빅데이터 적재 I

- 파일럿 실행

적재 파일럿 실행 1단계 - 적재 아키텍처

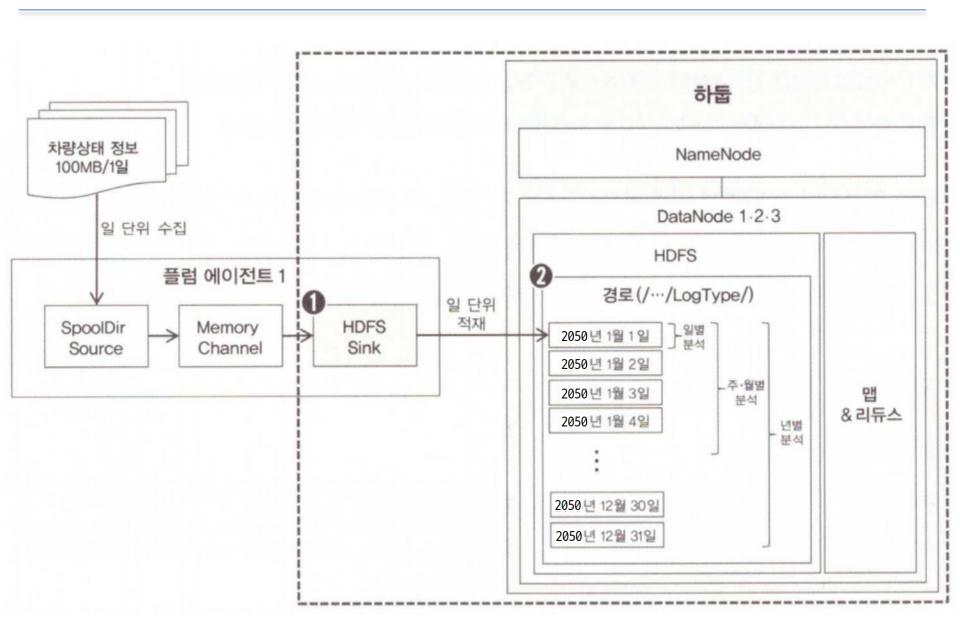
▶ 적재 요구 사항

- 요구사항1: 차량의 다양한 장치로부터 발생하는 로그 파일을 수집해서 기능별 상태를 점검.
- 요구사항2 : 운전자의 운행 정보가 담긴 로그를 실시간으로 수집해서 주행 패턴을 분석.

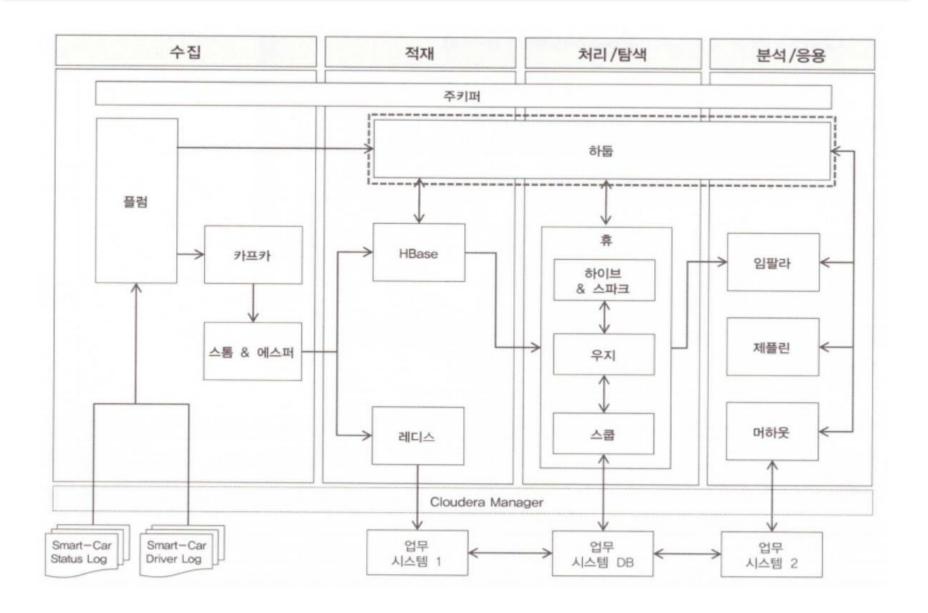
▶ 요구사항 구체화 및 분석(요구사항1)

