



E204 : S\$EN

삼성 SW 청년 아카데미 7기 부울경캠퍼스
특화프로젝트 (08.22 ~ 10.07)

포팅 매뉴얼

담당 컨설턴트 : 박종철
김유정(팀장), 배지우, 박주연, 안정현, 이근희, 이상진

Invoke Gradle script ?

Invoke Gradle ?

Gradle Version

gradle_6_7

Use Gradle Wrapper ?

Tasks ?

고급...

<<목차>>

프로젝트 기술 스택	2
빌드 상세내용	3
배포 특이사항	6
DB 접속 정보	17
외부 서비스 정보	18
데이터 수집 및 전처리	18

1. 프로젝트 기술 스택

- a. 이슈관리 : Jira
- b. 형상관리 : GitLab
- c. 커뮤니케이션 : Mattermost, Notion, Figma, Discord, Kakaotalk
- d. 테스트 : Swagger, Postman
- e. 개발환경
 - i. OS : Windows 10
 - ii. IDE
 1. IntelliJ
 2. Visual Studio Code : 1.70.0

- iii. Database : MySQL 5.7.39
- iv. Server : AWS EC2 (Ubuntu 20.04.4 LTS)
- v. BigData : Hadoop 3.2.1
- f. 상세 개발환경
 - i. Backend
 - 1. Java 1.8.0_192(Zulu 8.33.0.1-win64)
 - 2. Spring Boot Gradle 6.7
 - 3. lombok : 1.18.24, Querydsl-jpa : 4.4.0, swagger2 : 3.0.0
 - ii. FrontEnd
 - 1. HTML5, CSS3, JavaScript(ES6)
 - 2. React v18, React-Router-Dom v6, Redux v4
 - 3. Node.js 16.16.0
 - iii. DataBase
 - 1. MySQL 8.0.30
 - iv. BigData
 - 1. Hadoop 3.2.1
 - 2. Ui path(환율 데이터 수집)
 - 3. Python 3.7(뉴스 데이터 수집 및 전처리)

