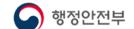


# 품질진단 기본교육 (범정부 품질진단기준에 따른 데이터 진단)

# Contents

- I. 데이터 품질진단과 평가지표의 이해
- Ⅱ. 데이터 품질진단 절차
- Ⅲ. 범정부 진단기준 및 오류유형







### 데이터 품질진단이란

준수되어야 하는 데이터 값의 유효성, 정확성, 표준 적용 여부 등이 도메인 속성에 위배되어 잠재적으로 업무에 영향을 줄 수 있는 오류데이터를 파악하기 위한 활동

#### 도메인 진단

컬럼의 고유한 특성을 진단

#### 업무규칙 진단

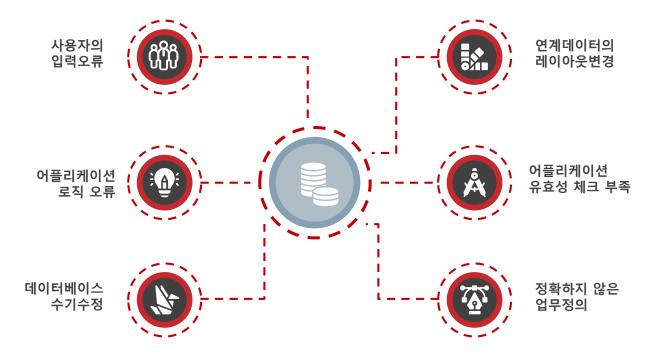
업무 요건 및 법령에 따른 규칙 진단

### 참조 관계 진단

자식/부모 테이블간 관계의 정합성을 진단

### 품질진단의 필요성

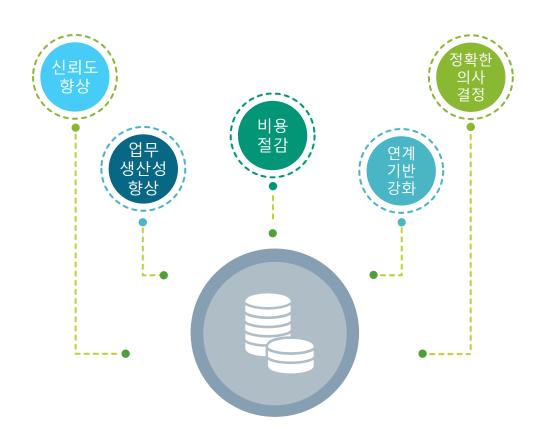
데이터 발생 및 가공 구간에서 충분한 검사가 이루어지지 않고 있기 때문에 업무적 또는 시스템적인 문제 발생 후 또는 데이터를 사용하려고 하는 시점에 오류데이터를 발견하는 경우가 대부분



### 품질진단수행의 기대효과

데이터의 신뢰도와 일관성 확보 및 향상 시켜 데이터 기반의 경영이 가능해지고, 다양한 분석정보를 이용하여 통합, 연계, 대국민 서비스 등의 토대를 마련

신규업무 개발 시 비용이 절감되며, 업무 및 IT부서의 생산성 및 운영 효율화를 통한 비용 절감



NIA 한국지능정보사회진흥원

### 2021년 공공데이터 품질관리 수준평가

공공데이터의 품질을 확보하기 위해 2개영역 11개 지표를 기준으로 평가 진행

#### 2021년 수준평가 영역

데이터 관리체계

#### 2021년 수준평가 지표

# 품질관리계획수립 데이터 표준확산 데이터 품질진단 오류신고 요구사항 분석 및 개선 예방적 품질관리 진단 데이터 구조 안정화 품질진단 결과 조치 개방데이터셋 오류 조치 데이터 연계 관리 데이터 오류율 개방표준 데이터셋 현행화 관리

#### 2021년 수준평가 11개 지표

| 영역       | 지표                  | 세부지표   | 배점 | 총괄      | DB담당자 |
|----------|---------------------|--|----|---------|-------|
|          | ① 품질관리 계획 수립        | 기관 데이터 품질관리 계획 수립 및 이행                         | 2  | 0       |       |
|          | ② 예방적 품질관리 진단       | 예방적 품질관리 진단<br>예방적 품질관리 진단결과 반영                | 10 | 0       |       |
| 데이터 관리체계 | ③ 데이터 표준 확산         | 데이터 표준 정의<br>데이터 표준 적용률                        | 10 |         | 0     |
| 32점      | ④ 데이터 구조 안정화        | 데이터 구조 정의 및 산출물 관리<br>데이터 구조 산출물 현행화           | 5  |         | 0     |
|          | ⑤ 데이터 연계 관리         | 연계데이터 관리 정보 정의<br>연계데이터 송수신 내역 관리              | 5  |         | 0     |
|          | ⑥ 데이터 품질진단          | 품질진단기준 정의 및 진단<br>업무규정에 따른 품질진단 규칙(BR) 정의 및 진단 | 15 |         | 0     |
|          | ⑦ 품질진단 결과 조치        | 품질진단에 따른 개선계획 수립 및 개선과제 정의<br>개선조치 이행률         | 23 |         | 0     |
| 데이터 값 관리 | ⑧ 데이터 오류율           | 데이터 오류율  | 15 |         | 0     |
| 68점      | ⑨ 오류신고 요구사항 분석 및 개선 | 데이터 품질 오류 분석 및 문제점 개선                          | 4  | $\circ$ |       |
|          | ⑩ 개방데이터셋 오류 조치      | 개방 데이터셋 품질오류 조치결과<br>오픈API서비스 품질오류 조치결과        | 4  | 0       |       |
|          | ⑪ 개방표준 데이터셋 현행화 관리  | 개방표준 데이터셋 현행화 관리                               | 7  | 0       |       |

#### 범정부 진단기준

| 순번                         | 진단기준   | 평가여부                      |
|----------------------------|--|---------------------------|
| 1                          | 시간순서 일관성                                       | 0                         |
| 2                          | 선후관계 정확성                                       | 0                         |
| 3                          | 컬럼 간 논리관계 일관성                                  | 0                         |
| 4                          | 계산식  | 0                         |
| 5                          | 참조관계   | 0                         |
| 6                          | 글자깨짐   |                           |
| 7                          | 공백, 특수문자                                       | DB<br>오류율,                |
| 8                          | 필수값  | 충족 <del>률</del><br>산정시 제외 |
|                            |  | 선정시 제외                    |
| 9                          | 중복데이터  | 선정시 제외                    |
| 9                          | 중복데이터<br>날짜 도메인                                | 산정시 제외                    |
|                            |  |                           |
| 10                         | 날짜 도메인   | 0                         |
| 10<br>11                   | 날짜 도메인<br>번호 도메인                               | 0                         |
| 10<br>11<br>12             | 날짜 도메인<br>번호 도메인<br>여부 도메인                     | O<br>O                    |
| 10<br>11<br>12<br>13       | 날짜 도메인<br>번호 도메인<br>여부 도메인<br>코드 도메인           | O<br>O<br>O               |
| 10<br>11<br>12<br>13<br>14 | 날짜 도메인<br>번호 도메인<br>여부 도메인<br>코드 도메인<br>금액 도메인 | O<br>O<br>O<br>O          |

#### 16개 진단기준

기존 12개 범정부 품질 진단기준에 4개 기준이 추가되어 16개로 변경 되면서 2021년 진단도구에서는 추가된 4개의 기준도 진단이 가능하도록 지원

#### DB는 16개 기준 중 12개 기준이 평가 대상

'글자깨짐, 공백/특수문자, 필수값, 중복데이터' 4가지 기준은 진단도구를 이용하여 진단은 가능하지만 지표 점수를 산정하기 위한 평가 기준에서는 제외

#### 업무규정·법령·지침 등에 위배되는 데이터 검증

업무규정·법령지침 등에 위배되는 데이터 검증을 위하여 업무규칙을 활용하여 데이터 정합성 점검을 수행하였는지 여부를 평가

### 평가기준 12개에 해당하는 컬럼만 진단대상임을 유의

선정DB의 진단 대상 테이블 전체 컬럼 중 평가기준 12개에 부합하는 컬럼을 대상으로 진단기준을 설정한다. 평가기준에 부합하지 않는 컬럼을 진단할 경우, 결과보고서의 정확도를 저하 시키므로 반드시 제외가 필요

| 테이블명 | 컬럼명  | 데이터 예시        | 진단대상 | 비고                        |
|------|------|---------------|------|---------------------------|
|      | 사원번호 | A001          | Υ    | 번호 도메인(생성규칙 존재)           |
|      | 사원명  | 홍길동           | N    | 명칭성 도메인으로 범정부 기준에 해당하지 않음 |
| 1101 | 부서코드 | 20            | Υ    | 코드 도메인                    |
| 사원   | 핸드폰  | 010-1234-5678 | Υ    | 번호 도메인(생성규칙 존재)           |
|      | 입사일  | 2020-05-04    | Υ    | 날짜 도메인(YYYY-MM-DD'로 관리)   |
|      | 비고   | 2019년 1월 개명함  | N    | 비정형 텍스트성 데이터로 해당 없음       |

## 품질점검 진단 대상 테이블 제외 및 범위 조정 가능

과거 데이터를 단순보존, 법령이나 관련 업무 지침 변경으로 데이터 진단기준 변경 등 진단범위 조정 가능임시, 백업, 로그 기록, 뷰, 시스템관리 테이블 등에 대해 품질진단 제외 가능

| 구분           | 예시   |
|--------------|--|
|              | ▶ 과거 데이터를 기록 차원에서 단순보존하고 있는 경우 (예시) 과거 수기 기록 행정데이터의 단순보존 등                               |
| 데이터 범위<br>조정 | ▶ 법령, 지침, 고시 등의 변경이나 기관 수행 업무의 변경 등으로 데이터를 더 이상 수집 및관리하지 않고 단순보존만 하는 경우                  |
|              | ▶ 삭제된 데이터를 단순 보존하는 경우 (예시) 물리적으로 데이터를 삭제하지 않지만 삭제 여부 칼럼으로 삭제된 데이터를 보존하는 경우               |
|              | ▶ 개발 등의 목적으로 임시 생성된 테이블 (예시)_TEMP,_TMP,_TEST 등   |
|              | ▶ 백업의 목적으로 데이터를 중복 생성한 테이블 (예시) _BAK, _BACKUP, 특정 날짜(_20200202) 등                        |
| 진단대상 테이블     | ▶ 시스템의 로그를 기록한 테이블 예시) 일반적으로 _LOG 등의 명칭을 포함  |
| 조정           | ▶ 사용자 접근이 허용된 자료만을 제한적으로 보여주기 위해 하나 이상의 기본 테이블로부터 유도된 가상 테이블 (예시)_VIEW,_VW 등             |
|              | ▶ 데이터베이스에 대한 메타데이터를 저장하는 등 시스템관리의 목적으로 생성된 테이블<br>(예시) SYS, DB 시스템 테이블, PLAN(실행계획) 테이블 등 |

#### 데이터 품질진단 지표 상세

| 지표   | 6. 데이터 품질진단                            | 영역                | 데이터 값 관리                  | 배점                    | 5   |
|------|--|-------------------|---------------------------|-----------------------|-----|
| 세부지표 | 품질진단기준 정의 및 진단                         |                   |                           |                       |     |
| 평가요지 | 범정부 진단기준(DB는 16개 항목 중 12개 적용)에 따라 품질진단 | 단대상을 선정           | 성하고 진단기 <del>준을</del> 설정하 | 였는지 여부 <mark>를</mark> | 를평가 |
| 점수   | 상세평                                    | <sup>]</sup> 가 기준 |                           |                       |     |
| 5점   | 범정부 진단기준 <del>충족률</del> 100%           |                   |                           |                       |     |
| 4점   | 범정부 진단기준 충족률 90% 이상                    |                   |                           |                       |     |
| 3점   | 범정부 진단기준 충족률 80% 이상                    |                   |                           |                       |     |
| 2점   | 범정부 진단기준 충족률 70% 이상                    |                   |                           |                       |     |
| 1점   | 범정부 진단기준 <del>충족률</del> 70% 미만         |                   |                           |                       |     |
| 0점   | 미제출                                    |                   |                           |                       |     |

