

빅데이터 플랫폼 및 센터

2019

# 데이터 품질관리 가이드

부록 산출물 서식

# Contents



## 부록

<b>1장</b>	데이터 품질관리 산출물 서식 .....	3
1.	데이터 품질진단 산출물 서식 .....	3
2.	데이터 품질관리 기준 산출물 서식 .....	34
<b>2장</b>	정량적 품질진단 규칙 예시 .....	40
1.	개요 .....	40
2.	진단 규칙 예시 .....	41



# V

## 부록

1장 데이터 품질관리 산출물 서식

2장 정량적 품질진단 규칙 예시





## 1 데이터 품질진단 산출물 서식

- 서식-1 품질관리 계획서
- 서식-2 품질진단 대상목록(데이터셋-테이블)
- 서식-3 품질진단 대상목록(항목-컬럼)
- 서식-4 품질진단 수행계획서
- 서식-5 프로파일링 결과내역
- 서식-6 품질진단 규칙 정의 내역
- 서식-7 품질진단 결과내역
- 서식-8 품질진단 결과서
- 서식-9 오류원인 분석 내역
- 서식-10 품질진단 결과 분석서
- 서식-11 품질개선 계획서
- 서식-12 품질개선 결과내역
- 서식-13 품질개선 결과서
- 서식-14 품질 통제 계획서

## 2 데이터 품질관리 기준 산출물 서식

- 서식-21 데이터 분류체계 기준 정의서
- 서식-22 핵심 품질 항목(CTQ) 정의서
- 서식-23 업무규칙 정의서
- 서식-24 품질 이슈 목록

I

II

III

IV

V

부  
록

## 품질진단 단계별 산출물 목록

아래는 데이터 품질진단 DMAIC 5단계 및 16개 세부 절차별로 작성해야 하는 14종의 품질진단 산출물 목록을 설명하고 있다.

품질진단 단계		품질진단 세부단계	산출물
Define	품질관리 계획	D0. 품질관리 계획	1. 품질관리 계획서
	품질진단 대상 정의	D1. 데이터셋(테이블) 단위 선정	2. 품질진단 대상목록 (데이터셋-테이블)
		D2. 항목(컬럼) 단위 선정	3. 품질진단 대상목록 (항목-컬럼)
Measure	품질진단 수행	M1. 품질진단 수행 계획	4. 품질진단 수행계획서
		M2. 품질진단 규칙 정의	5. 프로파일링 결과내역 6. 품질진단 규칙 정의내역
		M3. 품질진단 실시	7. 품질진단 결과내역
		M4. 품질진단 결과	8. 품질진단 결과서
Analyze	진단결과 분석	A1. 오류원인 분석	9. 오류원인 분석내역
		A2. 개선기회 도출	10. 품질진단 결과 분석서
Improve	개선 수행	I1. 품질개선과제정의	11. 품질개선 계획서
		I2. 품질개선계획수립	
		I3. 품질개선환경준비	
		I4. 품질개선수행	12. 품질개선 결과내역 13. 품질개선 결과서
Control	품질 통제	C1. 개선결과 평가	14. 품질통제 계획서
		C2. 품질목표 관리	
		C3. 품질통제 실시	

# 01

## 데이터 품질진단 산출물 서식



### 서식-1 품질관리 계획서

## 품질관리 계획서

2019.09

기관명

I

II

III

IV

V

부  
록

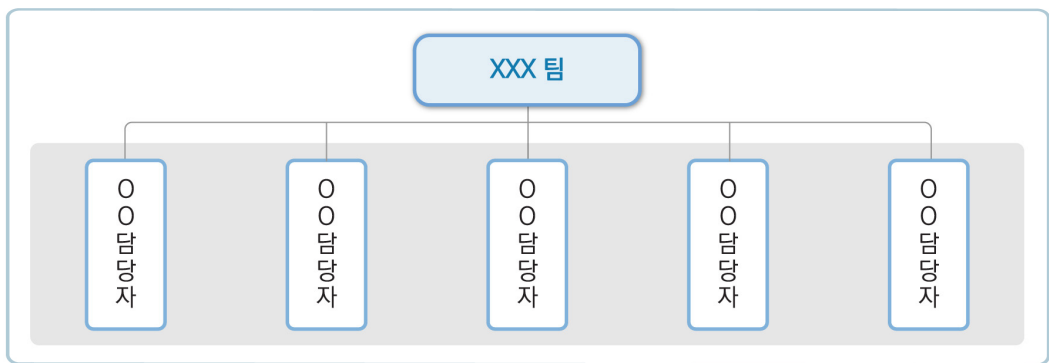


1 품질진단 및 개선 목적

작성방법 품질진단 및 개선 계획에 대한 목적을 기술한다.  
예) 품질진단 정기 1차 점검에 따른 품질진단 수행 및 오류 데이터 개선을 수행하고자 한다.

