hive存储过程

1、hive存储过程简介

```
1.x版本的hive中没有提供类似存储过程的功能，使用Hive做数据开发时候，一般是将一段一段的HQL语句封
装在Shell或者其他脚本中，然后以命令行的方式调用，完成一个业务或者一张报表的统计分析。好消息是，
现在已经有了Hive存储过程的解决方案（HPL/SQL –Procedural SQL on hadoop），并且在未来的
Hive的版本（2.0）中，会将该模块集成进来。该解决方案不仅支持Hive，还支持在SparkSQL，其他
NoSQL，甚至是RDBMS中使用类似于Oracle PL/SQL的功能，这将极大的方便数据开发者的工作，Hive中很
多之前比较难实现的功能，现在可以很方便的实现，比如自定义变量、基于一个结果集的游标、循环等等。
```

2、

下载：

```
官网下载地址：http://www.hplsql.org/download
0.3.31版本下载地址：http://www.hplsql.org/downloads/hplsql-0.3.31.tar.gz
```

安装hplsql：

```
[root@hadoop01 home]# tar -zxvf /home/hplsql-0.3.31.tar.gz -C /usr/local/


[root@hadoop01 home]# cd /usr/local/hplsql-0.3.31/
```

配置hplsql：

```
vi ./hplsql
内容如下:

#!/bin/bash

export "HADOOP_CLASSPATH=$HADOOP_CLASSPATH:$HADOOP_HOME/lib/*"
export "HADOOP_CLASSPATH=$HADOOP_CLASSPATH:$HADOOP_CONF_DIR/"

export "HADOOP_CLASSPATH=$HADOOP_CLASSPATH:$HADOOP_HOME/share/hadoop/mapreduce/*"
export "HADOOP_CLASSPATH=$HADOOP_CLASSPATH:$HADOOP_HOME/share/hadoop/mapreduce/lib/*"

export "HADOOP_CLASSPATH=$HADOOP_CLASSPATH:$HADOOP_HOME/share/hadoop/hdfs/*"
export "HADOOP_CLASSPATH=$HADOOP_CLASSPATH:$HADOOP_HOME/share/hadoop/hdfs/lib/*"

export "HADOOP_CLASSPATH=$HADOOP_CLASSPATH:$HADOOP_HOME/share/hadoop/yarn/*"
export "HADOOP_CLASSPATH=$HADOOP_CLASSPATH:$HADOOP_HOME/share/hadoop/yarn/lib/*"

export "HADOOP_CLASSPATH=$HADOOP_CLASSPATH:$HADOOP_HOME/share/hadoop/common/*"
```

```
export
"HADOOP_CLASSPATH=$HADOOP_CLASSPATH:$HADOOP_HOME/share/hadoop/common/lib/*"

export "HADOOP_CLASSPATH=$HADOOP_CLASSPATH:$HIVE_HOME/lib/*"
export "HADOOP_CLASSPATH=$HADOOP_CLASSPATH:$HIVE_HOME/conf"

export HADOOP_OPTS="$HADOOP_OPTS -Djava.library.path=$HADOOP_HOME/lib/native"

SCRIPTPATH=${0%/*}

java -cp $SCRIPTPATH:$HADOOP_CLASSPATH:$SCRIPTPATH/hplsql-
0.3.31.jar:$SCRIPTPATH/antlr-runtime-4.5.jar $HADOOP_OPTS
org.apache.hive.hplsql.Hplsql "$@"
```