#### 품질진단기준 정의 및 진단 - 충족률 산정방식

예) 진단대상 테이블 10개 중 9개 대상 선정, 진단대상 항목 100개 중 98개 진단기준 적용 및 수행

| 항목 상세 설명  |  |  |  |  |  |  |
|---|--|--|--|--|--|--|
| 기관에서 진단대상으로 선정한 테이블 수 + 기관에서 진단대상으로 선정하지 않은 테이블 중 지원단에서 '진단대상'으로 검토한 테이블 수  |  |  |  |  |  |  |
| 기관에서 품질지원단의 검토의견 확인 후 진단대상으로 선정이 확정된 최종 테이블 수   |  |  |  |  |  |  |
| ③ 기관에서 평가기준 12개 기준으로 도출한 진단 항목 수  | + 기관에서 도출되지 않은 진단 항목 중 지원단에서 '진단대상'으로 검토한 항목 수 |  |  |  |  |  |
| 기관에서 평가기군 12개 기군으로 도굴만 신한 영국 구  | - 기관에서 도출한 진단 항목 중 지원단에서 '진단 제외'로 검토한 항목 수     |  |  |  |  |  |
| (4) 기관에서 검토의견 확인 후 진단기준 적용을 확정하고 진단수행이 완료된 항목 수 - 기관에서 진단수행 한 항목 중 지원단에서 '진단 제외'로 검토한 항목 수 *진단기준이 적용된 모든 항목의 진단수행 상태가 완료 상태인 경우 충족률 산정이 가능, 실패 또는 미수행인 경우 충족률 산정 불가 |  |  |  |  |  |  |

#### 데이터 품질진단 지표 상세

| 지표         | 6. 데이터 품질진단                     | 영역                     | 데이터 값 관리      | 배점     | 10     |  |
|------------|---------------------------------|------------------------|---------------|--------|--------|--|
| 세부지표       | 업무규정에 따른 품질진단 규칙(BR) 정의 및 진단    |                        |               |        |        |  |
| 평가요지       | 업무규정 법령지침 등에 위배되는 데이터 검증을 위한 업목 | 무규칙을 설정하0              | 네이터 정합성 점검을 수 | -행하였는지 | 여부를 평가 |  |
| 점수         | 싱                               | ·세평가 기준                |               |        |        |  |
|            | 고유업무                            | 지원업무                   |               |        |        |  |
| 10점        | 업무규칙 8건 이상 설정 및 진단              | 업무규칙 5건 이상 설정진단        |               |        |        |  |
| <i>7</i> 점 | 업무규칙 5건 이상 7건 이하 설정 및 진단        | 업무규칙 3건 이상 4건 이하 설정 진단 |               |        |        |  |
| 4점         | 업무규칙 2건 이상 4건 이하 설정 및 진단        | 업무규칙 2건 설정 및 진단        |               |        |        |  |
| 1점         | 업무규칙 1건 설정 및 진단                 | 업무규칙 1건 설정 및 진단        |               |        |        |  |
| 0점         | 업무규칙 <mark>미적용</mark>           | 업무규칙 <mark>미적용</mark>  |               |        |        |  |

#### 데이터 오류율 지표 상세

| 지표   | 8. 데이터 오류율   | 영역      | 데이터 값 관리  | 배점 | 15 |
|------|--|---------|-----------|----|----|
| 세부지표 | 데이터 오류율  |         |           |    |    |
| 평가요지 | 범정부 데이터 품질진단기준(1 <mark>6개 중 12개</mark> )으로 측정한 오류율 | 에 따라 차등 | 적으로 점수 부여 |    |    |
| 점수   | 상세평  | 기 기준    |           |    |    |
| 15점  | 오류율 0.01% 미만                                       |         |           |    |    |
| 12점  | 오류율 0.1% 미만  |         |           |    |    |
| 9점   | 오류율 0.5% 미만  |         |           |    |    |
| 6점   | 오류율 1.0% 미만  |         |           |    |    |
| 3점   | 오류율 1.0% 이상  |         |           |    |    |
| 0점   | 미수행  |         |           |    |    |

진단기준 선정

#### 기관 자체 진단수행 절차 상세

진단항목 및 진단기준 변경을 허용하지 않음

# 진단항목 도출 및 진단기준 적용 DB담당자 진단기준 보고서 등록 품질관리수준평가관리시스템에 등록 진단기준보고서 검토 및 의견 작성 품질지원단 진단기준 보고서 검토 결과 등록 검토 결과 확인 범정부기준에부합하지 않는진단항목,기준이검토된내용 총괄 및 DB 담당자 검토내용은충족률및오류율지표점수에영향이있으니

반드시확인후 진행이필요

# 최초 진단결과 도출 검토내용반영및 추가 진단기준 도출 및 적용 진단수행 및 진단결과보고서 등록 품질관리수준평가관리시스템에 등록 최초 진단결과보고서 검토 및 의견 작성 최초 진단결과보고서 검토결과 등록 및 충족률 입력 전체진단항목의 수행상태가완료가이닌 경우 충족률 산정 불가 범정부 진단기준 충족률 확인 및 검토결과확인

충족률개선이필요한경우최종진단결과도출단계를 다시진행

충족률이적합하다판단되는경우,최초진단결과를 확정하고

최종진단결과 도출단계 진행

#### 최종 진단결과 도출

#### 오류데이터 개선 계획 수립 및 개선 활동

개선이어렵거나계획이없는경우최초진단결과를등록

#### 진단수행 및 진단결과보고서 등록

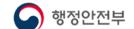
품질관리수준평가관리시스템에 등록

최초 진단결과보고서와 진단기준 및 진단항목 일치 여부 검토

최종 진단결과보고서 검토 결과 등록

최종결과확인

최종결과미감후에는이의제기가가능





# Ⅲ > 데이터 품질진단 절차

#### 계선계획이 없다면 보고서 작성 단계 시작 최종 결과보고서 작성 단계 진단대상 및 진단기준 도출 단계 진단수행 및 오류 개선 활동 단계 진단DB선정및 오류결과검토회의 진단수행 진단대상 테이블 및 범위 조건 확인 $\blacksquare$ V 테이블 · 컬럼 정보 수집 •-----최종 오류율 확정 오류데이터 확인 및 분석 (스키마수집) $\blacksquare$ 테이블 · 컬럼 데이터 분석 개선 계획 수립 품질진단 결과보고서 작성 (프로파일링) $\blacksquare$ 프로파일링 결과 상세 분석 개선 활동 도메인 업무 참조 개선이 반영된 오류 결과를 도출하기 위해 개선활동 완료후에 다시진단수행단계로이동 $\blacksquare$ 새롭게 도출된 오류 결과에서도 개선활동이 필요하다면 이 구간을계속반복 진단항목 도출 및 진단기준 적용

#### **SQL**

```
SELECT A.TABLE_NAME AS 테이블영문명,
      A.COMMENTS AS 테이블한글명,
      C.COLUMN ID AS 컬럼순서,
      B.COLUMN NAME AS 컬럼영문명,
      B.COMMENTS AS 컬럼한글명,
      C. DATA_TYPE AS 데이터타입,
      C.DATA_LENGTH AS 길이,
      C.NULLABLE AS 널여부,
      DECODE(D.CONSTRAINT TYPE, 'P', 'PK') AS PK
 FROM USER_TAB_COMMENTS A, USER_COL_COMMENTS B,
          USER_TAB_COLUMNS C,
       (SELECT X.TABLE_NAME, X.COLUMN_NAME,
              Y.CONSTRAINT TYPE
         FROM USER CONS COLUMNS X, USER CONSTRAINTS Y
        WHERE X.CONSTRAINT_NAME = Y.CONSTRAINT_NAME
          AND Y.CONSTRAINT_TYPE = 'P') D
WHERE A. TABLE NAME = B. TABLE NAME
   AND B.TABLE NAME = C.TABLE NAME
   AND B.COLUMN_NAME = C.COLUMN_NAME
  AND D.TABLE_NAME(+) = C.TABLE_NAME
  AND D.COLUMN_NAME(+) = C.COLUMN_NAME
   AND A.TABLE TYPE = 'TABLE'
  AND A.TABLE NAME IN ('테이블명') --필요시 테이블 지정
 ORDER BY 1, 3;
```

#### 테이블 · 컬럼 정보 수집 - 스키마수집

선정된 DB의 테이블과 컬럼 정보를 수집하여 정확한 진단대상을 도출

| 테이블명     | 컬럼영문명           | 컬럼명      | 데이터타입   | 길이 | 널여부 | PK |
|----------|-----------------|----------|---------|----|-----|----|
|          | CUST_ID         | 고객D      | VARCJAR | 5  | N   | PK |
|          | CUST_DVCD       | 고객구분코드   | VARCJAR | 5  | Y   |    |
|          | JUMIN_RGST_NO   | 고객주민등록번호 | VARCJAR | 15 | Υ   |    |
| CUSTOMER | BS_RGST_NO      | 사업자번호    | VARCIAR | 15 | Υ   |    |
| 고객       | CUST_STCD       | 고객상태코드   | VARCIAR | 5  | Υ   |    |
|          |                 |          | WYCIAD  | 10 | N   |    |
|          | VEND_CD         | 공급업체ID   | VARCJAR | 10 | Υ   |    |
|          | ITEM_NM         | 상품명      | VARCJAR | 15 | Υ   |    |
| ITEM_M   | ITEM_COLOR_CD   | 상품색상코드   | VARCJAR | 15 | Υ   |    |
| 상품       | STD_PRICE       | 기준단가     | NUMBER  | 22 | Υ   |    |
|          |                 |          |         | 8  | Y   |    |
|          | POPULAR_ITEM_YN | 인기상품여부   | VARCIAR | 1  | Υ   |    |
|          | SALE_RATE       | 할인율      | NUMBER  | 5  | Υ   |    |

품질진단 기본교육

```
SQL
-- 최소값 · 최대값 분석
SELECT '판매가격'
        , MIN(SAL_PRICE)
        , MAX(SAL_PRICE)
 FROM ORDER_DETAIL;
-- NULL · 공백 분석
SELECT COUNT(*) AS TOT CNT
     , COUNT(ETC_INFO) AS NOT_NULL_CNT
     , COUNT(CASE WHEN ETC_INFO IS NOT NULL
                 AND TRIM(ETC_INFO) IS NULL
              THEN '1' END) AS SPACE_CNT
FROM VENDER;
-- 단일값 . 중복값 분석
SELECT COUNT(*)
                              AS TOT_CNT
     , COUNT(COMB_RES_YN)
                                 AS NOT_NULL_CNT
     , COUNT(DISTINCT COMB_RES_YN) AS DISTINCT_CNT
     , (COUNT(COMB_RES_YN) - COUNT(DISTINCT COMB_RES_YN))
       AS DUPLICATE_CNT
FROM CUSTOMER;
```

