

COMP322002 Intro. to Databases (Fall 2020)

Instructor: Prof. Suh, Young-Kyoon

Phase 2: Relational Database 설계 및 구축

제출 기한: 2020년 10월 21일 (수) 08시 59분 59초

[개요]

이번 단계에서는 영상물 DB 사이트인, 'knuMovieDB'를 위한 관계형 데이터베이스 (Relational Database)를 구축하도록 한다. 1단계에서 생성한 개념 Schema를 바탕으로 데이터베이스를 구축하고, 영상물 DB 사이트의 현 상태에 대한 데이터를 생성하여 입력하는 것이 주요한 task이다.

1단계에서 생성한 결과물과 다소 차이가 있을 수 있으며, 해당 부분에 대해서는 감안을 하여 평가가 진행된다. 평가를 위하여 데이터베이스에 입력될 정보는 최소한의 제약 사항이 존재하며, 제약 사항 이외의 조건들은 스스로 판단을 하여야 한다. 생성할 데이터의 양이 제법 많으므로, 간단한 데이터 생성 및 적재 애플리케이션 프로그램을 작성해서 활용해도 좋다. (해당 데이터 생성/적재 프로그램은 제출할 필요가 없다. 단, 관련 제출물은 반드시 *.sql 파일이어야 한다.) 데이터 생성 후, 데이터베이스에 입력을 한 것을 확인하기 위하여 질의문을 구현하여야 한다. 구현한 질의문을 바탕으로 2단계가 평가된다.

[프로젝트 목적]

- 1) ER Schema가 관계형 Schema로 어떻게 매핑될 수 있는지 이해한다.
- 2) Oracle DBMS에 더 친숙해 질 수 있도록 한다.
- 3) 리눅스 환경에서 오픈 소스 DBMS인 PostgreSQL를 사용하여 과제를 제출할 경우, 추가 점수 20%를 부여한다. 과제 제출 시, PostgreSQL의 릴리즈 버전을 포함한 운영 환경 설정과 관련한 상세 사항을 readMe 파일 형태로 제공해 주어야 한다. 그리고 남은 프로젝트 단계 모두 PostgreSQL 기반으로 수행해야 함에 주의한다.

[관계형 데이터베이스 설계 - 테이블 생성]

Phase 1에서 제출한 ER Schema에 따라 관계형 데이터베이스 구축을 위한 테이블들을 생성한다. 수업 시간에 다룬 것처럼 ER-to-Relational 매핑 기법을 활용하도록 한다. 만약 매핑 과정에서 ER Schema의 문제점을 발견하였다면 수정하도록 한다. 단, 수정 내용이 반드시 매핑 기술 파일에 명시되어야 한다. **최종 수정된 ER Schema**에 근거하여, CREATE TABLE 문을 이용해 테이블을 생성한다. 테이블 생성 시, DDL에 **반드시 테이블 내, 혹은 테이블 간 제약 조건**도 명시할 것.

[관계형 데이터베이스 구축 - 데이터 생성 상세 사항 및 제약조건]

본 절에서는 실제 데이터베이스에서 사용될, 또한 **평가 기준**으로 이용될 데이터를 생성에 관한 내용이다. 평가를 위하여 최소한의 데이터가 필요하며, 이를 서술하였다. 아래의 내용을 토대로 데이터를 생성하여야 하며, 데이터 생성 프로그램을 만들어도 좋다. 아래의 내용에 명시되지 않은 사항에 대해서는 팀 별로 임의로 설정하되, readMe 파일에 해당 설정 내용을 기술하도록 한다.

※ 데이터베이스 구축 과정에서 한글을 쓰지 말 것.

