# #flume+kafka+inceptor实时日志采集

**创建topic**

--- create topic dgnsh

在kafka角色节点上的/usr/lib/kafka/bin下面：  
cd /usr/lib/kafka/bin  
./kafka-create-topic.sh --partition 3 --topic dgnsh --zookeeper poc-px28:2181,poc-px29:2181,poc-px30:2181 --broker poc-px28,poc-px29,poc-px30

--在kafka角色节点上的/usr/lib/kafka/bin下面：   
--- list the topic  
./kafka-list-topic.sh --zookeeper poc-px28:2181,poc-px29:2181,poc-px30:2181

**创建stream**

--- create stream with SQL in inceptor(在inceptor终端里使用streamsql语法建立一个stream)

create stream dwnsh\_stream(name string,age int) row format delimited fields terminated by ','  
tblproperties("topic"="dgnsh","kafka.zookeeper"="poc-px28,poc-px29,poc-px30");

--- show the created stream(查看建立的stream)

show streams;

--- create the table for result of stream(建立一张结果表，将stream中的数据存入结果表)

create table stream.dgnsh\_table(name string, age int);

--- trigger the stream(只有insert动作可以出发stream)

insert into dgnsh\_table select name,age from dwnsh\_stream;

--- check the streamjobs(查看正在运行的stream)

list streamjobs

安装配置flume

--安装flume服务  
yum install flume

--配置flume，将数据传输的kafka中  
vim /usr/lib/flume/bin/dgnsh.conf

---内容如下  
agent1.sources=source1  
agent1.sinks=sink1  
agent1.channels=channel1

**scource1**

agent1.sources.source1.type=spooldir  
agent1.sources.source1.spoolDir=/home/test/flume/dgnsh  
agent1.sources.source1.channels=channel1  
agent1.sources.source1.fileHeader=false

**sink1**

agent1.sinks.sink1.type=org.apache.flume.sink.kafka.KafkaSink  
agent1.sinks.sink1.channel=channel1  
agent1.sinks.sink1.brokerList=poc-px28:9092,poc-px29:9092,poc-px30:9092  
agent1.sinks.sink1.topic=dgnsh

**channel1**

agent1.channels.channel1.type=memory  
a1.channels.c1.capacity = 1000  
a1.channels.c1.transactionCapacity = 100

---启动flume-agent  
bin/flume-ng agent --conf conf -f dgnsh.conf -Dflume.root.logger=DEBUG,console -n agent1

**验证flume+kafka+inceptor**

------创建一个文件，并将改文件移动到flume监控的日志下  
vim test.txt  
zhangsan,20  
lisi,30  
wangwu,40

mv test.txt /home/test/flume/dgnsh

---在inceptor中查看结果表中的数据  
select \* from stream.dgnsh\_table;

**#####流中的滑动窗口**

--创建stream  
drop stream window\_stream;  
create stream window\_stream(id timestamp,name string,age int) row format delimited fields terminated by ','  
tblproperties("timefield"="id","topic"="dgnsh","kafka.zookeeper"="poc-px28,poc-px29,poc-px30");

--"timeformat"="yyyy-MM-dd HH:mm:ss",

--创建结果表  
create table window\_table(id timestamp,name string,age int);  
--触发job  
set streamsql.use.eventtime=true;  
insert into window\_table select \* from window\_stream STREAMWINDOW w1 as (length '20' second slide '20' second);  
--验证结果表  
select \* from window\_table;