

# Kaggle with Spark

라스칼라코딩단  
2016년 2월 26일 (금)

# 순서

- Spark, Zeppelin 설치하기
- Spark 맛보기
- Kaggle 시작하기

# Spark, Zeppelin 설치하기 (1)

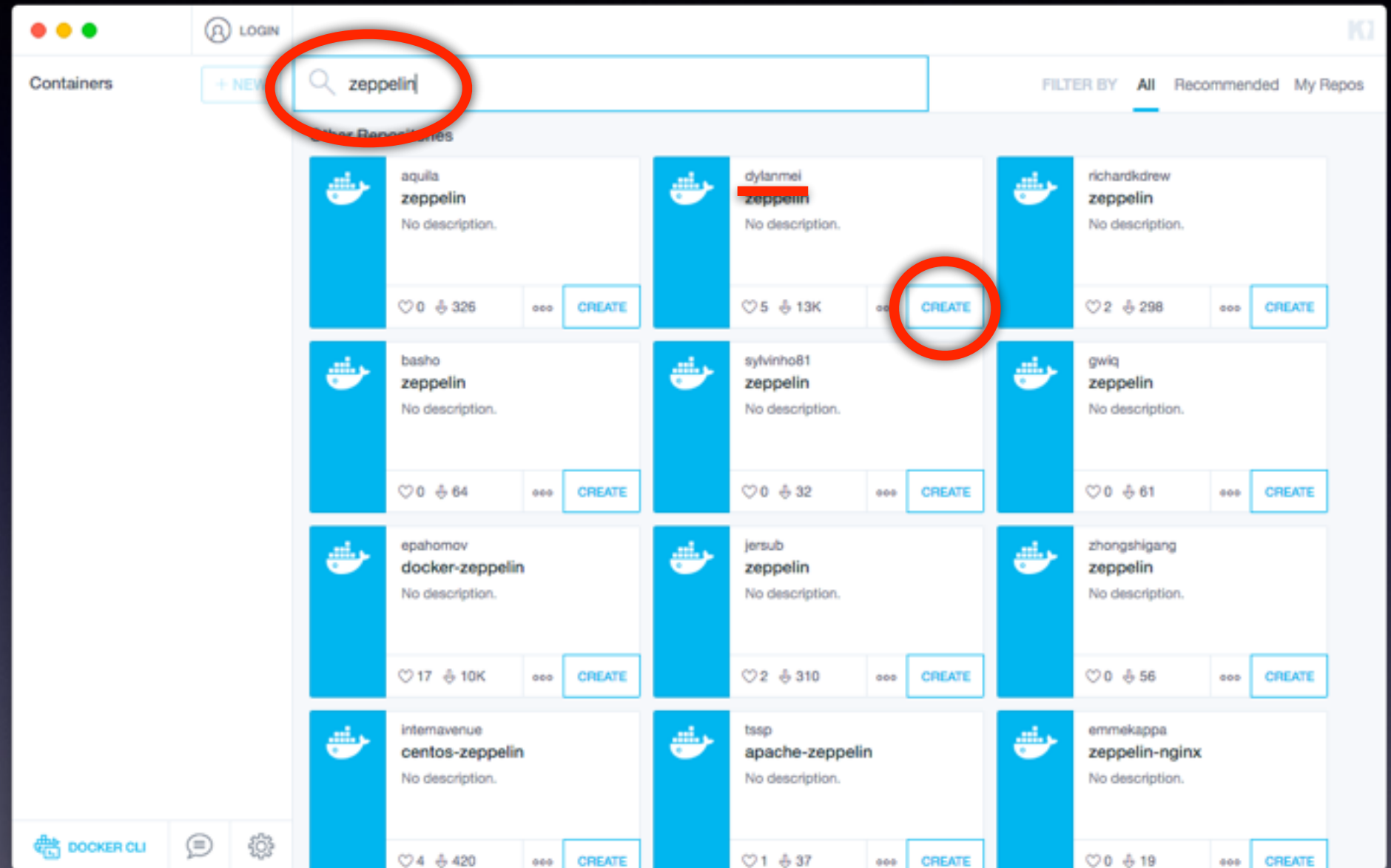
- (뭘 좀 아시는 분들을 위한) 한줄 요약
  - docker를 설치하고 <https://github.com/dylanmei/docker-zeppelin> 내용을 실행하세요!

# Spark, Zeppelin 설치하기 (2)

- Step by Step으로 따라해 봅시다.
- Step 1
  - 아래 URL에서 Docker Toolbox를 설치합니다  
<https://kitematic.com/>
  - 설치 중간에 Start Program을 둘 중 하나로 선택하는데 Kitematic을 선택합니다.