2 품질관리 조직

작성방법 품질관리 수행 조직에 대한 조직도/역할 내역을 기술한다.



수행주체	역할
품질관리 담당자	OOO 품질관리 계획수립, 추진, 기획·조정
품질진단 담당자	OOO 관련 시스템 운영 등 OOO 품질진단 업무
품질개선 담당자	OOO 품질개선 업무
...	...

3 품질진단 대상

수행주체	DBMS	DATA SIZE (GB)	전체 테이블수	진단대상영역 여부
수집/저장	HBASE	100	250	Y
가공/융합	ORACLE	50	100	Y
분석/활용	ORACLE	20	60	N
...	...	...	...	...

## 4 품질진단 일정

수행 TASK	수행 일정	수행 내용
1. 품질진단 대상 선정	2019/09/16 ~ 2019/09/20	품질진단 대상 테이블 및 컬럼 선정
2. 품질진단 환경구성	2019/09/23 ~ 2019/09/27	1.진단 데이터 적재 점검 2.진단DB, DQ TOOL 권한 구성
3. 품질진단 규칙 정의	2019/09/30 ~ 2019/10/08	* 지표 별 품질진단 규칙 정의
4. 품질진단 수행	2019/10/09 ~ 2019/10/18	* 지표 별 품질진단 수행
5. 품질진단 결과	2019/10/21 ~ 2019/10/23	* 품질진단 결과 보고서 작성
6. 품질진단 결과 분석	2019/10/24 ~ 2019/10/25	* 품질진단 결과 원인분석 및 개선기회 도출

## 5 품질개선 계획

작성방법 품질개선 범위 및 일정을 기술한다.

### ▶ 품질개선 범위

예) 품질진단 정기 1차 품질진단 수행 결과에 따라 최우선과제에 대한 개선을 수행하며, 단기 및 중장기 과제는 “품질개선 계획서”에 별도 일정을 수립할 예정이다.

### ▶ 품질개선 일정

수행TASK	수행일정	수행내용
1. 품질개선 과제 정의	2019/10/26~2019/10/26	품질개선 과제에 대한 우선순위 및 수행 시기 결정
2. 품질개선 환경 구성	2019/10/27~2019/11/01	1. 품질개선 수행 프로그램 구성 2. 협조체계 구성(원천데이터 재구성)
3. 품질개선 수행(우선)	2019/11/04~2019/11/05	* 개선 과제별 품질개선 수행
4. 품질개선 결과서 작성	2019/11/06~2019/11/08	* 품질개선 결과 보고서 작성
5. 품질개선 계획 (단기 및 중장기 과제)	2020/03~	* 단기 및 중장기 과제 개선 수행

I

II

III

IV

V

부  
록

## 서식-2 품질진단 대상목록(데이터셋-테이블)

### ▶ 품질진단 대상목록(데이터셋-테이블)

분류		테이블			
저장영역	DBMS	주제영역	테이블명	테이블 한글명	유형분류명
수집	ORACLE	고객	T_DD_ACDT_RT_LT	일일숙박대여내역	거래
...	...	...	...	...	...

테이블	테이블 단위 선정		업무담당자	비고
전수/샘플	진단대상선정여부	제외사유		
전수	Y		홍길동	
...	...	...	...	...

