

hive元数据：记录着hive中存储着哪些数据库，哪些表等信息，通常是使用mysql数据库来存储hive的元数据

hive元数据\*存储配置：利用mysql数据库存储hive元数据

因此要开启hive前要开启mysql

In [5]: 1 !echo "show databases;show tables;" | hive

SLF4J: Class path contains multiple SLF4J bindings.

SLF4J: Found binding in [jar:file:/root/bigdata/apache-hive-2.3.4-bin/lib/log4j-slf4j-impl-2.6.2.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: Found binding in [jar:file:/root/bigdata/hadoop-2.9.1/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: See [http://www.slf4j.org/codes.html#multiple\\_bindings](http://www.slf4j.org/codes.html#multiple_bindings) ([http://www.slf4j.org/code\\_s.html#multiple\\_bindings](http://www.slf4j.org/code_s.html#multiple_bindings)) for an explanation.

SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]

Logging initialized using configuration in jar:file:/root/bigdata/apache-hive-2.3.4-bin/lib/hive-common-2.3.4.jar!/hive-log4j2.properties Async: true

Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.

hive> show databases;show tables;

OK

article

default

profile

test2

toutiao

Time taken: 0.748 seconds, Fetched: 5 row(s)

OK

employee

sku\_detail

sku\_tag\_weights

student

tb\_goods

tb\_goods\_category

tb\_goods\_specification

tb\_sku

tb\_sku\_specification

tb\_specification\_option

u

u4

Time taken: 0.069 seconds, Fetched: 12 row(s)

hive>

mysqlToHive.sh

```
#!/bin/bash

array=(tb_goods tb_goods_category tb_goods_specification tb_sku tb_sku_specification tb_specification_option)

for table_name in ${array[@]};
do
    sqoop import \
        --connect jdbc:mysql://localhost/meiduo_mall \
        --username root \
        --password itcast \
        --table $table_name \
        --m 5 \
        --hive-home /root/bigdata/hive \
        --hive-import \
        --create-hive-table \
        --hive-drop-import-delims \
        --warehouse-dir /user/hive/warehouse \
        --hive-table $table_name
done
```

--hive-import 就是指明往hive里面导入

--warehouse-dir 指明hdfs上hive数据库的路径

重点注意 --hive-drop-import-delims 选项，会把如"\n \r"等字符自动处理掉，避免出现数据与字段不对应的问题

加可执行权限 chmod +x mysqlToHive.sh

执行 source mysqlToHive.sh

In [2]: 1 !hadoop fs -ls /user/hive/warehouse

```
Found 14 items
drwxr-xr-x  - root supergroup          0 2020-11-10 20:41 /user/hive/warehouse/article.db
drwxr-xr-x  - root supergroup          0 2020-12-10 16:32 /user/hive/warehouse/employee
drwxr-xr-x  - root supergroup          0 2020-11-11 22:14 /user/hive/warehouse/profile.db
drwxr-xr-x  - root supergroup          0 2020-11-10 12:38 /user/hive/warehouse/sku_detail
drwxr-xr-x  - root supergroup          0 2020-11-10 12:41 /user/hive/warehouse/sku_tag_weights
drwxr-xr-x  - root supergroup          0 2020-12-10 15:20 /user/hive/warehouse/student
drwxr-xr-x  - root supergroup          0 2020-11-09 14:02 /user/hive/warehouse/tb_goods
drwxr-xr-x  - root supergroup          0 2020-11-09 14:03 /user/hive/warehouse/tb_goods_category
drwxr-xr-x  - root supergroup          0 2020-11-09 14:04 /user/hive/warehouse/tb_goods_specification
drwxr-xr-x  - root supergroup          0 2020-11-09 14:04 /user/hive/warehouse/tb_sku
drwxr-xr-x  - root supergroup          0 2020-11-09 14:05 /user/hive/warehouse/tb_sku_specification
drwxr-xr-x  - root supergroup          0 2020-11-09 14:06 /user/hive/warehouse/tb_specification_option
drwxr-xr-x  - root supergroup          0 2020-11-10 19:51 /user/hive/warehouse/toutiao.db
drwxr-xr-x  - root supergroup          0 2020-12-10 17:21 /user/hive/warehouse/u
```

