

Mysql - Kafka_Connector - SparkStreaming

1. Docker- compose를 이용하여 MySQL과 Kafka를 설치

```
version: '3'
services:
  mysql:
    image: mysql:8.0
    container_name: mysql
    ports:
      - 3306:3306
    environment:
      MYSQL_ROOT_PASSWORD: admin
      MYSQL_USER: mysqluser
      MYSQL_PASSWORD: mysqlpw
    command:
      - --character-set-server=utf8mb4
      - --collation-server=utf8mb4_unicode_ci
    volumes:
      - D:/mysql/data:/var/lib/mysql

  zookeeper:
    container_name: zookeeper
    image: wurstmeister/zookeeper
    ports:
      - "2181:2181"

  kafka:
    container_name: kafka
    image: wurstmeister/kafka
    depends_on:
      - zookeeper
    ports:
      - "9092:9092"
    environment:
      KAFKA_ADVERTISED_HOST_NAME: 127.0.0.1
      KAFKA_ADVERTISED_PORT: 9092
      KAFKA_ZOOKEEPER_CONNECT: zookeeper:2181
    volumes:
      - /var/run/docker.sock:/var/run/docker.sock
```

2. MySQL 데이터 베이스 생성

```
mysql -u root -p

create database testdb;

use testdb;

CREATE TABLE accounts (
  Title VARCHAR(255),
  Score INT(10),
  Review VARCHAR(255),
  PRIMARY KEY (Review)
);
```

```
use mysql;
//mysqluser가 추가 되어 있는지 확인
select host, user from user;

//mysqluser 없으면 생성
CREATE USER 'mysqluser'@'%' IDENTIFIED BY 'mysqlpw';

// mysqluser 에게 권한 부여
GRANT ALL PRIVILEGES ON *.* TO 'mysqluser'@'%';

FLUSH PRIVILEGES;
```

<https://debezium.io/releases/1.5/>

```
// 다운로드
debezium-connector-mysql-1.5.4.Final-plugin.tar.gz

//위치에 압축 해제
/opt/kafka_2.13-2.7.1/connectors
```

3. 카프카 컨테이너에 접속, `/opt/kafka/config/connect-distributed.properties` 파일을 수정

```
// 원래 경로
#plugin.path=

// 수정 경로
plugin.path=/opt/kafka_2.13-2.8.1/connectors

// 카프카 컨테이너에서 실행
connect-distributed.sh /opt/kafka/config/connect-distributed.properties
```

```
// 8083 포트 열려있는지 확인
curl http://localhost:8083/

// 플러그인 조회
curl --location --request GET 'localhost:8083/connector-plugins'

//io.debezium.connector.mysql.MySqlConnector가 있어야함
```

