

## 데이터 특성 스캔 및 결합 시스템 구현

### 1. 개요

당신은 풀 스택 개발자로서 RDBMS 환경에서 테이블의 속성 도메인 스캔, 결합 키 속성의 관리, 이중 데이터의 결합 기능을 제공하는 [데이터 특성 스캔 및 결합 시스템]의 개발 의뢰를 받았다.

개발하고자 하는 시스템의 목적은 테이블의 속성별 도메인의 특성을 스캔하고, 다른 테이블과 결합 가능한 키 속성을 찾아 관리한다. 스캔이 완료된 테이블에 대하여 사용자가 속성의 조건에 맞는 테이블을 검색하고 결합하는 기능을 제공할 예정이다.

### 2. System Goals

당신이 구현해야 할 시스템은 [데이터 특성 스캔 및 결합 시스템]으로 단일 테이블의 속성별 도메인을 스캔하여 특성과 다른 테이블과 결합 가능한 결합 키 속성을 추천하고 대표 결합키를 매핑하여 관리한다. 또한 사용자가 지정한 조건에 맞는 테이블을 검색하고 대표 결합키를 사용하여 이중 데이터를 결합하는 기능을 제공한다. 해당 서비스를 구현하기 위해 당신은 서비스 이용을 위한 UI(클라이언트)와 서비스 로직을 구현하기 위한 백 앤드(Back-End) 서버, 속성 도메인 스캔 결과와 데이터 결합 결과를 저장, 관리하기 위한 데이터베이스를 구축해야 한다.

이와 같은 목적을 이루기 위해 시스템은 구체적으로 다음과 같은 조건을 만족해야 한다.

- ✓ RDBMS(관계형 데이터베이스 관리시스템)와 연동하고 단일 테이블 정보가 포함된 CSV 파일을 읽어 DB에 저장할 수 있어야 한다.
- ✓ 단일 테이블 속성의 도메인을 스캔하여 테이블에 포함된 속성들의 특성을 파악하고 결합 키로 사용할 수 있는 속성을 추천할 수 있어야 한다.
- ✓ 테이블에서 사용자가 지정한 속성을 삭제하거나 데이터 타입을 변경할 수 있어야 한다.
- ✓ 결합 키로 추천된 속성에 <대표 결합 키>를 매핑하여 관리할 수 있어야 한다.
- ✓ 사용자가 지정한 조건에 맞는 테이블을 검색할 수 있어야 한다.
- ✓ <대표 결합키>를 사용하여 이중 테이블을 결합할 수 있어야 한다.
- ✓ 테이블의 속성 스캔 결과, 데이터 결합 결과를 CSV로 내려 받을 수 있어야 한다.

### 3. Glossary Of Terms

이하 본문에서 사용하는 용어들은 아래의 용어 정의를 따라 사용된다.

- **사용자:** [결합 분석용 데이터 품질 정합성 측정 시스템] 서비스를 이용하는 유저를 지칭한다.
- **클라이언트:** 사용자가 서비스를 이용하기 위하여 사용하는 UI를 지칭한다.
- **백엔드(Back-End):** 클라이언트로부터 요청을 받아 데이터베이스의 생성/조회/갱신/삭제와 서비스에 관련된 기능을 수행하고 수행 결과를 응답으로 클라이언트에 전송한다.
- **DB:** 해당 시스템의 RDBMS(관계형 데이터베이스 관리 시스템)을 사용하여 구축된 데이터베이스를 의미한다.
- **속성 도메인 스캔:** 테이블의 속성을 <수치 속성>과 <범주 속성>으로 분류하고 각 속성의 도메인에 대하여 NULL 레코드 수, 최대값, 최소값과 같은 특성을 확인하는 과정을 의미한다.
- **수치 속성:** 관측된 값이 수치로 측정되는 속성을 의미한다. 데이터 타입이 INTEGER, FLOAT, DOUBLE과 같이 실수 또는 정수로 표현된다.
- **범주형 속성:** 관측 결과가 몇 개의 범주 또는 항목의 형태로 나타나는 속성을 의미한다. 수치형 속성의 데이터 타입을 제외한 나머지 데이터 타입을 가진다.
- **대표 속성:** 해당 테이블에서 대표되는 속성을 의미한다. 사용자가 데이터를 보고 판단하여 대표 속성을 지정한다.
- **표준 대표 속성 사전:** 속성의 성질을 표준화하여 사전과 같이 목록으로 관리하는 것으로 대표 속성과 매핑하는 대상이 된다. 표준 대표 속성 사전의 기본 값으로는 <금융정보>, <학업정보>, <회원정보>, <건강정보>가 있으며, 사용자가 필요에 따라 추가할 수 있다.
- **결합 키 속성:** 해당 데이터에서 중복되는 데이터의 값의 비율이 10% 미만이고 다른 테이블의 속성과 결합(Join)할 수 있는 후보군이 되는 속성이다.
- **표준 결합 키 사전:** 결합 키의 특성을 표준화하여 사전과 같이 목록으로 관리하는 것으로 대표 결합키와 매핑하는 대상이 된다. 표준 결합 키 사전의 기본 값으로는 <전화번호>, <이메일주소>, <IP주소>, <차량번호>, <주민등록번호>가 있으며, 사용자가 필요에 따라 추가할 수 있다.
- **대표 결합 키:** 테이블의 결합 키 속성에 표준 결합 키 사전의 항목이 매핑 된 결합 키를 의미한다

다. 예를 들어 테이블A에 "phone\_num"이라는 결합 키가 있을 때 표준 결합 키 사전의 <전화번호>와 매핑할 경우 "phone\_num"은 해당 테이블의 대표 결합 키가 된다. 한 테이블에 대표 결합 키는 여러 개 존재할 수 있다.