### ▶ 주요 항목 설명

항목명	설명
저장영역	<ul style="list-style-type: none"> <li>데이터 저장영역 명 (예 - 수집, 가공/분석, 활용)</li> </ul>
DBMS	<ul style="list-style-type: none"> <li>DBMS(Oracle, MySql, DB2 등), NoSQL(MongoDB, Neo4j, Redis) 등</li> </ul>
주제영역	<ul style="list-style-type: none"> <li>테이블의 주제영역</li> </ul>
유형 분류명	<ul style="list-style-type: none"> <li>데이터 유형 분류체계에 따른 데이터셋(테이블)의 유형 분류명 (예 - 메타, 기준, 거래, 로그, 분석, GIS, 문서, 미디어, 센서, 활동)</li> </ul>
전수/샘플	<ul style="list-style-type: none"> <li>데이터셋(테이블)에 대한 표본 추출 기준 참조(품질진단 및 개선 가이드) (예 - 전수, 샘플)</li> </ul>
진단대상 선정여부	<ul style="list-style-type: none"> <li>품질진단 대상 여부 (예- Y, N)</li> </ul>
제외사유	<ul style="list-style-type: none"> <li>품질진단 대상 제외 사유 (예- 그래프, 문서구조(JSON/XML), GIS데이터 등)</li> </ul>

(# 유형 분류명 항목의 상세 설명은 “[서식-21] 데이터 분류체계 기준 정의서” 참고)

### 서식-3 품질진단 대상목록(항목-컬럼)

#### ▶ 품질진단 대상목록(항목-컬럼)

분류		테이블					
저장영역	DBMS	주제영역	테이블명	테이블 한글명	유형분류명		
수집	ORACLE	고객	T_DD_ACDT_RT_LT	일일숙박대여내역	거래		
...	...		...	...			

컬럼							
컬럼명	컬럼 한글명	컬럼타입	컬럼길이	도메인명	PK여부	FK여부	NOT NULL여부
D_RT_CNT	일일대여건수	NUMBER	5				
...	...	...	...	...	...	...	...

컬럼 단위 선정		비고
CTQ여부	진단대상선정여부	
Y	Y	
...	...	...

#### ▶ 주요 항목 설명

항목명	설명
저장영역	<ul style="list-style-type: none"> <li>데이터 저장영역 명 (예 - 수집 , 가공/분석, 활용)</li> </ul>
DBMS	<ul style="list-style-type: none"> <li>DBMS(Oracle, MySql, DB2 등), NoSQL(MongGo DB, Neo4j, Redis) 등</li> </ul>
주제영역	<ul style="list-style-type: none"> <li>테이블의 주제영역</li> </ul>
유형 분류명	<ul style="list-style-type: none"> <li>데이터 유형 분류체계에 따른 데이터셋(테이블)의 유형 분류명 (예 - 메타, 기준, 거래, 로그, 분석, GIS, 문서, 미디어, 센서, 활동)</li> </ul>
CTQ여부	<ul style="list-style-type: none"> <li>CTQ 항목 여부 (예 - Y, N)</li> </ul>
진단대상선정여부	<ul style="list-style-type: none"> <li>품질진단 대상 여부 (예- Y, N)</li> </ul>

(# 유형 분류명 항목의 상세 설명은 “[서식-21] 데이터 분류체계 기준 정의서” 참고)

I

II

III

IV

V

부  
록

**서식-4** 품질진단 수행계획서

## 품질진단 수행계획서

2019.09

기관명

## 1 품질진단 목적

**작성방법** 품질진단을 수행하는 목적에 대하여 기술한다.

예) 품질진단 정책에 따른 월 정기 1차 데이터 품질진단을 수행한다.

## 2 품질진단 대상 현황

**작성방법** 품질진단 대상 현황에 대해 기술한다.

### ▶ 품질진단 대상 현황

저장영역	DBMS	주제영역	전체		진단 대상		제외 대상	
			테이블수	컬럼수	테이블수	컬럼수	테이블수	컬럼수
수집	ORACLE	고객	115	800	55	480	60	320
수집	HBASE	고객	50	600	30	350	20	250
가공/분석	HBASE	고객	100	400	60	150	40	250
활용	MYSQL	고객	90	300	60	190	30	110
...	...	...	...	...	...	...	...	...

(# 상세 내역: [서식-2]품질진단대상목록(데이터셋-테이블), [서식-3] 품질진단 대상목록(항목-컬럼) )

## 3 품질진단 환경구성

**작성방법** 품질진단 환경 구성에 대해 기술한다.

예) 저장영역 별, 진단의 환경(원천, 복제), 품질진단도구/SQL 사용 등 설명을 기술한다.