# Spark, Zeppelin 설치하기 (3)

- Step 2

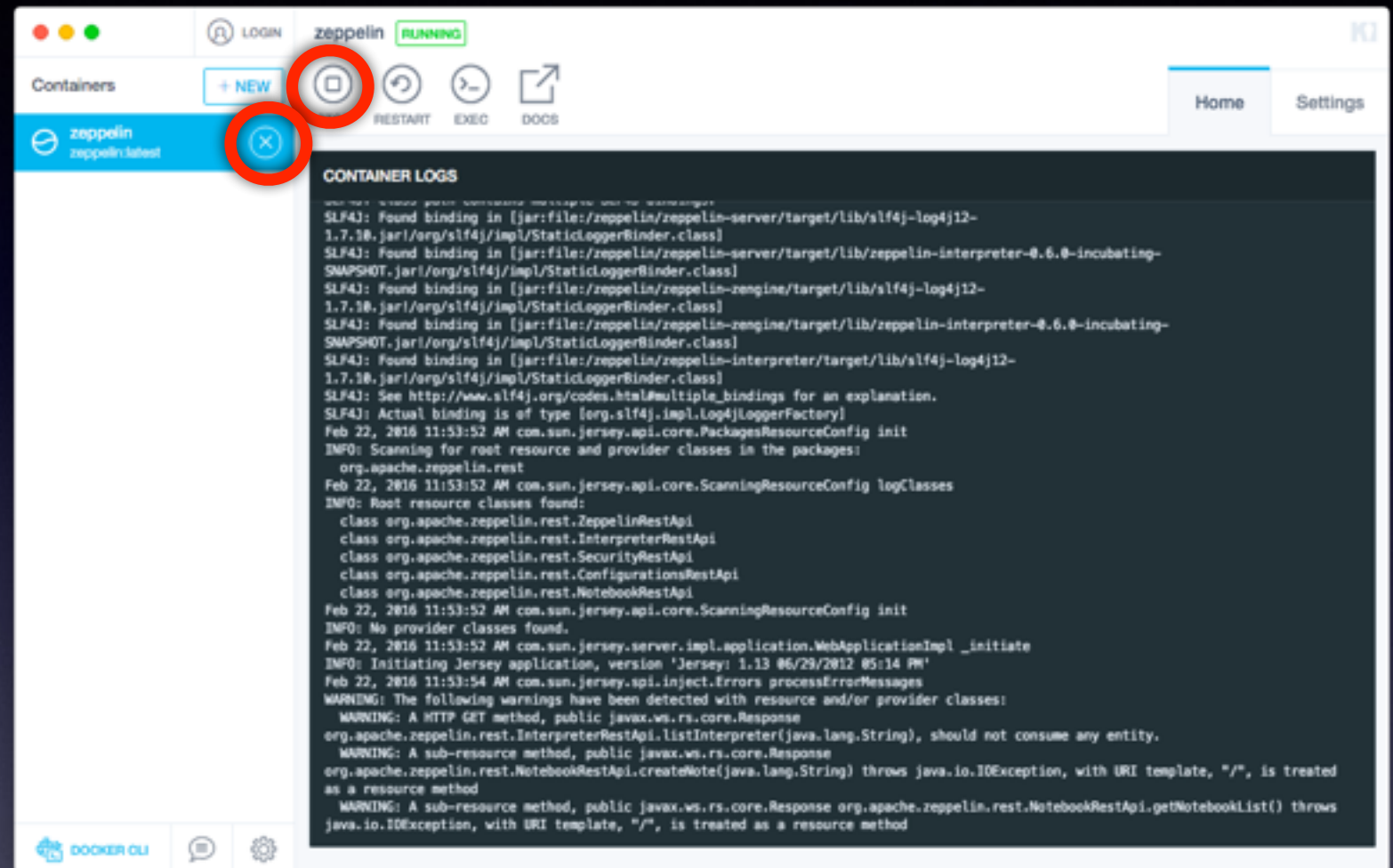


- 첫 화면의 검색창에서 “zeppelin”으로 검색하여 “dylanmei”의 Zeppelin 이미지를 생성합니다.



