

# 자바쌤~!

https://cafe.naver.com/javassem

검색

63기-자바 & 빅데이터 >

## [ 플룸 ] 실행



자바쌤 1:1 채팅

2020.06.23. 15:43 조회 10

댓글 0 URL 복사

```
1. 저장 폴더 및 파일생성 ( dn01에서 )
( 자바에서 저장되고 플룸에서 읽을 폴더 )
mkdir -p /home/hadoop/source_data/bitthumbbitCoin
# mkdir -p /home/hadoop/source_data/coinonebitCoin
# mkdir -p /home/hadoop/source_data/upbitbitCoin
( 파일 갯수를 저장할 파일들 )
mkdir -p /home/hadoop/crawler
vi /home/hadoop/crawler/bitThumb_fileCount.txt -> 0 입력 후 저장
# vi /home/hadoop/crawler/upBit_fileCount.txt -> 0 입력 후 저장
# vi /home/hadoop/crawler/coinOne_fileCount.txt -> 0 입력 후 저장
[주의] 숫자 0만 있어야 한다
( 플룸이 읽어서 HDFS에 저장될 폴더 )
hdfs dfs -mkdir -p /user/hadoop/testInput/bitThumb
#
#
2. 플룸 파일 생성하고 실행
cd $FLUME_HOME/conf
vi flume-bitcoin.properties 파일 생성하고 붙여넣기로 생성
#-----
# 기본적인 소스, 채널, 싱크를 변수(?) 선언
agent.sources = s1
agent.channels = m1
agent.sinks = h1
# 소스를 정해진 디렉토리와 파일로 설정
# For each one of the sources, the type is defined
agent.sources.s1.type = taildir
# The channel can be defined as follows.
agent.sources.s1.channels = m1
agent.sources.s1.filegroups = f1
agent.sources.s1.filegroups.f1 = /home/hadoop/source_data/bitthumbbitCoin/bitthumbbitCoin[0-9]{1}.csv
# Each sink's type must be defined
# 싱크가 hdfs인 경우는 플룸을 거쳐 hdfs로 보낸다는 것이다
agent.sinks.h1.type = hdfs
#Specify the channel the sink should use
agent.sinks.h1.channel = m1
agent.sinks.h1.hdfs.path = hdfs://nn01:9000/user/hadoop/testInput/bitThumb
agent.sinks.h1.hdfs.writeFormat = Text
agent.sinks.h1.hdfs.rollSize = 64000000
agent.sinks.h1.hdfs.rollInterval = 0
agent.sinks.h1.hdfs.rollCount = 0
agent.sinks.h1.hdfs.batchSize = 9900
agent.sinks.h1.hdfs.fileType = DataStream
# 채널 설정
# Each channel's type is defined.
agent.channels.m1.type = memory
# Other config values specific to each type of channel(sink or source)
# can be defined as well
```

카페정보 나의활동



매니저 자바쌤  
since 2013.03.15.  
카페소개

새싹2단계

437

초대하기

즐거찾는 멤버

64명

게시판 구독수

4회

우리카페업 수

0회

카페 글쓰기

카페 채팅

즐거찾는 게시판

전체글보기

4,047

KOSMO

63기-자바 & 빅데이터

팁

- 심심문제
- 기타등등
- JAVA
- Oracle/MySQL
- JSP
- WebUI
- 파이썬
- Spring
- 리눅스
- 하둡+eco
- 프로젝트

공유해요

- 알고리즘도전
- 기술면접 준비하기
- 실존 면접
- 기회는 준비된자에게
- 프로젝트 팁

쉬어가기

청년아카데미

전문가과정

공지사항  
카페회칙

카페탈퇴하기

```
# In this case, it specifies the capacity of the memory channel
agent.channels.m1.capacity = 10000
agent.channels.m1.transactionCapacity = 10000
-----

flume-ng agent --conf-file $FLUME_HOME/conf/flume-bitcoin.properties --name agent -
Dflume.root.logger=INFO,console
3. jar 파일 (dn01노드에)
   jar 파일을 /home/hadoop/source에 저장 (winscp 이용)
   [dn02노드] 현재 dn01에서 플룸이 구동되기에 [dn02] ssh dn01에 접속하여 실행
   $ java -jar /home/hadoop/source/bitcoin.jar temp.MainApp
   ( pom.xml에 지정했으면 temp.MainApp 생략 가능 )
4. 확인 ( 5분정도 지켜보자 )
   [nn01노드] 현재 dn01(플룸구동) + dn02(dn01 연결하여 자바실행) 중이기에
   [nn01] ssh dn01 접속하여 실행

$ ssh dn01
$ ls /home/hadoop/source_data/bitthumbbitCoin ( tmp 파일 생성확인-> 몇 분후 csv 파일 생성확인 )
csv 파일이 생기면 hdfs에서 확인
$ hdfs dfs -ls /user/hadoop/testInput/bitThumb ( tmp 파일 생성확인 -> 시간지나면 .tmp 없어짐)
` tmp로 임시로 저장하다가 완료되면 csv로 변경된 후에 HDFS에 올라간다.
` 웹 브라우저 http://192.168.56.101:50070 > Utilities > Browser the file system
/ > user > hadoop > testInput > bitThumb
[dn01] 플룸 콘솔에 complete 뜨면 웹 브라우저에서 새로고침으로 확인 ( 시간소요 )
( tmp 파일이 만들어지면 웹에서 보이긴 하지만 최종적인 csv 확인 )

6. 마무리
` HDFS에 쌓인 파일을 지운다
```

## 댓글

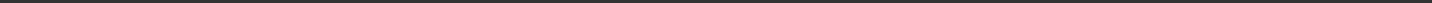
댓글알림

김용선

댓글을 남겨보세요



등록



자바샘~!    <https://cafe.naver.com/javassem>