

정보시스템 데이터 개발자 사전과제

과제 설명

고객의 최근 1개월 거래 금액과 이전 3개월의 평균 거래 금액을 비교하여 고객 능력에 대한 점수를 구하고, 이를 바탕으로 고객의 행동위험 등급을 구하고자 합니다.

- (조건1) 원천 데이터는 다음 2개의 테이블로 구성되며, 각 테이블은 HDFS 에 있습니다.

customer		transaction	
customer_id	ID	transaction_id	ID
name	고객명	transaction_date	거래일시
birthday	생년월일	customer_id	고객ID
nationality	국적	amount	거래금액

- customer

- 전체 3,000만 건 (샘플은 1만건)
- 샘플데이터 : 정보시스템 데이터 개발자_샘플데이터_customer.csv

customer_id	name	birthday	nationality
1	A	19901225	KR
2	B	19850505	KR
3	C	19500625	KP

- transaction

- 1일 1,000만건 적재되며, 4개월치 거래 적재 = 12억건 (샘플은 12만건)
- 샘플데이터 : 정보시스템 데이터 개발자_샘플데이터_transaction.csv

transaction_id	transaction_date	customer_id	amount
1	2019-01-01 00:00:01	3001	543500
2	2019-01-01 00:00:44	3002	1633600
3	2019-01-01 00:01:27	3003	620400

- (조건2) 결과는 mysql의 테이블에 저장되며, 결과를 구하기 위한 기준 테이블이 있습니다.

customer_risk_result	
id	ID
base_yymm	기준년월
customer_id	고객ID
risk_score	위험점수
risk_grade	위험등급

grade_base	
id	ID
grade	등급
min_score	최저점수
max_score	최고점수

- customer_risk_result
 - 매월 수행한 결과가 적재됨
 - 위험점수 : 0~100점의 정수
 - 위험등급 : 위험점수와 grade_base 를 이용하여 구한 위험등급 (A/B/C)
- grade_base
 - 데이터는 다음과 같습니다.

id	grade	min_score	max_score
1	A	71	100
2	B	31	70
3	C	0	30

- (조건3) 고객능력(위험점수)는 다음과 같이 정의 합니다.
 - 고객능력 = {(최근 1개월 거래금액) / (이전 3개월 거래금액 평균)} 의 누적확률
 - 누적확률 = (개별값 - 최소값) / (최대값 - 최소값)
 - 따라서, 고객능력은 0~1 사이의 값으로 구해지며, 소수점 이하 2째자리까지 반올림으로 구하여 100점으로 환산합니다.

과제 1

위와 같이 적재되는 데이터를 이용하여 통계를 구하는 시스템을 구축한다고 할 때, 각 단계별로 어떤 툴을 어느 요소에 사용할 것인지 아시는 범위 내에서 설계해주세요.

과제 2

위의 요건은 매월 1회 실행되어 결과가 적재되어야 하는 배치 어플리케이션 입니다. 사용 가능한 언어(ex. Spring boot batch 등)를 이용하여 어플리케이션을 작성해주세요.

참고사항

- 회신은 pdf, doc(word), ppt, google docs, GitHub 등 자유롭게 작성하여 주시면 됩니다.
- 예쁘거나 화려한 것 보다는 핵심 내용을 잘 담아주세요.
- 과제 2의 배치 어플리케이션에는 README.md 파일을 추가하여 개발프레임워크, 문제해결 전략, 빌드 및 실행 방법 등을 기술하세요.
- 단위 테스트(Unit Test) 코드를 개발하여 각 기능을 검증하세요.
- 데이터의 건수가 대용량인 점을 고려하여 설계와 개발이 되어야 하며, 중간 테이블이나 멀티 배치 등 구현 방법에는 제약이 없습니다.