配置hive2的服务：

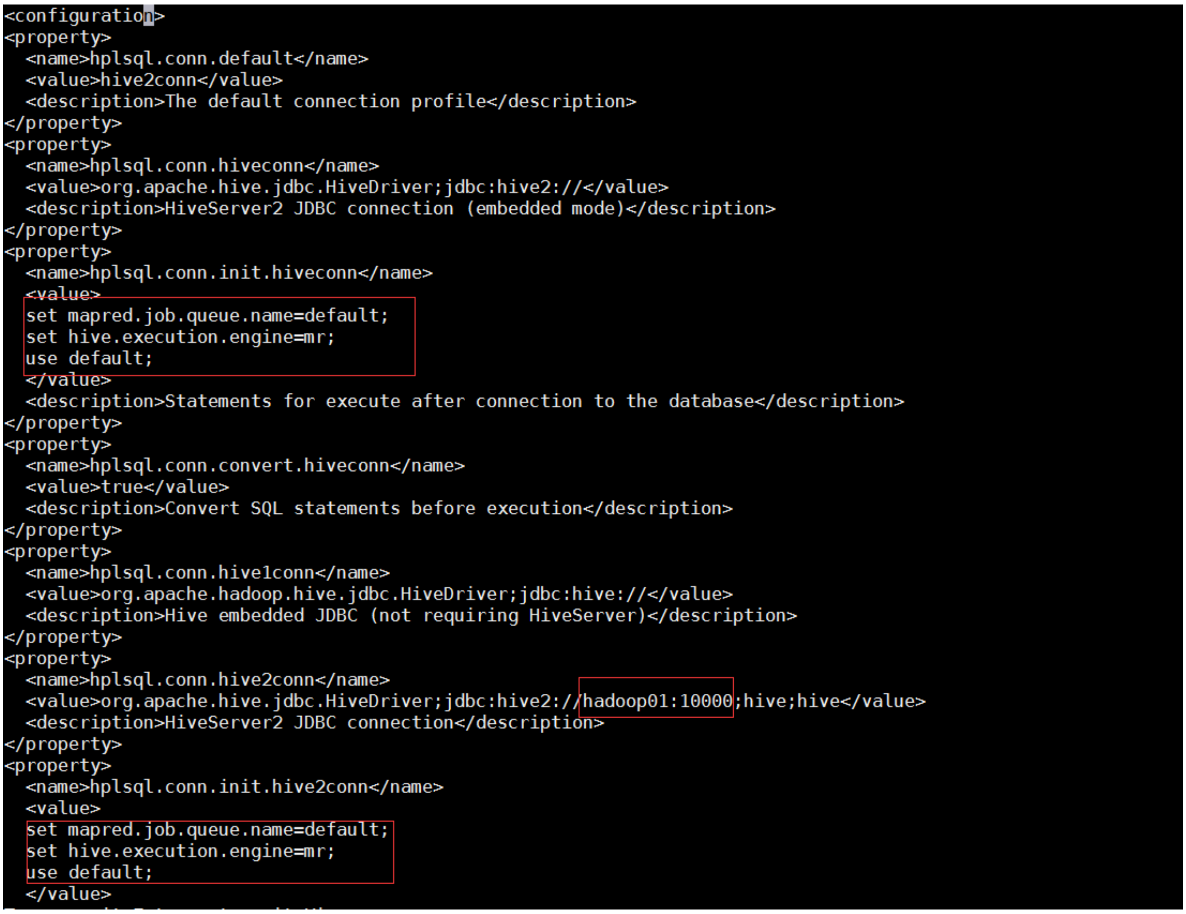
```
vi ./conf/hive-site.xml
追加内容如下：

  <property>
    <name>hive.server2.thrift.bind.host</name>
    <value>hadoop01</value>
  </property>
  <property>
    <name>hive.server2.thrift.port</name>
    <value>10000</value>
  </property>

重启hiveserver2：
nohup hive --service hiveserver2 > hiveserver2.log 2>&1 &
```

配置HPL/SQL与Hive的连接：

```
vi ./hplsql-site.xml
```



```
<configuration>
<property>
  <name>hplsql.conn.default</name>
  <value>hive2conn</value>
  <description>The default connection profile</description>
</property>
<property>
  <name>hplsql.conn.hiveconn</name>
  <value>org.apache.hive.jdbc.HiveDriver;jdbc:hive2://</value>
  <description>HiveServer2 JDBC connection (embedded mode)</description>
</property>
<property>
  <name>hplsql.conn.init.hiveconn</name>
  <value>
  set mapred.job.queue.name=default;
  set hive.execution.engine=mr;
  use default;
  </value>
  <description>Statements for execute after connection to the database</description>
</property>
<property>
  <name>hplsql.conn.convert.hiveconn</name>
  <value>true</value>
  <description>Convert SQL statements before execution</description>
</property>
<property>
  <name>hplsql.conn.hive1conn</name>
  <value>org.apache.hadoop.hive.jdbc.HiveDriver;jdbc:hive://</value>
  <description>Hive embedded JDBC (not requiring HiveServer)</description>
</property>
<property>
  <name>hplsql.conn.hive2conn</name>
  <value>org.apache.hive.jdbc.HiveDriver;jdbc:hive2://hadoop01:10000;hive;hive</value>
  <description>HiveServer2 JDBC connection</description>
</property>
<property>
  <name>hplsql.conn.init.hive2conn</name>
  <value>
  set mapred.job.queue.name=default;
  set hive.execution.engine=mr;
  use default;
  </value>
```

使用hplsql -e执行命令：

```
[root@hadoop01 hplsql-0.3.31]# ./hplsql -e "CURRENT_DATE + 1"
2019-09-18


[root@hadoop01 hplsql-0.3.31]# ./hplsql -e "for i in 1 .. 10 loop print i;end loop;"
1
2
3
4
5
6
7
8
9
10
```

hplsql -f执行脚本：

```
测试表：

hive>select * from qf24.u5;
OK
7	yy
3	cc
2	bb

创建脚本hp.sql
```

```
create function hello(text string)
returnS string
BEGIN
RETRUEN 'Hello,' || test || '!';
END;

FOR item IN(
SELECT id,name FROM qf24.u5 limit 3
)
loop
        println item.id || '|' || item.name || '|' || hello(item.name);
end loop;


测试：
[root@hadoop01 hplsql-0.3.31]# ./hplsql -f /home/hivedata/hp.sql

7|yy|Hello,!
3|cc|Hello,!
2|bb|Hello,!
```

创建存储过程的格式：

```
use database;
create procedure
begin
......
end;
```

调用存储过程:

include path/sp name

call sp name;

案例：

```
创建存储过程：

use qf24;
create procedure select_u5()
begin
select * from qf24.u5;
end;

调用：
include 文件路径   (在代码文件中引入)

执行执行文件：
[root@hadoop01 hplsql-0.3.31]# ./hplsql -f /home/hivedata/ps.sql

调用指定存储过程或者方法：
[root@hadoop01 hplsql-0.3.31]# ./hplsql -f /home/hivedata/ps.sql -main select_u5
```

参考：

https://cwiki.apache.org/confluence/pages/viewpage.action?pageId=59690156

http://www.hplsql.org/home

http://www.hplsql.org/doc

https://issues.apache.org/jira/browse/HIVE-11055

https://www.cnblogs.com/kinginme/p/7267638.html