적재 요구사항 구체화	분석 및 해결 방안
1. 100대에 달하는 스마트카들의 상태 정보가 일 단위로 취합 되어 제공된다.	플럼에서 수집 발생 시점의 날짜를 HdfsSink에 전달해서 해당 날짜 단위로 적재
2. 매일 100대의 스마트카 상태 정보는 약 100MB 정도이며, 220만 건의 상태 정보가 발생한다.	1년 적재 시 8억 건 이상의 데이터가 적재되며, 연 단위 분석에 하둡의 분산 병렬 처리 사용
3. 스마트카의 상태 정보 데이터의 발생일과 수집/적재되는 날짜가 다를 수 있다.	수집/적재되는 모든 데이터마다 데이터 발생일 외에 수집/적자 처리돼야 하는 처리일을 추가
4. 적재된 스마트카들의 상태 정보를 일 · 월 · 년 단위로 분석 할 수 있어야 한다.	HDFS에 수집 일자별로 디렉터리 경로를 만들어서 적재
5. 적재 및 생성되는 파일은 HDFS의 특성을 잘 고려해야 한다.	플럼의 HdfsSink의 옵션을 파일럿 프로젝트의 HDFS에 최적 화해서 설정
6. 적재가 완료된 후에는 원천 파일이 삭제돼야 한다.	플럼의 Source 컴포넌트 중 SpoolDir의 DeletePolicy 옵션을 활용

