

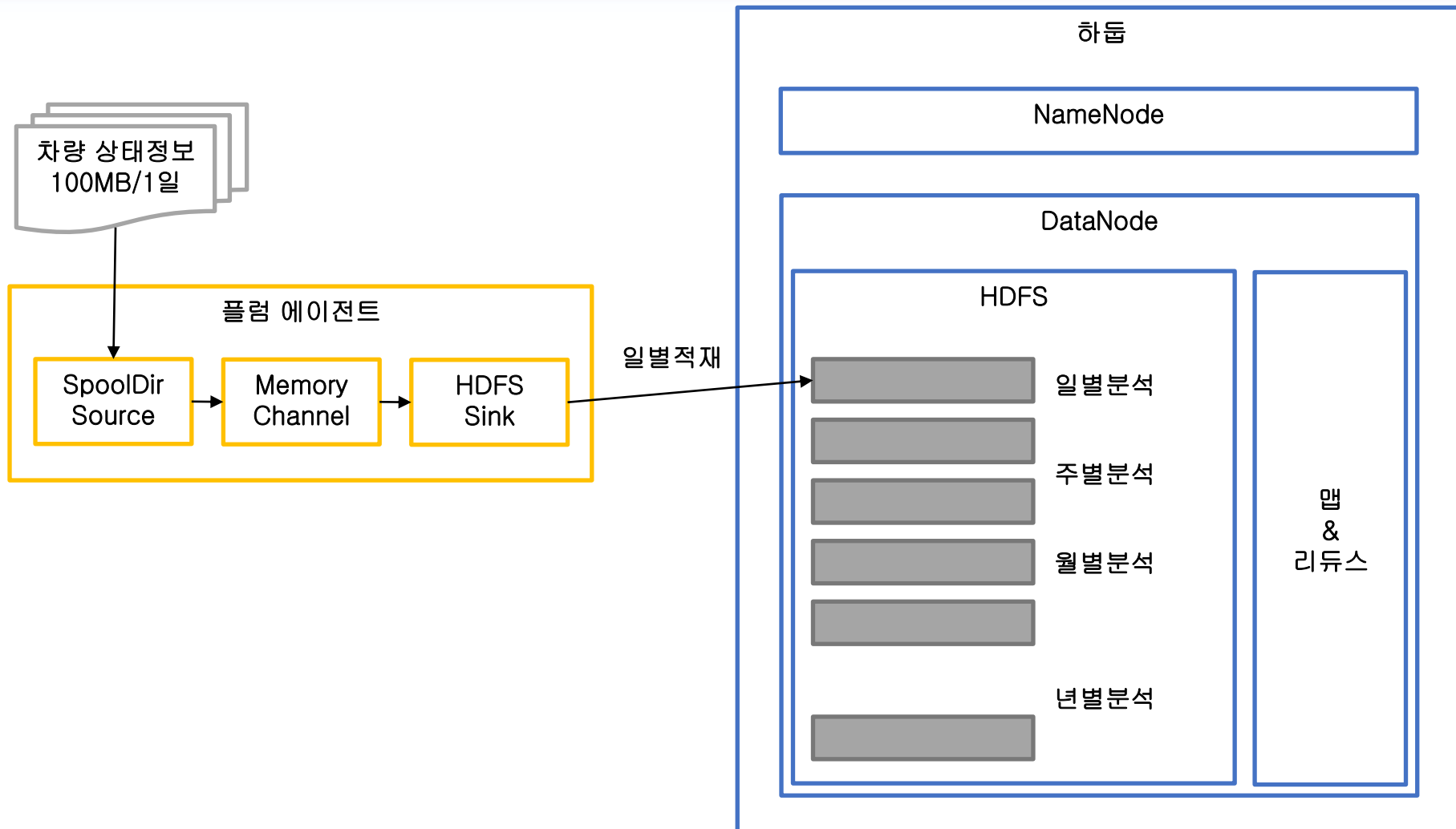
4강

빅데이터 적재



적재 아키텍처

아키텍처



적재 기능 구현

HDFS Sink 정보 설정

SmartCar_Agent.sinks = SmartCarInfo_LoggerSink DriverCarInfo_KafkaSink

SmartCar_Agent.sinks = SmartCarInfo_HdfsSink DriverCarInfo_KafkaSink

인터셉터 추가

- ❖ `SmartCar_Agent.sources.SmartCarInfo_SpoolSource.interceptors = timeInterceptor typeInterceptor collectDayInterceptor filterInterceptor`
 - 타임스탬프 활용을 위한 `timeInterceptor`
 - 로그 유형 상수값 지정을 위한 `typeInterceptor`
 - 수집일자 추가를 위한 `collectDayInterceptor`
- ❖ **timeInterceptor 인터셉터 설정**
 - 플럼 이벤트 헤더에 현재 타임스탬프 설정, 필요시 헤더에서 읽어 활용
 - `SmartCar_Agent.sources.SmartCarInfo_SpoolSource.interceptors.timeInterceptor.type = timestamp`
 - `SmartCar_Agent.sources.SmartCarInfo_SpoolSource.interceptors.timeInterceptor.preserveExisting = true`

인터셉터 추가

❖ typeInterceptor 인터셉터 설정

- logType 상수 선언 , 값은 car-batch-log
- SmartCar_Agent.sources.SmartCarInfo_SpoolSource.interceptors.typeInterceptor.type = static
- SmartCar_Agent.sources.SmartCarInfo_SpoolSource.interceptors.typeInterceptor.key = logType
- SmartCar_Agent.sources.SmartCarInfo_SpoolSource.interceptors.typeInterceptor.value = car-batch-log

❖ collectDayInterceptor 인터셉터 설정

- 이벤트 Body에 수집된 당일 작업 날짜 추가
- 기본 제공 아닌 추가 개발 사용자 정의 Interceptor => CollectDayInterceptor.java
- SmartCar_Agent.sources.SmartCarInfo_SpoolSource.interceptors.collectDayInterceptor.type =
com.wikibook.bigdata.smartcar.flume.CollectDayInterceptor\$Builder

HDFS Sink 상세 옵션 추가

❖ 설명

- 외부 수행 명령 결과를 플럼 Event로 가져와 수집
- 동적으로 path 설정
