69,029	AAAAAAAFIXABAAA	595	Pine	Ln	Suite 440	Centerville	Clinton County	PA	
69,029	AAAAAAAFIXABAAA	595	Pine	Ln	Suite 440	Centerville	Clinton County	PA	
69,029	AAAAAAAFIXABAAA	595	Pine	Ln	Suite 440	Centerville	Clinton County	PA	
69,029	AAAAAAAFIXABAAA	595	Pine	Ln	Suite 440	Centerville	Clinton County	PA	
69,029	AAAAAAAFIXABAAA	595	Pine	Ln	Suite 440	Centerville	Clinton County	PA	
69,029	AAAAAAAFIXABAAA	595	Pine	Ln	Suite 440	Centerville	Clinton County	PA	
69,029	AAAAAAAFIXABAAA	595	Pine	Ln	Suite 440	Centerville	Clinton County	PA	
69,029	AAAAAAAFIXABAAA	595	Pine	Ln	Suite 440	Centerville	Clinton County	PA	
69,029	AAAAAAAFIXABAAA	595	Pine	Ln	Suite 440	Centerville	Clinton County	PA	

数据源

工作表 1

FAQ

- 找不到C:\ProgramData\MIT\Kerberos5文件夹

C:\ProgramData一般属于隐藏文件夹，设置文件夹隐藏可见或者使用搜索功能即可解决问题。

- 连接成功无数据库权限

连接所使用的用户需要有数据库的权限，否则将导致ODBC连接成功却无法读取数据库内容。

- ODBC连接失败

常见情况是Host(s)、Port、Host FQDN等的输入数据有误，请根据实际情况进行输入