※ IMDB dataset을 활용하여 데이터를 population할 것을 권장하나, 필수 사항은 아님

1. 영상물 (MOVIE)

속성	설명	제약 사항
제목 (TITLE)	영상물의 제목	-
종류 (TYPE)	Movie	최소 100개 이상
	TV Series	최소 20개 이상
	KnuMovieDB Original	최소 100개 이상
재생시간 (RUNTIME)	60분 미만	최소 80개 이상
	60분 이상 120분 미만	최소 80개 이상
	120분 이상	최소 60개 이상
상영 년도 (START_YEAR)	1980년 이상 2000년 미만	최소 60개 이상
	2000년 이상 2020년 미만	최소 60개 이상
	2020년 이상	최소 60개 이상
장르 (GENRE)	각 '장르'별 영상물 수	최소 40개 이상
평가 (RATING)	평점이 10점	최소 50개 이상
	평점이 8점 이상 10점 미만	최소 50개 이상
	평점이 6점 이상 8점 미만	최소 50개 이상
	평점이 6점 미만	최소 50개 이상
	총 영상물 수	최소 200개 이상

2. 장르 (GENRE)

설명	제약 사항
반드시 포함되어야 하는 장르	'Action', 'Comedy', 'Romance'
총 장르 수	최소 5개 이상

3. 평가 (RATING)

설명	제약 사항
각 영상물별 평가 수	최소 0개 이상
총 평가 수	최소 200개 이상

4. 배우 (ACTOR)

설명	제약 사항
각 영상물별 배우 수	최소 2명 이상
총 배우 수	최소 100명 이상

5. 에피소드 (EPISODE)

설명	제약 사항
각 TV Series의 Episode 수	최소 5개 이상
총 에피소드 수	최소 100개 이상

6. 버전 (VERSION)

설명	제약 사항
각 영상물별 버전 수	최소 3개 이상
반드시 포함되어야 하는 지역	'UA', 'KR'
총 버전 수	최소 3종류

7. 계정 (ACCOUNT)

설명	제약 사항
기본 제공 정보만 입력한 고객 수	최소 10명 이상
성별, 생년월일을 추가로 입력한 고객 수	최소 30명 이상
모든 정보를 입력한 고객 수	최소 50명 이상
멤버십이 베이직인 고객 수	최소 20명 이상
멤버십이 프리미엄인 고객 수	최소 50명 이상
멤버십이 프라임인 고객 수	최소 20명 이상
관리자 계정 수	최소 1명 이상
총 계정 수	최소 100명 이상

※ 제약조건에 명시된 테이블 이름을 사용할 것.

[필수 구현 질의문]

본 절은 설계한 관계형 데이터베이스가 옳게 정의되고 적절히 Population 되었는지 검수하기 위한 질의문을 서술한다. 각 팀은 아래의 질의문들을 반드시 SQL로 구현해야 한다.

※ 질의문의 결과가 없는(나오지 않는) 질의문은 무조건 오답처리.

(질의문의 결과가 출력될 수 있도록 데이터를 생성)

※ 질의문마다 다른 배점이 적용될 수 있음

Q1	가입한 고객은 몇 명인가?
Q2	관리자의 이름과 주소는 무엇인가?
Q3	평가를 5 개이상 한 고객은 몇 명인가?
Q4	2020 년에 상영한 'Romance' 장르의 영상물은 몇 개인가?
Q5	2018/10/07 로부터 2 년동안 상영된 영상물은 몇 개인가?
Q6	2020/10/07 을 기준으로 지난 5 년간 상영된 영상물의 평균 평점은 몇 점인가?
Q7	모든 영상물을 대상으로 재생시간이 100 분 이상인 영상물은 몇 개인가?
Q8	모든 영상물을 대상으로 장르가 'Action' 또는 'Comedy'인 영상물 중에 평점이 가장 높은 영상물의 제목은 무엇인가?
Q9	상영된 'TV Series'를 대상으로 Episode 가 10 개 이상인 TV Series 는 몇 개인가?
Q10	평가를 한 번도 하지 않은 고객 중 가장 나이가 많은 고객의 이름과 나이는 무엇인가?
Q11	평가 내역이 1 건만 있는 고객 중에 직업 정보를 입력한 고객은 몇 명인가?
Q12	모든 배우를 대상으로 작업한 모든 영상물의 평균 평점이 8 점 이상인 배우의 이름은 무엇인가?
Q13	장르가 'Action'이 아니고 평점이 6 점 미만이며 상영 년도가 2010 년 이상인 영상물 중에 종류가 'Movie'에 출연한 배우의 이름과 영화의 제목은 무엇인가?
Q14	멤버십이 '프라임'인 고객 중에 가장 젊은 고객을 대상으로 그 사람의 생년월일 중에 '년'과 일치하는 상영 년도를 가진 영상물은 몇 개 있는가?

