

**CTSDB数据导入TBDS的示例说明**

**文档版本变更记录：**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **版本号** | **\*变化状态** | **变更说明** | **变更人** | **日  期** |
| 1 | C | 文档新建 | 胡习远 | 2018/12/20 |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

**\*修改状态：C - 创建，A - 增加，M - 修改，D - 删除**

**目 录**

[1. 概述 4](#_Toc533087769)

[2. 环境搭建与代码配置说明 4](#_Toc533087770)

[2.1. 环境搭建 4](#_Toc533087771)

[2.1.1. 代码加入TBDShadoop认证 4](#_Toc533087772)

[2.1.2. JSON相关依赖 5](#_Toc533087773)

[2.2. 配置说明 6](#_Toc533087774)

[3. 示例使用方法 7](#_Toc533087775)

[3.1. 后台执行步骤 7](#_Toc533087776)

[3.1.1. 代码打包 7](#_Toc533087777)

[3.1.2. 拷贝相关JSON包 7](#_Toc533087778)

[3.1.3. 运行demo程序 7](#_Toc533087779)

[3.2. 工作流任务调度 8](#_Toc533087780)

[3.2.1. 工作流任务参数配置 8](#_Toc533087781)

[3.2.2. 工作流任务运行 8](#_Toc533087782)

[3.2.3. 查看运行结果 9](#_Toc533087783)

1. 概述

CTSDB作为时序数据库提供了相关API用于数据查询，本文档提供了一种基于CTSDB的API取数，然后批量导入TBDS的HDFS的说明。CTSDB的相关API说明参考：

https://cloud.tencent.com/document/product/652/13611

1. 环境搭建与代码配置说明

此部分为基于java语言开发的批量提取CTSDB数据的示例，完整的代码参见：

https://github.com/TBDSUDC/TBDSDemo/tree/master/src/main/java/com/tencent/export

* 1. 环境搭建

### 代码加入TBDS hadoop认证

由于需要将CTSDB表的数据导入hdfs，需要加入hadoop的相关依赖及权限认证：

A：在pom.xml文件加入相关依赖，如下：

<dependency>

<groupId>org.apache.hadoop</groupId>

<artifactId>hadoop-client</artifactId>

<version>2.7.2-TBDS-4.0.3.3</version>

</dependency>

<dependency>

<groupId>org.apache.hadoop</groupId>

<artifactId>hadoop-common</artifactId>

<version>2.7.2-TBDS-4.0.3.3</version>

</dependency>

<dependency>

<groupId>org.apache.hadoop</groupId>

<artifactId>hadoop-hdfs</artifactId>

<version>2.7.2-TBDS-4.0.3.3</version>

</dependency>

B：在程序代码里面把四个参数添加到conf里面，如下：

conf.addResource(new Path("/etc/hadoop/conf/core-site.xml"));

conf.addResource(new Path("/etc/hadoop/conf/hdfs-site.xml"));

conf.set("hadoop.security.authentication","tbds");

conf.set("hadoop\_security\_authentication\_tbds\_secureid","a7KNzj5GhQvQBifzgshlTtyfCfUrPIisaQdb");

conf.set("hadoop\_security\_authentication\_tbds\_username","testuser");

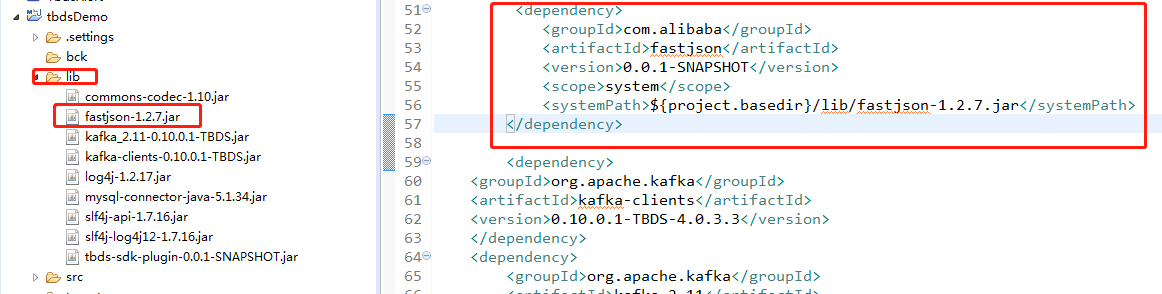
conf.set("hadoop\_security\_authentication\_tbds\_securekey","OuKjeECY9bmIZk2LVai5uRlGMoWew4lQ");

