

第 7 章 函数

7.1 系统内置函数

- 1) 查看系统内置函数

```
hive> show functions;
```

- 2) 显示内置函数用法

```
hive> desc function upper;
```

- 3) 详细显示内置函数用法

```
hive> desc function extended upper;
```

7.2 自定义函数

- 1) Hive 自带了一些函数，比如：max/min 等，但是数量有限，自己可以通过自定义 UDF 来方便的扩展。

- 2) 当 Hive 提供的内置函数无法满足你的业务处理需要时，此时就可以考虑使用用户自定义函数（UDF：user-defined function）。

- 3) 根据用户自定义函数类别分为以下三种：

(1) UDF (User-Defined-Function)

一进一出

(2) UDAF (User-Defined Aggregation Function)

聚集函数，多进一出

类似于：count/max/min

(3) UDTF (User-Defined Table-Generating Functions)

一进多出

如 lateral view explore()

- 4) 官方文档地址

<https://cwiki.apache.org/confluence/display/Hive/HivePlugins>

- 5) 编程步骤：

(1) 继承 org.apache.hadoop.hive.ql.UDF

(2) 需要实现 evaluate 函数；evaluate 函数支持重载；

(3) 在 hive 的命令行窗口创建函数

a) 添加 jar

```
add jar linux_jar_path
```

b) 创建 function,

```
create [temporary] function [dbname.]function_name AS class_name;
```

(4) 在 hive 的命令行窗口删除函数

```
Drop [temporary] function [if exists] [dbname.]function_name;
```

6) 注意事项

(1) UDF 必须要有返回类型, 可以返回 null, 但是返回类型不能为 void;

7.3 自定义 UDF 函数

1) 创建一个 Maven 工程 Hive

2) 导入依赖

```
<dependencies>
  <!-- https://mvnrepository.com/artifact/org.apache.hive/hive-exec -->
  <dependency>
    <groupId>org.apache.hive</groupId>
    <artifactId>hive-exec</artifactId>
    <version>1.2.1</version>
  </dependency>
</dependencies>
```

3) 创建一个类

```
package com.atguigu.hive;
import org.apache.hadoop.hive.ql.exec.UDF;

public class Lower extends UDF {

    public String evaluate (final String s) {

        if (s == null) {
            return null;
        }

        return s.toString().toLowerCase();
    }
}
```

4) 打成 jar 包上传到服务器/opt/module/jars/udf.jar

5) 将 jar 包添加到 hive 的 classpath

```
hive (default)> add jar /opt/module/datas/udf.jar;
```

6) 创建临时函数与开发好的 java class 关联

```
hive (default)> create temporary function udf_lower as "com.atguigu.hive.Lower";
```

7) 即可在 hql 中使用自定义的函数 strip

```
hive (default)> select ename, udf_lower(ename) lowername from emp;
```