

[Hortonworks Sandbox III – Sqoop]

※ Hortonworks Sandbox에는 mySql과 sqoop이 설치되어 있다.

1. Sqoop에 액세스할 데이터 mySql에 넣기

```
root@sandbox ~]# mysql -u root -p # mySql 접속
```

```
mysql> show databases;
```

```
mysql> use test; # 호튼웍스에서는 test만 스콧에서 액세스하도록 config
```

```
mysql> create table salaries( # 테이블 생성
```

```
    gender varchar(1),
```

```
    age int,
```

```
    salary double,
```

```
    zipcode int);
```

```
mysql> alter table salaries add column `id` int(10) unsigned primary key auto_increment
```

id컬럼을 추가하여 주키로 설정 (스콧에서 액세스 하기 위해서는 반드시 주키(primary key)가 있어야 한다.)

```
mysql> insert into salaries (gender, age, salary, zipcode) values ('m', 40, 76000,95102); # 데이터 추가
```

2. Sqoop 실행

```
root@sandbox ~]# sudo sqoop import --connect jdbc:mysql://localhost/test --table salaries --username root
--target-dir /sqoop_out # 스콧 실행
```

```
root@sandbox ~]# hadoop fs -ls /sqoop_out
```

```
sqoop_out/_SUCCESS
```

```
sqoop_out/part-m-00000
```

```
sqoop_out/part-m-00001
```

```
sqoop_out/part-m-00002 # m은 map단계에서 만들어짐. r은 reduce단계에서 만들어진 파일
```

```
root@sandbox ~]# hadoop fs -cat /sqoop_out/*
```

```
M, 40,76000.0, 95102, 1
```