저장영역	DBMS	진단 대상 데이터 정보	품질진단방식	환경구성 필요내역
수집/저장	HBASE	복제 데이터 진단 (DQ전용저장소-ORACLE)	SQL 작성, DQ TOOL	1. 진단DB 권한 구성 2. DQ TOOL 설치 및 환경구성
가공/융합	HBASE	복제 데이터 진단 (DQ전용저장소-ORACLE)	DQ TOOL	1. 진단DB 권한 구성 2. DQ TOOL 설치 및 환경구성
분석/활용	MYSQL	운영 데이터 진단	SQL 작성	진단DB 권한 구성

## 4 품질진단 일정

**작성방법** 품질진단 환경 구성에 대해 기술한다.

수행TASK	수행일정	수행내용
1. 품질진단환경구성	2019/09/24 ~ 2019/09/27	1. 진단 데이터 적재 점검 2. 진단DB, DQ TOOL 권한 구성
2. 품질진단규칙정의	2019/09/30 ~ 2019/10/08	* 지표 별 품질진단 규칙 정의
3. 품질진단수행	2019/10/09 ~ 2019/10/18	* 지표 별 품질진단 수행
4. 품질진단결과확인	2019/10/21 ~ 2019/10/23	* 품질진단 결과 보고서 작성

## 서식-5 프로파일링 결과내역

### ▶ 프로파일링 결과내역

분류				테이블	
저장영역	DBMS	주제영역	유형분류명	테이블명	테이블 한글명
수집	ORACLE	숙박	거래	T_DD_ACDT_RT_LT	일일숙박대여내역
...	...	...	...	...	...

컬럼							
컬럼명	컬럼 한글명	컬럼타입	컬럼길이	도메인명	PK여부	FK여부	NOT NULL 여부
D_RT_CNT	일일대여건수	NUMBER	5				

프로파일링 결과					비고
패턴 값	건수	MIN 값	MAX 값	총건수	
999	300	0	550	10,000,000,000	
...	...	...	...	...	...

### ▶ 주요 항목 설명

항목명	설명
저장영역	• 데이터 저장영역 명 (예 - 수집 , 가공/분석 , 활용)
DBMS	• DBMS(Oracle, MySql, DB2 등), NoSQL(MongGo DB, Neo4j, Redis) 등
주제영역	• 테이블의 주제영역
유형분류명	• 데이터 유형 분류체계에 따른 데이터셋(테이블)의 유형 분류명 (예 - 메타, 기준, 거래, 로그, 분석, GIS, 문서, 미디어, 센서, 활동)
(프로파일링 결과)	• 패턴 값, 패턴별 건수, MIN, MAX, 총 건수 등 프로파일링 의도에 맞게 항목 변경 가능
패턴 값	• 패턴 분석이 의미 있는 항목에 대해 DISTINCT 등을 수행한 결과의 값 종류 (예 - 여부 항목 인 경우, Y, N, 0, 1 )

(# 유형 분류명 항목의 상세 설명은 “[서식-21] 데이터 분류체계 기준 정의서” 참고)

## 서식-6 품질진단 규칙 정의 내역

### ▶ 품질진단 규칙 정의 내역

분류				테이블	
저장영역	DBMS	주제영역	유형분류명	테이블명	테이블 한글명
수집	ORACLE	숙박	거래	T_DD_ACDT_RT_LT	일일숙박대여내역
...	...	...	...	...	...

분류		테이블				
저장영역	DBMS	주제영역	테이블명	테이블 한글명	전수/샘플	유형분류명
활용	ORACLE	숙박	T_DD_ACDT_RT_LT	일일숙박대여내역	전수	거래

컬럼							
컬럼명	컬럼 한글명	컬럼타입	컬럼길이	도메인명	PK 여부	FK 여부	NOT NULL여부
D_RT_CNT	일일대여건수	NUMBER	5				

품질진단 규칙 정의						
CTQ여부	진단대상선택여부	품질지표(DQI)	품질상세지표	진단규칙내용	진단기법	업무규칙ID
Y	Y	정확성	논리관계 정확성		업무규칙	BR0001

품질진단 규칙 정의 (업무규칙인 경우 생략)			비고
총건수	오류건수	오류내역	
SELECT COUNT(*) ...	SELECT COUNT(*) ....	SELECT 숙박대여일자 ...	