4. REST API 호출하여 Connector를 생성

```
curl --location --request POST 'http://localhost:8083/connectors' \
--header 'Content-Type: application/json' \
--data-raw '{
  "name": "source-test-connector",
  "config": {
    "connector.class": "io.debezium.connector.mysql.MySqlConnector",
    "tasks.max": "1",
    "database.hostname": "mysql",
    "database.port": "3306",
    "database.user": "mysqluser",
    "database.password": "mysqlpw",
    "database.server.id": "184054",
    "database.server.name": "dbserver1",
    "database.allowPublicKeyRetrieval": "true",
    "database.include.list": "testdb",
    "database.history.kafka.bootstrap.servers": "kafka:9092",
    "database.history.kafka.topic": "dbhistory.testdb",
    "key.converter": "org.apache.kafka.connect.json.JsonConverter",
    "key.converter.schemas.enable": "true",
    "value.converter": "org.apache.kafka.connect.json.JsonConverter",
    "value.converter.schemas.enable": "true",
    "transforms": "unwrap,addTopicPrefix",
    "transforms.unwrap.type": "io.debezium.transforms.ExtractNewRecordState",
    "transforms.addTopicPrefix.type": "org.apache.kafka.connect.transforms.RegexRouter",
    "transforms.addTopicPrefix.regex": "(.*)",
    "transforms.addTopicPrefix.replacement": "$1"
  }
}'
```

```
//커넥터의 Java 클래스
connector.class

//커넥터에 대해 생성되어야 할 태스크의 최대 수
task.max

//DB엔드포인트
database.hostname

//MySQL 인스턴스를 고유하게 식별하는데 사용하는 문자열
database.server.name

//지정한 서버에서 호스팅 하는 데이터베이스의 목록
database.include.list

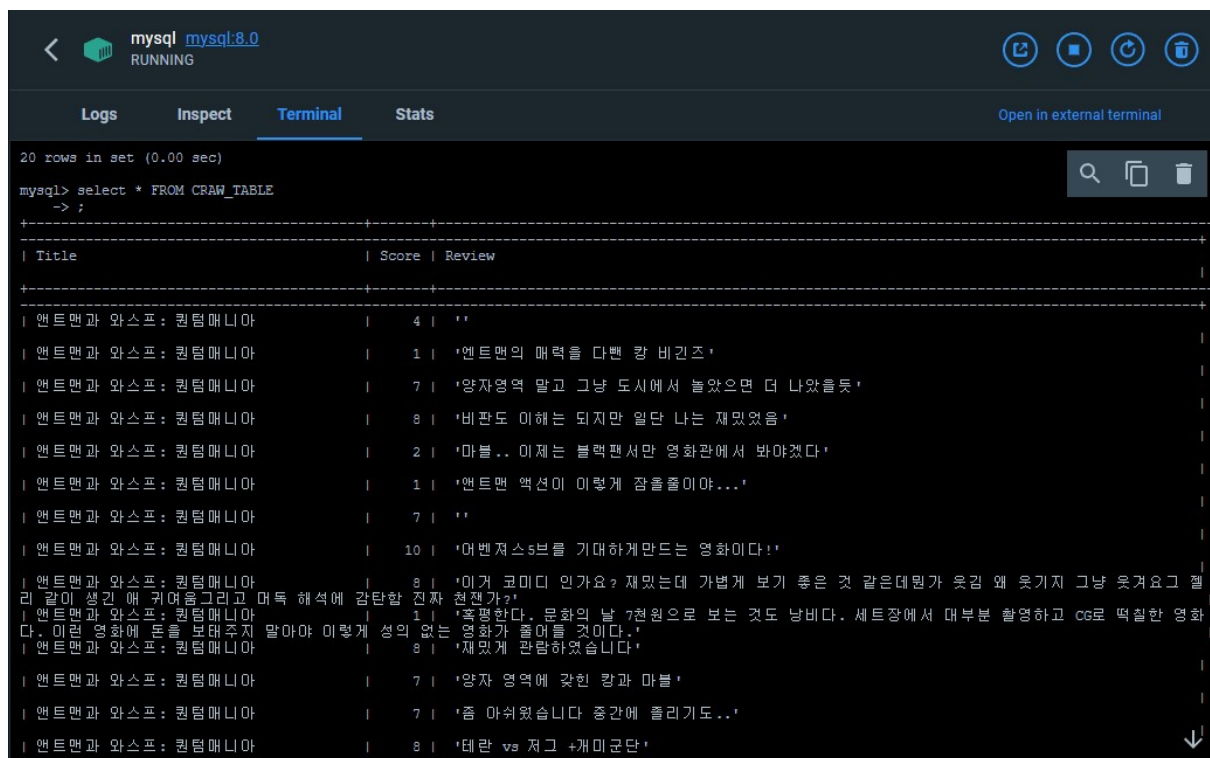
//부트스트랩의 서버 주소
```

```
database.history.kafka.bootstrap.servers
```

```
// 데이터베이스 스키마 변경을 추적하기 위해 Debezium에서 내부적으로 사용하는 Kafka 주제  
database.history.kafka.topic
```

```
# 목록 확인  
kafka-topics.sh --list --bootstrap-server localhost:9092  
  
## 분산 모드로 카프카 커넥트 실행 후 생성되는 토픽  
__consumer_offsets  
connect-configs  
connect-offsets  
connect-status  
  
## 커넥터 생성 후 생성되는 토픽  
dbhistory.testdb  
dbserver1 // DDL 정보  
dbserver1.testdb.accounts // 테이블 정보
```

5. 크롤러 작동 이후



mysql mysql:8.0
RUNNING

Logs Inspect **Terminal** Stats [Open in external terminal](#)