적재 아키텍처

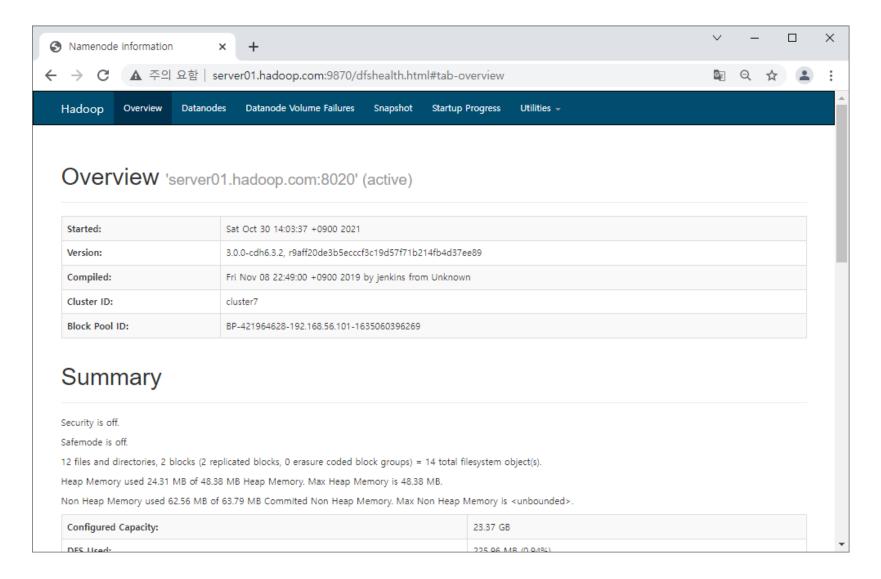


적재 파일럿 실행 2단계 – 적재 환경 구성



하둡의 상태 정보 확인

URL: http://server01.hadoop.com:9870



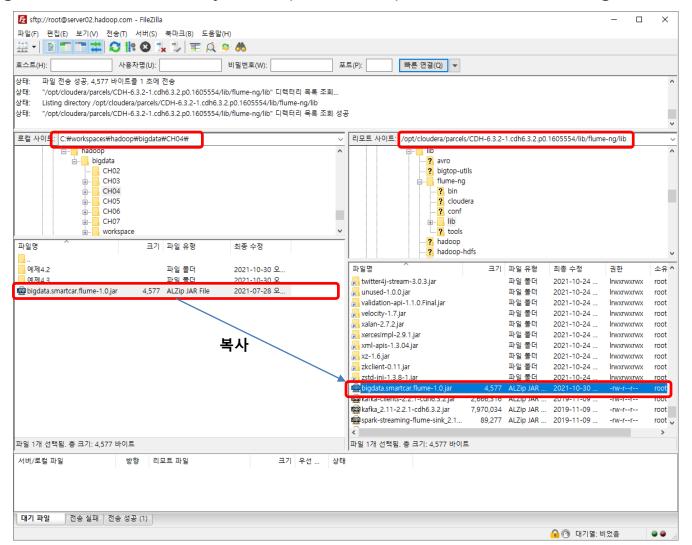
적재 파일럿 실행 3단계 - 적재 기능 구현

```
SmartCar_Agent.sources = SmartCarInfo_SpoolSource DriverCarInfo TailSource
SmartCar_Agent.channels = SmartCarInfo_Channel DriverCarInfo_Channel
SmartCar_Agent.sinks
                    = SmartCarInfo_LoggerSink DriverCarInfo_KafkaSink
                                         ① SmartCarInfo_HdfsSink로 교체
SmartCar_Agent.sinks.SmartCarInfo_LoggerSink.type = logger
                                                      2 HdfsSink로 교체
SmartCar_Agent.sources.SmartCarInfo_SpoolSource.channels = SmartCarInfo_Channel

    HdfsSink

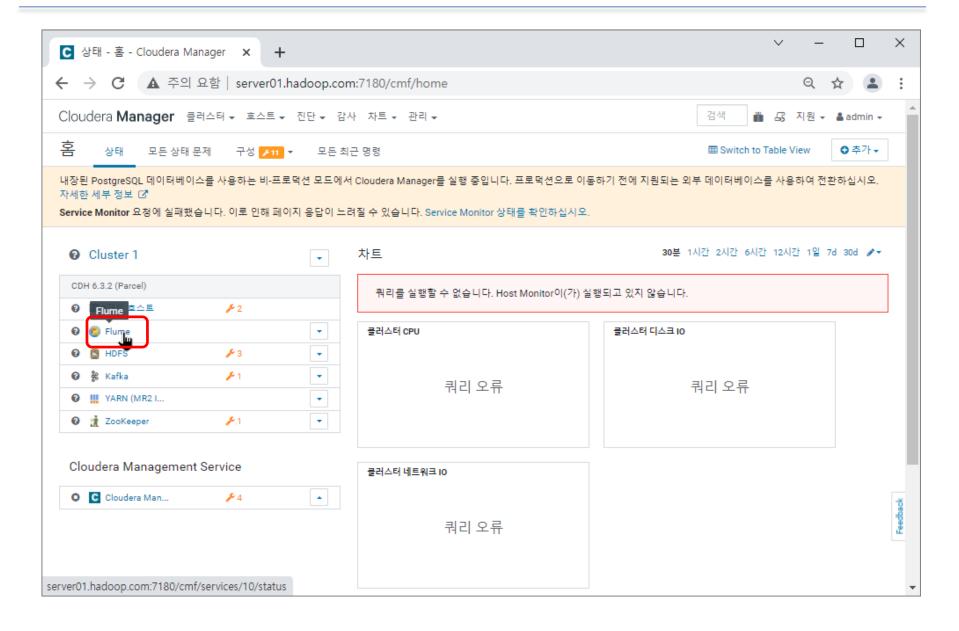
SmartCar_Agent.sinks.SmartCarInfo_LoggerSink.channel = SmartCarInfo_Channel
                                                                       Channel로 교체
```

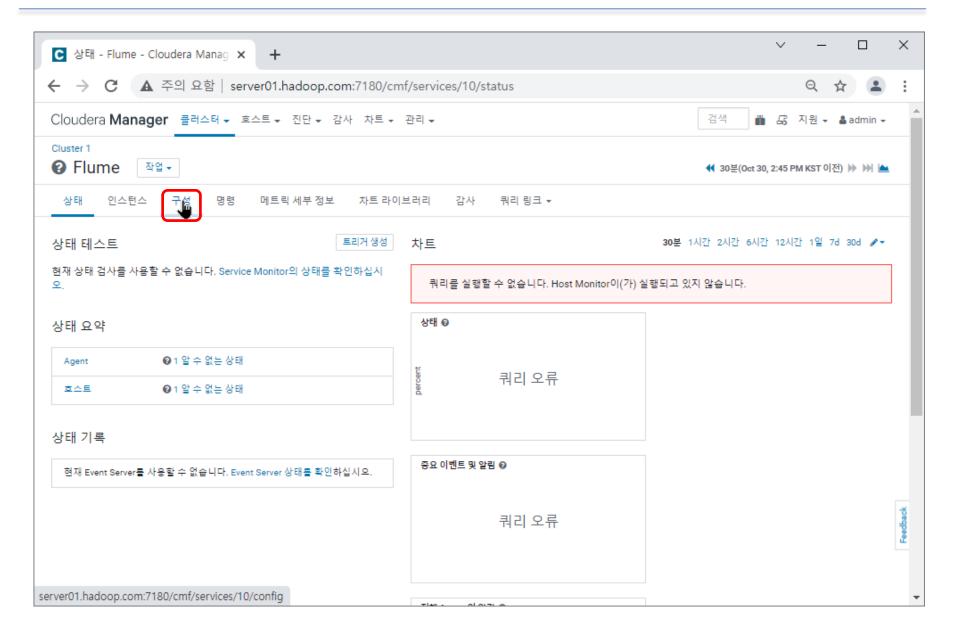
- ➤ 플럼의 사용자 정의 Interceptor 추가
 - 플럼 라이브러리 경로 : sftp://server02.hadoop.com/opt/cloudera/parcels/CDH/lib/flume-ng/lib
 - bigdata.smartcar.flume-1.0.jar -> /opt/cloudera/parcels/CDH/lib/flume-ng/lib 경로에 업로드

