

- Discussion

Discussion 1-1

1-1. List four applications you have used that most likely employed a database system to store persistent data. (other than the ones mentioned in class).

Facebook: 많은 사용자 정보, 게시물, 좋아요 정보, 팔로우 등 다양한 정보를 저장하기 위해

Youtube: 전세계에서 업로드되는 동영상, 동영상 별로 좋아요와 싫어요, 댓글, 또 댓글의 좋아요 싫어요, 댓글의 답글 등을 저장. 사용자들의 구독 채널, 알림 설정 등 사용자 정보도 저장.

Instagram: facebook 과 비슷하게 많은 사용자와 그 사용자들이 올린 사진을 저장.

Kakaotalk: 사용자 정보, 채팅방, 대화 내역, 미디어들을 저장해야 함.

Discussion 1-2

1-2. Explain the concept of *physical data independence*, and its importance in database systems.

데이터베이스 시스템은 View level - Logical level - physical level 의 세 단계로 나뉘어지는데, 다른 레벨의 변경 없이 각각의 레벨의 변경을 할 수 있어야 한다. 즉 아래 레벨에서의 데이터가 어떻게 저장되어 있는지에 대해 신경쓰지 않아도 됨을 보장하기 위해 각각 level 의 independence 가 필요하다. Physical level independence 는 logical level 에서 physical level 내부를 몰라도 됨을 보장한다.

Discussion 1-3

1-3-A. Faculty table 이므로 학부의 연구실에 대한 정보인 것 같다. ID, name, salary, dept_name, building, budget 의 column 들이 있다. Id 는 고유 번호로 데이터를 구분할 수 있는 정보이고 나머지 column 들은 그렇지 못하다.

1-3-B. 데이터를 구분할 수 있는 ID 가 5 자리 number 로 되어있어서 10 만명을 넘어가면 ID 값이 고유함을 보장하지 못할 것 같습니다. 또 동일한 학부, 빌딩을 사용하는 유저가 있을 경우 중복된 데이터가 저장될 수 있습니다.

Discussion 1-4

1-4-A. 카카오톡 같은 앱의 테이블 5 개

유저 정보 테이블 / 친구 목록 테이블 / 각 친구들의 상태메시지, 프로필사진 테이블 / 채팅방 목록 테이블 / 각 채팅방의 대화와 미디어 테이블

1-4-B. 우선 유저 정보를 통해 다른 정보를 접근해야 하므로 유저 정보를 retrieve 하는 쿼리가 top query 일 것 같습니다.

Discussion 1-5

데이터가 파일 시스템에 저장될 경우 생길 가능성이 있는 문제들.

윈도우, OS X, 리눅스 등 운영체제마다 다른 파일 시스템을 사용하므로 각각의 파일 시스템에 맞춰 다르게 처리하는 작업이 필요. 그리고 수많은 사용자들이 같은 시간대에 동일한 DB 를 접속하는데 파일 시스템을 사용하면 데이터의 consistency 보장하기가 어려울 것 같다. 그리고 atomic transaction 이 작업이 실패할 경우 보장되지 않는다. 예를 들어 카카오톡 메시지 보내기 같은 경우, A 에서 B 로 메시지를 보내면 A 와 B 의 파일 시스템에 모두 채팅 내역이 저장되어 있는데, 중간에 실패할 경우 A 는 메시지를 보냈다고 나오겠지만, B 는 메시지를 받지 못하는 경우가 생길 수 있음.

Discussion 1-6

검색엔진에서 사용되는 Keyword 쿼리와 데이터베이스 쿼리의 다른 점들. - 쿼리가 어떻게 정의되고 쿼리의 결과의 다른 점

DB 쿼리는 미리 정의되어 있는 column 들의 정확한 값을 받지만 keyword 쿼리는 column 이 미리 정의되어 있지 않다. 따라서 정확한 값이 아니라, 특정한 알고리즘을 이용해(ex: 구글의 PageRank) 우선 순위 순으로 데이터를 뽑는다. 결과는 DB 쿼리는 콘텐츠가 같으면 항상 같은 실행 결과가 나오지만 keyword 쿼리는 콘텐츠가 같아도 알고리즘을 수정하면 다른 실행 결과가 도출될 수도 있음.