2. 빌드 상세내용

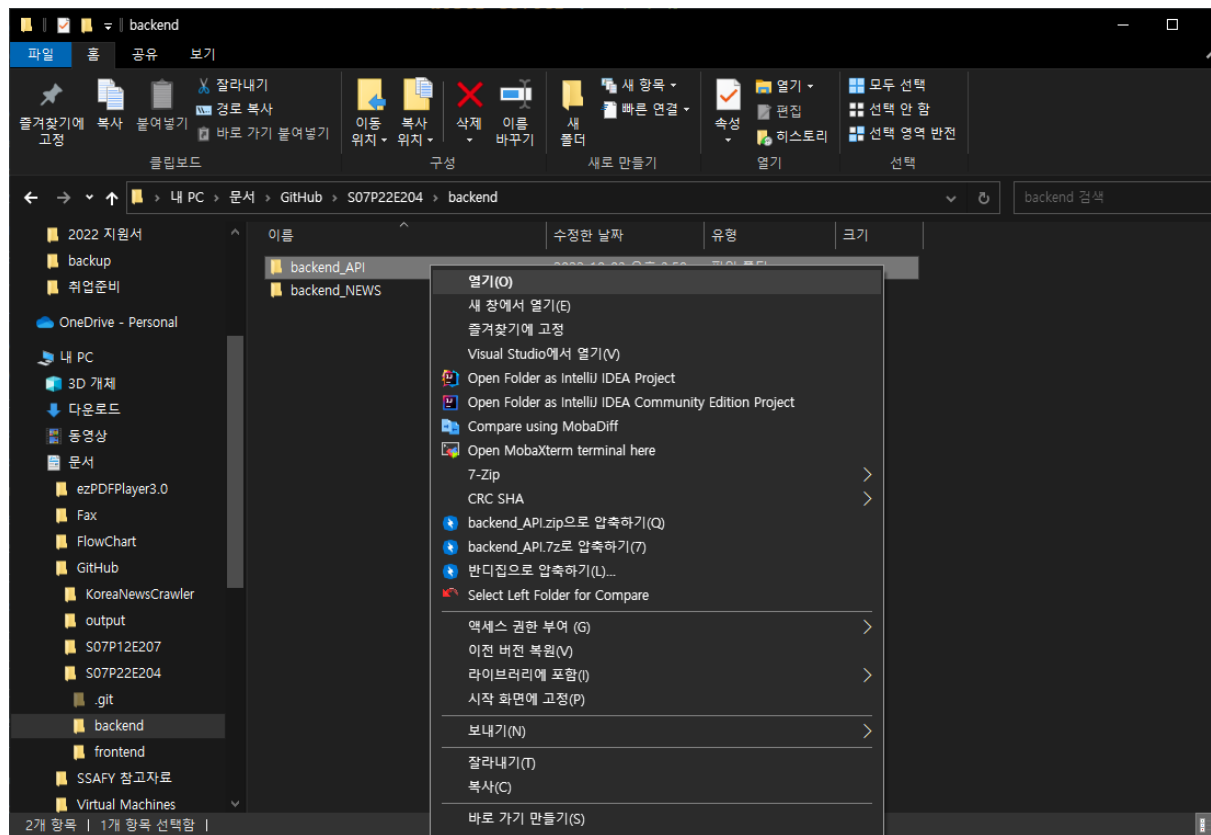
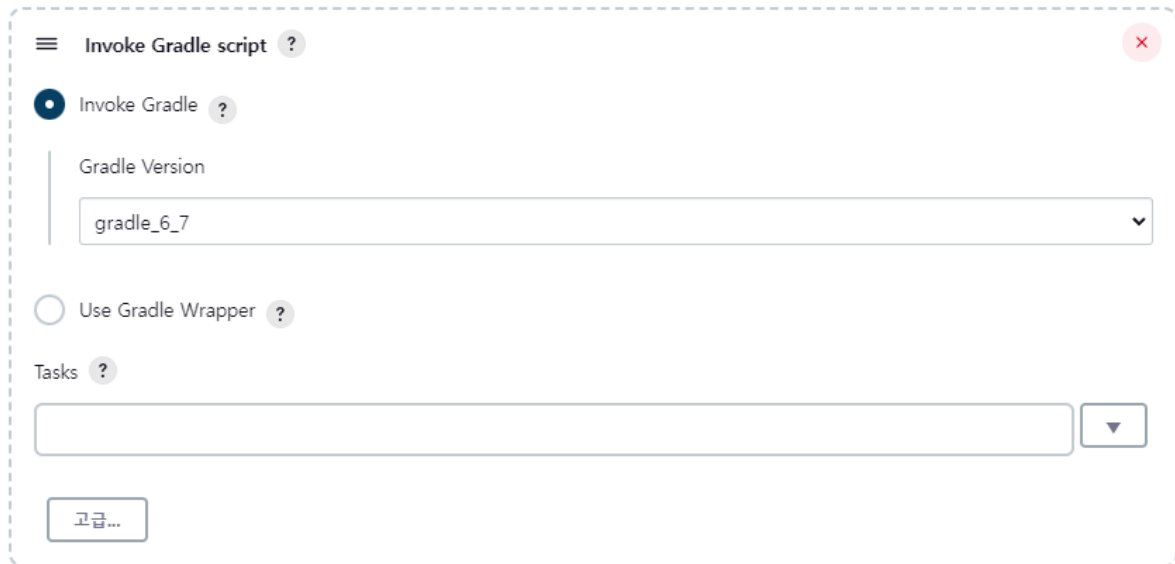
a. DB

먼저, 배포할 EC2 혹은 로컬 호스트에서 MySQL 8.0.30 을 설치합니다.
그 후, e204 계정을 만들고 접속 한 다음 데이터베이스를 생성합니다.

```
mysql -u root -p create database IF NOT EXISTS `SSEN` collate utf8mb4_general_ci;
```

