Lab3. Using Amazon RDS

1. MySQL RDS 생성하기
   1)[서비스] > [데이터베이스] > [RDS]
   2)[Amazon RDS] > [대시보드] 페이지에서 [데이터베이스 생성] 클릭
   3)[데이터베이스 생성] 페이지에서
      -[데이터베이스 생성 방식 선택] : "표준 생성"
      -[엔진 옵션] > [엔진 유형] : MySQL
      -[엔진 옵션[> [엔진 버전] : MySQL 8.0.32

   4)[템플릿] : "프리 티어"
   5)[설정]
      -[DB 인스턴스 식별자] : {계정}-db
      -[마스터 사용자 이름] : admin
      -[마스터 암호], [마스터 암호 확인] : datalakemysql

   6)[인스턴스 구성]
      -[DB 인스턴스 클래스] : "버스터블 클래스(t 클래스 포함)
      -db.t3.micro

   7)[스토리지]
      -[스토리지 유형] : 범용 SSD(gp2)
      -[할당된 스토리지] : 20GiB

   8)[연결]
      -[컴퓨팅 리소스] : EC2 컴퓨팅 리소스에 연결 안 함
      -[Virtual Private Cloud(VPC) : {계정}-datalake-vpc
      -[DB 서브넷 그룹] : 새 DB 서브넷 그룹 생성
      -[퍼블릭 액세스] : 예
      -[VPC 보안 그룹(방화벽)] : 새로 생성
      -[새 VPC 보안 그룹 이름] : {계정}-db-sg
      -[가용 영역] : ap-northeast-2a

   9)[데이터베이스 인증] : 암호 인증

   10)[추가 구성]에서
      -[초기 데이터베이스 이름] : newyork_taxi
      -[백업] : "자동 백업을 활성화합니다" 체크 해제
      -[암호화] : "암호화 활성화" 체크 해제

   11)[데이터베이스 생성] 버튼 클릭
   12)만일 "DB 인스턴스 henry-db 생성 요청이 실패했습니다." 에러 발생하면 그 이유는, {계정}-datalake-vpc는 한 개의 subnet({계정}-datalake-subnet-2a)만 있기 때문

   13)[VPC] > [서브넷]으로 이동하여 [서브넷 생성] 클릭
      -[서브넷 생성] 페이지에서
         --[VPC ID] : {계정}-datalake-vpc
         --[서브넷 이름] : {계정}-datalake-subnet-2b
         --[가용 영역] : 아시아 태평양(서울)/ap-northeast-2b
         --[IPv4 CIDR 블록] : 172.16.2.0/24
         --[서브넷 생성] 버튼 클릭

   14)다시 RDS 페이지로 돌아와서 [데이터베이스 생성] 버튼 클릭
   15)만일 "DB 인스턴스 henry-db 생성 요청이 실패했습니다." 에러 발생하면 그 이유는, DNS 문제이다.
   16)VPC 페이지로 돌아와서 {계정}-datalake-vpc의 상세페이지에서 [작업] > [VPC 설정 편집] 클릭
   17)[DNS 설정] > "DNS 호스트 이름 활성화" 체크 > [저장] 버튼 클릭
   18)다시 RDS 페이지로 돌아와서 [데이터베이스 생성] 버튼 클릭

2. 생성 후 확인
   1)[RDS] > [데이터베이스] > {계정}-db의 상세 페이지로 이동
   2)[VPC 보안 그룹]을 클릭하여 [인바운드 규칙]에 3306 포트 확인
   3)[퍼블릭 액세스 가능]이 "예"인지 확인
   4)[엔드포인트]와 [포트] 확인

3. MySQL Workbench(https://dev.mysql.com/downloads/workbench/) 또는 HeidiSQL(Only Windows, https://www.heidisql.com/download.php)을 각 OS에 맞게 설치

4. [Open Data on AWS] 페이지에서 csv 파일 다운로드
   1)https://aws.amazon.com/opendata/?wwps-cards.sort-by=item.additionalFields.sortDate&wwps-cards.sort-order=desc
   2)[Find publicly available data on AWS] 버튼 클릭
   3)Search 창에서 taxi로 검색 > "New York City Taxi and Limousine Commission (TLC) Trip Record Data" 링크 클릭
   4)[Documentation]의 "https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page" 링크 클릭
   5)TLC Trip Record Data 페이지에서 "Taxi Zone Maps and Lookup Tables" 섹션의 "Taxi Zone Lookup Table (CSV)"를 클릭하여 csv 파일 다운로드
   6)파일을 열어서 "LocationID", "Borough", "Zone", "service_zone" 4개의 컬럼 확인

5. HeidiSQL에서 연결하여 csv 데이터 가져오기
   1)HeidiSQL 프로그램 로딩 후
   2)[세션 관리자] > [신규] 클릭
   3)[호스트명 / IP] : 생성한 RDS의 엔드포인트 값
   4)[사용자] : admin

5)[암호] : datalakemysql
6)[포트] : 3306
7)[데이터베이스] : 목록에서 newyork_taxi
8)[열기] 클릭
9)변경된 사항을 저장하시겠습니까? : [예]
※연결에 실패할 때 보안 그룹에서 3306 포트의 [소스]를 0.0.0.0/0으로 변경
10)[도구] > [CSV 파일 가져오기]
11)[문서 파일 가져오기] 창에서
-[파일명] : taxi+_zone_lookup.csv
-[인코딩] : utf8mb3:UTF-8 Unicode
-[목적지] > [데이터베이스] : newyork_taxi
-[목적지] > [테이블] : <새 테이블> > [예, 테이블을 생성합니다]
-[가져오기] 클릭
※혹시 데이터를 가져올 수 없다는 경고창이 발생하면 CSV 파일을 열어서 제일 마지막 두개의 행을 삭제한다.
※두개의 행을 삭제하는 이유는 값이 Unknown과 N/A이기 때문이다.