#### 테이블 · 컬럼 데이터 분석 - 프로파일링

수집한 테이블과 컬럼 정보를 기반으로 프로파일링을 수행하여 실제 데이터의 통계 및 패턴을 분석

• 최소 값과 최대 값을 통하여 수치 값의 범위를 파악, 개별 빈도 값을 추출하지 않고 빠른 시간 내에 오류 징후를 파악 가능

| 테이블명 | 컬럼명  | 최소값     | 최대값        |
|------|------|---------|------------|
| 주문상세 | 판매가격 | -999999 | 9999999999 |

- NULL과 공백 건수를 통하여 값의 유무를 파악
- 컬럼의 사용 유무, 필수 입력 컬럼 등에 대한 오류 징후 파악 가능

| 테이블명 | 컬럼명  | 총건수 | NULL | 공백 |
|------|------|-----|------|----|
| 공급업체 | 기타정보 | 168 | 131  | 37 |

- 단일 값 및 중복 값의 건수를 통해 값의 분포를 파악
- 여부 컬럼 등 값의 범위가 정해진 컬럼에서 오류 징후 파악 가능

| 테이블명 | 컬럼명    | 총건수 | 단일값 건수 | 중복값 건수 |
|------|--------|-----|--------|--------|
| 고객   | 통합고객여부 | 10  | 3      | 7      |

#### 프로파일링 결과 상세 분석 - 도메인

고객 테이블 진단대상 도출 및 분석 예시

| 테이블명     | 컬럼명                 | 총건수 | 공백건수 | NULL건수 | NOT NULL 건수 | 최소값         | 최대값           | 유일값 건수 | 중복값 건수 |
|----------|---------------------|-----|------|--------|-------------|-------------|---------------|--------|--------|
|          | 고객D                 | 52  | 0    | 0      | 52          | 00001       | 00105         | 52     | 0      |
|          | 고객구분코드              | 52  | 0    | 0      | 52          | C01         | C02           | 2      | 50     |
|          | 고객주민등록번호            | 52  | 0    | 4      | 48          | 8007811228  | 8904161550616 | 47     | 1      |
| CUSTOMER | 사업자번호               | 52  | 0    | 48     | 4           | 10181397590 | 6038113200    | 4      | 0      |
| 고객       | 고객상태코드              | 52  | 0    | 0      | 52          | ST01        | 법인            | 6      | 46     |
|          | 고객 <del>등록</del> 일자 | 52  | 0    | 0      | 52          | 20000330    | 20161206      | 15     | 37     |
|          | 고객명                 | 52  | 0    | 1      | 51          |             | 황정민           | 50     | 1      |
|          | 통합고객등록여부            | 52  | 1    | 0      | 52          | 0           | Υ             | 6      | 46     |
|          |                     |     |      |        |             |             |               |        |        |

| 고객주민등록번호            | 13자리가 아닌 주민등록번호가 존재하며, 중복된 주민등록번호가 1건 존재 하는 것을 확인                 |
|---------------------|---|
| 사업자번호               | 사업자 번호가 모두 알맞은 번호 형식으로 발생되어 있을지에 대한 의문                            |
| 고객상태코드              | 코드아이디가 아닌 '법인' 이라는 코드 값 형태로 데이터가 발생 된 것을 확인. 다른 코드 값들의 유효성에 대한 의문 |
| 고객 <del>등록</del> 일자 | 고객등록일자에 날짜 포멧이 YYYMMDD 형태로 일관성 있게 생성되고 있는지 에 대한 의문                |
| 통합고객등록여부            | 통합고객등록여부에 Y,N이 아닌 0이라는 데이터가 존재하며, 6개 유형의 값이 존재하는 것을 확인            |

#### 프로파일링 결과 상세 분석 - 도메인

상품 테이블 진단대상 도출 및 분석 예시

| 테이블명   | 컬럼명      | 총건수 | 공백건수 | NULL건수 | NOT NULL 건수 | 최소값     | 최대값     | 단일값 건수 | 중 <del>복</del> 값 건수 |
|--------|----------|-----|------|--------|-------------|---------|---------|--------|---------------------|
|        | 상품 아이디   | 384 | 0    | 0      | 384         | A547    | EIK5961 | 384    | 0                   |
|        | 공급업체 아이디 | 384 | 0    | 0      | 384         | 15476   | 32088   | 96     | 288                 |
|        | 상품명      | 384 | 0    | 0      | 384         | LGU군인카드 | 환타오렌지   | 48     | 336                 |
| ITEM_M | 상품색상코드   | 384 | 0    | 310    | 74          | BL      | YE      | 4      | 70                  |
| 상품     | 생상년도     | 384 | 0    | 0      | 384         | 0000    | 9999    | 19     | 365                 |
|        | 판매여부     | 384 | 0    | 0      | 384         | N       | Υ       | 2      | 382                 |
|        | 기준단가     | 384 | 0    | 0      | 384         | -1      | 8500    | 334    | 50                  |
|        | 할인율      | 384 | 0    | 0      | 52          | -10     | 120     | 4      | 384                 |
|        |          |     |      |        |             |         |         |        |                     |

| 상품아이디               | 아이디의 데이터 생성 패턴이 다른 것을 확인                                    |
|---------------------|---|
| 공급업체 아이디,<br>상품색상코드 | 공급업체 아이디, 상품색상코드가 공통코드 테이블에서 관리되고 있는 유효한 값들로 발생되고 있는 것인지 의문 |
| 생산년도                | YYYY형태 포멧의 유효한 년도 데이터가 발생되고 있는 것인지 확인 필요                    |
| 판매여부                | 판매여부에 유일 값 건수가 2개이고 최소값 N, 최대값 Y로 이상이 없음을 확인                |
| 기준단가, 할인율           | 기준단가 및 할인율에 비정상적인 범위 값을 확인                                  |

#### 프로파일링 결과 상세 분석 - 도메인

주문상세 테이블 진단대상 도출 및 분석 예시

| 테이블명         | 컬럼명  | 총건수 | 공백건수 | NULL건수 | NOT NULL 건수 | 최소값        | 최대값        | 유일값 건수 | 중 <del>복</del> 값 건수 |
|--------------|------|-----|------|--------|-------------|------------|------------|--------|---------------------|
|              | 주문번호 | 55  | 0    | 0      | 55          | 00001AAAAA | E596       | 55     | 0                   |
|              | 고객번호 | 55  | 0    | 0      | 55          | 00001      | 00053      | 45     | 10                  |
|              | 상품ID | 55  | 0    | 0      | 55          | A547       | C994       | 4      | 51                  |
| ORDER_DETAIL | 주문일시 | 55  | 0    | 0      | 55          | 20200110   | 2020-08-12 | 54     | 1                   |
| 주문상세         | 배송일자 | 55  | 0    | 0      | 55          | 202005     | 20200102   | 41     | 14                  |
|              | 취소일자 | 55  | 0    | 54     | 1           | 20160803   | 20160803   | 1      | 0                   |
|              | 판매가격 | 55  | 0    | 0      | 55          | -9999      | 999999     | 49     | 6                   |
|              | 주문개수 | 55  | 0    | 0      | 55          | -10        | 98         | 45     | 10                  |

| 주문번호       | 주문번호의 최소값 최대값의 패턴이 큰 차이가 있 <del>음을</del> 확인 |
|------------|---|
| 주문일시       | 날짜 포멧이 통일되어 있지 않음을 확인                       |
| 배송일자       | 배송일자에 년월(YYYYMM) 포멧이 존재 하는 것을 확인            |
| 취소일자       | 취소일자에 다른 포멧이 존재하는지 확인 할 필요성이 보임             |
| 판매가격, 주문개수 | 판매가격 및 주문개수에 비정상적인 범위 값을 확인                 |

#### 도메인 분류

상세 분석 데이터를 기준으로 진단이 필요한 도메인을 분류 및 정의

| 통합고객등록여부            | Y,N이 아닌 0이라는 데이터가 존재.6개 유형의 값이 존재하는 것을 확인     | 여부   |                | ○○여부, ○○유무 등 2개의 유효 값을 진단        |  |  |
|---------------------|---|------|----------------|----------------------------------|--|--|
| 판매여부                | 판매여부에 유일 값 건수가 2개이고 최소값, N, 최대값 Y로 이상이 없음을 확인 |      |                | 선구, ○○ㅠ구 중 선택 규포 없을 선진           |  |  |
|                     |   |      | •              |                                  |  |  |
| 주문번호                | 주문번호의 최소값 최대값의 패턴이 큰 차이가 있음을 확인               |      |                |                                  |  |  |
| 고객주민등록번호            | 13자리가 아닌 주민등록번호가 존재, 중복된 주민등록번호가 존재 하는 것을 확인  | 비송   | ᄆᇺᄉᇺ           | 트이크 그성된 펜턴은 가느 가이 이과서 되다         |  |  |
| 사업자번호               | 사업자 번호가 모두 알맞은 번호 형식으로 발생되어 있을지에 대한 의문        | 번호   | · 군시, 굿시 :<br> | 등으로 구성된 패턴을 갖는 값의 일관성 진단         |  |  |
| 상품아이디               | 아이디의 데이터 생성 패턴이 다른 것을 확인                      |      |                |                                  |  |  |
|                     |   |      |                |                                  |  |  |
| 주문일시                | 날짜 포멧이 통일되어 있지 않음을 확인                         |      |                |                                  |  |  |
| 배송일자                | 배송일자에 년월(YYYMM) 포멧이 존재 하는 것을 확인               | Lŀπl | OITL OLL       | 시 값에 대한 포멧의 유일함과 유효함을 진단         |  |  |
| 취소일자                | 취소일자에 다른 포멧이 존재하는지 확인 할 필요성이 보임               | 날짜   | 일시: 일시         |                                  |  |  |
| 고객 <del>등록</del> 일자 | 날짜 포멧이 YYYMMDD 형태로 일관성 있게 생성되고 있는지 에 대한 의문    |      |                |                                  |  |  |
|                     |   |      | •              |                                  |  |  |
| 판매가격,주문개수           | 판매가격 및 주문개수에 비정상적인 범위 값을 확인                   | дон  | 수량 율           | <br>  정해진 범위 값의 유효 범위를 벗어난 값을 진단 |  |  |
| 기준단가, 할인율           | 기준단가 및 할인율에 비정상적인 범위 값을 확인                    | - H  | 19 E           | 영애선 남자 따라 규모 남자를 것이던 없을 선단<br>   |  |  |
|                     |   |      |                | •                                |  |  |
| 고객상태코드              | 코드아이디가 아닌 '법인' 이라는 코드 값 형태로 데이터가 발생 된 것을 확인   |      |                |                                  |  |  |
| 공급업체 아이디            | 공급업체 아이디가 유효한 업체 코드로 발생하는 것인지 의문              | 코드   | 개별코드·          | 공통코드 사용 컬럼의 유효성 진단               |  |  |
| 상품색상코드              | 상품색상코드가 유효한 값들도 발생 하는 것인지 의문                  |      |                |                                  |  |  |