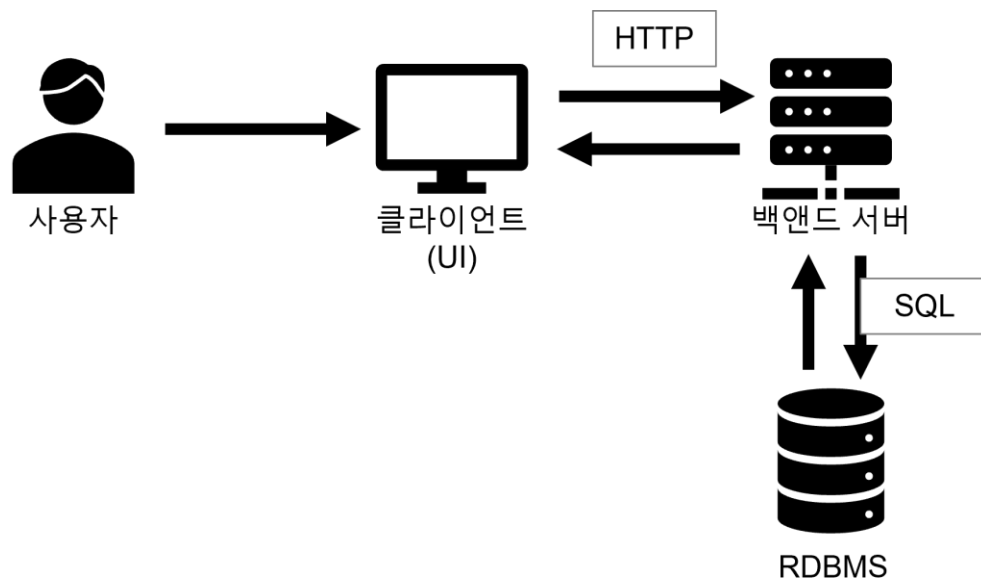
- **데이터 결합:** 대표 결합 키를 사용하여 복수의 테이블을 결합(Inner Join)하여 하나의 테이블로 합치는 것을 의미한다.
- **결합 성공률:** 복수의 테이블을 결합하였을 때 생성되는 레코드의 비율을 의미한다.

## 4. System Requirements Specification

### 4.1 시스템 구조 명세

- ✓ 시스템은 상용 RDBMS 사용을 기반으로 테이블 정보가 포함된 CSV 파일 또는 테이블 정보를 읽어 속성별 데이터를 분석할 수 있어야 한다.
- ✓ 상용 브라우저(크롬, 사파리, IE, 파이어폭스 등)를 이용하거나 독립적인 소프트웨어를 사용하여 UI를 제공해야한다.
- ✓ 시스템은 클라이언트, 백엔드, DB로 구성되며 클라이언트와 백엔드 서버의 통신은 HTTP 기반의 REST API 사용을 권장한다. 백엔드 서버의 programming 언어는 제한하지 않는다.