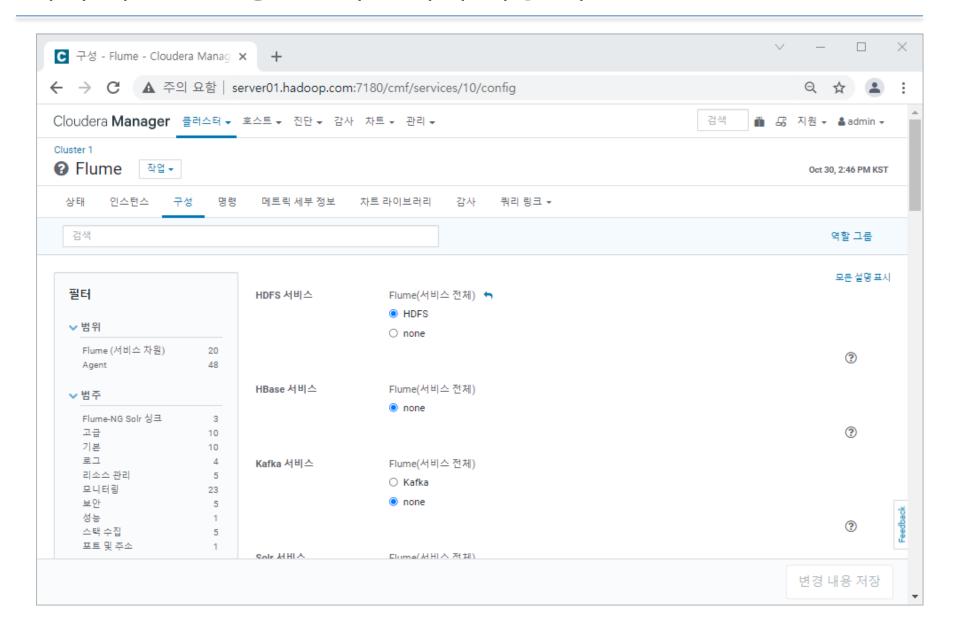


➤ 플럼의 Conf 파일 수정

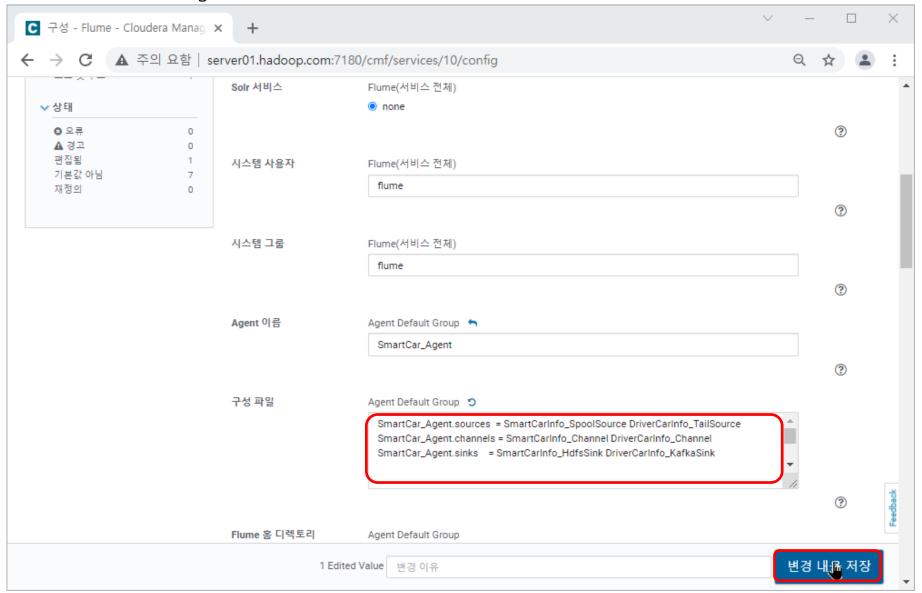
- CM 홈 -> [Flume] -> [구성] 을 선택해서, 예제4.2 SmartCar_Agent.conf 파일을 열어 전체 내용을 복사 후 "구성파일" 항목에 붙여 넣기를 한다.
- 플럼의 [변경 내용 저장] 버튼 클릭.
- CM 홈으로 다시 이동해서 플럼 재시작.

