

# Sqoop Ecosystem

Kim Hye Kyung

topickim@naver.com

# | Sqoop

# 목차

---

- 스쿱(Sqoop) 개요
- 스쿱(Sqoop) 데이터 가져오기
- 스쿱(Sqoop) 데이터 내보내기

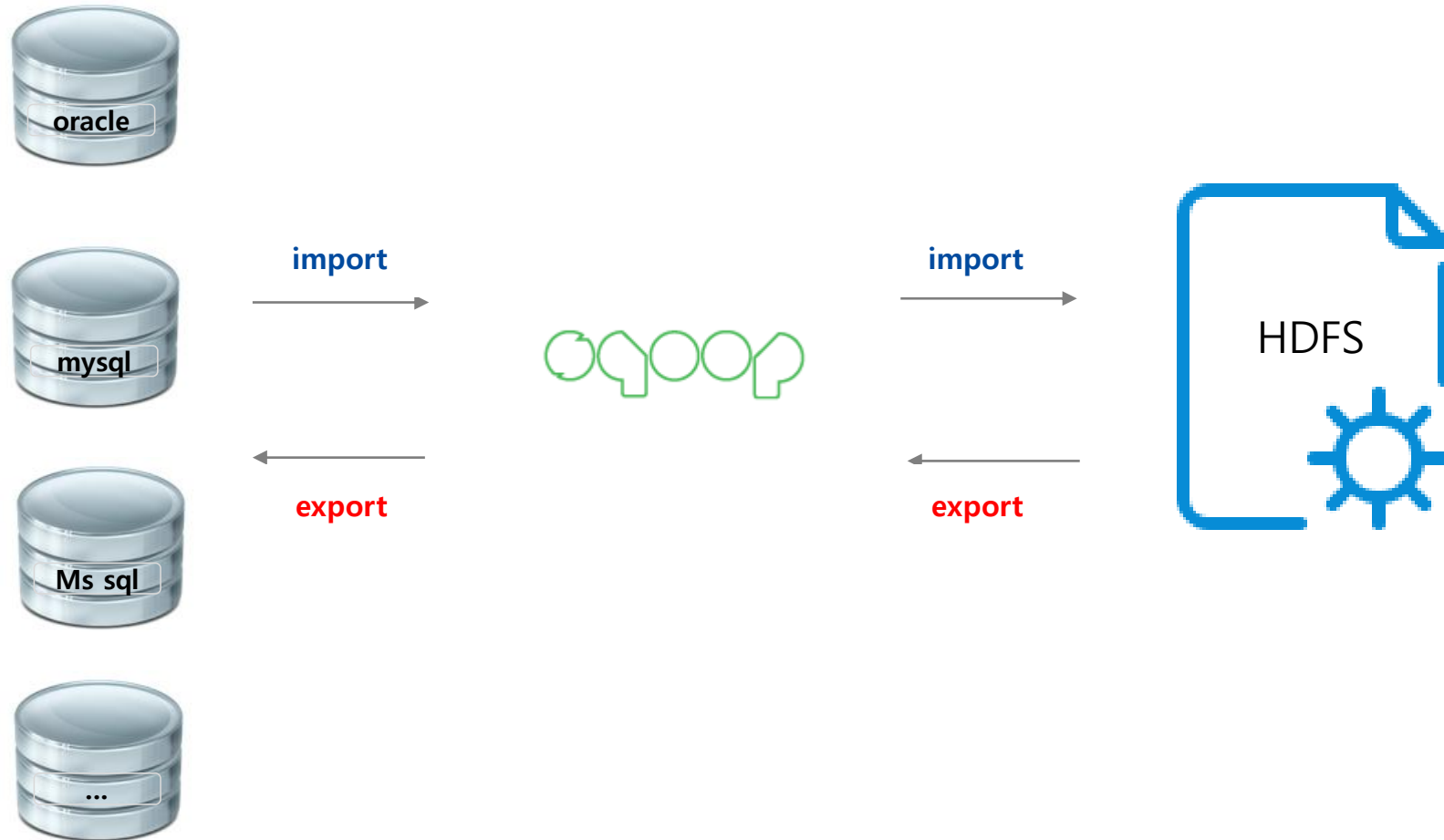
# 스쿱(Sqoop) 개요



- RDBMS와 HDFS 간의 효율적인 대용량 데이터 전송하는 도구
- Hadoop 에코 시스템 수집 Software
- 외부 시스템의 데이터를 HDFS로 가져와서 Hive 테이블, Hbase 테이블 등 Hadoop의 다양한 파일 형태로 저장 가능
- JDBC와 호환되는 모든 RDBMS에 사용 가능
- <http://sqoop.apache.org/>