b. Backend

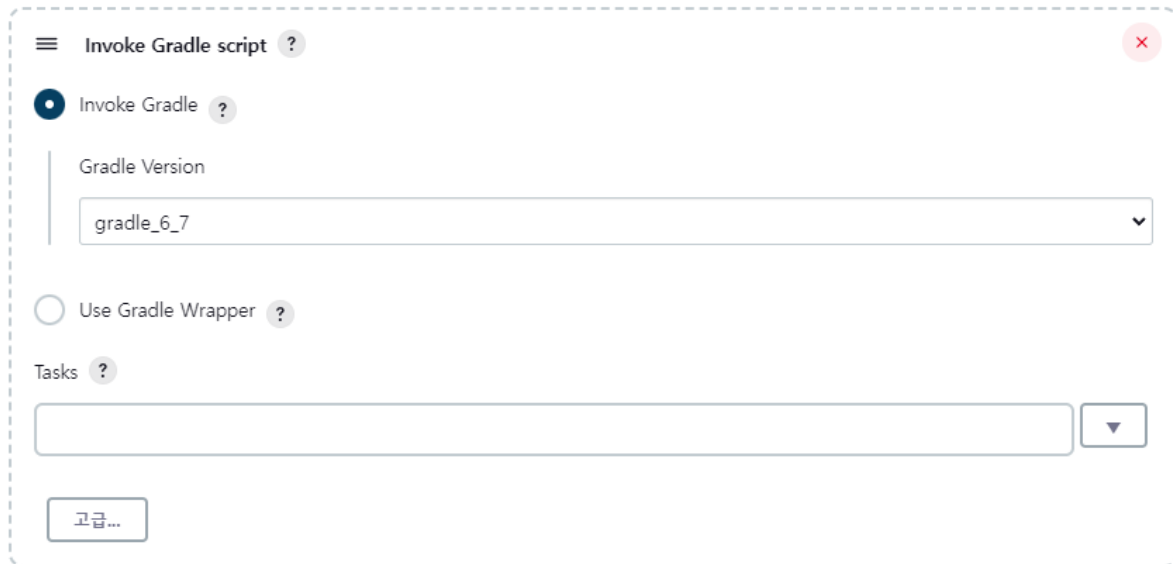
인텔리제이에서 프로젝트 실행



백엔드 프로젝트의 경우 두가지가 존재합니다. 실행 방법은 동일하므로 하나만 기재하겠습니다. `/backend/backend_API` 폴더를 `Open Folder as IntelliJ IDEA Project`를 클릭해서 엽니다.

```
spring.datasource.url=jdbc:mysql://j7e204.p.ssafy.io:3306/$SEN?serverTimezone=Asia/Seoul&characterEncoding=UTF-8
spring.datasource.hikari.username=e204
spring.datasource.hikari.password=your_password
```

`src/main/resources` 폴더의 `application.properties` 파일의 35~37 line에서 연결할 DB와의 설정이 맞는지 확인합니다.



Invoke Gradle script ?

☒ Invoke Gradle ?

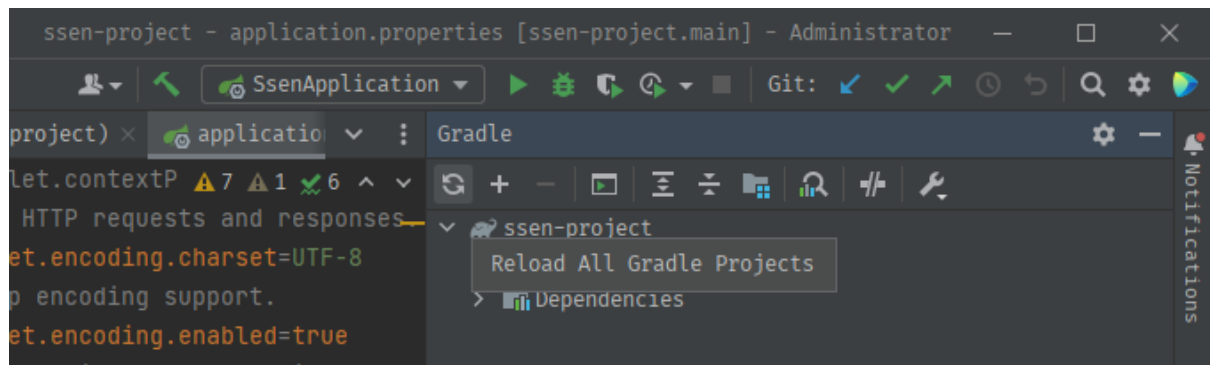
Gradle Version

gradle_6_7

☐ Use Gradle Wrapper ?

Tasks ?

고급...



이후, Gradle 탭에서 Reload All Gradle Project 버튼을 누른 다음, Run Ssen Application 버튼을 눌러 스프링 프로젝트가 정상적으로 실행되는지 확인합니다. 정상적으로 실행이 되었다면, Swagger 링크인 <http://localhost:8080/swagger-ui/#/> 로 접속이 가능합니다.