## pyspark中从Hive数据库进行数据加载

注意配置：

- 配置中加入 ("hive.metastore.uris", "thrift://localhost:9083") 配置项，告诉spark使用hive已有的metastore，否则spark会尝试自己建立一个新的metastore
- 创建SparkSession时，调用 enableHiveSupport 方法开启对Hive的操作接口

```

In [1]: 1 import os
2 # 配置pyspark和spark driver运行时 使用的python解释器
3 JAVA_HOME = '/root/bigdata/jdk'
4 PYSARK_PYTHON = '/miniconda2/envs/py365/bin/python'
5 # 当存在多个版本时, 不指定很可能会导致出错
6 os.environ['PYSARK_PYTHON'] = PYSARK_PYTHON
7 os.environ['PYSARK_DRIVER_PYTHON'] = PYSARK_PYTHON
8 os.environ['JAVA_HOME'] = JAVA_HOME
9 # 配置spark信息
10 from pyspark import SparkConf
11 from pyspark.sql import SparkSession
12
13 SPARK_APP_NAME = "transMeiDuoData"
14 SPARK_URL = "yarn"
15
16 conf = SparkConf() # 创建spark config对象
17 config = (
18     ("spark.app.name", SPARK_APP_NAME), # 设置启动的spark的app名称, 没有提供, 将随
19     ("spark.executor.memory", "2g"), # 设置该app启动时占用的内存用量, 默认1g, 指一
20     ("spark.master", SPARK_URL), # spark master的地址
21     ("spark.executor.cores", "2"), # 设置spark executor使用的CPU核心数, 指一台虚拟
22     ("spark.executor.instances", "1"), # yarn模式下配置
23     ("hive.metastore.uris", "thrift://hadoop-master:9083"), # 配置hive元数据的访问,
24     # 以下三项配置, 可以控制执行器数量
25     # ("spark.dynamicAllocation.enabled", True),
26     # ("spark.dynamicAllocation.initialExecutors", 1), # 1个执行器
27     # ("spark.shuffle.service.enabled", True)
28     # ('spark.sql.pivotMaxValues', '99999'), # 当需要pivot DF, 且值很多时, 需要修改, 默
29 )
30 # 查看更详细配置及说明: https://spark.apache.org/docs/latest/configuration.html
31
32 conf.setAll(config)
33
34 # 利用config对象, 创建spark session
35 spark = SparkSession.builder.config(conf=conf).enableHiveSupport().getOrCreate()

```

```

In [2]: 1 spark.catalog.listDatabases()

```

```

Out[2]: [Database(name='article', description='artcile information', locationUri='hdfs://hadoop-
-master:9000/user/hive/warehouse/article.db'),
Database(name='default', description='Default Hive database', locationUri='hdfs://hado
op-master:9000/user/hive/warehouse'),
Database(name='profile', description='use action', locationUri='hdfs://hadoop-master:9
000/user/hive/warehouse/profile.db'),
Database(name='test2', description='', locationUri='hdfs://hadoop-master:9000/user/hiv
e/warehouse/test2.db'),
Database(name='toutiao', description='user,news information of 136 mysql', locationUri
='hdfs://hadoop-master:9000/user/hive/warehouse/toutiao.db')]

```

```
In [3]: 1 spark.catalog.listTables("default")
```

```
Out[3]: [Table(name='employee', database='default', description=None, tableType='MANAGED', isTemporary=False),
Table(name='sku_detail', database='default', description=None, tableType='MANAGED', isTemporary=False),
Table(name='sku_tag_weights', database='default', description=None, tableType='MANAGED', isTemporary=False),
Table(name='student', database='default', description=None, tableType='MANAGED', isTemporary=False),
Table(name='tb_goods', database='default', description='Imported by sqoop on 2020/11/09 14:02:49', tableType='MANAGED', isTemporary=False),
Table(name='tb_goods_category', database='default', description='Imported by sqoop on 2020/11/09 14:03:21', tableType='MANAGED', isTemporary=False),
Table(name='tb_goods_specification', database='default', description='Imported by sqoop on 2020/11/09 14:04:11', tableType='MANAGED', isTemporary=False),
Table(name='tb_sku', database='default', description='Imported by sqoop on 2020/11/09 14:04:53', tableType='MANAGED', isTemporary=False),
Table(name='tb_sku_specification', database='default', description='Imported by sqoop on 2020/11/09 14:05:57', tableType='MANAGED', isTemporary=False),
Table(name='tb_specification_option', database='default', description='Imported by sqoop on 2020/11/09 14:07:00', tableType='MANAGED', isTemporary=False),
Table(name='u', database='default', description=None, tableType='MANAGED', isTemporary=False),
Table(name='u4', database='default', description=None, tableType='EXTERNAL', isTemporary=False)]
```