20 rows in set (0.00 sec)

```
mysql> select * FROM CRAW_TABLE  
-> ;
```

Title	Score	Review
엔트맨과 와스프: 퀴텀매니아	4	''
엔트맨과 와스프: 퀴텀매니아	1	'엔트맨의 매력을 다룬 칸 비긴즈'
엔트맨과 와스프: 퀴텀매니아	7	'양자영역 말고 그냥 도시에서 놀았으면 더 나았을듯'
엔트맨과 와스프: 퀴텀매니아	8	'비판도 이해는 되지만 일단 나는 재밌었음'
엔트맨과 와스프: 퀴텀매니아	2	'마블.. 이제는 블랙팬서만 영화관에서 봐야겠다'
엔트맨과 와스프: 퀴텀매니아	1	'엔트맨 액션이 이렇게 잠들줄이야...'
엔트맨과 와스프: 퀴텀매니아	7	''
엔트맨과 와스프: 퀴텀매니아	10	'어벤저스5를 기대하게 만드는 영화이다!'
엔트맨과 와스프: 퀴텀매니아	8	'이거 코미디 인가요? 재밌는데 가법게 보기 좋은 것 같은데뭔가 웃김 왜 웃기지 그냥 웃겨요그 찰리 같이 생긴 애 귀여움그리고 머독 해석에 감탄할 진짜 천재가?'
엔트맨과 와스프: 퀴텀매니아	1	'죽평한다. 문화의 날 7천원으로 보는 것도 낭비다. 세트장에서 대부분 촬영하고 CG로 찍칠한 영화다. 이런 영화에 돈을 보태주지 말아야 이렇게 성의 없는 영화가 줄어들 것이다.'
엔트맨과 와스프: 퀴텀매니아	8	'재밌게 관람하였습니다'
엔트맨과 와스프: 퀴텀매니아	7	'양자 영역에 갇힌 칸과 마블'
엔트맨과 와스프: 퀴텀매니아	7	'좀 아쉬웠습니다 중간에 졸리기도...'
엔트맨과 와스프: 퀴텀매니아	8	'데란 vs 저그 +개미군단'

BS4에서 넘어온 데이터 - MySQL

Mysql에 저장되면 Kafka_Connector로 읽어와 토픽에 저장됨

SparkStreaming으로 연결하려 하였으나...

스파크에서 실행은 잘 되는데 데이터가 넘어오질 않았다.

카프카에서 위와 같은 알람이 넘어왔는데 해결책이

- 3-node cluster
- 2 consumers
- try to perform a group coordinator change (eg kafka maintenance upgrade, or zookeeper losing connectivity to kafka node) by stopping and starting the kafka nodes
- consumer will usually get reassigned to the new group coordinator, but not always
- repeat the process above until 1 of the consumers is not assigned

[대충 이유]

그렇기 때문에 카프카에서 **Flatfile Sink**로 드랍하기로 했다.

```

echo '{
  "name" : "Flatfile_Sink",
  "config" : {
    "connector.class" : "org.apache.kafka.connect.file.FileStreamSinkConnector",
    "file" : "/root/test.txt",
    "topics" : "dbserver1.testdb.accounts"
  }
}' | curl -X POST -d @- http://localhost:8083/connectors --header "content-Type:application/json"