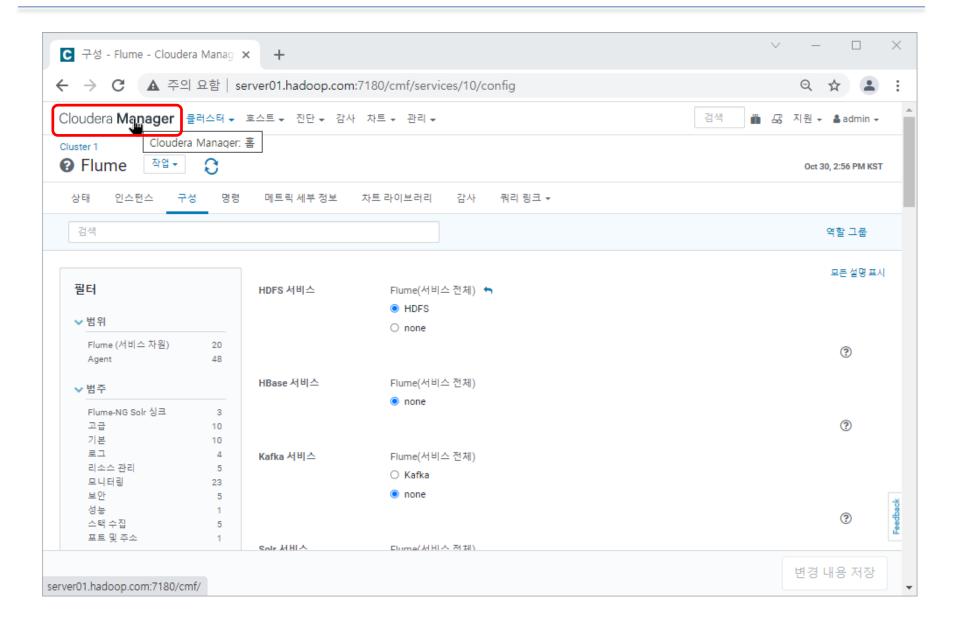


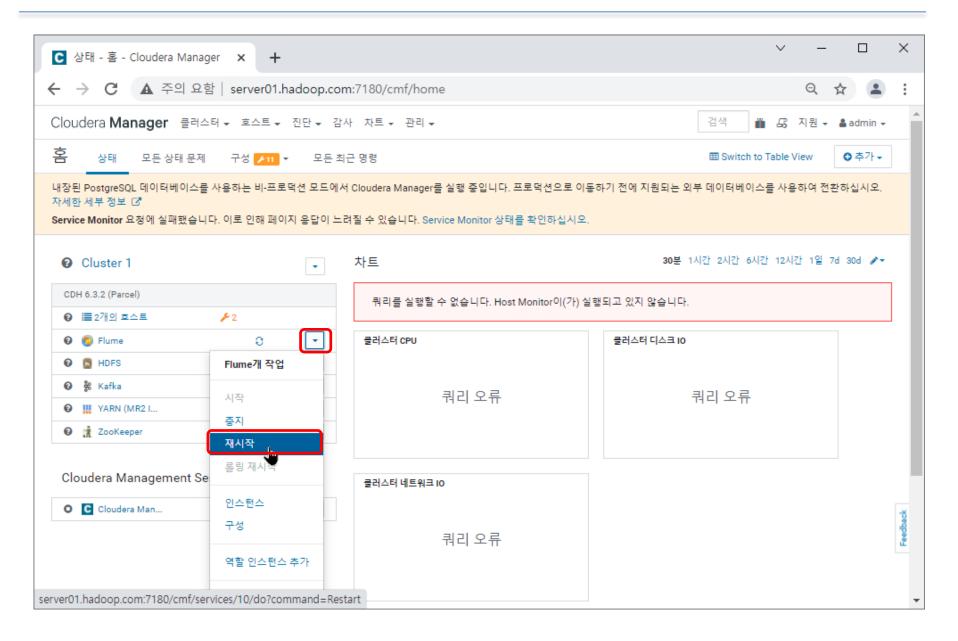


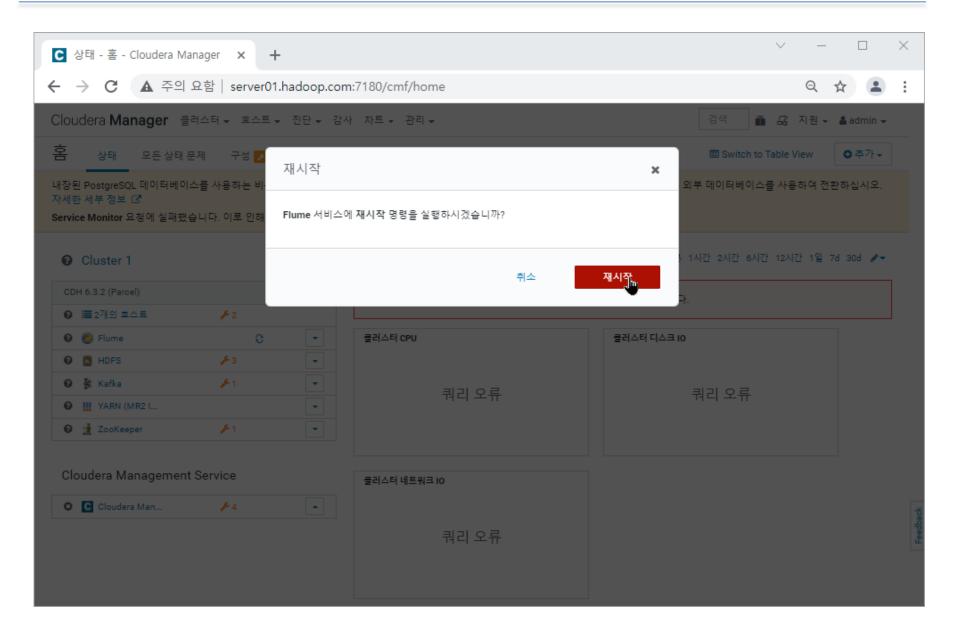


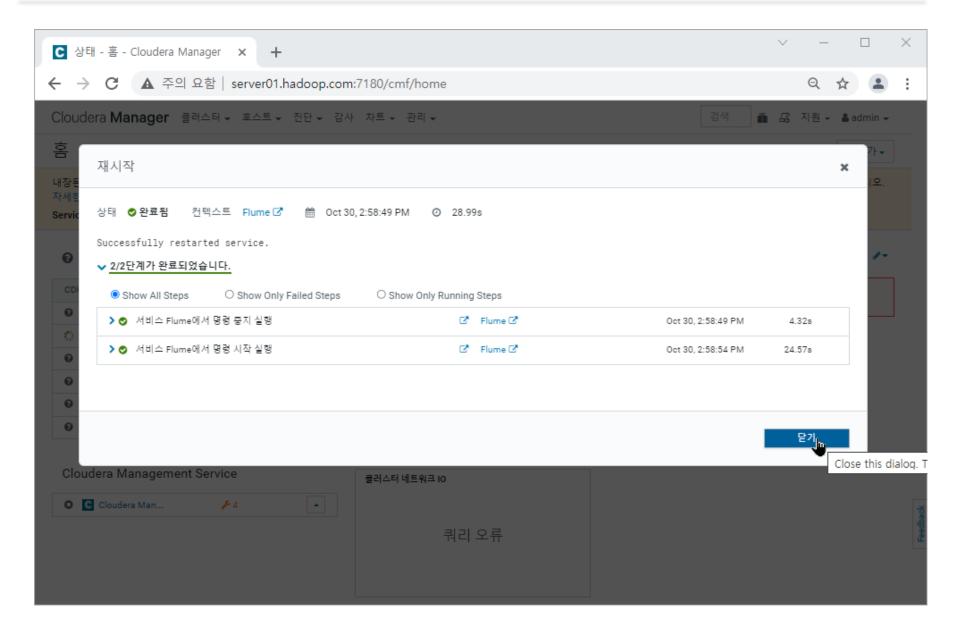
- 예제4.2 SmartCar_Agent.conf 파일을 열어 전체 내용을 복사 후, "구성파일" 항목에 붙여 넣기