0 >= 값 <=100

## 데이터 품질진단 절차

#### 도메인 진단기준 도출 및 적용

도메인 사상에 맞는 진단기준을 생성 하고 생성된 기준을 컬럼 도메인에 알맞게 적용 진단 적용시 진단수행이 필연적으로 발생하게 되고 그에 따른 오류율과 데이터를 확인하면서 적용한 진단기준에 대한 적합성을 판단해야 함

| 테이블명     | 컬럼영문명           | 컬럼명                 | 도메인 | 진단기준               |    |          |             |
|----------|-----------------|---------------------|-----|--------------------|----|----------|-------------|
|          | CUST_ID         | 고객D                 |     |                    |    | OH VAI   |             |
|          | CUST_DVCD       | 고객구분코드              | 코드  | 공통코드               | 여부 | 여부 Y/N   |             |
|          | JUMIN_RGST_NO   | 고객주민등록번호            | 번호  | 주민번호999999-9999999 |    | 여부 1/2   |             |
| CUSTOMER | BS_RGST_NO      | 사업자번호               | 번호  | 사업자번호 999-99-99999 |    | !<br>]   |             |
| 고객       | CUST_STCD       | 고객상태코드              | 코드  | 공통코드               | 번호 | 주민번호 999 | 999-9999999 |
|          | CUST_RGST_DT    | 고객 <del>등록</del> 일자 | 날짜  | 년월일 YYYYYMMDD      | 건오 | 사업자번호 9  | 99-99-99999 |
|          | CUST_NM         | 고객명                 |     |                    |    |          |             |
|          | INTE_CUST_YN    | 통합고객등록여부            | 여부  | 여부 Y/N             |    | 년월일 YYYY | MMDD        |
|          | ITEM_CD         | 상품ID                |     |                    | 날짜 | 년 YYYY   |             |
|          | VEND_CD         | 공급업체ID              |     |                    |    | 21111    |             |
|          | ITEM_NM         | 상품명                 |     |                    |    |          | 0 >= 값      |
| ITEM_M   | ITEM_COLOR_CD   | 상품색상코드              | 코드  | 상품색상 개별코드          | 금액 | · 수량 · 율 | 0 7h        |
| 상품       | STD_PRICE       | 기준단가                | 금액  | 0 >= 값             |    |          | 0 >= 값 <=   |
|          | CUST_RGST_DT    | 생상년도                | 날짜  | 년 YYYY             |    | 공통코드     |             |
|          | POPULAR_ITEM_YN | 인기상품여부              | 여부  | 여부 1/2             | 코드 | 884      |             |
|          | SALE_RATE       | 할인율                 | 율   | 0 >= 값 <=100       |    | 상품색상 개'  | 별코드         |
|          |                 |                     |     |                    |    |          |             |

### 업무규칙은 업무데이터와 관련된 모든 것에 내재되어 있다.



데이터 관리문서

데이터 관리문서에 명시된 데이터 규칙



법령

국가에서 시행하는 법령에 의한 규칙



업무담당자

업무담당자가 경험 지식으로 알고 있는 규칙



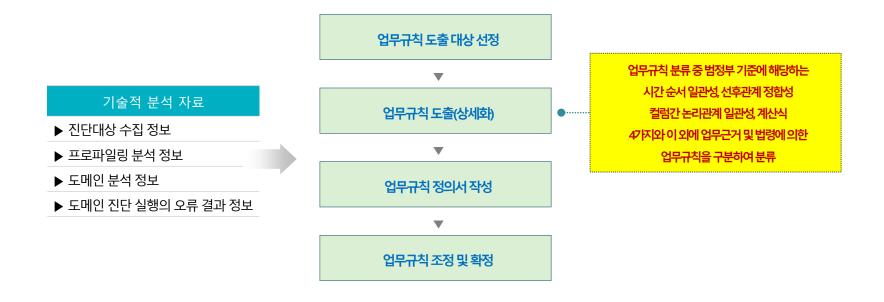
개발자

개발자의 경험 및 어플리케이션 코드에 존재하는 규칙

# 일반적으로 INSIDE-OUT 방식을 통해 업무규칙을 도출한다.

#### **INSIDE-OUT**

- 데이터 프로파일링으로 파악된 품질 문제를 통해 도출하는 방식
- 해당 기업의 비지니스 지식이 부족해도 가능한 기술적 분석



## 분석 자료를 토대로 업무규칙으로 관리할 컬럼 후보를 결정한다.

프로파일링 및 도메인 진단 결과를 참고하여, 복수 컬럼간의 선순위 파악 등 값이 정확하기 위한 조건을 비고란에 기재하여 복잡한 데이터 규칙을 도출하는데 활용 한다.

| 테이블명        | 컬럼영문명          | 컬럼한글명      | 도메인 | 진단기준       | 진단결과       |             |       | 업무규칙 관리대상 |  |  |
|-------------|----------------|------------|-----|------------|------------|-------------|-------|-----------|--|--|
| 네이글링        | 얼립청단정          | 28228      | 포메인 | 선인기판       | 총건수        | 오류          | 오류율   | 대상        | 비고                                       |  |
|             | STND_Y         | 기준년도       | 날짜  | YYYY       | 99,526,169 | 0           | 0.00% |           |  |  |
|             | IDV_ID         | 가입자일련번호    | 번호  | 8자리 숫자     | 99,526,169 | 0           | 0.00% |           |  |  |
|             | KEY_SEQ        | 진료내역일련번호   | 번호  | 8자리 숫자     | 99,526,169 | 0           | 0.00% |           |  |  |
|             | SEX            | 성별코드       | 코드  | 1(남), 2(여) | 99,526,169 | 0           | 0.00% |           |  |  |
|             | AGE_GROUP      | 연령대코드      | 코드  | 코드값 : 1~28 | 99,526,169 | 0           | 0.00% |           |  |  |
|             | SIDO           | 시도코드       | 코드  | 코드값 : 시도코드 | 99,526,169 | 0           | 0.00% |           |  |  |
| 진료내역정보      | RECU_FR_DT     | 요양개시일자     | 날짜  | YYYYMMDD   | 99,526,169 | 14850       | 0.01% | 0         | 니자이오 조크이지 이저                             |  |
| HSMT_TBGJME | RECU_LS_DT     | 요양종료일자     | 날짜  | YYYYMMDD   | 99,526,169 | 24875       | 0.02% | 0         | 시작일은 종료일자 이전                             |  |
|             | FORM_CD        | 서식코드       | 코드  | 서식코드       | 99,526,169 | 0           | 0.00% |           |  |  |
|             | MAIN_SICK      | 주상병코드      | 코드  | 주상병코드      | 99,526,168 | 56,618      | 0.06% | 0         | 심사불능, 반송에 대한 데이터는                        |  |
|             | SUB_SICK       | 부상병코드      | 코드  | 부상병코드      | 67,144,983 | 6,653       | 0.01% | 0         | 잘못된 코드가 존재 할 수 있음                        |  |
|             | EDEC_TRAMT     | 심결요양급여비용총액 | 금액  | 0이상의 숫자    | 99,526,169 | 308,69<br>1 | 0.31% | 0         | 심결본인부담금+심결보험자부담금<br>심사불능, 반송 데이터는 금액이 0원 |  |
|             | EDEC_SBRDN_AMT | 심결본인부담금    | 금액  | 0이상의 숫자    | 99,526,169 | 3           | 0.01% | Ο         |  |  |
|             | EDEC_JBRDN_AMT | 심결보험자부담금   | 금액  | 0이상의 숫자    | 99,526,169 | 20          | 0.01% | Ο         |  |  |

NIA 한국지능정보사회진흥원 27

# 프로파일링 및 도메인 진단결과의 오류데이터 분석을 통해 업무규칙을 상세화

프로파일링 결과

오류추정 데이터 분석

업무규칙 도출

주상병코드에 정의되지 않은 코드 값으로 오류가 발생함 심사불능 반송의 사유로 주상병코드 값이 잘못될 수 있음 주상병코드는 심사결과 코드가 '심사불능, 반송' 상태가 아닌 경우 유효함

#### 도메인 진단 SQL

SELECT 주상병코드,

COUNT(\*)

FROM 진료내역정보

WHERE 주상병코드 NOT IN (SELECT 주상병코드 FROM 상병코드정보)

GROUP BY 주상병코드;



#### 업무규칙 진단 SQL

SELECT A.주상병코드,

COUNT(\*)

FROM 진료내역정보 A, 진료내역상세 B

WHERE A.주상병코드 NOT IN (SELECT 주상병코드

FROM 상병코드정보)

AND A.진료내역일련번호 = B.진료내역일련번호

AND B.심사결과코드 NOT IN ('심사불능', '반송')

GROUP BY A.주상병코드;

NIA 한국지능정보사회진흥원

#### 업무 진단기준 도출 및 적용

업무요건을 테이블에 매핑하고 분석영역을 범정부 기준에 부합하도록 분류 범정부 기준 외에 업무규정과, 법령에 근거한 기준은 별도의 업무규칙으로 분류

| 테이블명              | 컬럼영문명           | 컬럼명      |  |  |
|-------------------|-----------------|----------|--|--|
|                   | CUST_ID         | 고객D      |  |  |
|                   | CUST_DVCD       | 고객구분코드   |  |  |
| CUSTOMER          | JUMIN_RGST_NO   | 고객주민등록번호 |  |  |
| 고객                | BS_RGST_NO      | 사업자번호    |  |  |
|                   | CUST_STCD       | 고객상태코드   |  |  |
|                   | CUST_RGST_DT    | 고객등록일자   |  |  |
|                   | ITEM_CD         | 상품ID     |  |  |
|                   | VEND_CD         | 공급업체ID   |  |  |
| ITEM_M            | STD_PRICE       | 기준단가     |  |  |
| 상품                | CUST_RGST_DT    | 생상년도     |  |  |
|                   | POPULAR_ITEM_YN | 인기상품여부   |  |  |
|                   | SALE_RATE       | 할인율      |  |  |
|                   | ORDER_DTM       | 주문일시     |  |  |
|                   | SAL_PRICE       | 판매가격     |  |  |
| ORDER_DTL<br>주문상세 | DELIVERY_DT     | 배송일자     |  |  |
| ᆍᇉᆼᄱ              | ORDER_QTY       | 주문개수     |  |  |
|                   | CANCEL_DT       | 취소일자     |  |  |

| ㆍ 컬럼간 논리관계 | - 고객상태코드가 활동인 경우 가입 일자가 반드시 존재해야 한다. |
|------------|--------------------------------------|
| • 컬럼간 논리관계 | - 고객상태코드가 탈퇴인 경우 탈퇴 일자가 반드시 존재해야 한다. |
| • 선후관계 정확성 | - 탈퇴자일가 있으면 반드시 가입일자가 존재해야 한다.       |
| • 시간순서 일관성 | - 가입일자는 탈퇴일자 보다 과거 시점이여야 한다.         |
|            |                                      |
| • 시간순서일관성  | 상품의 생산년도는 판매일자보다 미래 일수 없다.           |
| • 업무규칙     | - 당해년도 이전에 생산된 상품은 10프로 할인율을적용한다.    |
|            |                                      |
| • 계산식      | - 할인 가격은 판매금액을 초과할 수 없다.             |
| • 계산식      | - 월간 판매금액은 통계테이블의 월간 매출액과 일치해야 한다.   |
| • 시간순서일관성  | - 배송일자는 주문일자보다 미래일수 없다.              |
|            |                                      |

### 기관의 작성지침에 따라 일관되고 정형화된 양식으로 작성한다

"기관의 업무규칙 작성지침을 기반으로 업무담당자 및 IT담당자 모두가 이해하기 쉽게 작성 해야 하며 데이터의 개정, 변경 등의 관리를 위해 이력을 관리 해야 한다."

| 업무규칙 ID | BR-XXXX-001  | 작성자   | 김수영           | 작성일자    | 2016-12-13 |  |  |  |  |
|---------|--|-------|---------------|---------|------------|--|--|--|--|
| 시스템 명   | 건강정보시스템  | DB    | 건강정보DB        | 담당자     | 홍길동        |  |  |  |  |
| 업무규칙 명  | 진료내역정보의 주상병코드에 대한 코  | 크드 검증 |               |         |            |  |  |  |  |
| 업무규칙 내용 | 진료내역정보의 주상병코드는 심사불능 및 반송의 사유를 제외하고 상병코드정보의 주상병코드를 참조하여야 한다.  |       |               |         |            |  |  |  |  |
| 근거 규정   | 급여비용심사청구 제12조 1항   |       |               |         |            |  |  |  |  |
| 테이블     | 진료내역정보, 진료내역상세   | 컬럼    | 주상병코드, 심사결과코드 | 품질지표    | 정합성        |  |  |  |  |
| 스크립트    | SELECT A.주상병코드, COUNT(*) FROM 진료내역정보 A, 진료내역상세 B WHERE A.주상병코드 NOT IN ( SELECT 주상병코드 FROM 상병코드정보) AND A.진료내역일련번호 = B.진료내역일련번호 AND B.심사결과코드 NOT IN ('심사불능', '반송') GROUP A.주상병코드 |       |               |         |            |  |  |  |  |
| 원인분석    | 데이터 연계 시 심사결과코드 누락데(   | 기터 발생 | 언무규칙          | 직 정의서   | 예시         |  |  |  |  |
| 업무영향    | 심사불능 건에 대한 진료내역 포함 오   | 류     | 8111          | 7 0-111 |            |  |  |  |  |

**NIA 한국지능정보사회진흥원** 38

### 다수의 이해관계자와의 협의를 통해 보완하고 공식화 한다.

작성된 업무규칙을 확정하기 위해, 업무담당자, 품질담당자 등 관련자의 리뷰와 보완이 필요하며, 결재권자의 승인을 얻어 공식화 한다.