UserGroupInformation.setConfiguration( conf );

UserGroupInformation.loginUserFromSubject(null);

### JSON相关依赖

加入解析时序数据库CTSDB查询数据的JSON依赖包，fastjson-1.2.7.jar，pom.xml如下图：



* 1. 配置说明

配置参数说明如下：

##tbds 权限认证参数配置

hadoop.security.authentication=tbds

##注释掉hadoop\_security\_authentication\_tbds\_username配置时

##不走权限认证（在TBDS未开启权限认证时使用）

#hadoop\_security\_authentication\_tbds\_username=hdfs

hadoop\_security\_authentication\_tbds\_secureid=xxxx

hadoop\_security\_authentication\_tbds\_securekey=xxxx

##CTSDB数据库访问用户名

ctsdb.user=root

##CTSDB数据库访问密码

ctsdb.password=tbds123456

##CTSDB数据库访问地址

ctsdb.url=http://10.0.\*.\*:9200

##CTSDB数据导入HDFS存储的临时目录，目录不存在，会自动创建

ctsdb.dataInput=/root/kafkatest/ctsdb/

##CTSDB数据导入HDFS存储的目录，需要提前创建好

ctsdb.hdfsOutput=/ctsdb/output/

##CTSDB需要导入的数据表名称（导入HDFS）

ctsdb.tableName=ctsdb\_tbds

##CTSDB需要导入数据表的列名称

ctsdb.colNames=0001,0002

#n\*60000 一个批次查询60000条，设置多少个批次生成一个文件

#设置成5，表示5\*600000条数据生成一个文件

ctsdb.fileDataSum=5

#test 测试用，运行多少个批次查询，示例采集26\*60000条数据

ctsdb.runBatchSum=26

#10m 数据10分钟后会过期，过期后 scroll\_id将变得无效，后续查询会失败

ctsdb.validDataTime=10m

1. 示例使用方法
   1. 后台执行步骤

### 代码打包

基于之前的代码，修改完整相关配置后，打包为tbdsDemo-0.0.1-SNAPSHOT.jar

### 拷贝相关JSON包

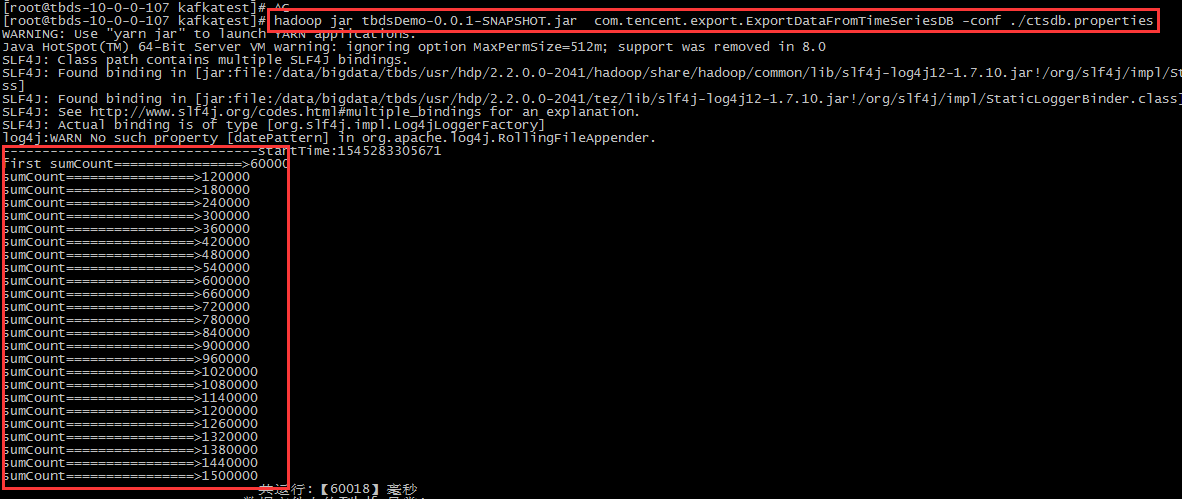
在TBDS节点上将fastjson-1.2.7.jar拷贝到/usr/hdp/2.2.0.0-2041/hadoop/share/hadoop/common/。可以利用bootstrap脚本批量进行

