

데이터 과제

환경

- 언어
 - Python, SQL
- DB
 - Oracle
 - Mysql

과제 내용

1. DB 설계

- 필수 필드 외에는 가급적 다양한 정보를 포함하도록 임의로 설계

- a. 상품 정보를 관리하기 위한 N 개 테이블 설계
 - i. 상품은 유니크한 코드로 관리됨 (상품 종류 + 색상)
 - ii. 필수 데이터
 1. 옵션, 사이즈, 재고, 브랜드
 2. 브랜드별 배송비
- b. 주문 정보를 관리하기 위한 M 개 테이블 설계
 - i. 한 번의 주문에는 X 개의 상품별 주문이 포함될 수 있음
예) 결제금액 200,000원 → 상품 1 (블랙, XL) x 3 + 상품 1 (레드, L) x 1 + 상품 2 (화이트, S)
 - ii. 필수 필드
 1. 주문자, 배송지
 2. 주문 상태 (결제전, 후, 배송 상태 등)
 3. 주문 금액, 할인 금액, 배송비
- c. 회원 정보를 관리하기 위한 L 개 테이블 설계
 - i. ID/PW 등 기본적인 정보를 포함하여, 다수의 암호화된 필드를 사용하여 구성할 것
- d. 회원의 페이지 뷰 이력을 담은 로그 테이블 생성
 - i. access_timestamp (%Y-%m-%d %H:%M:%S)
 - ii. user_id (user_x)
 - iii. page (page_x)

2. 데이터 생성 스크립트 개발 (with python)

- JSON 포맷의 회원 활동 로그는 제공되는 파일을 사용

- a. 회원 정보를 생성하는 스크립트
 - i. 활동 로그에서 추출한 회원 ID 수 만큼 회원 정보 생성
 - ii. ID 외 정보는 임의 입력
- b. 상품 정보 100개 랜덤 생성
 - i. 브랜드 A~Z 까지 분산하여 생성할 것
 - ii. 상품 정보가 0건은 브랜드는 없어야 함
 - iii. 브랜드별로 배송비는 모두 달라야 함
- c. 주문 정보 1000개 랜덤 생성
 - i. 단, 주문일자를 2024-01-01 ~ 현재까지로 분산하여 생성할 것
 - ii. 주문이 0건인 일자는 없어야 함

3. 분석 SQL 작성 (반드시 2개 DB 타입에 대해서 각각 모두 작성할 것)

- statement 1개로 작성할 것 (with, sub query 허용되나 테이블 생성/업데이트는 X)

- a. 월별 주문 회원 수, 주문 수, 주문 금액, 할인 금액, 배송비
- b. 매월 1회이상 주문한 회원 수
 - i. 이 중 월별 주문 금액의 최근 3개월간 평균이 100만원이 넘는 회원 수
- c. 3월에 A 브랜드를 처음으로 구매한 회원들의 주문 건
 - i. 회원 수, 주문수, 주문금액, 할인금액, 배송비 합계
- d. 회원 페이지 뷰 로그를 의 access time 기준으로 30분 단위 유저 세션 그룹핑 (해당 과제는 테이블 생성 허용)
 - 세션의 개념은 아래 예시를 참고할 것 (평균 값은 소수점 셋째 자리에서 반올림)
 - i. 일별 세션 수 = 지표1
 - ii. 월별 평균 세션 유지 시간 = 지표2
 - 세션 유지 시간 = 세션 내 "마지막 로그의 access_timestamp" - "최초 로그의 access_timestamp"
 - 단, 평균을 구할때 그 값이 0 인 경우는 모수에서 제외
 - iii. 세션별 평균 총 페이지 (+유니크 페이지) 수 (전 기간) = 지표3
- iv. 샘플 로그

```

1 {"access_timestamp": "2024-07-01 12:01:00", "user_id": "A", "page": "a"}
2 {"access_timestamp": "2024-07-01 12:21:00", "user_id": "B", "page": "c"}
3 {"access_timestamp": "2024-07-01 12:31:00", "user_id": "A", "page": "b"}
4 {"access_timestamp": "2024-07-01 12:40:00", "user_id": "B", "page": "a"}
5 {"access_timestamp": "2024-07-01 12:45:00", "user_id": "B", "page": "c"}

```

1. 지표 1 : 3
2. 지표 2 : $(0 + 0 + 24m) / (3 - 2) = 24m$
3. 지표 3 : $(1 + 1 + 3) / 3 = 1.67, (1 + 1 + 2) / 3 = 1.33$

제출 방법

1. github 에 올리고 repository url 제공
2. db 는 dump 하여 파일로 업로드