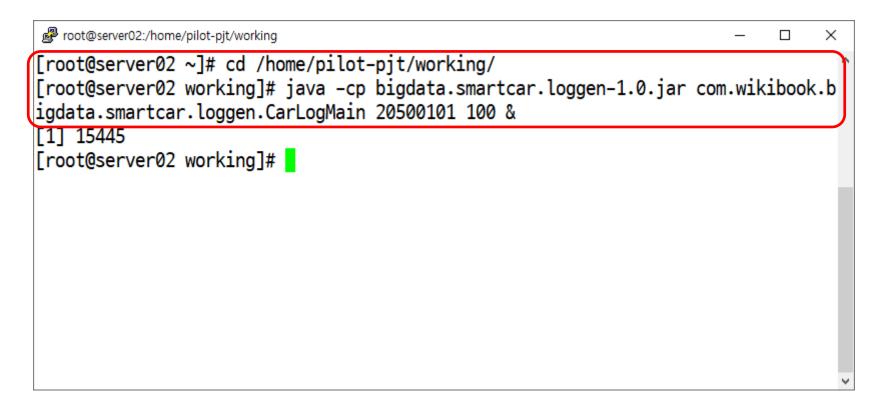








- ▶ SmartCar 로그 시뮬레이터 작동
 - # cd /home/pilot-pjt/working
 - 2. # java –cp bigdata.smartcar.loggen-1.0.jar com.wikibook.bigdata.smartcar.loggen.CarLogMain 20xxxxxx(실행일자) 100 &



