

Spark Web UI

출처

<https://spark.apache.org/docs/latest/monitoring.html>



spark standalone mode를 기준으로 설명

Spark web history UI를 통한 모니터링

- 기본 스파크 Web UI
 - 모든 SparkContext는 기본적으로 포트 4040에서 애플리케이션에 대한 정보를 표시(<http://<driver-node>:4040> 로 확인 가능)
 - standalone 모드의 경우 대부분 <http://localhost:4040> 이용 가능
 - Web UI의 경우 디폴트로 **응용 프로그램이 실행되는 동안에만** 실행 가능

- Spark web history UI
 - incomplete/completed application들과 attempt들을 기록된 로그를 통해 보여줌
 - 사용 방법

1. 이벤트 로그가 저장될 디렉토리 생성
(예시)

```
mkdir /home/user/spark-events
```

2. spark-defaults.conf 파일 하단에 다음을 추가

- 로컬에 로그 파일이 저장됨

```
spark.eventLog.enabled      true
spark.eventLog.dir          file:/home/user/spark-events # 파일 event 로그가 저장되는 디렉토리
spark.history.fs.logDirectory file:/home/user/spark-events # spark.eventLog.dir와 같은 디렉토리여야 함
```

- (보류) hdfs에 로그 파일이 저장됨

- 현재 connection refuse 오류가 계속 남

```
spark.eventLog.enabled      true
spark.eventLog.dir          hdfs://localhost:8020/eventLogging # 파일 event 로그가 저장되는 디렉토리
spark.history.fs.logDirectory hdfs://localhost:8020/eventLogging # spark.eventLog.dir와 같은 디렉토리여야 함
```

3. spark/sbin/start-all.sh 로 스파크 실행
4. spark/sbin/start-history-server.sh 로 스파크 히스토리 서버 실행
 - localhost:18080 에서 확인 가능