 - `.hdfs.path = /pilot-pjt/collect/{logType}/wrk_date=%Y%m%d`
- 적재시 파일명 규칙, 파일 크기등 정의

❖ 코드

- `SmartCar_Agent.sinks.SmartCarInfo_HdfsSink.hdfs.path = /pilot-pjt/collect/{logType}/wrk_date=%Y%m%d`
- `SmartCar_Agent.sinks.SmartCarInfo_HdfsSink.hdfs.filePrefix = {logType}`
- `SmartCar_Agent.sinks.SmartCarInfo_HdfsSink.hdfs.fileSuffix = .log`
- `SmartCar_Agent.sinks.SmartCarInfo_HdfsSink.hdfs.fileType = DataStream`
- `SmartCar_Agent.sinks.SmartCarInfo_HdfsSink.hdfs.writeFormat = Text`
- `SmartCar_Agent.sinks.SmartCarInfo_HdfsSink.hdfs.batchSize = 10000`
- `SmartCar_Agent.sinks.SmartCarInfo_HdfsSink.hdfs.rollInterval = 0`
- `SmartCar_Agent.sinks.SmartCarInfo_HdfsSink.hdfs.rollCount = 0`
- `SmartCar_Agent.sinks.SmartCarInfo_HdfsSink.hdfs.idleTimeout = 100`
- `SmartCar_Agent.sinks.SmartCarInfo_HdfsSink.hdfs.callTimeout = 600000`
- `SmartCar_Agent.sinks.SmartCarInfo_HdfsSink.hdfs.rollSize = 67108864`
- `SmartCar_Agent.sinks.SmartCarInfo_HdfsSink.hdfs.threadsPoolSize = 10`

Source와 Channel Sink 연결

- ❖ SmartCar_Agent.sources.SmartCarInfo_SpoolSource.channels = SmartCarInfo_Channel
- ❖ SmartCar_Agent.sinks.SmartCarInfo_HdfsSink.channel = SmartCarInfo_Channel

SmartCar_Agent.conf 추가

Cluster 1

● Flume

작업 ▼

상태

인스턴스

구성

명령

메트릭 세부 정보

차트 라이브러리

감사

쿼리 링크 ▼

구성

필터

▼ 범위

Flume (서비스 자원)	4
Agent	15

▼ 범주

Flume-NG Solr 싱크	0
고급	6
기본	2
로그	0

Agent 이름

Agent Default Group ↶

SmartCar_Agent

구성 파일

Agent Default Group ↶

SmartCar_Agent.sources = Smart
SmartCar_Agent.channels = Smar
SmartCar_Agent.sinks = SmartC