<그림 ㄴ> 데이터 특성 스캔 및 결합 시스템 구조

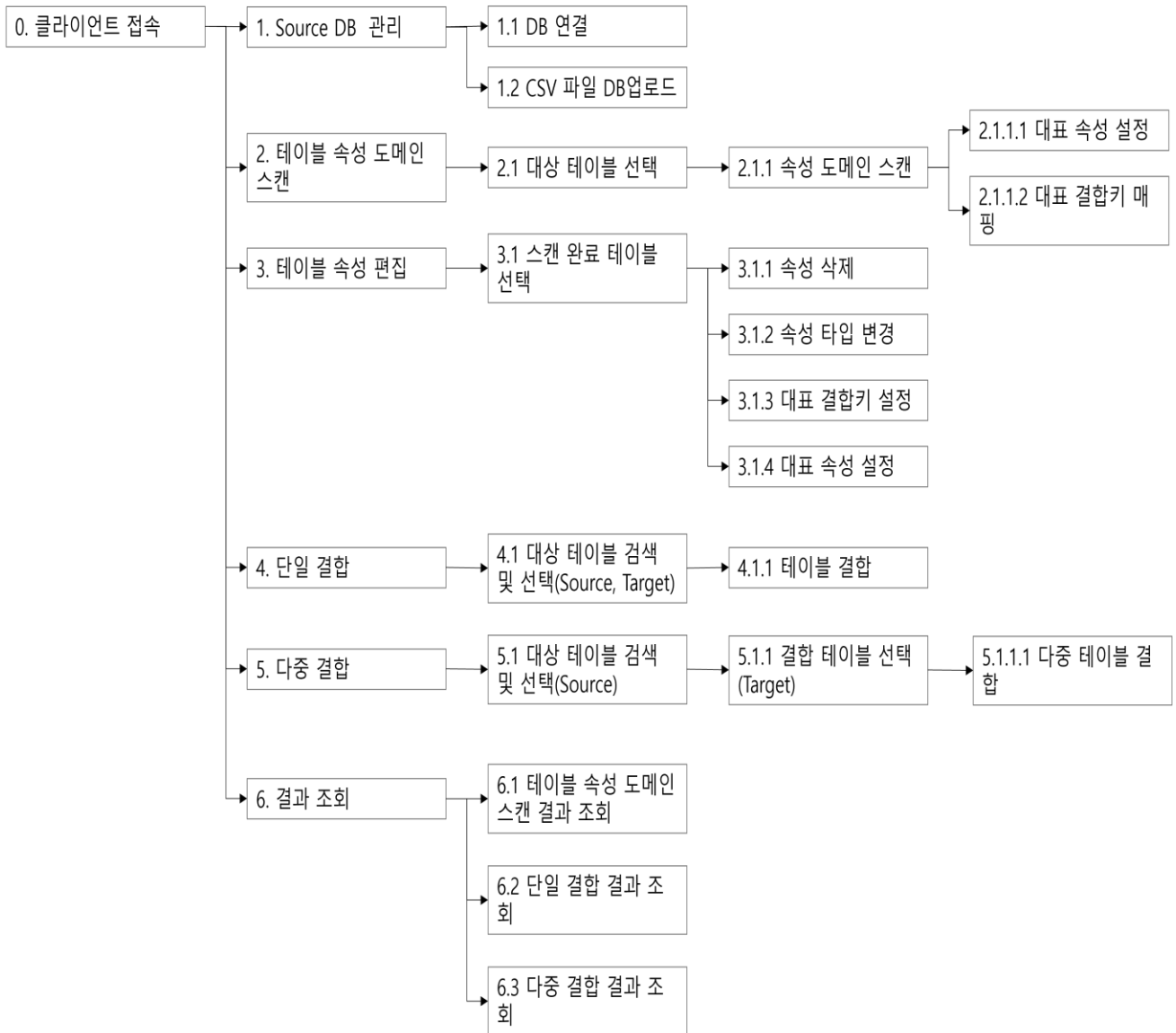


- ✓ 구현과 관련하여 언어나 프레임워크에 대한 제한조건은 없으나 ORM(Object-Relation Mapping)과 같은 구현에 SQL을 직접적으로 사용하지 않는 툴은 사용할 수 없다. (ex: Python의 sqlalchemy, Rails의 Active Record 등)

## 5. 기능 설계

본 시스템은 최소 다음 기능을 만족하도록 설계되어야 한다.

<그림 1> 시스템 기능 설계도



각 번호에 해당하는 기능에 대한 설명은 다음과 같다.

## 1 DB 연결 관리

시스템을 사용하기 위하여 데이터를 포함하고 있는 데이터베이스(이하 DB)의 연결을 관리한다. 사용자로부터 DB 정보를 입력 받아 연결 상태를 관리하고, 테이블 정보가 포함된 CSV 파일을 테이블로 생성하는 기능을 한다.

### 1.1 Source DB 접속정보 입력

사용자로부터 테이블 정보가 포함된 DB의 정보를 입력 받고 연결을 수행한다. 입력 받는 정보는 다음과 같다.

- **사용자 DB 입력정보:** DB 종류(e.g. MariaDB, MySQL, MSSQL 등...), Host/IP 정보, 연결 PORT 정보, 사용자 계정명, 사용자 비밀번호, 데이터베이스명

### 1.2 CSV 파일 DB업로드

DB가 연결된 상태에서 사용자로부터 테이블 정보가 포함된 CSV를 업로드 받아 현재 연결된 DB의 테이블로 생성한다. 생성되는 테이블의 이름은 확장자(.csv)를 제외한 파일명이 된다.

## 2 테이블 속성 도메인 스캔

현재 연결된 DB에 포함된 테이블 중 하나를 선택하여 해당 테이블의 속성별 도메인을 스캔하여 지표로 나타내고 결합 키 속성 추천과 대표 결합키 매핑을 수행한다.

### 2.1 대상 테이블 선택

사용자에게 현재 DB에 포함된 테이블 목록을 보여주고 속성 도메인 스캔을 수행할 대상 테이블을 선택할 수 있도록 한다.

#### 2.1.1 속성 도메인 스캔

2.1에서 사용자가 선택한 테이블의 속성을 데이터 타입에 따라 <수치 속성>과 <범주 속성>으로 분류하고 각 속성에 대하여 특성을 지표로 나타내고 결합 키를 추천한다.

<그림 2> 원본 테이블 예시 (고객정보)

고객번호	고객명	연락처	성별	나이	직업	월소득
C_001	고객A	010-1234-5678	M	52	개인사업자	1,000
C_002	고객B	010-2345-6789	M	NULL	NULL	200
C_003	고객C	010-3214-8589	F	34	변호사	2,000
C_004	고객D	010-7594-3821	M	21	유튜버	100

다음과 같은 원본 테이블이 있을 때, 속성의 값이 수치(numeric)인 "나이", "월소득"은 <수치 속성>이 되며, 수치 속성을 제외한 나머지 속성인 "고객번호", "고객명", "연락처", "성별", "직업"은 <범주 속성>이 된다.

<수치 속성>과 <범주 속성>에 대하여 도메인 스캔을 수행하고 각 분류에 따라 다음 지표를 구하고 사용자에게 보여준다.