# Spark, Zeppelin 설치하기 (4)

- Step 3

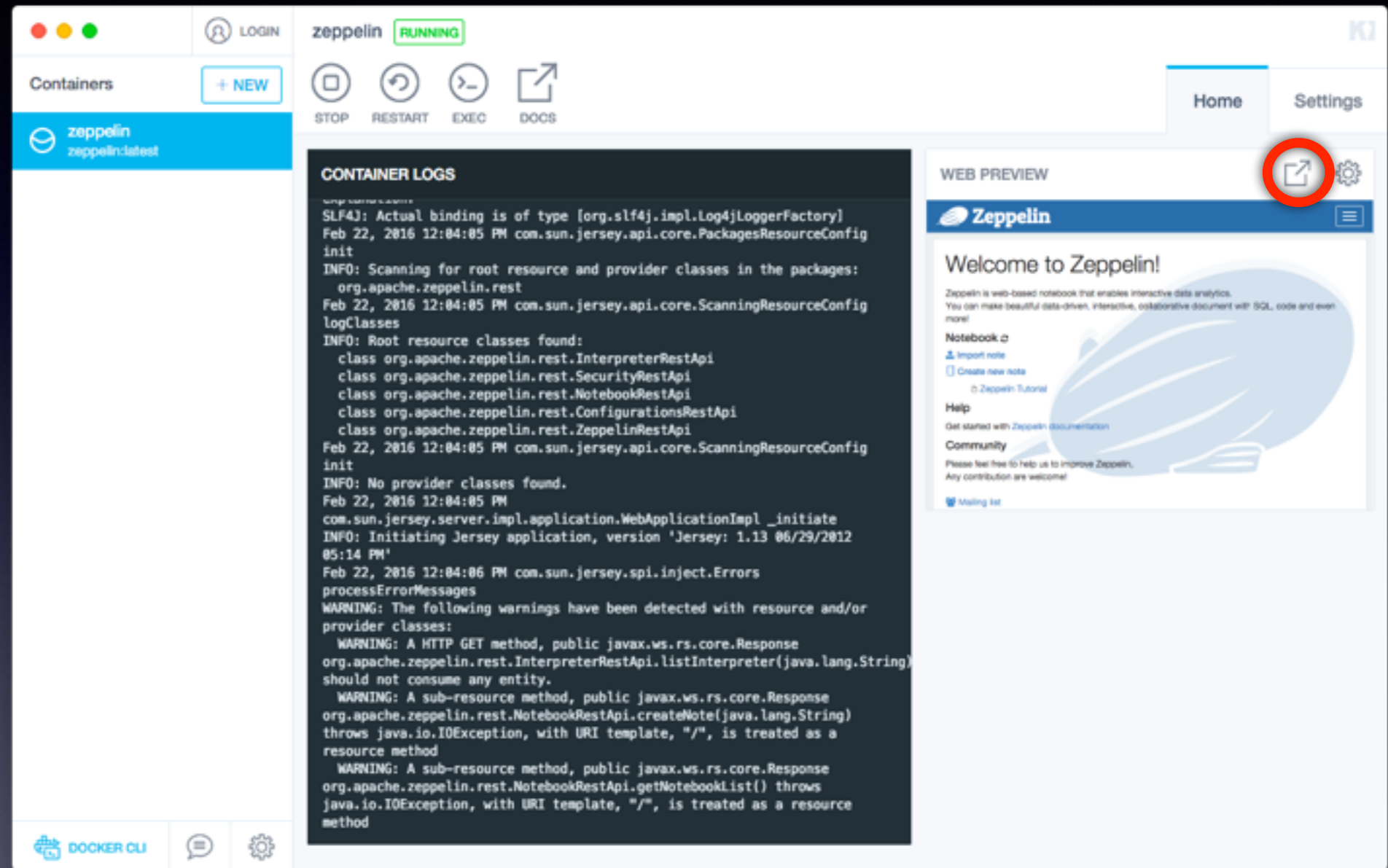


- 위와 같은 최종 화면이 나온다면 아직 정상적인 부팅 상태가 아닙니다.
- 이미지를 멈추고 왼쪽 탭에서 제거한 다음, 터미널에서 아래 명령으로 최신 상태로 업데이트 한 후 다시 시작합니다.

```
docker pull dylanmei/zeppelin:latest
docker run --name zeppelin -p 8080:8080 -p 8081:8081 dylanmei/zeppelin:latest
```