Command 버전 배포 방법

배포할 때는 Windows command 창에서 다음과 같이 빌드를 진행합니다.

```
cd C:\Users\SSAFY\<....>\backend_API
gradlew clean build
```

빌드에 성공하면, /buid/libs 경로에 ssen-project-1.0-SNAPSHOT 파일이 생성되는 걸 확인할 수 있습니다. 해당 파일을 아래의 커맨드로 인텔리제이 내에서와 동일하게 실행할 수 있습니다.

```
java -jar ssafy-web-project-1.0-SNAPSHOT.jar
```

또 다른 스프링 프로젝트인 backend_NEWS의 경우도 이와 동일하게 실행합니다. (다만 포트는 8081번을 사용합니다.)

c. Frontend

로컬환경에서 실행

npm install
npm start

[Node.js 16.16.0 버전](#)을 설치한 뒤, /frontend 폴더에서 **npm install**로 **node_modules**를 설치해줍니다. **npm run build** 명령어로 **build**파일을 생성합니다. **npm start** 명령어로 **localhost**환경에서 테스트해 볼 수 있습니다.

배포 환경에서 사용할 **build** 폴더 빌드

npm update
npm run build

/frontend 폴더에서 **npm update**로 **dependency**를 업데이트 해줍니다. 그 다음, **npm run build** 명령어로 **build**파일을 생성합니다.

3. 배포 특이사항

이 프로젝트는 빅데이터 분산 도메인으로 하둡을 사용합니다. 하둡은 클러스터가 아닌 **EC2** 내부적으로 **standalone** 버전을 설치하는 것으로 구성하였습니다. 배포시에는 백엔드와 프론트 엔드 폴더에서 개별 도커파일을 이용하여 배포하였고, 젠킨스를 통해 이를 자동화 하였습니다.

EC2에 하둡 설치 및 기본 프로그램 설치 방법, 도커 파일 작성과 수동 배포 방법 이후 젠킨스를 이용한 자동 배포 방법 순으로 설명하겠습니다.

a. **EC2**에 하둡 설치 및 프로젝트 파일 업로드

EC2에 하둡을 **standalone** 모드로 설치합니다. 설치 시에는 해당 [블로그](#)를 참고하였습니다. 다음 명령어를 통해 **hdfs**를 시작하고 구동 여부를 확인합니다.

```
start-dfs.sh
jps
```

이후, EC2의 `/home/ubuntu` 경로에 `mapreduce.tar.gz` 를 업로드 후, 압축을 풀어줍니다. `mapreduce` 경로에서, 사용할 `hdfs` 디렉토리를 만든 다음, 해당 디렉토리에 데이터들을 넣어줍니다.

```
tar -xvzf mapreduce .
cd ~/mapreduce
hdfs dfs -mkdir news_in topksearch
hdfs dfs -put topksearch_data.txt topksearch
hdfs dfs -put ./input/weekend/ news_in
```

`weekend` 폴더의 경우 최종 배포 날짜를 기준으로 과거 일주일간의 뉴스가 저장되어 있습니다. 이후 포팅 시 변경이 필요합니다.

pem 키 업로드

백엔드에서 EC2에 접속하여 맵리듀스를 사용하기 위해서, SSH 연결을 하는것이 필요합니다. EC2의 `/home/ubuntu` 경로에 `pem` 키를 저장해줍니다.

Invoke Gradle script ?

Invoke Gradle ?

Gradle Version

gradle_6_7

☐

Use Gradle Wrapper ?

Tasks ?

고급...

/home/ubuntu/				
Name	Size (KB)	Last modified	Owner	Group
..				
backend_API		2022-09-30...	ubuntu	ubuntu
frontend		2022-09-30...	ubuntu	ubuntu
mapreduce		2022-10-04...	ubuntu	ubuntu
J7E204T.pem	2	2022-09-26...	ubuntu	ubuntu
nohup.out	11 219	2022-10-04...	ubuntu	ubuntu
ssn-news-project-1.0-SNAP...	63 836	2022-10-03...	ubuntu	ubuntu
ssn.log	503	2022-10-04...	ubuntu	ubuntu
ssn.log.2022-10-03.0.gz	107	2022-10-04...	ubuntu	ubuntu

b. EC2에 기본 프로그램 설치

java openjdk 8, mysql 8.0.30, docker 최신버전을 설치해줍니다.

letsencrypt 설치 및 SSL 인증키 획득

서비스에 https 연결을 제공하기 위해, 아래의 명령어를 통해서 SSL 인증서를 발급받는 도구인 letsencrypt를 설치하고 인증서를 발급 받습니다.

```
sudo apt-get install letsencrypt
sudo letsencrypt certonly --standalone -d <i7e204.p.ssafy.io>
```

이후 등장하는 약관에 동의하면
/etc/letsencrypt/live/i7e204.p.ssafy.io/ 경로에 fullchain.pem과
privkey.pem이 생성된걸 확인할 수 있습니다.

