

보고서

데이터베이스 구축 강의 실기 과제

빅데이터 9기 전동인

문제 이해

조건

1. 2022년 분기별 (총 4 분기) 콘텐츠

⇒ 각 테이블은 yyyy와 mm이 존재하는데 월 별 회원의 학습 콘텐츠 사용 이력 집계 데이터를 담고 있는 e_study 테이블에서 mm을 3개월씩 조정하며 분기를 바꿀 수 있음. 연도는 2022년으로 고정

2. 초등 3학년, 4학년, 5학년, 6학년

⇒ 학년은 e_member와 e_content_meta에만 존재했고 콘텐츠에 해당 하는 학생의 학년이 아닌 실제 이용한 학생 정보를 가져오기 위해 e_member 테이블에서 3학년 부터 6학년까지를 이용.

3. 영상강의 + 문제풀이가 함께 서비스되는 콘텐츠

⇒ e_study 테이블의 subcontent_name_sequence에서는 각 콘텐츠가 어떤 구성으로 되어있는지 배열을 통해 나타나 있는데

[Game]

[인트로, 호기심 쑥쑥, 개념 배우기, 교과서 읽기, 교과서 활동, 한눈에 정리하기, 문제]

[실력 탄탄]

[인트로, 개념 다지기, 기본문제]

이런 식으로 볼 수 있고 여기서 인트로는 영상 강의의 인트로라고 생각했고 같은 시퀀스 내에 문제는 문제풀이라고 생각했다. 즉 한 배열내에서 인트로와 문제가 같이 들어 있다면 이는 영상강의 + 문제풀이가 함께 서비스되는 콘텐츠이라 판단했고 배열을 먼저 문

자열로 바꾼뒤 인트로 , 문자가 들어간 콘텐츠에 대해 정규표현식을 사용하여 조건에 맞는 콘텐츠를 추출.

접근

[초기]

처음에는 with문을 통해 4개의 분기를 나누고 이를 quarter1~4로 지정한 뒤 각 quarter별로 inner join을 하여 마지막 select 문을 통해 조회하는 코드로 작성함.

[최종]

하지만 시간이 너무 오래 걸려서 필요한 테이블별로 join을 여러 번하는 것을 최대한 방지하고자 나누어 with 문을 작성. 마지막 select문에서 필요한 테이블들을 join하여 결과 출력

쿼리

[with]

- filtered_study : text_biz_dw.e_study
 - mcode, userid, 학습시간, [인트로, 문제] 글자 추출을 위해 이용
- filtered_members : text_biz_dw.e_member
 - grade를 이용하기 위함
- user_data :
 - filtered_study와 filtered_members를 inner join
- content_meta : text_biz_dw.e_content_meta
 - mcode를 가져오기 위함
- study_data_with_content
 - user_data와 content_meta를 inner join
- test_data : text_biz_dw.e_test
 - 문제수, 정답수, 평균 점수 집계를 위함
 - 콘텐츠별 조회를 위해 mcode로 group by

[조회]

study_data_with_content와 test_data를 left join하여 최종 조회.