<그림 3> 수치 속성 도메인 스캔 예시

속성명	데이터 타입	NULL 레코드 수	NULL 레코드 비율	상이 수치 값	최대 값	최소 값	0 레코드 수	0 레코드 비율	대표 속성	결합키 후보	대표 결합키
나이	INTEGER	1	0.25	4	52	21	0	0	-	O	-
월소득	DOUBLE	0	0	4	2,000	100	0	0	금융정보	O	-

<수치 속성>의 도메인 스캔 결과 지표의 설명은 다음과 같다.

- **NULL 레코드 수:** 레코드의 속성 값이 공백(NULL)인 레코드의 수
- **NULL 레코드 비율:** NULL 레코드 수 / 전체 레코드 수
- **상이 수치 값:** 해당 속성의 distinct 값
- **최대 값:** 해당 속성의 MAX 값
- **최소 값:** 해당 속성의 MIN 값
- **0 레코드 수:** 속성의 값이 0인 레코드의 수
- **0 레코드 비율:** 0 레코드 수 / 전체 레코드 수

<그림 4> 범주 속성 도메인 스캔 예시

속성명	데이터 타입	NULL 레코드 수	NULL 레코드 비율	상이 범주 값	특수문자 포함 레코드 수	특수문자 포함 레코드 비율	대표 속성	결합키 후보	대표 결합키
고객번호	VARCHAR	0	0	4	4	1.0	-	O	-
고객명	VARCHAR	0	0	4	0	0	-	O	-
연락처	TEXT	0	0	4	4	1.0	-	O	전화번호
성별	CHAR	0	0	2	0	0	-	X	-
직업	VARCHAR	1	0.25	4	0	0	-	O	-

<범주 속성>의 도메인 스캔 결과 지표의 설명은 다음과 같다.

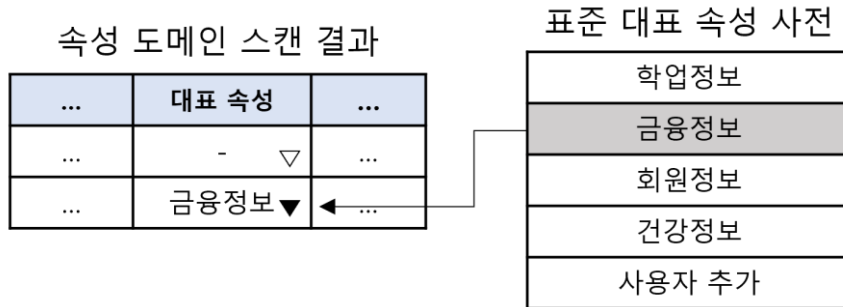
- **NULL 레코드 수:** 레코드의 속성 값이 공백(NULL)인 레코드의 수
- **NULL 레코드 비율:** NULL 레코드 수 / 전체 레코드 수
- **상이 범주 값:** 해당 속성의 distinct 값
- **특수문자 포함 레코드 수:** 한글, 영문, 숫자를 제외한 모든 특수문자를 포함한 레코드 수
- **특수문자 포함 레코드 비율:** 특수문자 포함 레코드 수 / 전체 레코드 수

<수치 속성>, <범주 속성>에서 전체 레코드 수에 대하여 해당 속성의 상이 수치(또는 범주)값의 비율이 0.9보다 클 경우 "결합키 추천"의 값을 추천으로 한다.

2.1.1.1 대표 속성 설정

해당 테이블에서 사용자가 대표될 수 있다고 생각하는 속성에 대하여 표준 대표 속성 사전의 값과 매핑한다. 하나의 속성은 하나의 표준 대표 속성 사전의 값과 매핑할 수 있으며, 필요에 따라 사용자가 값을 추가할 수 있다.

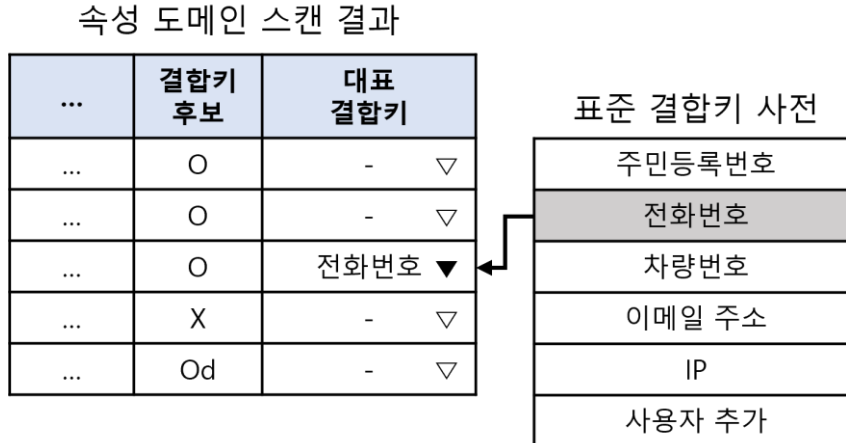
<그림 5> 대표 속성과 표준 대표 속성 사전 매핑 예시



2.1.1.2 대표 결합키 매핑(Mapping)

속성 도메인 스캔 결과 결합키로 추천된 속성에 대하여 사용자가 대표 결합키 매핑을 수행한다. 대표 결합키를 매핑할 때 표준 결합키 목록에서 하나를 선택하여 설정할 수 있으며, 표준 결합키는 필요에 따라 사용자가 추가할 수 있다.