➤ SmartCar 로그 시뮬레이터 작동

: /home/pilot-pjt/working/SmartCar 경로에 SmartCarStatusInfo_실행일자.txt 파일이 생성됐는지 확인

- 1. # cd /home/pilot-pjt/working/SmartCar
- 2. # tail -f SmartCarStatusInfo_실행일자.txt

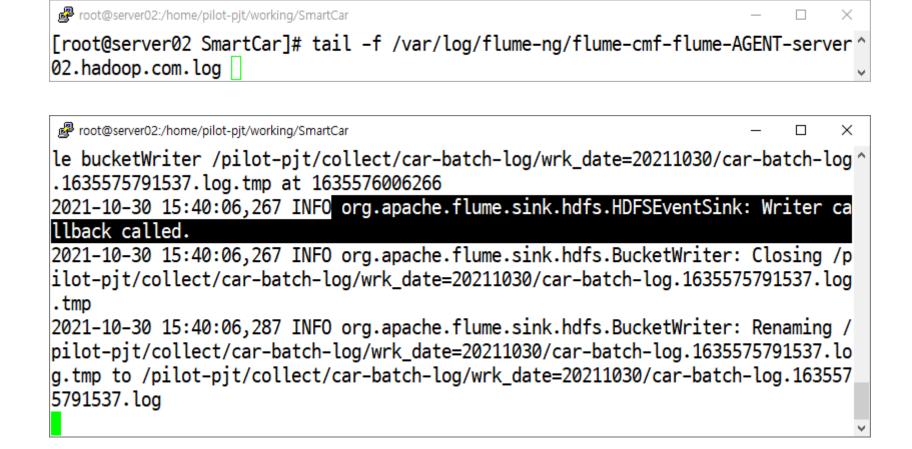
```
root@server02:/home/pilot-pit/working/SmartCar
                                                                                 ×
[root@server02 ~]# cd /home/pilot-pjt/working/
[root@server02 working]# java -cp bigdata.smartcar.loggen-1.0.jar com.wikibook.b
igdata.smartcar.loggen.CarLogMain 20500101 100 &
Γ11 15445
[root@server02 working]# cd /home/pilot-pjt/working/SmartCar/
[root@server02 SmartCar]# ls]
SmartCarStatusInfo 20500101.txt
[root@server02 SmartCar]# tail -f SmartCarStatusInfo_20500101.txt]
20500101235924,X0097,76,80,89,93,1,1,1,1,A,A,100
20500101235928,X0097,84,91,87,79,1,1,1,1,A,A,57
20500101235932,X0097,89,93,93,84,1,1,1,1,A,A,82
20500101235936,X0097,83,79,92,100,1,1,1,1,A,A,81
20500101235940,X0097,94,92,81,99,1,1,1,1,B,A,82
20500101235944,X0097,75,98,87,76,1,1,1,1,A,A,81
20500101235948,X0097,70,97,96,95,1,1,1,1,A,B,91
20500101235952,X0097,94,78,74,93,1,1,1,1,A,A,95
20500101235956,X0097,90,87,98,97,1,1,1,1,B,A,93
20500101240000,X0097,79,99,94,90,1,1,1,1,A,B,90
```

▶ 플럼 이벤트 작동

- /home/pilot-pjt/working/SmartCar 경로에 만들어진 SmartCarStatusInfo_실행일자.txt 파일을 플럼의 SmartCarInfo의 SpoolDir 경로인 /home/pilot-pjt/working/car-batch-log로 옮겨서 플럼의 File 이벤트가 작동되게 하자.
 - mv /home/pilot-pjt/working/SmartCar/SmartCarStatusInfo_ 실행일자.txt /home/pilotpjt/working/car-batch-log/

```
root@server02:/home/pilot-pjt/working/SmartCar
20500101235944,X0097,75,98,87,76,1,1,1,1,A,A,81
20500101235948,X0097,70,97,96,95,1,1,1,1,A,B,91
20500101235952,X0097,94,78,74,93,1,1,1,1,A,A,95
20500101235956,X0097,90,87,98,97,1,1,1,1,B,A,93
20500101240000,X0097,79,99,94,90,1,1,1,1,A,B,90
^C
[root@server02 SmartCar]# kill -9 15445
[root@server02 SmartCar]#
[1]+ Killed
                               java -cp bigdata.smartcar.loggen-1.0.jar com.wikib
ook.bigdata.smartcar.loggen.CarLogMain 20500101 100 (wd: /home/pilot-pjt/workin
g)
(wd now: /home/pilot-pjt/working/SmartCar)
[root@server02 SmartCar]# mv /home/pilot-pjt/working/SmartCar/SmartCarStatusInfo
_20500101.txt /home/pilot-pjt/working/car-batch-log/
[root@server02 SmartCar]#
```

