1. 介绍

Hive自定义函数包括三种UDF、UDAF、UDTF

1. UDF(User-Defined-Function) 一进一出
2. UDAF(User- Defined Aggregation Funcation) 聚集函数，多进一出。Count/max/min
3. UDTF(User-Defined Table-Generating Functions) 一进多出，如lateral view explore)

2. 函数类型

2.1. UDF(一进一出)

1. 继承UDF类
2. 重写evaluate方法
3. 将该java文件编译成jar

2.2. UDAF(多进一出)

实现方法:

1. 用户的UDAF必须继承了org.apache.hadoop.hive.ql.exec.UDAF;

2. 用户的UDAF必须包含至少一个实现了org.apache.hadoop.hive.ql.exec的静态类，诸如实现了 UDAFEvaluator

3. 一个计算函数必须实现的5个方法的具体含义如下：

init()：主要是负责初始化计算函数并且重设其内部状态，一般就是重设其内部字段。一般在静态类中定义一个内部字段来存放最终的结果。

iterate()：每一次对一个新值进行聚集计算时候都会调用该方法，计算函数会根据聚集计算结果更新内部状态。当输 入值合法或者正确计算了，则 就返回true。

terminatePartial()：Hive需要部分聚集结果的时候会调用该方法，必须要返回一个封装了聚集计算当前状态的对象。

merge()：Hive进行合并一个部分聚集和另一个部分聚集的时候会调用该方法。

terminate()：Hive最终聚集结果的时候就会调用该方法。计算函数需要把状态作为一个值返回给用户。

4. 部分聚集结果的数据类型和最终结果的数据类型可以不同。

2.3. UDTF(一进多出)

1. 继承org.apache.hadoop.hive.ql.udf.generic.GenericUDTF

2.initialize()：UDTF首先会调用initialize方法，此方法返回UDTF的返回行的信息（返回个数，类型）

3.process：初始化完成后，会调用process方法,真正的处理过程在process函数中，在process中，每一次forward() 调用产生一行；如果产生多列 可以将多个列的值放在一个数组中，然后将该数组传入到forward()函数

4.最后close()方法调用，对需要清理的方法进行清理

> 来自: https://blog.csdn.net/weixin_42181917/article/details/82865140
> https://www.cnblogs.com/lrxvx/p/10974341.html

3. 加载函数的方式

3.1 使用add jar [classPath]语句(临时加载)

```
#加载jar
add jar [classPath]

#创建函数
create temporary function [functionName] as [calssPath]
```

这种方式不建议在生产环境中使用，通过该方式添加的jar文件只存在于当前会话中，当会话关闭后不能够继续使用该jar文件，最常见的问题是创建了永久函数到metastore中，再次使用该函数时却提示ClassNotFoundException。所以使用该方式每次都要使用add jar [classPath]语句添加相关的jar文件到classPath中。倒是可以用在临时使用函数的情况

3.2 修改hive-site.xml文件

修改参数hive.aux.jars.path的值指向UDF文件所在的路径。该参数需要手动添加到hive-site.xml文件中。

```
<property>
   <name>hive.aux.jars.path</name>
   <value>file:///[path],file:///[path]</value>
</property>
```

3.3 将jar包放入hive的jar目录(永久加载)

是在${HIVE_HOME}下创建auxlib目录，将UDF文件放到该目录中，这样hive在启动时会将其中的jar文件加载到classpath中。

> 或者 **指定jar所在目录(这种反而更管理一点),实现:**
> 可以拷贝${HIVE_HOME}/conf中的hive-env.sh.template 为 hive-env.sh 文件，并修改最后一行:

```
export HIVE_AUX_JARS_PATH=[classPath]
```

> 或者在系统中直接添加HIVE_AUX_JARS_PATH环境变量。

> 来自: https://blog.csdn.net/snail_bing/article/details/82869435