<그림 6> 대표 결합키와 표준 결합키 사전 매핑 예시



3 테이블 속성 편집

도메인 스캔 및 프로파일링이 완료된 테이블의 속성을 삭제하거나 데이터 타입을 변경하고, 키 속성으로 사용 가능하다고 판단한 속성의 대표 결합키를 설정한다.

3.1 스캔 완료 테이블 선택

속성 도메인 스캔이 완료된 테이블의 목록을 보여주고 하위 작업을 수행할 테이블을 선택할 수 있도록 한다.



**3.1.1 속성 삭제**

3.1에서 선택한 테이블의 속성 중 사용자가 지정한 속성을 삭제(drop column)할 수 있는 기능을 제공한다.

**3.1.2 속성 타입 변경**

3.1에서 선택한 테이블의 속성의 데이터 타입을 변경(modify)할 수 있는 기능을 제공한다. 단, <수치 속성>의 데이터 타입을 문자열 데이터 타입(e.g. varchar, text, longtext 등...)으로 변경할 수는 있으나, <범주 속성>의 도메인에 문자 또는 특수문자가 포함되어 있을 경우, 정수형 또는 실수형 데이터 타입으로는 변환할 수 없다.

**3.1.3 대표 결합 키 설정**

도메인 스캔이 완료된 테이블에 대하여 2.1.1.1과 동일한 대표 결합키 설정을 수행할 수 있다. 도메인 스캔 완료 후 대표 결합키를 설정하지 않았거나 표준 결합키를 변경해야 할 경우 해당 기능을 사용한다.

**4 단일 결합**

속성 도메인 스캔이 완료된 테이블 중 대표 결합키가 설정된 테이블을 두개 선택하여 결합(inner join)을 수행한다.

**4.1 대상 테이블 검색**

속성 도메인 스캔 완료 후 대표 결합키 설정이 완료된 테이블의 목록을 보여주고, 사용자가 입력한 조건에 맞는 테이블을 검색하여 선택할 수 있는 기능을 제공한다.

<그림 7> 결합 대상 테이블 검색 조건 예시

테이블명	고객정보	표준 결합키	주민등록번호	대표속성	학업정보	속성명	직업
			전화번호		금융정보		
			차량번호		회원정보		
			이메일 주소		건강정보		
			IP				
			사용자 추가				

테이블명 일부 또는 전체, 표준 결합키, 속성명을 입력하여 결합하고자 하는 테이블을 검색할 수 있다. 결합에 사용할 테이블은 2개 선택할 수 있으며, 첫번째 테이블(Source)을 선택하면 해당 테이블과 같은 표준 결합키를 가지는 테이블의 목록을 보여주고 해당 테이블 목록에서 결합을 수행할 테이블(Target)을 선택하여 결합을 수행한다.

**4.1.1 테이블 결합**

4.1에서 선택한 첫번째 테이블(Source)과 두번째 테이블(Target) 테이블을 대표 결합키를 사용하여 결합(inner join)을 수행한다

<그림 8> 다중 테이블의 결합 예시

대상 테이블A(고객정보)

고객번호	고객명	연락처	성별	나이	직업	월소득
C_001	고객A	010-1234-5678	M	52	개인사업자	1,000
C_002	고객B	010-2345-6789	M	NULL	NULL	200
C_003	고객C	010-3214-8589	F	34	변호사	2,000
C_004	고객D	010-7594-3821	M	21	유튜버	100

대상 테이블B(건강정보)

고객명	폰번호	신장	체중
고객A	010-1234-5678	180	70
고객B	010-2345-6789	173	63
고객C	010-3214-8589	172	55

테이블A + 테이블B 결합결과

고객번호	고객명	연락처	성별	나이	직업	월소득	고객명	폰번호	신장	체중
C_001	고객A	010-1234-5678	M	52	개인사업자	1,000	고객A	010-1234-5678	180	70
C_002	고객B	010-2345-6789	M	NULL	NULL	200	고객B	010-2345-6789	173	63
C_003	고객C	010-3214-8589	F	34	변호사	2,000	고객C	010-3214-8589	172	55

Sample Query: select \* from 고객정보 join 건강정보 on 고객정보.연락처 = 건강정보.폰번호

다음 예시와 같이 결합 키를 사용하여 서로 다른 두 테이블을 결합하고 결합 결과를 사용자에게 보여준다.

<그림 9> 테이블 결합결과 예시

테이블A	테이블A 레코드 수	결합 키 속성A	테이블B	테이블B 레코드 수	결합 키 속성B	대표 결합키	결과 레코드 수	결합 성공률(W1)	결합 성공률(W2)	결합 진행상황	결합 테이블명
고객정보	4	연락처	건강정보	3	폰번호	전화번호	3	0.75	1.0	진행중	테이블A_테이블B_결합

테이블 결합 후 사용자에게 보여주는 결합결과의 지표에 대한 설명은 다음과 같다.