- ▶ 플럼 이벤트 작동
 - 플럼의 실행 로그를 통해 SmartCarInfo_Agent가 정상적으로 작동하는지 확인.
 - tail -f /var/log/flume-ng/flume-cmf-flume-AGENT-server02.hadoop.com.log



▶ HDFS 명령어 확인

- # hdfs dfs –ls –R /pilot-pjt/collect/car-batch-log/

```
[root@server02 SmartCar] hdfs dfs -ls -R /pilot-pjt/collect/car-batch-log/drwxr-xr-x - flume supergroup 0 2021-10-30 16:08 /pilot-pjt/collect/car-batch-log/wrk_date=20211030 -rw-r--r- 1 flume supergroup 68303553 2021-10-30 15:37 /pilot-pjt/collect/car-batch-log/wrk_date=20211030/car-batch-log.1635575791536.log -rw-r--r- 1 flume supergroup 55167400 2021-10-30 15:40 /pilot-pjt/collect/car-batch-log/wrk_date=20211030/car-batch-log.1635575791537.log [root@server02 SmartCar] hdfs dfs -cat /pilot-pjt/collect/car-batch-log/wrk_date=20211030/car-batch-log.1635575791536.log
```

```
20500101165540,W0015,99,96,94,89,1,1,1,1,A,A,93,20211030
20500101165544,W0015,77,97,95,96,1,1,1,1,A,A,89,20211030
20500101165548,W0015,95,99,85,100,1,1,1,1,A,A,100,20211030
20500101165552,W0015,71,96,76,82,1,1,1,1,A,A,99,20211030
20500101165556,W0015,95,73,92,75,1,1,1,1,A,B,81,20211030
20500101165600,W0015,71,80,86,90,1,1,1,1,A,B,99,20211030
20500101165604,W0015,75,98,94,94,1,1,1,1,A,A,87,20211030
20500101165608,W0015,83,94,85,95,1,1,1,1,A,B,81,20211030
20500101165610,W0015,83,79,95,77,100,88,1,1,1,1,B,A,70,20211030
20500101165610,W0015,83,72,95,72,1,1,1,1,B,A,91,20211030
20500101165610,W0015,88,72,95,72,1,1,1,1,B,A,91,20211030
[root@server02 SmartCar]#
```

➤ HDFS 명령어 확인

```
root@server02:/home/pilot-pjt/working/SmartCar
                                                                                                                                      Х
[root@server02 SmartCar]# hdfs dfs -tail /pilot-pjt/collect/car-batch-log/wrk_date=20211030/car-batch-log.1635575791536.log
1165508, W0015, 70, 71, 88, 98, 1, 1, 1, 1, B, A, 88, 20211030
20500101165512,W0015,80,80,95,88,1,1,1,1,A,A,80,20211030
20500101165516,W0015,91,84,76,91,1,1,1,1,A,B,98,20211030
20500101165520,W0015,100,97,77,99,1,1,1,1,A,A,95,20211030
20500101165524,W0015,89,100,95,90,1,1,1,1,A,A,91,20211030
20500101165528,W0015,83,83,86,88,1,1,1,1,A,B,84,20211030
20500101165532,W0015,90,91,73,91,1,1,1,1,A,A,99,20211030
20500101165536,W0015,99,96,92,89,1,1,1,1,A,A,93,20211030
20500101165540,W0015,93,96,94,89,1,1,1,1,A,B,90,20211030
20500101165544,W0015,77,97,95,96,1,1,1,1,A,A,89,20211030
20500101165548,W0015,95,99,85,100,1,1,1,1,A,A,100,20211030
20500101165552,W0015,71,96,76,82,1,1,1,1,A,A,99,20211030
20500101165556,W0015,96,73,92,75,1,1,1,1,A,B,81,20211030
20500101165600,W0015,71,80,86,90,1,1,1,1,A,B,99,20211030
20500101165604,W0015,75,98,94,94,1,1,1,1,A,A,87,20211030
20500101165608,W0015,83,94,85,95,1,1,1,1,A,B,81,20211030
20500101165612,W0015,99,77,100,88,1,1,1,1,B,A,70,20211030
20500101165616, W0015, 88, 72, 95, 72, 1, 1, 1, 1, B, A, 91, 20211030
[root@server02 SmartCar]#
```