JDIP 使用手册

jdip （java dip loader）依靠 dip tserver， 接收 tserver 的数据，并将数据装载近 datahub, hdfs, hbase, kafka，activeMQ达到 dip 与大数据平台融合的功能。

jdip 配置文件分为两类：

注意：配置 xml 文件时，注意XML特殊字符的转义，修改时，请参照原有配置文件进行修改。

1. 基础配置类：

路径：jdip/config/base/config.xml

配置项：

|  |  |  |
| --- | --- | --- |
| **名称** | **示例** | **简介** |
| socketServerIp | 172.16.1.100 | tserver 地址  (dip tserver 地址) |
| socketServerPort | 7007 | tserver 端口  (dip tserver 端口) |
| datahubVersion | 1.0 | （不用修改） |
| srcId | =172.16.1.25:1521 | （不用修改） |
| dip\_xml\_compress | 1 | 是否压缩  （不用修改） |
| dip\_xml\_group | dipgroup | dip 配置链路 group  （依 dip 配置） |
| dip\_xml\_queue\_name | etlque | dip 中 etl 队列名称，配置时必须将etl 字符集改为 utf-8  （依dip 配置） |
| loader\_class\_name | com.r7data.jdip.kafka.loader.DipKafkaLoader | dip loader 装载类名称，  datahub:  com.r7data.jdip.datahub.loader  hdfs:  com.r7data.jdip.hdfs.loader.DipHdfsSingleLoader  hbase:  com.r7data.jdip.hbase.loader. DipHbaseLoader  kafka:  com.r7data.jdip.kafka.loader.DipKafkaLoader  activemq:  com.r7data.jdip.activemq.loader.DipActivemqLoader  （该配置用来设置当前使用的 loader 名称，将该项配置为对应的名称即可，同时只能配置一个loader.） |

注：无论使用哪个 loader ，该配置文件必须配置。

1. loader 配置类
   1. datahub-loader (略)
   2. hdfs-loader

配置文件1：jdip/config/hdfs/core-site.xml

该文件为 hdfs 集群的 core-site.xml 文件，直接从集群中复制到该路径下即可。

配置文件2：jdip/config/hdfs/hdfs-config.xml

|  |  |  |
| --- | --- | --- |
| **名称** | **示例** | **简介** |
| cache\_file\_base\_dir | D:/testFile/hadoop/cache | 用来缓存数据的本地目录 |
| hdfs\_file\_base\_dir | /groupData | 该路径为保存到hdfs 的目录路径 |

* 1. hbase-loader

配置文件1：jdip/config/hbase/hbase-site.xml

该文件为 hbase 集群的 hbase-site.xml 配置文件，直接从集群中赋值到该路径下即可

配置文件2：jdip/config/hbase/hbase-config.xml

暂时无需配置

* 1. kafka-loader

配置文件：jdip/config/kafka/kafka-config.xml

|  |  |  |
| --- | --- | --- |
| **名称** | **示例** | **简介** |
| dip\_kafka\_topic | r7dip | 要写入的kafka topic 名称 |
| zk\_hosts | hap02:2181,hap03:2181,hap04:2181 | zookeeper 集群地址 |
| kafka\_brokers | hap02:9092,hap03:9092,hap04:9092 | kafka 集群 borkers 地址 |
| kafka\_group\_id | testGroup | 测试用消费者 groupId |
| json\_type | 0 | kafka 写入json 格式   1. 全数据 2. 主要数据 |
| kafka\_external\_config | {  "kdc":"gaj-manager11:9093,gaj-manager12:9093",  "principal":"hdfs",  "use.global.authentication":"false",  "realm":"TDH",  "keytab":"/dip/hdfs.keytab"  } | kafka 附加参数，使用 json 格式的 key:value 对 |

* 1. activeMq

配置文件：jdip/config/activemq/activemq-config.xml

|  |  |  |
| --- | --- | --- |
| **名称** | **示例** | **简介** |
| activemq\_broker\_url | tcp://172.16.1.100:61616 | 要写入的activemq broker url 地址 |
| queue\_name | active\_dip\_que | activemq 队列名称 |
| json\_type | 0 | 同kafka |

1. 启动脚本
   1. jdip/bin/hbase-table-create.sh

启动 hbase 之前，需要创建对应的 hbase 表，执行该脚本，初始化 hbase table。

注意：如果有必要，创建表时，需设计hbase分区策略，或手动创建 hbase table.

* 1. jdip/bin/kafka-scan.sh

该启动脚本可以扫描 kafka topic 中所有的数据，从而可以监测 kafka 数据写入情况。

* 1. jdip/bin/start-client.sh

当需要启动的 loader 相关信息配置好后，就可以执行该脚本启动 loader 程序