# Spark, Zeppelin 설치하기 (5)

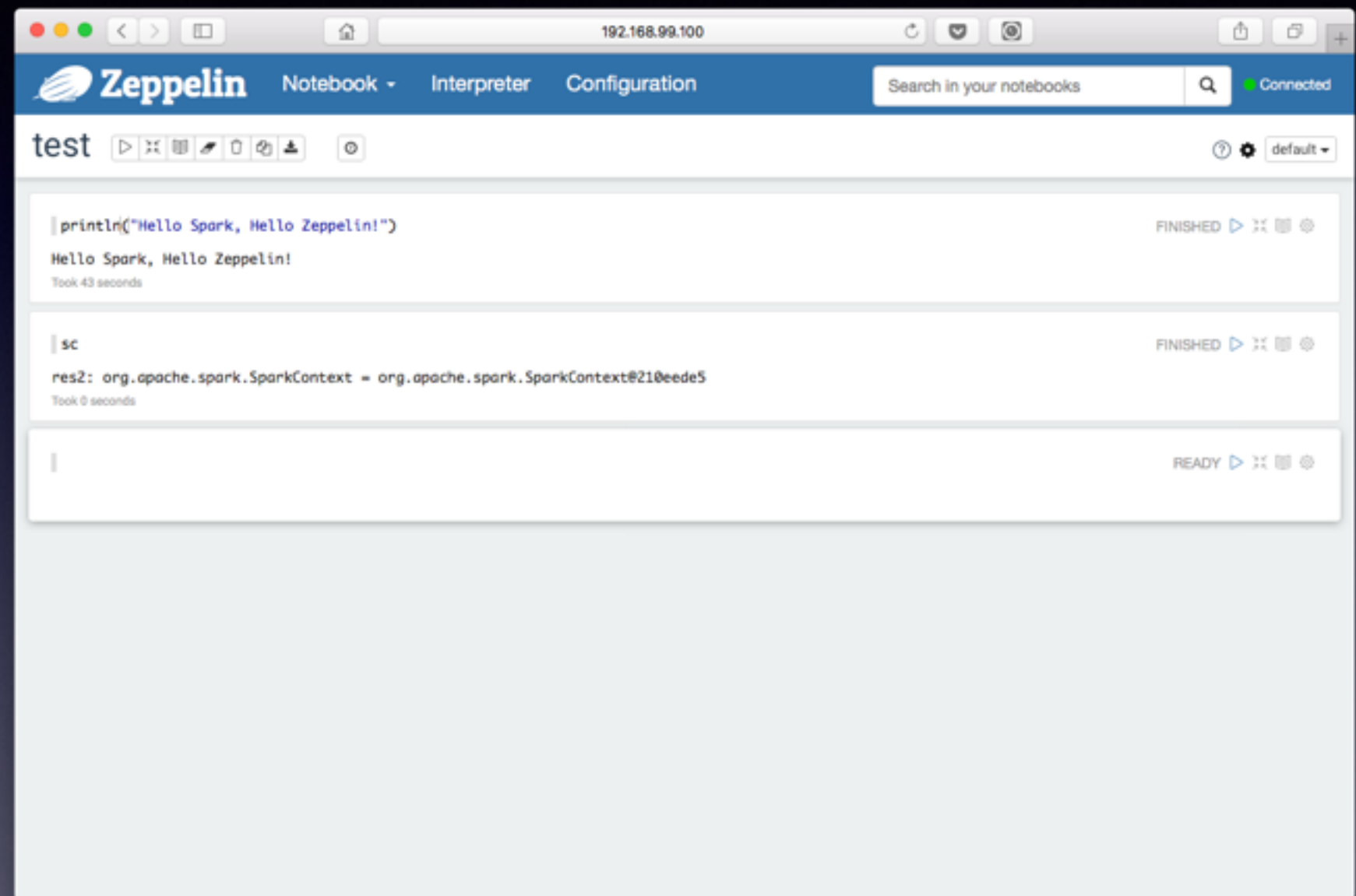
- Step 4



- 가운데 콘솔 창 오른 편에 Zeppelin 화면을 클릭하면 브라우저가 뜹니다.

# Spark, Zeppelin 설치하기 (6)

- Step 5



- “sc + ”를 입력해서 SparkContext 객체가 잘 나온다면 성공입니다.