.p12 파일 획득 및 설정

배포 버전에서 웹소켓을 활용하기 위해서, 백엔드에도 SSL 인증서가 필요합니다. openssl을 이용해서 pem 파일을 활용해 p12 파일을 얻습니다. 이후 사용할 비밀번호를 설정합니다.

```
cd /etc/letsencrypt/live/j7e204.p.ssafy.io/

openssl pkcs12 -export -inkey privkey.pem -in cert.pem -out
/<resources 폴더경로>/keystore.p12
```

keystore.p12 파일을 다운로드 후,
backend_API/src/main/resources/ssl 경로에 저장합니다. 그 다음,
application.properties 파일에 다음을 추가합니다.

```
server.ssl.key-store=classpath:ssl/keystore.p12
server.ssl.key-store-type=PKCS12
server.ssl.key-store-password=your_password
```

nginx 설치 및 설정

이 프로젝트에서는 사용하는 백엔드 서버의 포트 번호를 감추기 위한 리버스 프록시 기능을 위해 nginx 를 사용하였습니다. 먼저 nginx를 설치해줍니다.

```
sudo apt-get install nginx
nginx -v # nginx/1.18.0 (ubuntu)
```

그 다음, nginx의 환경 설정 파일인
/etc/nginx/sites-available/default를 수정해줍니다. 아래는 리버스 프록시 및 SSL 인증이 포함된 default 파일의 전체 내용입니다.

```
server{
```

Invoke Gradle script ?

Invoke Gradle ?

Gradle Version

gradle_6_7

Use Gradle Wrapper ?

Tasks ?

고급...

```

location /{
    proxy_pass http://localhost:3000;
}

location /api {
    proxy_pass https://localhost:8080/api;
}

location /news {
    proxy_pass http://localhost:8081/news;
}

location /ssen {
    proxy_pass http://localhost:8080/ssen;
    proxy_redirect off;
    charset utf-8;

    proxy_set_header X-Real-IP $remote_addr;
    proxy_set_header X-Forwarded-for
@proxy_add_x_forwarded_for;
    proxy_set_header X-Forwarded-Proto $scheme;
    proxy_set_header X-NginX-Proxy true;
}

listen 443 ssl; # managed by Certbot
ssl_certificate
/etc/letsencrypt/live/j7e204.p.ssafy.io/fullchain.pem; #
managed by Certb>
ssl_certificate_key
/etc/letsencrypt/live/j7e204.p.ssafy.io/privkey.pem; # managed
by Cer>
# include /etc/letsencrypt/options-ssl-nginx.conf; #
managed by Certbot
# ssl_dhparam /etc/letsencrypt/ssl-dhparams.pem; #
managed by Certbot
}

```

Invoke Gradle script ?

Invoke Gradle ?

Gradle Version

gradle_6_7

Use Gradle Wrapper ?

Tasks ?

고급...

```
server {
    if ($host = j7e204.p.ssafy.io) {
        return 301 https://$host$request_uri;
    } # managed by Certbot

    listen 80;
    server_name j7e204.p.ssafy.io;
    return 404; # managed by Certbot
}
```

방화벽 설정

프로젝트에서는 http/https 접속을 위한 80, 443, MySQL을 위한 3306, 백엔드 서버를 위한 8080, 8081 외에도 젠킨스를 위한 9090등을 열어주었습니다. 포트를 여는 명령어와 전체 포트 상태 확인은 아래와 같습니다.

22	ALLOW	Anywhere
3306	ALLOW	Anywhere
9090	ALLOW	Anywhere
8080	ALLOW	Anywhere
80	ALLOW	Anywhere
443	ALLOW	Anywhere
3000	ALLOW	Anywhere
8081	ALLOW	Anywhere
8443	ALLOW	Anywhere
22 (v6)	ALLOW	Anywhere (v6)
3306 (v6)	ALLOW	Anywhere (v6)
9090 (v6)	ALLOW	Anywhere (v6)
8080 (v6)	ALLOW	Anywhere (v6)
80 (v6)	ALLOW	Anywhere (v6)
443 (v6)	ALLOW	Anywhere (v6)

```
cd backend_API
sudo chmod +x gradlew
./gradlew clean build
docker build -t ssen_api_back:0.1 .
```

```
docker run --name ssen_api_back -d -p 8080:8080  
ssen_api_back:0.1
```

d. 프론트엔드 **Dockerfile** 작성 및 수동 배포
마찬가지로, **frontend** 폴더 내에서 아래와 같은 도커 파일을 생성합니다.

