

课堂实验

一、基本信息

文档编号		文档版本	1.0
实验名称	StreamSQL 基本用法		
所属课程	第 13 讲 实时流处理引擎 Slipstream	认证等级	数据工程师（初级）
授课形式	上机实验	实验批次	第 9 次 / 共 11 次

二、实验目的

- 掌握 StreamSQL 的基本用法。
- 掌握 Slipstream 流处理的基本过程。
- 理解输入流表、衍生流表、结果表和流任务的关系。

三、实验准备

- 下载并安装 TDH Client。
- 下载并安装 Waterdrop，StreamSQL 语句在 Waterdrop 中运行。
- 实验目录与命名规划
 - (1) 本地目录
工作目录：/mnt/disk1/{student_name}
 - (2) Kafka Topic: {topic_log_student_name}
 - (3) 输入流表名: {log_student_name}
 - (4) 衍生流表名: {wlog_student_name}
 - (5) Inceptor 数据库名: {db_student_name}
 - (6) {student_name} 为变量，代表学员姓名全拼

四、实验内容

1、创建 Kafka Topic

- 任务：创建 Kafka Topic，作为输入流表的数据源。
- 步骤

Linux:

```
// 创建 Kafka Topic, 分区数为 4, 副本数为 2
1. cd {TDH_Client_install_dir}/kafka/bin
2. ./kafka-topics.sh --create --topic {topic_log_student_name} --zookeeper
   172.16.140.85:2181,172.16.140.86:2181,172.16.140.87:2181 --partitions 4 --replication-
   factor 2
// 查看 Topic 是否创建成功
3. ./kafka-topics.sh --list --zookeeper
   172.16.140.85:2181,172.16.140.86:2181,172.16.140.87:2181
```

2、创建输入流表

- 任务: 基于 Kafka Topic 创建输入流表。
- 步骤

SQL:

```
// 基于 Kafka Topic 创建输入流表 {log_student_name}
1. create stream {log_student_name} (ip string, url string, time timestamp) row format
   delimited fields terminated by ',' tblproperties( "topic"="{topic_log_student_name}",
   "kafka.zookeeper"="172.16.140.85:2181,172.16.140.86:2181,172.16.140.87:2181",
   "kafka.broker.list"="172.16.140.85:9092,172.16.140.86:9092,172.16.140.87:9092");
```

3、创建衍生流表

- 任务: 基于输入流表创建衍生窗口流表。
- 步骤

SQL:

```
// 基于输入流表创建衍生流表 {wlog_student_name}, 窗口 Length 和 Slide 为 10 秒
1. create stream {wlog_student_name} as select * from {log_student_name} streamwindow
   sw as (length '10' second slide '10' second);
```

4、创建结果表

- 任务: 创建 Inceptor 结果表, 用于存储流处理的结果。
- 步骤

SQL:

```
// 创建 Inceptor 表, 用于存储流处理结果
1. use {db_student_name};
2. create table result_table(url string, count int);
```

5、启动流任务

- 任务：将基于衍生流的统计结果写入 Inceptor 结果表，该操作启动流任务，触发流处理过程。
- 步骤

SQL:

```
1. insert into result_table select url, count(*) from {wlog_student_name} group by url;
```

6、向 Kafka Topic 发送数据

- 任务：使用 Console Producer 向 Kafka Topic 发送数据。
- 步骤

Linux:

```
1. cd {TDH_Client_install_dir}/kafka/bin
2. ./kafka-console-producer.sh --topic {topic_log_student_name} --broker-list
   172.16.140.85:9092,172.16.140.86:9092,172.16.140.87:9092
```

发送给 Kafka 的实验数据:

```
27.0.1.125,www.transwarp.io/home.html,2016-8-14 20:12:31.132
54.231.66.16,www.transwarp.io/product.html,2016-8-14 20:12:32.132
72.21.203.5,www.transwarp.io/case.html,2016-8-14 20:12:33.132
207.241.224.2,www.transwarp.io/product.html,2016-8-14 20:12:34.132
12.129.206.133,www.transwarp.io/product.html,2016-8-14 20:12:51.132
208.111.148.7,www.transwarp.io/home.html,2016-8-14 20:12:52.132
```

7、查看流处理结果

- 任务：查看 Inceptor 结果表中的流处理结果。
- 步骤

SQL:

```
1. select * from result_table;
```