- **테이블A:** 결합 대상이 되는 테이블A 명칭
- **테이블A 레코드 수:** 결합 대상이 되는 테이블A의 레코드 수
- **결합키 속성A:** 테이블A에서 결합 키로 사용된 속성의 명칭
- **테이블B:** 결합 대상이 되는 테이블B 명칭
- **테이블B 레코드 수:** 결합 대상이 되는 테이블B의 레코드 수
- **결합키 속성B:** 테이블B에서 결합 키로 사용된 속성의 명칭
- **대표 결합키:** 결합키 속성A와 결합키 속성B에 매핑 된 대표 결합키
- **결과 레코드 수:** 테이블A와 테이블B 결합 결과 테이블의 레코드 수
- **결합 성공률(W1):**  $\frac{\text{테이블A} \cap \text{테이블B}}{\text{테이블A}}$
- **결합 성공률(W2):**  $\frac{\text{테이블A} \cap \text{테이블B}}{\text{테이블B}}$
- **결합 진행상황:** 결합의 진행 상황에 대하여 나타낸다. 결합이 진행중일 경우 결과 레코드 수, 결합 성공률(W1,W2)은 표시되지 않는다.
- **결합 테이블명:** 테이블A와 테이블B의 결합 결과를 저장하는 테이블의 명칭

첫번째 테이블(Source)과 두번째 테이블(Target)의 결합 결과를 사용자가 CSV로 내려 받을 수 있어야한다.

## 5 다중 결합

사용자가 선택한 대상 테이블과 결합이 가능한 모든 테이블을 찾아 결합을 수행한다.

### 5.1 대상 테이블 검색 및 선택(Source)

속성 도메인 스캔이 완료된 테이블 목록을 보여주고 4.1과 같은 조건으로 사용자가 테이블을 검색할 수 있는 기능을 제공한다.

테이블 목록 또는 검색 결과 테이블 목록에서 결합 대상이 되는 테이블을 1개 선택한다.

#### 5.1.1 결합 테이블 선택(Target)

5.1에서 선택한 테이블(Source)의 대표 결합키와 같은 값을 가지는 테이블을 모두 사용자에게 보여준다. 해당 테이블 목록 중 5.1에서 선택한 테이블(Source)과 결합을 수행할 테이블을 1개 이상 선택한다.

##### 5.1.1.1 다중 테이블 결합

5.1에서 선택한 대상 테이블(Source)과 5.1.1에서 선택한 테이블(Target)의 결합을 수행한다. 5.1에서 선택한 테이블(Source)을 테이블A, 5.1.1에서 선택한 테이블(Target)을 테이블B, 테이블C를 선택하였다고 가정하였을 때, 테이블A+테이블B, 테이블A+테이블C 결합을 수행한다.

해당 작업은 실시간으로 처리 후 결과를 보여주는 것이 아니라, 배치 작업으로 수행하여 결합 결과는 6.3에서 조회할 수 있도록 한다.

## 6 결과 조회

본 시스템을 사용하여 수행한 속성 도메인 스캔, 다중 테이블 결합, 단일 테이블 결합의 결과를 조회한다.

### 6.1 테이블 속성 도메인 스캔 결과 조회

사용자에게 속성 도메인 스캔이 완료된 테이블의 목록을 보여주고 사용자가 선택한 테이블의 2.1.1에 해당하는 속성 도메인 스캔 결과를 보여준다.

<범주 속성>과 <수치 속성>의 속성 도메인 스캔 결과를 CSV로 내려 받을 수 있는 기능을 제공한다.

### 6.2 단일 결합 결과 조회

4.1.1에서 수행한 단일 결합 결과를 사용자에게 보여주고 결합한 테이블을 CSV로 내려 받을 수 있는 기능을 제공한다.

단일 테이블 결합 목록 결과 또한 CSV로 내려받는 기능을 제공한다.

<그림 10> 다중 테이블 결합 결과 목록 예시

테이블A	테이블A 레코드 수	결합 키 속성A	테이블B	테이블B 레코드 수	결합 키 속성B	대표 결합키	결과 레코드 수	결합 성공률(W1)	결합 성공률(W2)	결합 진행상황	결합 테이블 명	CSV 저장
고객정보	4	연락처	건강정보	3	폰번호	전화번호	3	0.75	1.0	완료	고객정보-건강정보-결합	다운로드
판매내역	10	이메일	고객정보	4	메일주소	이메일주소	4	0.4	1.0	완료	판매내역-고객정보-결합	다운로드
...	...	...	...	...	...	...	...	...	...	...	...	...

### 6.3 다중 결합 결과 조회

5.1.1에서 수행한 다중 결합 결과의 목록을 사용자에게 보여주고 사용자가 선택한 결합 결과 테이블을 CSV로 내려 받을 수 있는 기능을 제공한다.

다중 테이블 결합 목록 결과 또한 CSV로 내려받는 기능을 제공한다.

<그림 11> 단일 테이블 결합 결과 목록 예시

테이블A	테이블A 레코드 수	결합 키 속성A	테이블B	테이블B 레코드 수	결합 키 속성B	대표 결합키	결과 레코드 수	결합 성공률(W1)	결합 성공률(W2)	결합 진행상황	결합 테이블 명	CSV 저장
고객정보	4	연락처	건강정보	3	폰번호	전화번호	3	0.75	1.0	완료 (1/2)	고객정보-건강정보-결합	다운로드
고객정보	4	연락처	판매내역	10	핸드폰번호	전화번호	4	1.0	0.4	진행중 (2/2)	고객정보-판매내역-결합	다운로드
...	...	...	...	...	...	...	...	...	...	...	...	...