```

```

1 Struct{Title=TESdT3,Score=3,Review=TEST143}
2 Struct{Title=TESdT3,Score=3,Review=TEST1443}
3 Struct{Title=엔트맨과 와스프: 권텔메니아,Score=2,Review= 'imax로 본 사람인데 이게 찡찡 10점 줄 정도인가? 재미하나 못이기는 세계관 칭찬자라.. 설정파괴에 개연성 부족하고 심지어 이상한 개
4 Struct{Title=엔트맨과 와스프: 권텔메니아,Score=6,Review= '마블에 대한 기대치가 있는데 ... 납 기대하고 라닐타일이 간편이 아닌데 지루해요'}
5 Struct{Title=엔트맨과 와스프: 권텔메니아,Score=10,Review= '마리 큰 아저씨 나올때 둥겨 죽을뻔한 원상의 영화!'}
6 Struct{Title=엔트맨과 와스프: 권텔메니아,Score=8,Review= '다 죽던 마블을 개미가 살릴'}
7 Struct{Title=엔트맨과 와스프: 권텔메니아,Score=10,Review= '산왕 고달라할 쿠팡왕 크기 비례'}
8 Struct{Title=엔트맨과 와스프: 권텔메니아,Score=10,Review= '1점도아깝 나날뻔할진심노골'}
9 Struct{Title=엔트맨과 와스프: 권텔메니아,Score=8,Review= '개미나 키워보고싶다..'}
10 Struct{Title=엔트맨과 와스프: 권텔메니아,Score=6,Review= '화려한 배경 말고는 글썽요'}
11 Struct{Title=엔트맨과 와스프: 권텔메니아,Score=10,Review= '난 재미있는데.....'}
12 Struct{Title=엔트맨과 와스프: 권텔메니아,Score=8,Review= '그럭 저럭 볼만해요'}
13 Struct{Title=엔트맨과 와스프: 권텔메니아,Score=10,Review= '엔트맨 웃겨서 재미있음 ㅋㅋ'}
14 Struct{Title=엔트맨과 와스프: 권텔메니아,Score=1,Review= '재미없고 졸릴 뻔한전개 스토리 액션'}
15 Struct{Title=엔트맨과 와스프: 권텔메니아,Score=7,Review= '앞으로 마블 영화가 어떻게 나아갈지 궁금해졌다그러나 점점 어려워지는 내용과 등장인물 때문인지대중성을 예전만큼 기대하던 어려울듯
16 Struct{Title=엔트맨과 와스프: 권텔메니아,Score=10,Review= '평 안중았는데 진짜 블랙팬서의 100배정도 재미는듯'}
17 Struct{Title=엔트맨과 와스프: 권텔메니아,Score=2,Review= '그날 sf영화화판본거같~음'}
18 Struct{Title=엔트맨과 와스프: 권텔메니아,Score=10,Review= '오랜만에 재미는 마블 영화..'}
19 Struct{Title=엔트맨과 와스프: 권텔메니아,Score=10,Review= '재미있다 영화가 재미있다'}
20 Struct{Title=엔트맨과 와스프: 권텔메니아,Score=1,Review= '아이언맨부터 황제해운 마블 이제 보낼때가되건가싶다 '}
21 Struct{Title=엔트맨과 와스프: 권텔메니아,Score=10,Review= '너무 재미있고 캐릭터들도 개성있었습니다 ㅋㅋ'}
22 Struct{Title=엔트맨과 와스프: 권텔메니아,Score=6,Review= '빌런은 약하고 영화는 지루하다'}
23 Struct{Title=엔트맨과 와스프: 권텔메니아,Score=1,Review= '진짜 천원도 아깝다 최악'}
24 Struct{Title=엔트맨과 와스프: 권텔메니아,Score=8,Review= '재밌었어요'}
25 Struct{Title=엔트맨과 와스프: 권텔메니아,Score=5,Review= '예전의 마블은 이게 없나보다'}
26 Struct{Title=엔트맨과 와스프: 권텔메니아,Score=10,Review= '처음에 기대 안하고 봤는데 생각보다 재미있었다뒤로 갈수록 흥미진진하고 결말이 이겨서 통쾌했다! 결말 개인적으로 정말 마음에 들었다'}
27 Struct{Title=엔트맨과 와스프: 권텔메니아,Score=10,Review= '이정도면 마블의 평균치는 충분히 해냈다고 본다 최소한 블랙팬서2보다 3000만큼 더 재미있음'}
28 Struct{Title=엔트맨과 와스프: 권텔메니아,Score=10,Review= '재미가있다재미가있다'}
29 Struct{Title=엔트맨과 와스프: 권텔메니아,Score=8,Review= '같은 타노스만큼 임팩트가 크지 않았다원작에선 어마어마한 빌런인데...'}
30 Struct{Title=엔트맨과 와스프: 권텔메니아,Score=7,Review= '기준 작품의 매력이 떨어진 시리즈'}
31 Struct{Title=엔트맨과 와스프: 권텔메니아,Score=7,Review= '찰이 원래 썬니까? 이영화에선 글썽요알자 세계 또다른 우주를 보는듯 해 열심마는 훌륭했어요'}
32 Struct{Title=엔트맨과 와스프: 권텔메니아,Score=10,Review= '블랙팬서 토르보다 훨씬 재미있고 벌써 어벤져스5가 기대되게 만드는 영화'}
33

```

mysql에 데이터가 들어갈때 마다 잘 기록되고있다.