#### 확인사항

- ▶ 중요 업무규칙의 누락이 없는지 파악한다.
- ▶ 업무규칙의 중복성이 없는지 정확하게 도출되었는지 확인한다.
- ▶ 동일 항목이나 업무별로 다르게 작성되지 않았는지 확인한다.
- ▶ 업무규칙 표현이 업무규칙 작성지침과 일치하고 일관된 표현을 사용하는지 확인한다.
- ▶ 업무규칙의 개정, 변경 등으로 인해 변경이 요구될 때 이력을 모두 관리할 수 있도록 한다

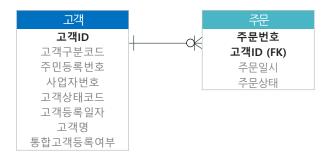
#### 프로파일링 결과 상세 분석 - 참조

진단DB의 ERD를 통해 테이블 간 관계정보를 추출 하는 것이 일반적이지만, ERD가 존재하지 않거나 를 확보할 수 없는 경우, 또는 ERD는 현행화가 이루어지고 있지 않은 경우 업무에 대한 지식과 프로파일링 통해 분석된 컬럼 속성 및 데이터 등에 근거하여 관계를 도출



#### 참조관계 진단기준 도출 및 적용

고객 테이블과 주문 테이블의 무결성을 점검 하기 위한 정의

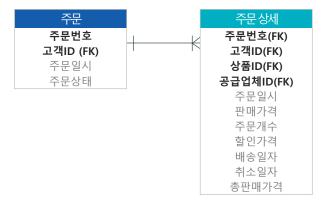


#### 진단 관계정의

| 진단대상 테이블 - 자식 |         |     | 관계   | 진단기준 테이블 - 부모 |         |                |
|---------------|---------|-----|------|---------------|---------|----------------|
| 테이블명          | 컬럼명     | 한글명 | 조인순서 | 한글명           | 컬럼명     | 테이블명           |
| ORDER<br>주문   | CUST_ID | 고객D | 1    | 고객D           | CUST_ID | CUSTOMER<br>고객 |

#### 참조관계 진단기준 도출 및 적용

주문 테이블과 주문상세 테이블의 무결성을 점검 하기 위한 정의



#### 진단 관계정의

| 진단대상 테이블 - 자식     |          |      | 관계   | 진단기준 테이블 - 부모 |          |             |
|-------------------|----------|------|------|---------------|----------|-------------|
| 테이블명              | 컬럼명      | 한글명  | 조인순서 | 한글명           | 컬럼명      | 테이블명        |
| ORDER_DTL<br>주문상세 | CUST_ID  | 고객D  | 1    | 고객D           | CUST_ID  | ORDER<br>주문 |
|                   | ORDER_NO | 주문번호 | 2    | 주문번호          | ORDER_NO |             |

#### 참조관계 진단기준 도출 및 적용

주문상세 테이블과 상품 테이블의 무결성을 점검 하기 위한 정의



#### 진단 관계정의

| 진단대상 테이블 - 자식     |         |                    | 관계   | 진단기준 테이블 - 부모 |         |            |
|-------------------|---------|--------------------|------|---------------|---------|------------|
| 테이블명              | 컬럼명     | 한글명                | 조인순서 | 한글명           | 컬럼명     | 테이블명       |
| ORDER_DTL<br>주문상세 | ITEM_ID | 상품ID               | 1    | 상품ID          | ITEM_ID | ITEM<br>상품 |
|                   | VEND_ID | <del>공급</del> 업체ID | 2    | 공급업체ID        | VEND_ID |            |

#### 참조관계 진단기준 도출 및 적용

상품 테이블과 공급업체 테이블의 무결성을 점검 하기 위한 정의



#### 진단 관계정의

| 진단대        | <del>ll상</del> 테이블 - 지 | 식     | 관계   | 진단기준 테이블 - 부모     |         |            |
|------------|------------------------|-------|------|-------------------|---------|------------|
| 테이블명       | 컬럼명                    | 한글명   | 조인순서 | 한글명               | 컬럼명     | 테이블명       |
| ITEM<br>상품 | VEND_ID                | 공급업체D | 1    | <del>공급</del> 업체D | VEND_CD | ITEM<br>상품 |

### 진단실행은 어떤 방법으로?



#### 공식배포 진단도구를 이용하여 실행

- 공식 배포되는 진단도구를 이용하여 진단을 실행하는 방법
- 범정부 진단기준 및 오류 발생에 대한 문의 및 처리가 용이함



#### 자체 진단도구를 이용하여 실행

- 기존에 사용하고 있는 진단도구를 이용하여 실행하는 방법
- 기관에 알맞은 업무규칙이나 도메인 규칙을 분석하기가 용이함



#### 직접 작성한 SQL을 이용하여 실행

- 어떤 진단도구도 사용하지 않고 순수 SQL을 이용하여 실행하는 방법
- 진단도구를 이용하는 것보다 시간이 많이 소비됨

### 실행결과를 이해관계자와 검토 후 오류데이터를 수정·보완 한다.

오류로 추정되는 데이터를 오류로 확정하기 위해, 업무담당자, 품질담당자 등 관련자와의 결과 공유가 반드시 필요하며, 검토후 개선계획을 수립하고 보완한다.



#### 확인사항

- ▶ 진단대상 누락이 없는지 파악한다. (12개 평가기준에 해당하는 모든 컬럼)
- ▶ 진단대상 및 기준이 정확하게 도출되었는지 확인한다.
- ▶ 동일 항목이나 업무별로 다르게 작성되지 않았는지 확인한다
- ▶ 업무규칙 진단기준 도출 시 일관된 표현을 사용하는지 확인한다.
- ▶ 선정한 진단항목 전체가 정상적으로 실행이 완료 되었는지 확인한다.

| 테이블                   | ~  | 상태 | 수집일자           | <b>v</b> | 조건사용상태 | ▼   | 제외 및 미수집 사유 |
|-----------------------|----|----|----------------|----------|--------|-----|-------------|
| ESTUSER.DPSTD007      | 대상 | 20 | 20-05-21 19:13 |          |        |     |             |
| ESTUSER.DPSTD008      | 대상 | 20 | 20-05-21 19:13 |          |        |     |             |
| ESTUSER.DPSTD009      | 대상 | 20 | 20-05-21 19:13 |          |        |     |             |
| ESTUSER.DPSTD010      | 계외 |    |                |          |        | 진단대 | 상에서 제외합니다.  |
| ESTUSER.DPSTD011      | 대상 | 20 | 20-05-21 19:13 |          |        |     |             |
| ESTUSER.DPSTD012      | 대상 | 20 | 20-05-21 19:13 |          |        |     |             |
| ESTUSER.DPSTD013      | 대상 | 20 | 20-05-21 19:13 |          |        |     |             |
| ESTUSER.TEMP_TABLE_61 | 대상 | 20 | 20-05-21 19:15 |          |        |     |             |
| ESTUSER.TEMP_TABLE_62 | 대상 | 20 | 20-05-21 19:15 |          |        |     |             |
| ESTUSER.TEMP_TABLE_63 | 대상 | 20 | 20-05-21 19:15 |          |        |     |             |
| ESTUSER.TEMP_TABLE_64 | 대상 | 20 | 20-05-21 19:15 |          |        |     |             |
| ESTUSER.TEMP_TABLE_65 | 대상 | 20 | 20-05-21 19:15 |          |        |     |             |
| ESTUSER.TEMP_TABLE_66 | 대상 | 20 | 20-05-21 19:15 |          |        |     |             |
| ESTUSER.TEMP_TABLE_67 | 대상 | 20 | 20-05-21 19:15 |          |        |     |             |
| ESTUSER.TEMP_TABLE_68 | 대상 | 20 | 20-05-21 19:15 |          |        |     |             |
| ESTUSER.TEMP_TABLE_69 | 대상 | 20 | 20-05-21 19:15 |          |        |     |             |
| ESTUSER.TEMP_TABLE_70 | 대상 | 20 | 20-05-21 19:15 |          |        |     |             |
| ESTUSER.TEMP_TABLE_71 | 대상 | 20 | 20-05-21 19:15 |          |        |     |             |
| ESTUSER.TEMP_TABLE_72 | 대상 | 20 | 20-05-21 19:15 |          |        |     |             |
| ESTUSER.TEMP_TABLE_73 | 대상 | 20 | 20-05-21 19:15 |          |        |     |             |
| ESTUSER.TEMP_TABLE_74 | 대상 | 20 | 20-05-21 19:15 |          |        |     |             |
| ESTUSER.TEMP_TABLE_75 | 대상 | 20 | 20-05-21 19:15 |          |        |     |             |
| ESTUSER.TEMP_TABLE_76 | 대상 | 20 | 20-05-21 19:15 |          |        |     |             |
| ESTUSER.TEMP_TABLE_77 | 대상 | 20 | 20-05-21 19:16 |          |        |     |             |
| ESTUSER.TEMP_TABLE_78 | 대상 | 20 | 20-05-21 19:16 |          |        |     |             |
| ESTUSER.TEMP_TABLE_79 | 대상 | 20 | 20-05-21 19:16 |          |        |     |             |
| ESTUSER.TEMP_TABLE_80 | 대상 | 20 | 20-05-21 19:16 |          |        |     |             |
| ESTUSER.TEMP_TABLE_81 | 대상 | 20 | 20-05-21 19:16 |          |        |     |             |
| ESTUSER.TEMP_TABLE_82 | 대상 | 20 | 20-05-21 19:16 |          |        |     |             |
| ESTUSER.TEMP_TABLE_83 | 대상 | 20 | 20-05-21 19:16 |          |        |     |             |

#### 진단대상 테이블 목록

진단대상 전체 테이블 목록을 기재하고, '진단대상 미수집 및 제외 사유, 범위조건 사용' 등에 대한 이유가 명확하게 입력되어야 한다.