## 6. 평가 기준

System Requirement Specification과 기능 설계를 바탕으로 아래의 필수 기능들을 구현해야 하며 최종 시연 평가는 아래의 필수 기능 리스트를 기준으로 평가한다. 한 팀을 A그룹과 B그룹으로 나누고 기능단위로 점수를 부여하여 공동 평가, A그룹 평가, B그룹 평가로 진행한다.

### ■ 공동 평가 항목

번호	기능 구분	상세 기능	점수
1	Source DB 관리	사용자로부터 DB 접속정보를 입력 받아 DB에 접속(connect)한다.	5
2		(추가점수) 사용자로부터 CSV 파일을 업로드 받아 연결된 DB에 테이블로 생성한다	3
3		(추가점수) 속성 도메인 스캔 후 <범주 속성>에서 사용자가 지정한 속성의 도메인과 발생 횟수를 CSV로 내려 받을 수 있다.	5
4		(추가점수) <범주 속성>의 도메인 스캔 결과를 BOXPLOT으로 UI에 표시한다.	5
5	단일 테이블 결합 조회, 다중 테이블 결합 조회	(추가점수) 사용자로부터 결합에 사용된 테이블명을 입력 받아 결합 결과 목록에서 해당 테이블과 결합을 수행한 결과만 표시한다.	3
6		(추가점수) 사용자로부터 결합 성공률 임계 값을 입력 받아 결합 성공률이 임계 값 이상인 결과를 표시한다.	3
7		(추가점수) 사용자로부터 결합에 사용된 테이블명과 결합 성공률 임계 값을 함께 입력 받아 입력 받은 테이블과 결합한 결과 결합 성공률이 임계 값보다 큰 결과를 표시한다.	5
8		(추가점수) 사용자로부터 최소 결합 결과 레코드 수를 입력 받아 결합 결과 레코드 수가 최소 결합 레코드 수보다 많은 결과를 표시한다.	3
9		(추가점수) 결합 진행 상태를 상태 바(Progress Bar)로 표시한다.	3

■ A그룹 평가 항목

번호	기능 구분	상세 기능	점수
1	테이블 속성 도메인 스캔	현재 DB에 저장된 테이블 목록을 표시하고 속성 도메인 스캔할 테이블을 선택한다.	5
2		<수치 속성>의 도메인 스캔을 수행한다.	0
3		속성의 데이터 타입을 보여준다.	3
4		"NULL 레코드 수", "NULL 레코드 비율"을 보여준다.	3
5		"상이 수치 값(distinct count)"을 보여준다.	3
6		"최대 값", "최소 값"을 보여준다.	3
7		"0 레코드 수", "0 레코드 비율"을 보여준다.	3
8		값이 중복되는 레코드의 비율이 10% 미만인 속성을 결합 키 후보 속성으로 설정한다.	3
9		결합 키 후보 속성과 표준 결합키 사전의 값을 매핑한다.	5
10		<수치 속성> 도메인 스캔결과를 사용자의 PC에 CSV로 내려받는다.	5
11		<범주 속성>의 도메인 스캔을 수행한다.	0
12		속성의 데이터 타입을 보여준다.	3
13		"NULL 레코드 수", "NULL 레코드 비율"을 보여준다.	3
14		"상이 범주 값(distinct count)"을 보여준다.	3
15		"특수문자 포함 레코드 수", "특수문자 포함 레코드 비율"을 보여준다.	3
16		값이 중복되는 레코드의 비율이 10% 미만인 속성을 결합 키 후보 속성으로 설정한다.	3
17		결합 키 후보 속성과 표준 결합키 사전의 값을 매핑한다.	3
18		<수치 속성> 도메인 스캔결과를 사용자의 PC에 CSV로 내려 받는다.	5
19		표준 대표 속성 사전의 값과 대표 속성을 매핑한다.	3
20	테이블 속성 편집	속성 도메인 스캔이 완료된 테이블 목록을 표시하고 속성 편집을 수행할 테이블을 선택한다.	4
21		사용자가 지정한 속성을 삭제하고 속성 도메인 스캔 결과에 반영한다	3
22		사용자가 지정한 속성의 데이터 타입을 변경하고 속성 도메인 스캔 결과에 반영한다	3
23		결합 키 후보 속성과 표준 결합키 사전의 값을 매핑한다.	3