Q15	평가를 10 개 이상 한 고객들 중에 멤버십이 ‘프리미엄’인 고객이 평가한 영상물의 종류가 ‘TV Series’가 아니고 평점이 8 점 미만인 영상물 중 가장 최신의 영상물의 상영 년도 는 몇 년도인가?
Q16	가장 높은 평점을 가진 영상물을 평가한 고객들 중에 주소 정보를 입력한 고객들을 대상으로 그 주소의 길이가 가장 긴 고객의 전화번호 는 무엇인가?
Q17	영상물을 세 번째로 많이 촬영한 배우의 출연 영상물 중에 상영 년도가 가장 오래된 영상물의 평점과 +1, -1 점 차이의 평점을 가진 영상물은 몇 개 인가?
Q18	모든 영상물을 대상으로 버전이 5 개 이상 10 개 이하이고 종류가 ‘TV Series’ 또는 ‘Movie’이면서 출연한 배우의 수가 5 명 이하인 영상물의 평균 평점은 몇 점 인가?
Q19	가장 높은 평점을 가진 영상물에 출연한 배우들이 참여한 영상물 중에 평점이 두 번째로 높은 영화의 재생 시간과 일치하는 재생 시간을 가진 영상물의 제목 은 무엇인가?
Q20	모든 영상물을 대상으로 여성 고객들에게 가장 많은 평가를 받은 영상물을 평가한 남성 고객들이 가장 평가를 많이 한 영상물의 장르를 포함하는 영상물 중 평가가 가장 적은 영상물의 상영 년도와 일치하는 생년을 가진 고객의 이름 은 무엇인가?

[권장 개발 환경 (채점 환경)]

- DBMS: Oracle Database 12c Release 2(12.2.0.1.0)
 - <https://www.oracle.com/kr/database/technologies/oracle12c-windows-downloads.html>
- 리눅스 환경에서 PostgreSQL DBMS로 SQL 작성 시, 제출한 readMe 파일에 상세한 환경 설정 사항을 기술할 것.

[제출 정보]

- 제출기한: 2020년 10월 21일 (수) 08시 59분 59초
- 제출물: 크게 5가지 + 추가 파일들.
 - 1) **TeamX-Phase2-(1, 2, 3).sql**: 팀별 SQL 파일 3개.
[중요] 스크립트에는 다음의 내용을 반드시 포함해야 함.
 - 1. 해당 *.sql 스크립트 파일에 제약조건들을 포함한 DDL (15점)
 - 2. DDL을 통해 생성된 테이블에 대한 INSERT문들 (5점)
 - 3. Q1 ~ Q20까지의 Query문들 (30점)
 - 2) **TeamX-etr_mapping.txt**: 팀 별 ER Schema가 어떻게 관계형 Schema로 매핑 되었는지 관계 순으로 설명할 것.
ER Schema에 수정사항이 있으면 반드시 명시할 것
 - 3) **TeamX-README.txt**: SQL 동작 환경 및 실행 방법 등 설명하고자 하는 사항들.
 - 4) 수정된 ER Schema 파일
 - 5) 기타 팀별 추가 파일들, if any.
- 배점: (50점 / 140점)
- 주의 사항:
 - o 위에 기술된 권장 개발 환경 (채점 환경)을 준수하도록 한다.
만약 개발 환경이 다를 시 **TeamX-README.txt**에 반드시 기술할 것.
 - o 만약 팀이 Phase 1에서 제출한 ER Schema에 변동이 있을 시, 해당 사항을 **TeamX-etr_mapping.txt**에 반드시 기술하고 수정한 ER Schema를 제출할 것.
- 제출 방법: LMS 팀 프로젝트 메뉴에서 Phase 2 게시물에 팀별로 결과물 제출

[Q&A] TA: jyan@knu.ac.kr (석사과정 안준영)

[Late Submission] 하루 당 50%씩 감점. 단 상황에 따라 기간이 연장될 수 있음.

[Plagiarism] 절대 허용되지 않음. 발견되는 즉시 남은 결과에 상관없이 “F”학점 부여.