#### 품질진단 기본교육

### 데이터 품질진단 절차

| 테이블                  | ▼ 컬럼        | 데이터타입        | ▼ 검증물명   | 도메인 도메인 | 검증형식                      | <b>V</b> | 오류제외데이터 | v   | 수집데이터<br>(참고를 위한 10건 제공)   | V | 의견 | v |
|----------------------|-------------|--------------|----------|---------|---------------------------|----------|---------|-----|----------------------------|---|----|---|
| EDU.CODE_DTL         | CODE_ID     | VARCHAR2(50) | 테스트 개별코드 | 코드      | 02                        |          |         |     |                            |   |    |   |
| EDU.CODE_DTL         | CODE_VAL    | VARCHAR2(50) | 테스트 개별코드 | 코드      | 02                        |          |         |     |                            |   |    |   |
| EDU.CODE_DTL         | CODE_DLT_NM | VARCHAR2(50) | 테스트 개별코드 | 코드      | 02                        |          |         |     |                            |   |    |   |
| EDU.CODE_DTL         | USE_YN      | VARCHAR2(2)  | 테스트 개별코드 | 코드      | 02                        |          |         |     |                            |   |    |   |
| EDU.CODE_GRP         | CODE_ID     | VARCHAR2(50) | 여부YN     | 여부      | 0/1                       |          |         |     |                            |   |    |   |
| EDU.CODE_GRP         | CODE_GRP_NM | VARCHAR2(50) | 여부YN     | 여부      | 0/1                       |          |         |     |                            |   |    |   |
| EDU.CODE_TEST_TABLE  | ID          | VARCHAR2(42) | 가격       | 금액      | 컬럼 > 1000 AND 컬럼<br>10000 | <        |         | A01 | ,A010,A0100,A011,A012      |   |    |   |
| EDU.CODE_TEST_TABLE  | VAL         | NUMBER       | 가격       | 금액      | 컬럼 > 1000 AND 컬럼<br>10000 | <        |         | 406 | 40,44465,50814,63252,65346 |   |    |   |
| EDU.CODE_TEST_TABLE2 | ID          | VARCHAR2(42) | 우편번호     | 번호      | 999-999                   |          |         |     |                            |   |    |   |
| EDU.CODE_TEST_TABLE2 | VAL         | NUMBER       | 우편번호     | 번호      | 999-999                   |          |         |     |                            |   |    |   |

#### 도메인 규칙 설정 내역

평가기준 16개에 부합하는 컬럼을 도출하여 진단기준을 모두 등록 하였는지 확인한다. 진단을 제외한 컬럼에 대해서는 반드시 사유를 작성해야 한다.

행정안전부

품질진단 기본교육

| 분석영역 | 업무규칙명                                  | 테이블                  | 대상 전체건수 <b>SQL</b>           | 오류 데이터 추출 <b>SQL</b>               |
|------|--|----------------------|------------------------------|------------------------------------|
| 업무규칙 | 고객상태코드가 활동인 경우 가입 일자가 반드시 존재해야 한다      | EDU.CODE_TEST_TABLE  | SELECT * FROM DUAL           | SELECT * FROM DUAL WHERE 1=1       |
| 업무규칙 | 고객상태코드가 탈퇴인 경우 탈퇴 일자가 반드시 존재해야 한다.     | EDU.CODE_GRP         | SELECT * FROM DUAL WHERE 1=1 | SELECT * FROM DUAL WHERE 1=1       |
| 업무규칙 | 탈퇴자일가 있으면 반드시 가입일자가 존재해야 한다.           | EDU.CODE_GRP         | SELECT * FROM DUAL WHERE 1=1 | SELECT * FROM DUAL WHERE ROWNUM =2 |
| 업무규칙 | 상품의 생산년도는 판매일자보다 미래 일수 없다.             | EDU.CODE_TEST_TABLE2 | SELECT * FROM DUAL           | SELECT * FROM DUAL                 |
| 업무규칙 | 할인가격은 판매금액을 초과할 수 없다.                  | EDU.CODE_GRP         | SELECT * FROM DUAL WHERE 1=1 | SELECT * FROM DUAL WHERE ROWNUM =4 |
| 업무규칙 | 빨리끝나는 월간 판매금액은 통계테이블의 월간 매출액과 일치해야 한다. | EDU.CODE_GRP         | SELECT * FROM DUAL WHERE 1=1 | SELECT * FROM DUAL WHERE ROWNUM =4 |

#### 업무규칙 설정 내역

업무규칙 분석영역과 업무규칙명, 진단대상 테이블, 진단대상SQL을 기재한다.

| 진단테이블               | 진단컬럼명    | 조인순서 | 참조컬럼        | 참조테이블               | 진단테이블 추가조건 <b>SQL</b> |
|---------------------|----------|------|-------------|---------------------|-----------------------|
| OT.CODE_DTL         | CODE_ID  | 1    | CODE_ID     | OT.CODE_GRP         |                       |
| OT.CODE_DTL         | CODE_VAL | 2    | CODE_GRP_NM | OT.CODE_GRP         |                       |
| OT.ITEM_M_BACK      | ITEM_ID  | 1    | CODE_ID     | OT.CODE_GRP         |                       |
| OT.ORDER_M          | ORDER_NO | 1    | CODE_ID     | OT.CODE_DTL         |                       |
| EDU.DUMMY_TABLE     | SEQ      | 1    | CODE_ID     | EDU.CODE_GRP        |                       |
| EDU.CODE_TEST_TABLE | ID       | 1    | ID          | EDU.CODE_TEST_TABLE |                       |

#### 참조관계 설정 내역

자식테이블과 부모테이블의 조인 컬럼명과 순서를 기재한다. 진단테이블이 자식테이블이고, 부모테이블이 진단기준 테이블이 된다.

| 테이블                  | 컬럼             | 검증물  | 물명          | 실행상태      | 실행일시             |
|----------------------|----------------|------|-------------|-----------|------------------|
| EDU.CODE_TEST_TABLE  | ID             | 범위   | 가례          | STOPPING  | 2020-05-24 13:40 |
| OT.CODE_DTL          | CODE_ID        | 행식   | 금액수량        | READY     |                  |
| EDU.CODE_TEST_TABLE2 |                | 업무규칙 | 1           | READY     | 2020-04-23 21:02 |
| EDU.CODE_TEST_TABLE2 |                | 업무규칙 | 2           | READY     | 2020-04-23 20:50 |
| EDU.CODE_TEST_TABLE2 |                | 업무규칙 | 3           | READY     | 2020-04-23 20:50 |
| EDU.CODE_TEST_TABLE2 |                | 업무규칙 | 오래걸리는 업무규칙  | READY     | 2020-04-23 19:32 |
| EDU.DUMMY_TABLE      | EDU.CODE_GRP   | 활조   | 황조값         | READY     |                  |
| OT.ITEM_M_BACK       | OT.CODE_GRP    | 참조   | <u>참조값</u>  | READY     |                  |
| OT.ITEM_M_BACK       | ITEM_COLOR_CD  | 형식   | 금액수랑        | READY     |                  |
| OT.ORDER_DETAIL      | ORDER_NO       | 행식   | 금액수량        | READY     |                  |
| OT.ORDER_M           | OT.CODE_DTL    | 참조   | <u> 참조값</u> | READY     |                  |
| EDU.CODE_TEST_TABLE  | EDU.ORDER_M    | 참조   | <u>참조강</u>  | FAILED    | 2020-05-25 13:28 |
| EDU.ORDER_DETAIL     | SAL_PRICE      | 형식   | 금액수량        | FAILED    | 2020-05-24 13:40 |
| EDU.ORDER_DETAIL     | CANCEL_DT      | 형식   | 금액수량        | FAILED    | 2020-05-24 13:40 |
| EDU.ORDER_DETAIL     | CUST_ID        | 형식   | 금액수랑        | FAILED    | 2020-05-24 13:40 |
| EDU.ORDER_DETAIL     | DELY_DT        | 형식   | 급액수량        | FAILED    | 2020-05-24 13:40 |
| EDU.ORDER_DETAIL     | DISCOUNT_PRICE | 행식   | 금액수량        | FAILED    | 2020-05-24 13:40 |
| EDU.ORDER_DETAIL     | ITEM_ID        | 형식   | 금액수랑        | FAILED    | 2020-05-24 13:40 |
| EDU.ORDER_DETAIL     | ORDER_DTM      | 형식   | 급액수량        | FAILED    | 2020-05-24 13:40 |
| EDU.ORDER_DETAIL     | ORDER_NO       | 행식   | 금액수량        | FAILED    | 2020-05-24 13:40 |
| EDU.ORDER_DETAIL     | ORDER_QTY      | 형식   | 금액수량        | FAILED    | 2020-05-24 13:40 |
| EDU.ORDER_DETAIL     | VEND_ID        | 행식   | 금액수랑        | FAILED    | 2020-05-24 13:40 |
| EDU.CODE_DTL         |                | 중복   | 데이터중복       | COMPLETED | 2020-04-09 01:04 |
| OT.CODE_DTL          | OT.CODE_GRP    | 참조   | 참조 <i>값</i> | COMPLETED | 2020-05-25 13:28 |
| EDU.CODE_DTL         | CODE_ID        | 코드   | 데스트 개념코드    | COMPLETED | 2020-05-24 13:40 |
| EDU.CODE_DTL         | CODE_VAL       | 35   | 레스트 개별코드    | COMPLETED | 2020-05-24 13:40 |
| EDU.CODE_DTL         | CODE_DLT_NM    | 코드   | 데스트 개별코드    | COMPLETED | 2020-05-24 13:40 |
| EDU.CODE_DTL         | USE_YN         | 35   | 레스트 개별코드    | COMPLETED | 2020-05-24 13:40 |
| EDU.CODE_GRP         |                | 업무규칙 | <u> </u>    | COMPLETED | 2020-05-25 13:27 |

#### 진단항목 실행 내역

진단항목들의 실행상태를 확인한다. 실행상태가 모두 완료상태여야 한다. '미실행상태, 실패상태, 중지상태'가 있는 경우 반드시 재실행하여 결과를 확인한다.