사용자 정의 Interceptor 추가

라이브러리 가져오기

❖ /opt/cloudera/parcels/CDH-6.3.2-1.cdh6.3.2.p0.1605554/lib/flume-ng/lib

 flume-ng-configuration-1.9.0-cdh6.3.2

 flume-ng-core-1.9.0-cdh6.3.2

 flume-ng-sdk-1.9.0-cdh6.3.2

코드 생성

```
public class CollectDayInterceptor implements Interceptor{
    @Override
    public void close() {}
    @Override
    public void initialize() {}
    private String getToDate() {
        SimpleDateFormat day = new SimpleDateFormat("yyyyMMdd");

        return day.format(new Date(System.currentTimeMillis()));
    }
    @Override
    public Event intercept(Event event) {
        String eventBody = new String(event.getBody() + "," + getToDate());
        event.setBody(eventBody.getBytes());
        return event;
    }
    @Override
    public List<Event> intercept(List<Event> events) {
        for(Event e : events)
            intercept(e);
        return events;
    }
}
```

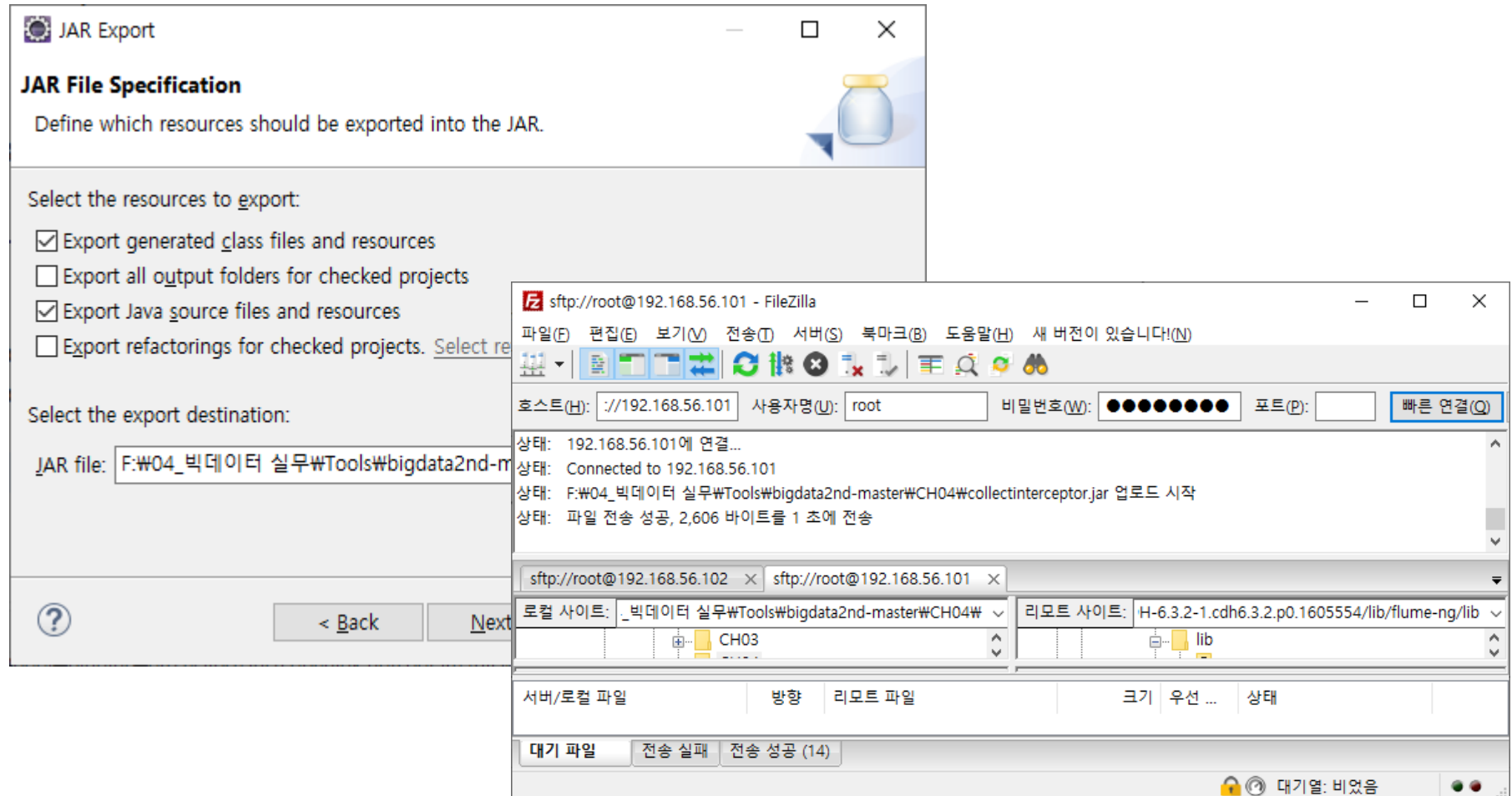
코드 생성

```
public static class Builder implements Interceptor.Builder{
    @Override
    public void configure(Context ctx) {}

    @Override
    public Interceptor build() {
        return new CollectDayInterceptor();
    }
}
```