# 스쿱(Sqoop) 개요

- JDBC와 호환되는 모든 RDBMS에 사용 가능

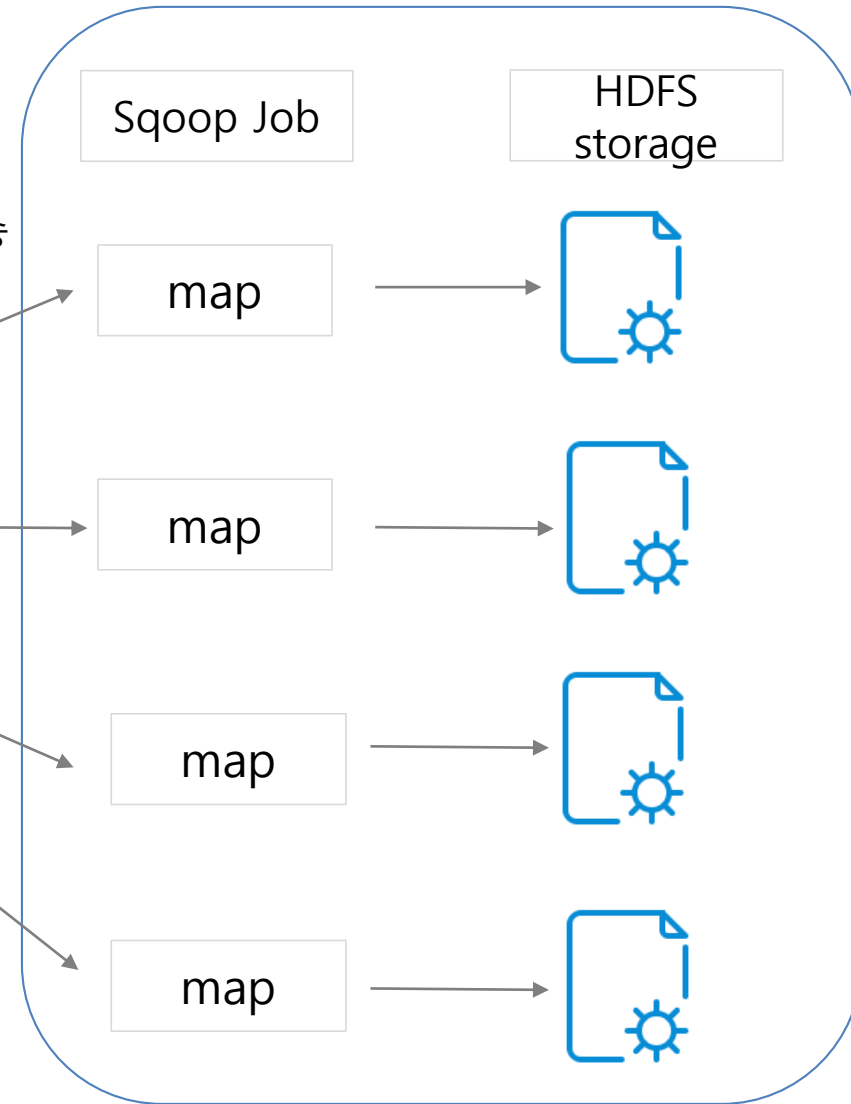


# 스쿱(Sqoop) 데이터 가져오기

- 스쿱으로 데이터 가져오기
- import
  - RDB 데이터의 메타 정보 획득
  - Map-Only Hadoop job으로 Hadoop Clouster로 데이터 전송



Table의 접속 권한등으로 sqoop이 필요한 만큼의 분량으로 DB로 부터 해당하는만큼의 블록을 가져와서 hdfs 파일 시스템에 저장



# 스쿱(Sqoop) 데이터 가져오기

- 스쿱 데이터 가져오기
  - 관계형 데이터베이스에서 메타데이터 수집
  - 필요 정보
    - 데이터베이스명, 테이블명, ID, PW등

```
sqoop list-databases
--connect jdbc:mysql://_HOST_NAME/_DATABASE_NAME_
--username ID --password PW
```

-----

테이블 list 검색됨

# 스쿱(Sqoop) 데이터 가져오기

- 스쿱 데이터 가져오기
  - 맵 적용 작업을 Hadoop에 요청
  - Sqoop import 명령어 특징
    - 리듀스 작업 없이 맵잡만 실행
  - Hadoop 클러스터의 각 노드는 데이터베이스 접근 권한을 미리 부여받아야 함
  - 옵션
    - -m : 몇 개의 map job을 만들것인지에 대한 개수
    - --target-dir : 검색해서 어느 파일 시스템에 넣을 것인가 명시
  - RDBSM의 table의 내용을 HDFS의 파일 시스템에 저장하는 명령어

```
sqoop import
--connect jdbc:mysql://localhost:3306/world
--username ID
--password PW
--table Country -m 1 --target-dir /user/sqoop/country
```