COPY ./nginx.conf /etc/nginx/conf.d

80 포트 오픈
EXPOSE 80
 # https 사용을 위한 443 포트 오픈
EXPOSE 443
 # container 실행 시 자동으로 실행할 command. nginx 시작함
CMD ["nginx", "-g", "daemon off;"]

이후, frontend 폴더를 EC2에 업로드한 다음 다음 명령어를 통해 수동 배포할 수 있습니다.

```
npm update
CI=false npm run build
docker build -t ssen_front:0.1 .
docker run --name ssen_front -d -p 3000:80 ssen_front:0.1
```

e. 젠킨스를 이용한 자동 배포 환경 설정

젠킨스 이미지를 받고 실행시키고, 젠킨스 admin 계정을 가입하고 깃랩을 연동합니다. Credential 설정 및 access token 발급 등의 일련의 과정은 SSAFY GIT 7기 계절학기 자기주도 프로젝트 필수 8번, 구글링 및 [블로그](#) 글을 참고하였습니다. 추가로 Plugin 매니저에서 NodeJS와 Gradle을 설치해줍니다.

이후 General > Build Steps를 도커를 이용한 수동 배포와 똑같은 방식으로 작성해줍니다. 아래는 순차적으로 실행하는 Execute Shell과 Gradle Script, Node.js 스크립트 캡처입니다.

```
docker stop ssen_front || true
docker rm ssen_front || true
docker stop ssen_api_back || true
docker rm ssen_api_back || true
docker stop ssen_news_back || true
```

≡ Invoke Gradle script ?

● Invoke Gradle ?

Gradle Version

gradle_6_7

☐ Use Gradle Wrapper ?

Tasks ?

고급...

```
docker rmi ssen_front:0.1 || true
docker rmi ssen_api_back:0.1 || true
```

```
cd backend
cd backend_API
```

```
cd backend
cd backend_API
ls
/var/lib/jenkins/tools/hudson.plugins.gradle.GradleInstallation/gr
adle_6_7/bin/gradle clean build
docker build -t ssen_api_back:0.1 .
docker run --name ssen_api_back -d -p 8080:8080
ssen_api_back:0.1
cd ..
cd ..
cd frontend
```

Invoke Gradle script ?

Invoke Gradle ?

Gradle Version

gradle_6_7

☐ Use Gradle Wrapper ?

Tasks ?

고급...

Execute NodeJS script ?

NodeJS Installation

Specify nodejs installation where npm installed packages to execute the script

nodejs-front

Script

See [the list of available environment variables](#) accessible by process.env.ENV_VARIABLE.

npmrc file

- use system default -

Cache location

Default (~/.npm or %APP_DATA%\npm-cache)

```

ls
cd frontend
npm update
CI=false npm run build
docker build -t ssen_front:0.1 .
docker run --name ssen_front -d -p 3000:80 ssen_front:0.1
  
```


Invoke Gradle script ?

Invoke Gradle ?

Gradle Version

gradle_6_7

Use Gradle Wrapper ?

Tasks ?

고급...

4.DB 접속 정보

- a. EC2 MySQL DB Connection 연결
로컬에서 EC2에 설치된 MySQL에 연결하기 위해서,
Workbench에서 새로운 connection을 생성합니다.

MySQL Connections

Local instance MySQL80

PJT2 E204

특화 EC2 E204

Connection Name: 특화 EC2 E204

Connection

Remote Management

System Profile

Connection Method: Standard (TCP/IP)

Method to use to connect to the RDBMS

Parameters

SSL

Advanced

Hostname: j7e204.p.ssafy.io

Port: 3306

Name or IP address of the server host - and TCP/IP port.

Username: e204

Name of the user to connect with.

Password:

Store in Vault ...

Clear

The user's password. Will be requested later if it's not set.

Default Schema:

The schema to use as default schema. Leave blank to select it later.