품질진단 기본교육

| 테이블                  | <b>▼</b> 컬럼       | <u>▼</u> 5 | 메인 물명    | <b>▼</b> 전체건수 | 오류건수       | 오류율    | 오류데이터 추출 SQL | 오류데이터<br>▼ (참고를 위한 10건 제공)   |
|----------------------|-------------------|------------|----------|---------------|------------|--------|--------------|--|
| EDU.CODE_GRP         | CODE_GRP_NM       | 여부         | 여부YN     | 4             | 4          | 100.00 | SELECT *     | 고객구분코드, 고객상태코드, 상품색상코<br>드, 주문상태코드   |
| EDU.CODE_GRP         | CODE_ID           | 여부         | 여부YN     | 4             | 4          | 100.00 | SELECT *     | CUST_DVCD, CUST_STCD,<br>ITEM_COLOR_CD, ORDER_ST_CD  |
| EDU.CODE_TEST_TABLE  | VAL               | 금액         | 가격       | 50,001,000    | 49,902,096 | 99.80  | SELECT<br>*  | 1007610, 1010521, 1011865, 10120,<br>1013431, 1014254, 1015322, 101677,<br>1017089 1017331, 701042, 10105, |
| EDU.CODE_TEST_TABLE2 | ID                | 변호         | 우편번호     | 931,000       | 931,000    | 100.00 | SELECT<br>*  | A01009, A01036, A01042, A0105,<br>A01050, A01056, A01057, A0106,   |
| EDU.CODE_TEST_TABLE2 | VAL               | 번호         | 우편번호     | 931,000       | 930,664    | 99.96  | SELECT *     | 100, 1000, 1003, 1009, 1011, 1013,<br>1036, 1041, 1098, 1110   |
| EDU.CODE_DTL         | CODE_DLT_NM       | 코드         | 테스트 개별코드 | 10            | 10         | 100.00 | SELECT *     | 개인, 레드, 블랙, 사업자, 주문, 주문취소,<br>탈퇴, 화이트, 활동중, 휴면상태   |
| EDU.CODE_DTL         | CODE_ID           | 코드         | 테스트 개별코드 | 10            | 10         | 100.00 | SELECT *     | CUST_DVCD, CUST_STCD,<br>ITEM_COLOR_CD, ORDER_ST_CD  |
| EDU.CODE_DTL         | CODE_VAL          | 코드         | 테스트 개별코드 | 10            | 0          | 0.00   | SELECT *     |  |
| EDU.CODE_DTL         | USE_YN            | 코드         | 테스트 개별코드 | 0             | 0          | 0      | SELECT *     |  |
| EDU.CODE_TEST_TABLE  | ID                | 참조         | 참조값      | 50,001,000    | 0          | 0.00   | SELECT *     |  |
| OT.CODE_DTL          | CODE_ID, CODE_VAL | 참조         | 참조값      | 0             | 0          | 0      | SELECT *     |  |

#### 진단항목 오류정보

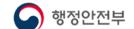
진단항목들의 실제 오류율과 오류데이터 정보, 오류 추출 SQL정보를 기재한다.



|          |                | 진단대상[                   | )B정보       |         |                |  |  |
|----------|----------------|-------------------------|------------|---------|----------------|--|--|
| 진단기간     | 2020-05-01 00: | 00:00 ~ 2020-05-25 23:5 | 9:59       |         |                |  |  |
| 기관명      | 비투엔            | 비투엔                     |            |         |                |  |  |
| 시스템      | 진단대상시스템        |                         |            |         |                |  |  |
| DB명      | Local Oracle   |                         | DB서비스명     | orcl    |                |  |  |
| DB종류     | oracle         |                         | 버젼         | 12.2    |                |  |  |
| DBoloIII | xxx.xxx.0.1    |                         | 포트         | 1521    |                |  |  |
|          |                |                         |            |         |                |  |  |
|          |                | 진단대상 테                  | 이블 현황      |         |                |  |  |
| 대상       | 3              | 323 / 329               | 대상율        |         | 98.18%         |  |  |
| 제외       |                | 3 / 329                 | 제외율        |         | 0.91%          |  |  |
| 미수집      |                | 3 / 329                 | 미수집율       | 0.91%   |                |  |  |
|          |                |                         |            |         |                |  |  |
|          |                | 진단실형                    | l<br>상태    |         |                |  |  |
| 완료       |                | 18 / 40                 | 완료율        | ,       | 45.0%          |  |  |
| 실패       |                | 11 / 40                 | 실패율        | ,       | 27.5%          |  |  |
| 대기       |                | 11 / 40                 | 대기율        |         | 27.5%          |  |  |
|          |                |                         |            |         |                |  |  |
|          |                | 값진단                     | 결과         |         |                |  |  |
| 진단항목     | 진단대상구분         | 진단건수                    | 오류건수       | 오류율     | 최근진단일자         |  |  |
| 여부       | 컬럼             | 8                       |            | 100.00% | 20.05.24 13:40 |  |  |
| 범위       | 컬럼             | 50,001,000              | 49,902,096 | 99.80%  | 20.05.24 13:40 |  |  |
| 형식       | 컬럼             | 1,862,000               | 1,861,664  | 99.98%  | 20.05.24 13:40 |  |  |
| 코드       | 컬럼             | 30                      | 20         | 66.67%  | 20.05.24 13:40 |  |  |
| 참조       | 테이블            | 50,001,000              | 0          | 0.00%   | 20.05.25 13:28 |  |  |
| 업무규칙     | 테이블            | 6                       | 3          | 50.00%  | 20.05.25 13:27 |  |  |
| 전체       |                | 101,864,044             | 51,763,791 | 50.82 % |                |  |  |

#### 종합결과보고서

진단대상 DB 정보·진단 대상 테이블의 현황·진단 실행 현황· 진단결과 현황 정보로 구성되어 있다. 평가기준 12개의 유형별 오류율 확인이 가능하고, 최종 오류율 확인이 가능하다.





# 범정부 진단기준 및 오류 유형

### ○○여부, ○○유무 등 2개의 유효 값을 진단

#### 진단방법

- 컬럼에 존재하는 2개의 유효 값 이외의 데이터는 오류로 판단
- 단일값 분석을 통해 오류 예측이 가능

| 테이블 | 컬럼     | 단일값 건수 | 중복값 건수 |
|-----|--------|--------|--------|
| 고객  | 통합고객여부 | 6      | 71     |

- 여부 가능 값이 혼재되어 있는 경우 유효 값 파악을 위해 건수 분포를 확인하여 유효값 예측
- Y·N·0·1 이 모두 유효한 값인 경우는 코드 도메인 진단을 수행
- YN으로 끝나는 컬럼은 대부분 여부 도메인 컬럼

| 테이블 | 컬럼             | 컬럽값 | 분석건수 |
|-----|----------------|-----|------|
|     |                | Υ   | 355  |
|     | 드쉬그게신니         | N   | 32   |
| 고객  | 통합고객여부         | 1   | 2    |
| 工品  | 진단기준<br>여부 Y/N | 0   | 4    |
|     | 311 7/13       | 0   | 1    |
|     |                | X   | 1    |

### ○○여부, ○○유무 등 2개의 유효 값을 진단

#### 오류원인

- 입력자의 수기 입력 오류
- 다수의 응용프로그램에서 데이터 입력 발생
- 기본 값 설정이 되어 있지 않은 경우

#### 개선방안 및 고려사항

- 여부 값에 대한 표준화
- 기존 값에 대한 정제
- 기본 값 설정

| 테이블 | 컬럼             | 컬럽값 | 분석건수 |
|-----|----------------|-----|------|
|     |                | Υ   | 355  |
|     |                | N   | 32   |
| 고객  | 통합고객여부         | 1   | 2    |
| 工品  | 진단기준<br>여부 Y/N | 0   | 4    |
|     | 911 1/10       | 0   | 1    |
|     |                | X   | 1    |

### 일자 · 일시 값에 대한 포멧의 유일함과 유효함을 진단

#### 진단방법

- 날짜 도메인 컬럼의 값은 유일한 포멧으로 발생 되어야 함
- DATE, TIMESTAPM 타입이 아닌 문자 타입의 날짜 값이 발생하는 컬럼 진단

| 생산년도   | VARCHAR   | 4  | $\rightarrow$ | 년도 (YYYY)      |
|--------|-----------|----|---------------|----------------|
| 고객등록일자 | VARCHAR   | 8  | <b>→</b>      | 년월일 (YYYYMMDD) |
| 주문일시   | TIMESTAMP | 12 | <b>→</b>      | 진단 불필요         |

• 최소 값, 최대 값 분석을 통하여 포멧이 혼재되어 있음을 파악

| 테이블  | 컬럼   | 최소값  | 최대값        |
|------|------|------|------------|
| 주문상세 | 배송일자 | 2015 | 2017-03-07 |

• DATE, TIMESTAPM 타입의 경우에는 시간 순서 일관성 기준으로 진단 가능

| 테이블  | 컬럼                             | 컬럽값        |
|------|--------------------------------|------------|
|      |                                | 20170232   |
|      |                                | 201506     |
| 주문상세 | 배송일자<br>진단기준<br>년월일 YYYY-MM-DD | 2015       |
|      |                                | 2017-03-07 |
|      |                                | 2021년1월1일  |
|      |                                | 2015-11-13 |

### 일자 · 일시 값에 대한 포멧의 유일함과 유효함을 진단

#### 오류원인

- 입력자의 수기 입력 오류
- 날짜 입력 시 날짜 형식에 대한 검증 부재
- 날짜 입력 시 날짜 범위 유효값에 대한 검증 부재

#### 개선방안 및 고려사항

- 날짜 입력 시 날짜를 선택하여 입력하도록 응용프로그램 개선
- 날짜 포멧 에 대한 표준화 수행

| 테이블       | 컬럼                             | 컬럽값        |
|-----------|--------------------------------|------------|
|           |                                | 20170232   |
|           |                                | 201506     |
| <b>* </b> | 배송일자<br>진단기준<br>년월일 YYYY-MM-DD | 2015       |
| 주문상세      |                                | 2017-03-07 |
|           |                                | 2021년1월1일  |
|           |                                | 2015-11-13 |

### 문자, 숫자 등으로 구성된 패턴을 갖는 값의 일관성 진단

#### 진단방법

- 정해진 생성 패턴에 어긋나는 값을 오류로 판단
- 내부 표준에 따른 번호 · 문자체계 기준으로 단순 패턴 검증을 수행

| 번호유형 | 번호 패턴                         |
|------|-------------------------------|
| 고객번호 | 고객번호는 5자리의 숫자로 구성된다           |
| 주문번호 | 주문번호는 5자리의 숫자와 5자리의 영문자로 구성된다 |
| 전화번호 | 전화번호는 하이픈 구분자를 갖는 숫자로 구성된다    |

| 테이블  | 컬럼                                    | 컬럽값           |
|------|---------------------------------------|---------------|
|      |                                       | 99-9999-***   |
|      | 전화번호                                  | 99B999B9999   |
| 공급업체 | 진단기준<br>숫자3-4-4<br>패턴 : 999-9999-9999 | 010-1234-1234 |
|      |                                       | 010-323-3231  |
|      |                                       | 01032393321   |
|      |                                       | 2015-11-13    |

### 문자, 숫자 등으로 구성된 패턴을 갖는 값의 일관성 진단

#### 오류원인

- 데이터 생성규칙 미정의
- 입력 시점에 생성규칙에 의한 패턴 검증기능 부재
- 같은 속성에 다른 구분자 사용

# 개선방안 및 고려사항 • 값입력시패턴검증개선

- 데이터 생성규칙 정의

| 테이블  | 컬럼                                    | 컬럽값           |
|------|---------------------------------------|---------------|
|      |                                       | 99-9999-***   |
|      | 전화번호                                  | 99B999B9999   |
| 공급업체 | 진단기준<br>숫자3-4-4<br>패턴 : 999-9999-9999 | 010-1234-1234 |
|      |                                       | 010-323-3231  |
|      |                                       | 01032393321   |
|      |                                       | 2015-11-13    |

### 정해진 범위 값의 유효 범위를 벗어난 값을 진단

#### 진단방법

- 숫자로 저장된 값이 진단 대상
- 건수, 횟수, 사업비, 가격, 비율, 산정율 등과 같이 숫자로 저장되는 값의 단위가 혼재되어 있거나, 유효 범위를 벗어나는 값을 오류로 판단
- 최소 값, 최대 값 분석을 통하여 수치 값의 범위 파악이 가능
  - ※ 일반적으로 금액, 수량과 관련된 컬럼은 음수(-)가 나올 확률이 적음

| 테이블명 | 컬럼명  | 최소값      | 최대값        |
|------|------|----------|------------|
| 주문상세 | 판매가격 | -9999999 | 9999999999 |

| 테이블  | 컨텀<br>                              | 컬럽값        |
|------|-------------------------------------|------------|
|      |                                     | -9999999   |
|      | 판매가격                                | -99999     |
| 주문상세 | 기준<br>컬럼 >=0 그리고<br>컬럼 <=10,000,000 | -1111      |
|      |                                     | 31000      |
|      |                                     | 500000     |
|      |                                     | 9999999999 |