# 스쿱(Sqoop) 데이터 가져오기

- 스쿱 데이터 가져오기
  - RDBMS에서 가져온 데이터는 HDFS 디렉터리에 저장됨
    - RDBMS의 table 구조의 데이터가 HDFS에 , 구분 형태로 저장됨
  - HDFS에 저장된 파일 확인 명령어

```
hdfs dfs -ls sqoop-mysql-import/country
```

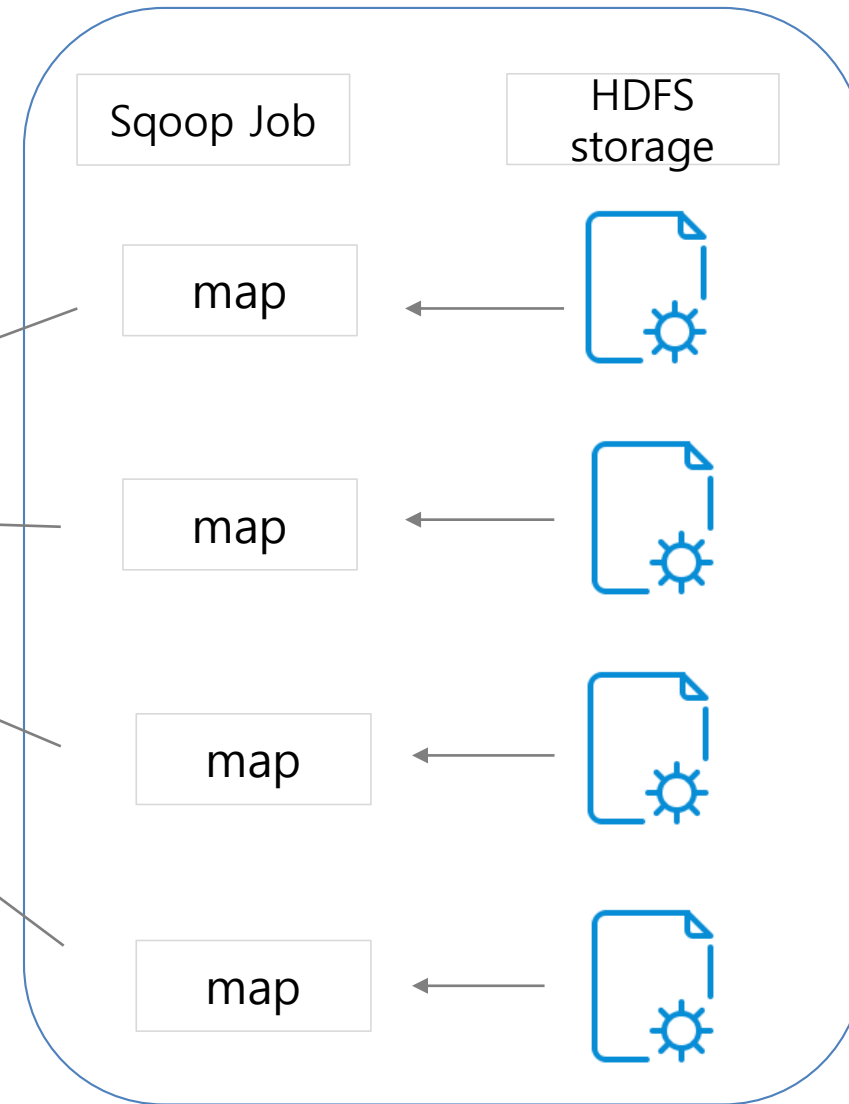
```
Hdfs dfs -cat sqoop-mysql-import/country/part-m-0000
```

# 스쿱(Sqoop) 데이터 내보내기

- 스쿱으로 데이터 내보내기
- export



메타데이터에 접근해서 table의 구성 정보를 확인하고 map job을 실행하고 map job이 파일들의 정보를 읽어서 RDBMS로 넣는 구조



# 스쿱(Sqoop) 데이터 내보내기

- 스쿱 데이터 내보내기
  - 관계형 데이터베이스의 메타데이터 수집
  - 작업 단계
    - RDBMS에 Table 생성
    - Export 명령어를 사용해서 HDFS에 내장된 데이터를 RDBMS에 export
    - 단, Hadoop 클러스터의 각 노드는 RDBMS 접근 권한을 미리 부여받아야 함

```
sqoop export
--connect jdbc:mysql://localhost:3306/world
--username sqoop
--password sqoop
--table CityExport
-m 4 --export-dir /user/sqoop/city3
```

# 스쿱(Sqoop) 실습

---