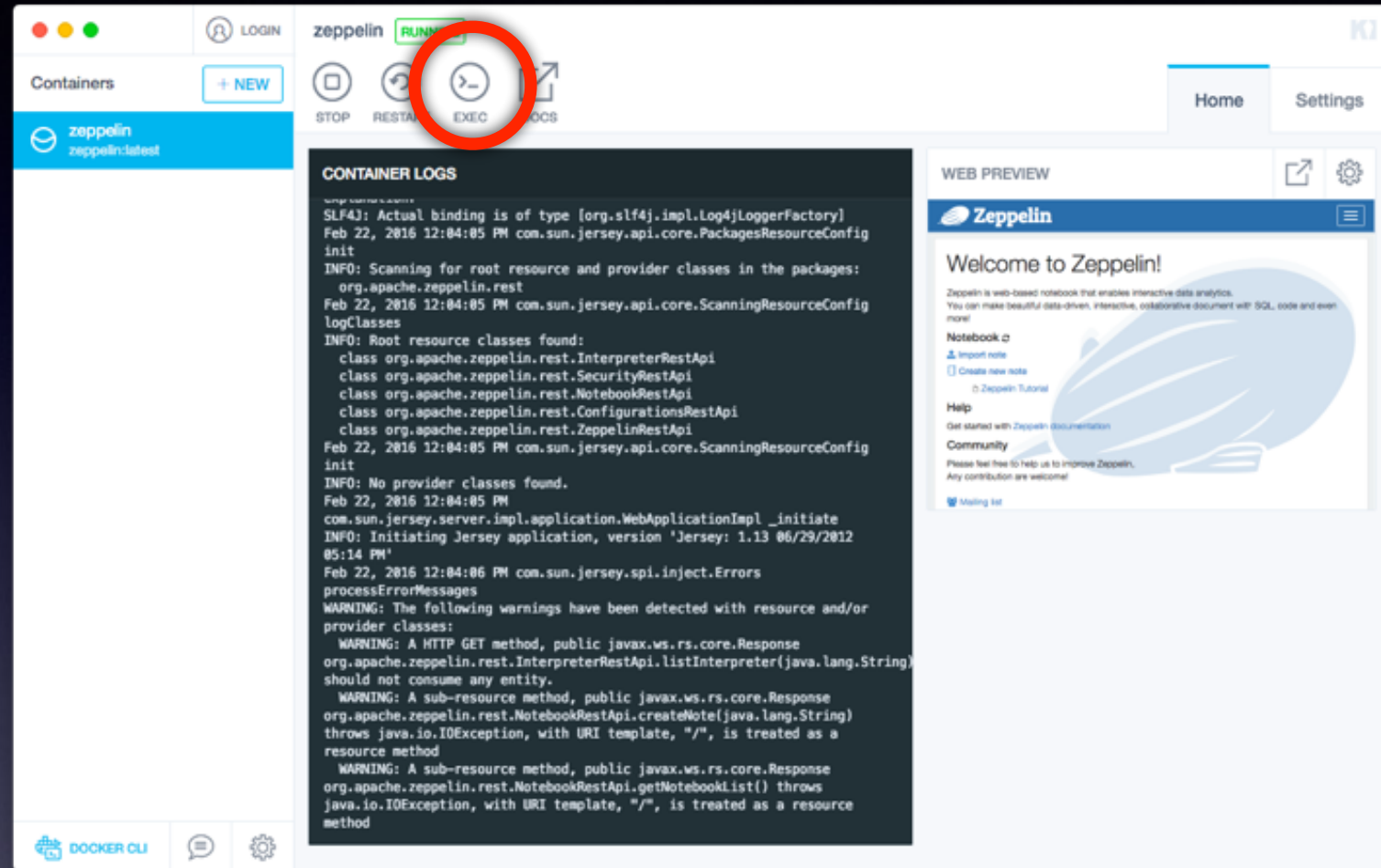


# Spark 맛보기

- “Welcome to Zeppelin!”이라고 쓰인 홈페이지에서 “Import note”를 눌러 아래 URL을 입력하여 임포트 합니다.
- [https://raw.githubusercontent.com/krikkit/kaggle\\_with\\_spark/master/zeppelin\\_spark.json](https://raw.githubusercontent.com/krikkit/kaggle_with_spark/master/zeppelin_spark.json)

# Kaggle 시작하기 (1)

- Step 1



- Kitematic 화면에서 “EXEC” 버튼을 눌러 쉘 프롬프트로 진입하여, 아래와 같이 파일을 복사합니다.

```
# mkdir /titanic
# cd /titanic
# scp <user>@192.168.99.1:~/Downloads/train.csv
# scp <user>@192.168.99.1:~/Downloads/test.csv
```

# Kaggle 시작하기 (2)

- Step 2
  - “Welcome to Zeppelin!”이라고 쓰인 홈페이지에서 “Import note”를 눌러 아래 URL을 입력하여 임포트합니다.
  - [https://raw.githubusercontent.com/krikkit/kaggle\\_with\\_spark/master/zeppelin\\_kaggle.json](https://raw.githubusercontent.com/krikkit/kaggle_with_spark/master/zeppelin_kaggle.json)