■ B그룹 평가 항목

번호	기능 구분	상세 기능	점수
----	-------	-------	----

1	단일 결합	사용자가 입력한 조건에 맞는 테이블을 검색하고 결합을 수행할 테이블을 선택한다.	0
2		사용자로부터 테이블명 전체 또는 일부를 입력 받아 테이블을 검색한다.	4
3		표준 결합키를 사용하여 해당 표준 결합키와 매핑 된 대표 결합키를 보유한 테이블의 목록을 표시한다.	4
4		표준 대표 속성 사전의 값으로 대표 속성이 설정된 테이블을 검색한다.	
5		사용자로부터 테이블의 속성명의 전체 또는 일부를 입력 받아 테이블을 검색한다.	4
6		결합을 수행할 대상 테이블A와 테이블B를 선택한다.	5
7		테이블A와 테이블B의 결합(inner join)을 수행한다.	0
8		대표 결합키가 같은 속성을 사용하여 결합을 수행하고 결과 테이블을 DB에 저장한다.	5
9		테이블A와 테이블B의 정보(테이블명, 레코드 수, 결합키 속성명, 대표 결합키)을 보여준다.	5
10		결합 결과(결과 레코드 수, 결합 성공률_w1, 결합 성공률_w2, 결합 테이블명)를 보여준다.	5
11		결합 진행 상황을 UI에 표시한다.	3
12		결합 결과 테이블을 사용자의 PC에 CSV로 내려 받는다.	5
13	다중 결합	사용자로부터 테이블명 전체 또는 일부를 입력 받아 테이블을 검색한다.	4
14		표준 결합키를 사용하여 해당 표준 결합키와 매핑 된 대표 결합키를 보유한 테이블의 목록을 표시한다.	4
15		사용자로부터 테이블의 속성명의 전체 또는 일부를 입력 받아 테이블을 검색한다.	4
16		검색된 테이블 목록에서 결합을 수행할 테이블을 2개 이상 선택하고 결합을 수행한다.	
17	결과 조회	속성 도메인 스캔이 완료된 테이블 목록을 보여주고, 사용자가 선택한 테이블의 도메인 스캔 결과를 보여준다.	4
18		테이블의 도메인 스캔 결과를 CSV로 내려 받을 수 있다.	4
19		다중 테이블 결합 결과와 단일 테이블 결합 결과를 CSV로 내려 받을 수 있다.	4
20		다중 테이블 결합 수행 결과를 모두 보여주고 결합 테이블을 CSV로 내려 받을 수 있다.	4
21		단일 테이블 결합 수행 결과를 모두 보여주고 결합 테이블을 CSV로 내려 받을 수 있다.	4





## 7. 프로젝트 일정

프로젝트 일정은 다음과 같으며 수업 진행 상황에 따라 일부 변경될 수 있다.

(\* 각 보고서 별 필수 항목은 빨간색으로 표기)

구분	기한	설명
1차 보고서	~11/16(수) 23:59	<ul style="list-style-type: none"> <li>■ <b>ER-Diagram &amp; Mapping ERD to Relational Schema</b> <ul style="list-style-type: none"> <li>- 시스템의 기초가 되는 ER-Diagram을 작성하고, 이를 relational schema 로 매핑한다.</li> <li>- 이 시점에서 정한 내용들은 추후 변경 가능하지만, 전체적인 일정을 고려하여 최대한 최적화된 분석을 하도록 노력한다.</li> <li>- 팀원의 역할 분담 내용 및 일정, ERD 작성과 관련하여 가정한 내용을 기록한 보고서를 제출한다.</li> </ul> </li>   <li>■ <b>시스템 설계 문서 및 중간 구현내용</b> <ul style="list-style-type: none"> <li>- 시스템 설계 문서는 State Diagram, Sequence Diagram, Application Flowchart 등 다양한 표현 방식을 자유롭게 포함할 수 있으며, 각 팀의 아이디어를 최대한 나타낼 수 있는 것으로 한다.</li> </ul> </li>   <li>■ <b>Database table 생성 script 명기</b> <ul style="list-style-type: none"> <li>- 또한, 이 보고서에는 시스템이 사용하는 Database table들의 생성 script를 명기해야 한다.</li> </ul> </li> </ul>
최종 보고서	~12/21(수) 23:59	<ul style="list-style-type: none"> <li>■ <b>최종 결과 보고서 및 구현 내용</b> <ul style="list-style-type: none"> <li>- 완성된 시스템에 대한 최종적인 보고서는 결과물뿐 아니라 단계별 프로젝트 진행 과정을 포함하는 것이 좋다.</li> <li>- 완성도 평가를 위하여 기능별 구체적 구현 내용을 명기한다.</li> <li>- 시스템에 사용된 DB 전체를 dump하여 제출한다.</li> <li>- 보고서의 가독성 또한 평가의 요소가 된다.</li> <li>- 1차 보고서에서 변경/추가된 내용을 반드시 명기한다.</li> </ul> </li> </ul>
발표 및 시연	~12/21(수)	<ul style="list-style-type: none"> <li>■ <b>모든 조를 대상으로 시연 평가</b> <ul style="list-style-type: none"> <li>- 추후 공지되는 장소에서 모든 조 시연</li> </ul> </li> </ul>
SW 설치	추후 공지	<ul style="list-style-type: none"> <li>■ <b>개발된 SW를 연구실 서버에 설치</b></li> </ul>

## 8. 주의 사항

- ✓ 팀 구성은 6인 1조 원칙으로 하고 3명으로 구성된 그룹A와 그룹B로 나누어 평가를 진행한다.
- ✓ 각 단계별 제출기한을 엄수, 제출기한 이후의 결과물은 인정하지 않는다.
- ✓ 보고서와 프로그램을 [팀명]\_project1.zip / [팀명]\_project2.zip / [팀명]\_project3.zip 으로 압축하여 제출한다. 제출 방법은 추후 별도 공지한다.
- ✓ 각 단계별 수정사항이나 차별화를 위한 추가 기능 및 부각하고자 하는 장점이 있다면 보고서에 명시적으로 작성해야 한다.
- ✓ 과제의 목표는 관계형 데이터베이스를 이해하고 SQL을 사용하는 것이다. 따라서 본 목표에 부합하지 않은 구현물에 대해서는 감점이 있을 수 있으니 유의한다. 관계형 데이터베이스, SQL을 사용하지 않을 시 감점한다.