### 정해진 범위 값의 유효 범위를 벗어난 값을 진단

#### 오류원인

- 입력자의 수기 입력 오류
- 값 입력 시 수치 허용범위 검증기능 부재
- 수량, 금액의 단위가 혼재하는 경우

#### 개선방안 및 고려사항

- 값 입력시 이상치 범위 값에 대한 응용프로그램 개선
- 수량, 금액 등 단위가 존재하는 값에 대한 표준화 수행

| 테이블  | 컬럼<br>                              | 컬럽값        |
|------|-------------------------------------|------------|
|      |                                     | -9999999   |
|      | 판매가격                                | -99999     |
| 주문상세 | 기준<br>컬럼 >=0 그리고<br>컬럼 <=10,000,000 | -1111      |
|      |                                     | 31000      |
|      |                                     | 500000     |
|      |                                     | 9999999999 |

### 개별코드 · 공통코드 사용 컬럼의 유효성 진단

#### 진단방법

- 개별 · 공통 코드를 사용하는 컬럼이 정의된 코드로 구성되어 있는지 분석
- 공통코드 테이블에 의한 분석과 개별코드 테이블 의한 분석으로 분류

| 유형   | 설명   |
|------|--|
| 공통코드 | <ul><li>코드, 유효값, 유효값명 등 정해진 정보만 관리</li><li>일반적으로 공통 모델에서 한 엔터티에서 통합 관리</li></ul>                         |
| 개별코드 | <ul> <li>코드, 유효값, 유효값명 이외의 다양한 부가 정보를 관리</li> <li>부가 정보가 다양해서 하나의 엔터티에서 관리하기 힘들어, 개별 엔터티에서 관리</li> </ul> |

| 테이블 | 컬럼명  | 컬럽 <b>값</b> |
|-----|--|-------------|
|     |  | ST01        |
|     | CUST_ST_CD   | ST1         |
| 고객  | 정의된 코드 값<br>- ST01 : 휴면<br>- ST02 : 활 <del>동중</del><br>- ST03 : 탈퇴 | 법인          |
|     |  | ST02        |
|     |  | ST03        |
|     |  | ST01        |

### 개별코드 · 공통코드 사용 컬럼의 유효성 진단

#### 오류원인

- 코드의 신규, 변경 건에 대한 현행화 관리 미흡
- 응용프로그램에서 코드 값을 하드 코딩하여 사용

#### 개선방안 및 고려사항

- 코드 신규, 변경 시 공통코드에 현행화 관리 및 배포 통제
- 프로그램 상에서 코드 값 매핑 시 하드코딩 금지

| 테이블 | 컬럼                          | 컬럽값  |
|-----|-----------------------------|------|
|     |                             | ST01 |
|     | CUST_ST_CD                  | ST1  |
| 고객  | 정의된 코드 값<br>- ST01 : 휴면     | 법인   |
|     |                             | ST02 |
|     | - ST02 : 활동중<br>- ST03 : 탈퇴 | ST03 |
|     |                             | ST01 |

2017-05-22

### 컬럼에 시간 순서 및 선 · 후 관계가 존재하는 경우를 검증

#### 진단방법

- 날짜 도메인 컬럼 간의 관계
- 날짜 도메인 컬럼 간 검증하려고 하는 목적에 따라 시간 순서 일관성 이거나 선후관계 정확성으로 분류
- 주문 일자가 있다고 해서 반품 일자가 반드시 존재해야하는 컬럼 간의 시간 순서를 진단해야 하는 것은 아님
- 단순히 시작일과 종료일의 선후관계 뿐만 아니라 누락된 날짜 값을 통해 다르게 해석가능

| 유형   | 검증기준                             |
|------|----------------------------------|
| 시간순서 | 반품일자는 주문일자보다 미래여야 한다.            |
| 선후관계 | 반품일자가 있다면 반드시 주문일자가 존재해야 한다.     |
| 선후관계 | 프로젝트 시작일이 있다면 반드시 종료일이 등록되어야 한다. |

#### 오류 추정 데이터 예시

프로젝트계약

|      | 주문일자       | 반품일자       |
|------|------------|------------|
|      | 2019-04-03 | 2020-04-01 |
|      | 2019-05-21 | 2017-05-22 |
| 주문상세 | 2020-05-23 | 2019-07-23 |
|      | 2020-06-17 |            |
|      | 2021-06-01 | 2021-05-30 |
|      | 2021-03-10 |            |
|      |            |            |
|      | 프로젝트시작일    | 프로젝트종료일    |

2019-04-03

### 날짜 컬럼에 시간 순서 및 선/후 관계가 존재하는 경우를 검증

#### 오류원인

- 입력자의 수기 입력 오류
- 값 입력 시 수치 유효성 검증기능 부재

#### 개선방안 및 고려사항

- 기존 값에 대한 정제 필요
- 입력시 유효 값 검증 대한 응용프로그램 개선

|                   | 주문일자       | 반품일자       |
|-------------------|------------|------------|
|                   | 2019-04-03 | 2020-04-01 |
|                   | 2019-05-21 | 2017-05-22 |
| 주 <del>문</del> 상세 | 2020-05-23 | 2019-07-23 |
|                   | 2020-06-17 |            |
|                   | 2021-06-01 | 2021-05-30 |
|                   | 2021-03-10 |            |
|                   |            |            |
|                   | 프로젝트시작일    | 프로젝트종료일    |
| 프로젝트계약            | 2019-04-03 |            |
|                   |            | 2017-05-22 |

### 복수의 컬럼에 논리적인 관계가 존재하는 경우를 검증

#### 진단방법

- 컬럼 간에 논리적인 관계가 존재하는 경우 진단
- 선후관계 정확성과 비슷하지만 날짜 도메인 간의 컬럼이 아닌 컬럼 간의 선후관계 검증

| 검증기준                          |
|-------------------------------|
| 등록여부가 Y인 경우 등록일자는 반드시 존재해야 한다 |
| 결혼여부가 Y인 경우 결혼일자는 반드시 존재해야 한다 |
| 결혼여부가 N인 경우 결혼일자는 존재할 수 없다.   |

|      | 성명  | 등록여부 | 등록일자     | 결혼여부 | 결혼일자     |
|------|-----|------|----------|------|----------|
|      | 안웅수 | Υ    | 20160307 | Υ    | 20180901 |
|      | 서영무 | Υ    | 20170101 | Υ    | 20170827 |
| 고객정보 | 석영준 | Υ    | 20170221 | Ν    |          |
|      | 심우창 | Υ    |          | Υ    |          |
|      | 정광균 | Υ    |          | N    | 20210607 |
|      | 조용범 | Υ    |          | N    |          |

### 복수의 컬럼에 논리적인 관계가 존재하는 경우를 검증

#### 오류원인

- 입력자의 수기 입력 오류
- 값 입력 시 수치 유효성 검증기능 부재

#### 개선방안 및 고려사항

- 기존 값에 대한 정제 필요
- 데이터 입력 시 유효 값 에 대한 응용프로그램 개선

|      | 성명  | 등록여부 | 등록일자     | 결혼여부 | 결혼일자     |
|------|-----|------|----------|------|----------|
|      | 안웅수 | Υ    | 20160307 | Υ    | 20180825 |
|      | 서영무 | Υ    | 20170101 | Υ    | 20170827 |
| 고객정보 | 석영준 | Υ    | 20170221 | Ν    |          |
|      | 심우창 | Υ    |          | Υ    |          |
|      | 정광균 | Υ    |          | N    | 20210607 |
|      | 조용범 | Υ    |          | N    |          |

### 컬럼에 계산된 값이 존재하는 경우를 검증

#### 진단방법

• 특정 통계테이블 등에 계산값이 존재하는 경우를 진단

#### 검증기준

월간 통계테이블 매출액은 주문테이블의 해당 년월의 총금액 합계와 비교하여 일치해야 한다

|     | 주문일자       | 매출액       | 거래완료여부 |
|-----|------------|-----------|--------|
|     | 2021-05-10 | 700,000   | Υ      |
| 주문  | 2021-05-28 | 1,000,000 | Υ      |
| 1 = | 2021-04-02 | 1,000,000 | Υ      |
|     | 2021-04-08 | 3,000,000 | Υ      |
|     | 2021-04-29 | 1,000,000 | Ν      |

|                    | 년    | 월 | 매출액       |
|--------------------|------|---|-----------|
| 월매 <del>출</del> 통계 | 2021 | 5 | 1,800,000 |
|                    | 2021 | 4 | 5,000,000 |

### 컬럼에 계산된 값이 존재하는 경우를 검증

#### 오류원인

- 입력자의 수기 입력 오류
- 값 입력 시 수치 유효성 검증기능 부재

#### 개선방안 및 고려사항

- 기존 값에 대한 정제 필요
- 데이터 입력 시 유효 값 에 대한 응용프로그램 개선

|     | 주문일자       | 매출액       | 거래완료여부 |
|-----|------------|-----------|--------|
|     | 2021-05-10 | 700,000   | Υ      |
| 주문  | 2021-05-28 | 1,000,000 | Υ      |
| 1 = | 2021-04-02 | 1,000,000 | Υ      |
|     | 2021-04-08 | 3,000,000 | Υ      |
|     | 2021-04-29 | 1,000,000 | N      |

|                    | 년    | 월 | 매출액       |
|--------------------|------|---|-----------|
| 월매 <del>출</del> 통계 | 2021 | 5 | 1,800,000 |
|                    | 2021 | 4 | 5,000,000 |

### 부모-자식 관계 데이터의 참조 관계 진단

#### 진단방법

- 부모-자식간의 관계 정합성을 검증하여 부모가 없는 자식데이터가 존재하는지 진단
- 데이터를 관리하는 구조의 설계, 구축의 문제로 인해 데이터 값이 일관되게 관리되지 못하거나 정합성을 확보하지 못하는 경우 발생
- 데이터 모델(ERD)상 식별관계로 정의하였으나, 데이터베이스의 물리제약(FK)으로 설정하지 않은 경우 발생함



|      | 주문번호     | 고객번호   | 주문일시     | 주문상태 |
|------|----------|--------|----------|------|
|      | ORD001AB | CUST01 | 20160307 | Υ    |
|      | ORD062AC | CUST01 | 20170101 | Υ    |
| 주문번호 | ORD093AD | ORD032 | 20170221 | N    |
|      | ORD032BO | CUST05 |          | Υ    |
|      | ORD011ZE | CUST07 |          | N    |
|      | ORD014IC | CUSZD0 |          | N    |

### 부모-자식 관계 데이터의 참조 관계 진단

#### 오류원인

- 부모데이터 삭제 시, 자식 데이터가 그대로 남아 있는 경우
- 프로그램 개발, 성능 확보 등을 위해 RK 설정을 하지 않은 경우
- DB가 아닌 응용프로그램에서 무결성을 확보하려 하는 경우

#### 개선방안 및 고려사항

- 물리적인 RK를 설정
- 응용프로그램 개선을 통해 참조무결성이 보장되도록 개선
- 데이터 모델(RD) 현행화 및 변경관리 수행

|      | 주문번호     | 고객번호   | 주문일시     | 주문상태 |
|------|----------|--------|----------|------|
|      | ORD001AB | CUST01 | 20160307 | Υ    |
|      | ORD062AC | CUST01 | 20170101 | Υ    |
| 주문번호 | ORD093AD | ORD032 | 20170221 | N    |
|      | ORD032BO | CUST05 |          | Υ    |
|      | ORD011ZE | CUST07 |          | N    |
|      | ORD014IC | CUSZD0 |          | N    |
|      |          |        |          |      |