In [4]: 1 spark.sql("select \* from tb\_goods").show()

```

+-----+-----+-----+-----+-----+-----+-----+-----+
| id|      create_time|      update_time|      name|sales|comments|bran
d_id|category1_id|category2_id|category3_id|      desc_detail|      desc_pack|
desc_service|
+-----+-----+-----+-----+-----+-----+-----+-----+
| 1|2018-04-11 16:01:...|2018-04-25 12:09:...|Apple MacBook Pro...|1|1|
1|4|45|157|<h1 style="text-a...|<h2>包装清单</h2><p>M...|
<p>&nbsp;<strong>...|
| 2|2018-04-14 02:09:...|2018-04-25 11:51:...|Apple iPhone 8 Plus|3|1|
1|1|38|115|<p><img alt="" sr...|<h3>包装清单</h3><p>采...
|<p>&nbsp;<strong>...|
| 3|2018-04-14 03:03:...|2018-04-25 11:51:...|华为 HUAWEI P10 Plus|1|8|
2|1|38|115|<p><img alt="" sr...|<h3>包装清单</h3><p>手...
|<p>&nbsp;<strong>...|
| 4|2018-11-01 10:43:...|2018-11-01 10:43:...|【次日达】泰火薄充电宝手机壳无线...|
0|0|1|1|39|126|
|
| 5|2018-11-01 10:43:...|2018-11-01 10:43:...|GoPro hero7运动相机水下...|0|
0|1|2|40|132|
|
| 6|2018-11-01 10:43:...|2018-11-01 10:43:...|川宇 USB3.0高速多功能合一T...|0|
0|1|3|41|140|
|
| 7|2018-11-01 10:43:...|2018-11-01 10:43:...|腾讯听听 9420 智能音箱/音...|0|
0|1|3|42|143|
|
| 8|2018-11-01 10:44:...|2018-11-01 10:44:...|苹果XS Max无线充电器iPho...|0|
0|1|1|39|123|
|
| 9|2018-11-01 10:44:...|2018-11-01 10:44:...|behringer 百灵达X32数...|0|0|
1|3|42|146|
|
| 10|2018-11-01 10:44:...|2018-11-01 10:44:...|ESCASE 荣耀畅玩8c手机壳 ...|0|
0|1|1|39|119|
|
| 11|2018-11-01 10:44:...|2018-11-01 10:44:...|FDY FW3C 5C 蓝牙无线W...|0|0|
1|3|42|143|
|
| 12|2018-11-01 10:44:...|2018-11-01 10:44:...|摩托罗拉（Motorola）威泰...|0|
0|1|1|38|118|
|
| 13|2018-11-01 10:44:...|2018-11-01 10:44:...|沃丁（WODING）亚马逊 Am...|0|
0|1|3|42|143|
|
| 14|2018-11-01 10:44:...|2018-11-01 10:44:...|科凌（keling）F8全波段收...|0|
0|1|3|42|144|
|
| 15|2018-11-01 10:44:...|2018-11-01 10:44:...|科威盛（kevsen）Q3对讲机...|0|
0|1|1|38|118|
|
| 16|2018-11-01 10:44:...|2018-11-01 10:44:...|小米（MI）小米手环3代nfc版...|0|

```

0		1		3		43		147			
		17		2018-11-01 10:44:...		2018-11-01 10:44:...		摩托罗拉 (Motorola) MA...		0	
0		1		1		38		118			
		18		2018-11-01 10:44:...		2018-11-01 10:44:...		C&C 儿童相机数码卡通照相机玩具...		0	
0		1		2		40		128			
		19		2018-11-01 10:44:...		2018-11-01 10:44:...		全景看房婚礼旅行! 理光 THETA...		0	
0		1		2		40		128			
		20		2018-11-01 10:44:...		2018-11-01 10:44:...		欧兴 KLG5.0 对讲机 待机1...		0	
0		1		1		38		118			
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+											
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+											
only showing top 20 rows											

```
In [5]: 1 spark.sql("select * from tb_goods").select("name", "desc_detail").show()
```

		name		desc_detail	
		Apple MacBook Pro...		<h1 style="text-a...	
		Apple iPhone 8 Plus		<p><img alt="" sr...	
		华为 HUAWEI P10 Plus		<p><img alt="" sr...	
		【次日达】泰火薄快充充电宝手机壳无线...			
		GoPro hero7运动相机水下...			
		川宇 USB3.0高速多功能合一T...			
		腾讯听听 9420 智能音箱/音...			
		苹果XS Max无线充电器iPho...			
		behringer 百灵达X32数...			
		ESCASE 荣耀畅玩8c手机壳 ...			
		FDY FW3C 5C 蓝牙无线W...			
		摩托罗拉 (Motorola) 威泰...			
		沃丁 (WODING) 亚马逊 Am...			
		科凌 (keling) F8全波段收...			
		科威盛 (kevsen) Q3对讲机...			
		小米 (MI) 小米手环3代nfc版...			
		摩托罗拉 (Motorola) MA...			
		C&C 儿童相机数码卡通照相机玩具...			
		全景看房婚礼旅行! 理光 THETA...			
		欧兴 KLG5.0 对讲机 待机1...			
+-----+-----+-----+-----+-----+-----+					
only showing top 20 rows					