

## Lab4. Tableau를 이용한 Data Source 연동하기

### 1. Tableau 설치하기

- 1) Tableau를 사용하기 위해서는 먼저 PC나 노트북에 Tableau Desktop이라는 프로그램을 설치해야 한다.
- 2) <https://www.tableau.com/ko-kr> 사이트를 방문한다.
- 3) 페이지 오른쪽 상단의 [지금 체험] 버튼을 클릭한다.
- 4) 페이지 하단에서 14일 [무료 평가판 다운로드]를 클릭한다.
- 5) 학생이나 강사인 경우 <https://www.tableau.com/ko-kr/community/academic>을 방문하여 무료 버전을 이용할 수 있다.
- 6) [Tableau Desktop: 14일 무료 평가판 시작] 페이지에서 각각의 값을 입력한 후, [무료 평가판 다운로드] 버튼을 클릭한다.
- 7) 다운로드 받은 TableauDesktop-64bit-2023-1-2.exe를 더블클릭하여 실행한다.
- 8) 설치가 끝나면 자동으로 Tableau 프로그램이 시작되고, [Tableau 등록]이라는 팝업창이 나타난다.
- 9) 필요한 정보를 입력한 후, [지금 평가판 시작] 버튼을 클릭한다.

### 2. MySQL Driver 설치하기

- 1) 설치 및 등록 작업을 마치면 첫 화면으로 [연결]이 나온다.
- 2) 화면의 좌측 메뉴 중 [서버에 연결] > [자세히...]에서 MySQL을 검색 혹은 직접 찾아서 연결할 수 있고, 또는 MySQL은 자주 사용하기 때문에 바로 [서버에 연결] > [MySQL]로 연결해도 된다.
- 3) 첫번째 방법
  - [MySQL] "이 커넥터를 사용하려면 먼저 드라이버를 다운로드하여 설치해야 합니다." 팝업창에서 [드라이버 다운로드 ▶]를 클릭한다.
  - [드라이버 다운로드] 페이지에서 먼저 로그인을 수행하고, 로그인 후 다음과 같이 입력한다.
    - [데이터 원본] : MySQL
    - [운영 체제] : Windows or Mac
    - [비트 버전] : 64-bit
  - [MySQL] 섹션에서, [Tableau Desktop: 2019.3 - 2023.1.2]의 순서대로 진행한다.
- 4) 두번째 방법
  - \* 첫번째 방법은 결국 MySQL 홈페이지에서 MySQL 드라이버를 다운로드하기 때문에 권장하지 않는다.
  - <https://dev.mysql.com/downloads/> 페이지로 이동한다.
  - 목록에서 [Connector/ODBC]를 클릭한다.
  - [MySQL Community Downloads Connector/ODBC] 페이지에서
    - [Select Operating System:] : Microsoft Windows or macOS
    - [Select OS Version:] : Windows (x86, 64-bit) or macOS 13 (x86, 64-bit) or macOS 13 (ARM, 64-bit)
    - Windows의 경우 MSI Installer의 [Download] 버튼을 클릭한다.
    - macOS의 경우 DMG Archive or Compressed TAR Archive [Download]를 클릭하여 다운로드한다.
  - [Login Now or Sign Up for a free account] 페이지에서, [No. thanks, just start my download.] 링크를 클릭한다.
  - 다운로드 받은 파일을 실행한다. (Windows의 경우 mysql-connector-odbc-8.0.33-winx64.msi)

### 3. MySQL Data Source 준비하기

- 1) Github의 4주차 실습예제 중 [Lab4. Running Spark Script on AWS Glue Studio.pdf]의 "5. AWS Glue Studio에서 spark-rds-demo3.py 실행하기"를 실행한다.
- 2) 권한문제 때문에 [Run]에 오류가 발생하면 바로 위의 "2. Parameter Store에 새로운 패스워드 설정하기"와 "3. AWS Glue Studio에서 spark-rds-demo1.py 실행하기"를 참조한다.
- 3) [Run]이 성공적으로 마치면, Amazon RDS의 {계정}-db의 newyork\_taxi 데이터베이스에 2개의 테이블 즉, taxi\_zone\_lookup과 taxi\_output 테이블이 생성된다.
- 4) Database 연결 프로그램(예: HeidiSQL or MySQL Workbench)를 이용하여 해당 데이터베이스의 taxi\_output 테이블을 오픈한다.
- 5) 해당 테이블의 컬럼 중 do\_location\_id의 [데이터 유형]을 TEXT에서 INT로 변경한 후, 저장한다.

### 4. MySQL Data Source 연동하기

- 1) MySQL 드라이버 설치 후, 다시 [연결] > [서버에 연결] > [MySQL]을 선택한다.
- 2) [MySQL] 팝업창에서 다음과 같이 Amazon RDS와 연결한다.
  - [일반]
    - [서버] : Amazon RDS {계정}-db의 엔드포인트, ex)henry-db.cx1hah81fdfdf.ap-northeast-2.rds.amazonaws.com
    - [포트] : 3306
    - [데이터베이스] : newyork\_taxi
    - [사용자 이름] : admin
    - [비밀 번호] : datalakemysql
  - [로그인] 버튼 클릭
- 3) 로그인 성공하면, 화면 좌측에 taxi\_zone\_lookup 테이블이 확인된다.
- 4) 즉시 화면 오른쪽에 테이블이 보이지 않으면 화면 왼쪽의 [테이블] > taxi\_zone\_lookup을 Drag & Drop 하면 된다.
- 5) 테이블 로딩 후, [지금 업데이트] 버튼을 클릭한다.
- 6) 데이터가 로딩 후, 화면 좌측의 테이블 중, taxi\_output 테이블을 Drag & Drop 한다.
- 7) 두 개의 테이블이 서로 연결되었지만, 조인 조건을 넣지 않았기 때문에 주황색 느낌표가 있는 라인이다.
- 8) 주황색 라인에 마우스를 올리면, "카디널리티: 다대다(기본값), 이 관계에는 일치하는 필드가 없습니다. 관계를 편집하여 일치하는 필드를 선택하십시오." 메시지를 확인할 수 있다.
- 9) 화면의 아래에 "관계를 생성하는 것과 조인하는 것의 차이점은 무엇입니까?"의 섹션에서, taxi\_zone\_lookup 테이블에서는 #Location ID를 선택하고, taxi\_output 테이블에서는 Do Location Id를 선택한다.
- 10) 그리고 [지금 업데이트]를 클릭한다.
- 11) 데이터를 확인한다.

### 5. sales\_data\_sample csv 파일 연동하기

- 1) Tableau Desktop을 실행한다.
- 2) [연결] 페이지에서 [파일에 연결] > [자세히...]를 클릭한다.
- 3) [열기] 팝업창에서 Lab2에서 다운로드했던 sales\_data\_sample.csv 파일을 선택한다.