### 运行demo程序

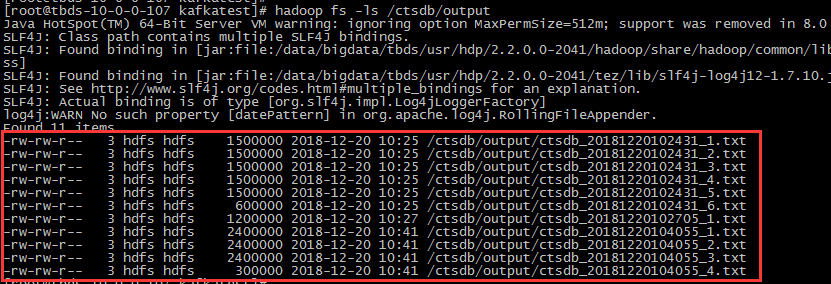
首先确认ctsdb.properties已按照配置说明修改完成，将tbdsDemo-0.0.1-SNAPSHOT.jar和ctsdb.properties拷贝到一台linux的一个目录下面，cd到这个目录，执行如下脚本：

注：如tbds未开启权限认证，需在运行脚本前导入：export HADOOP\_USER\_NAME=hdfs

运行示例如下图：



HDFS存储如下：



* 1. 工作流任务调度

### 工作流任务参数配置

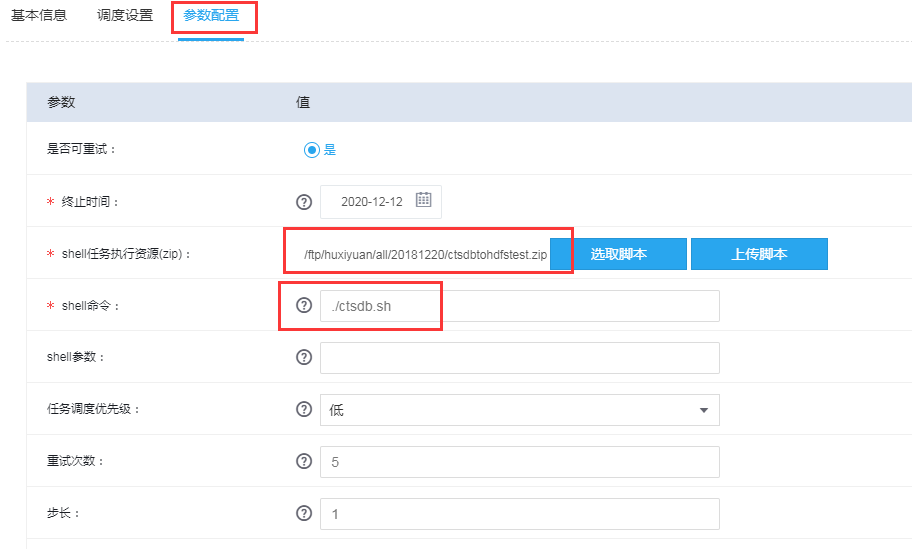
将ctsdb.properties、ctsdb.sh、tbdsDemo-0.0.1-SNAPSHOT.jar打成zip包，不要打包目录，打包文件，然后配置工作流。

注意：为了避免执行权限问题，数据目录参数ctsdb.dataInput设置在/tmp目录下，如：

ctsdb.dataInput=/tmp/kafkatest/ctsdb/

hdfs目录设置运行前设置好，如：ctsdb.hdfsOutput=/project/tbds\_test/ctsdb/output/

如下图：



### 工作流任务运行

配置工作流完成后，提交审核，审核通过后，向TBDS的工作流调度系统提交shell命令进行调度执行。

### 查看运行结果

直接在运维中心-文件管理-基础文件管理，找到目标目录查看文件如下：