Jar 업로드

/opt/cloudera/parcels/CDH-6.3.2-1.cdh6.3.2.p0.1605554/lib/flume-ng/lib



시뮬레이터 동작

```
[root@server02 ~]# cd /home/pilot-pjt/working/
[root@server02 working]# java -cp bigdata.smartcar.loggen-1.0.jar
com.wikibook.bigdata.smartcar.loggen.CarLogMain 20200101 100 &
[root@server02 working]# ls -l /home/pilot-pjt/working/SmartCar/
합계 81004
-rw-r--r-- 1 root root 82945641 2020-10-19 22:08
SmartCarStatusInfo_20200101.txt
[root@server02 working]# cd SmartCar/
[root@server02 SmartCar]# tail -f SmartCarStatusInfo_20200101.txt
20200101235924,Y0094,77,74,99,74,1,1,1,1,B,A,81
20200101235928,Y0094,95,87,97,85,1,1,1,1,A,A,83
20200101235932,Y0094,71,75,95,100,1,1,1,1,A,A,83
```


Flume 동작

❖ 2개의 putty 창을 띄움

❖ 1 putty

- [root@server02 ~]# tail -f /var/log/flume-ng/flume-cmf-flume-AGENT-server02.hadoop.com.log

❖ 2 putty

- [root@server02 SmartCar]# mv SmartCarStatusInfo_20200101.txt ../car-batch-log/

❖ 동작

- 2 putty를 통해 log를 이동하면 flume에서 인식하여 동작 시작
- 1 putty의 log 동작
- 다음과 같은 메시지 나오면 완료됨
- 2020-10-19 22:15:18,682 INFO org.apache.flume.sink.hdfs.BucketWriter:
Closing idle bucketWriter /pilot-pjt/collect/car-batch-log/wrk_date=20201019/car-batch-log.1603113175609.log.tmp at 1603113318682
- 2020-10-19 22:15:18,682 INFO org.apache.flume.sink.hdfs.HDFSEventSink:
Writer callback called

동작 확인

```
[root@server02 SmartCar]# hdfs dfs -ls -R /pilot-pjt/collect/car-batch-log
drwxr-xr-x  - flume supergroup      0 2020-10-19 22:15 /pilot-
pjt/collect/car-batch-log/wrk_date=20201019
-rw-r--r--  2 flume supergroup  68303145 2020-10-19 22:13 /pilot-
pjt/collect/car-batch-log/wrk_date=20201019/car-batch-
log.1603113175608.log
-rw-r--r--  2 flume supergroup  55237515 2020-10-19 22:15 /pilot-
pjt/collect/car-batch-log/wrk_date=20201019/car-batch-
log.1603113175609.log
[root@server02 SmartCar]# hdfs dfs -tail /pilot-pjt/collect/car-batch-
log/wrk_date=20201019/car-batch-log.1603113175608.log
0101143324,S0025,40,90,90,92,1,1,1,2,A,A,87,20201019
20200101143328,S0025,40,82,75,86,1,1,1,2,A,A,81,20201019
20200101143332,S0025,40,92,81,93,1,1,1,2,A,A,51,20201019
20200101143336,S0025,40,95,81,95,1,1,1,2,A,A,93,20201019
```

종료

```
[root@server02 SmartCar]# ps -ef | grep smartcar.log  
root    19199 19154  3 22:08 pts/1    00:00:23 java -cp  
bigdata.smartcar.loggen-1.0.jar  
com.wikibook.bigdata.smartcar.loggen.CarLogMain 20200101 100  
root    19930 19154  0 22:19 pts/1    00:00:00 grep smartcar.log  
[root@server02 SmartCar]# kill -9 19199
```