Store Password For Connection

Please enter password for the following service:

Service: Mysql@j7e204.p.ssafy.io:3306

User: e204

Password:

OK

Cancel

New

Delete

Duplicate

Move Up

Move Down

Test Connection

Close

b. 스키마 생성

```
create database IF NOT EXISTS `SSEN` collate
utf8mb4_general_ci;
```

프로젝트에서 사용하는 스키마를 생성합니다. 인코딩을 **UTF-8**으로 설정해줍니다.

c. **Dump 파일 실행**

프로젝트의 덤프 파일을 실행해줍니다.

(/backend_API/src/main/resources/application.properties에는 hibernate-ddl-auto=none으로 설정되어 있습니다.)

5. 외부 서비스 정보

a. [KoreaNewsCrawler](#)

뉴스 데이터 수집 시에는 **Github KoreaNewsCrawler**를 활용하여 네이버 뉴스의 글로벌 경제, 네이버 금융의 환율 카테고리의 뉴스를 2004.2월부터 2022.10까지 수집하였습니다.

6. 데이터 수집 및 전처리

최신 환율 데이터와 뉴스 데이터를 수집하는 방법은 다음과 같습니다.

a. 환율 데이터

‘investing.com’에서 제공하는 환율 데이터를 수집합니다.

(RPA를 전용 툴인 **UiPath**를 활용해 매일, 그리고 실시간으로 환율 정보를 크롤링 합니다.

차트를 그리기 위해 일별로 시가, 종가, 고가, 저가, 변동폭, 기준 날짜를 **DB**에 저장합니다.

실시간 환율 처리를 위해 약 7초마다 한번씩 크롤링을 합니다.)

b. 뉴스 데이터

KoreaNewsCrawler를 활용하여 원하는 카테고리의 뉴스를 수집합니다.

(네이버 금융의 환율 뉴스의 URL은

'https://finance.naver.com/news/news_list.naver?mode=LSS3D§ion_id=101§ion_id2=258§ion_id3=429&date=' 입니다)

Invoke Gradle script ?

Invoke Gradle ?

Gradle Version

gradle_6_7

Use Gradle Wrapper ?

Tasks ?

고급...

이후, `/data/news/concat.py`를 이용하여 이를 하나의 파일로 합치고, `keyword2.py`를 이용하여 키워드 추출을 한 다음 새로운 `.csv`파일을 얻을 수 있습니다. 최종 날짜별 `.csv` 파일은 아래와 같습니다.

	A	C	D	E	F	G	H
16	2022-01-06 11:10:35	연합뉴스	중 인민은행 산하 매체 위안화	미 금리인상-중 수출둔화 충격 대비	https://n.news.naver.com	금융시장 경고 중국 트럼프 신홍시장	
17	2022-01-09 23:19:57	뉴시스	대만 코로나 신규 60명-본토 1	타이베이 AP 뉴시스 대만 수도 타이	https://n.news.naver.com	타이난시 유행병 항공기 소빙 독일	
18	2022-01-09 21:14:01	매일경제	AI-로봇이 노동자 대체해도 전	인공지능 경제학 세션 애쓰모글루	https://n.news.naver.com	로봇 불평등 경쟁 전자상거래 고용	
19	2022-01-09 20:42:01	이코노미	고란 코인도란 비트코인 개미	코인 거래량 줄자 NFT 키우는 업비	https://n.news.naver.com	방탄소년단 공포 인플레이션 캐나다	
20	2022-01-09 19:43:02	한겨레	전직 삼성전자 특허담당 임원	삼성 쪽 "내부 출신의 소 제기는 처	https://n.news.naver.com	삼성 공격 영업비밀 특허법 텍사스	
21	2022-01-09 19:28:03	파이낸셜	미국 석학을 돈벌기 비판...경제	전미경제학회 연례총회 래리 서머스	https://n.news.naver.com	재정정책 노벨 우려 성장	
22	2022-01-09 18:29:03	매일경제	미 통화정책 안통할 땐...1970	세계 석학들의 인플레이 경고 스티글	https://n.news.naver.com	초인플레이션 도덕 탐욕 수단 마녀	
23	2022-01-09 18:29:02	매일경제	핵심인력 빼오기-기업간 채용	지난해 노벨경제학상 데이비드 카드	https://n.news.naver.com	노동시장 맥도널드 뉴저지주 노벨경제학상 우려	