I

II

III

IV

V

부록



## ▶ 주요 항목 설명

항목명	설명
저장영역	<ul style="list-style-type: none"> <li>데이터 저장영역 명 (예 - 수집 , 가공/분석 , 활용)</li> </ul>
DBMS	<ul style="list-style-type: none"> <li>DBMS(Oracle, MySql, DB2 등), NoSQL(MongGo DB, Neo4j, Redis) 등</li> </ul>
주제영역	<ul style="list-style-type: none"> <li>테이블의 주제영역</li> </ul>
유형 분류명	<ul style="list-style-type: none"> <li>데이터 유형 분류체계에 따른 데이터셋(테이블)의 유형 분류명 (예 - 메타, 기준, 거래, 로그, 분석, GIS, 문서, 미디어, 센서, 활동)</li> </ul>
CTQ여부	<ul style="list-style-type: none"> <li>CTQ 항목 여부 (예 - Y, N)</li> </ul>
품질지표(DQI)	<ul style="list-style-type: none"> <li>데이터 품질관리 지표(유효성, 정확성, 완전성, 일관성 등)</li> </ul>
품질상세지표	<ul style="list-style-type: none"> <li>데이터 품질관리 상세지표(참조무결성, 중복일관성, 범위유효성, 논리관계정확성, 단독완전성 등)</li> </ul>
진단기법	<ul style="list-style-type: none"> <li>품질진단의 기법 종류 (예 - 메타정보, 프로파일링, 업무규칙)</li> </ul>
진단규칙내용	<ul style="list-style-type: none"> <li>품질진단을 수행할 내용을 기반으로 진단 스크립트(sql)를 작성</li> </ul>
업무규칙ID	<ul style="list-style-type: none"> <li>업무규칙 생성시 업무규칙정의서에 정의 된 ID명</li> </ul>
총건수	<ul style="list-style-type: none"> <li>총건수를 구할 수 있는 스크립트(sql문)</li> </ul>
오류건수	<ul style="list-style-type: none"> <li>오류건수를 구할 수 있는 스크립트(sql문)</li> </ul>
오류내역	<ul style="list-style-type: none"> <li>품질진단 규칙 내용에 해당하는 오류내역을 구하는 스크립트(sql문)</li> </ul>

(# 유형 분류명 항목의 상세 설명은 “[서식-21] 데이터 분류체계 기준 정의서” 참고)

## 서식-7 품질진단 결과내역

### ▶ 품질진단 결과내역

분류		테이블				
저장영역	DBMS	주제영역	테이블명	테이블 한글명	전수/샘플	유형분류명
분석	ORACLE	숙박	T_DD_ACDT_RT_LT	일일숙박대여내역	전수	거래
컬럼						
컬럼명	컬럼 한글명	컬럼타입	컬럼길이	도메인명	PK 여부	FK 여부
D_RT_CNT	일일대여건수	NUMBER	5			
품질진단 규칙 정의						
CTQ여부	진단대상선택여부	품질지표(DQI)	품질상세지표	진단규칙내용	진단기법	업무규칙ID
Y	Y	정확성	논리관계 정확성		업무규칙	BR0001
품질진단 규칙 정의 (업무규칙인 경우 생략)				품질진단 결과		
총건수	오류건수	오류내역	전체건수	오류건수	오류율(%)	
SELECT COUNT(*) ...	SELECT COUNT(*) ....	SELECT 숙박대여일자 ...	10,000,000	98	0.001	

### ▶ 주요 항목 설명

항목명	설명
저장영역	• 데이터 저장영역 명 (예 - 수집 , 가공/분석 , 활용)
DBMS	• DBMS(Oracle, MySql, DB2 등), NoSQL(MongGo DB, Neo4j, Redis) 등
주제영역	• 테이블의 주제영역
유형분류명	• 데이터 유형 분류체계에 따른 데이터셋(테이블)의 유형분류명 (예 - 메타, 기준, 거래, 로그, 분석, GIS, 문서, 미디어, 센서, 활동)
CTQ여부	• CTQ 항목 여부 (예 - Y, N)
품질지표(DQI)	• 데이터 품질관리 지표(유효성, 정확성, 완전성, 일관성 등)
품질상세지표	• 데이터 품질관리 상세지표( 참조무결성, 중복일관성, 범위유효성, 논리관계정확성, 단독완전성 등)
진단규칙내용	• 품질진단을 수행할 내용을 기반으로 진단스크립트(sql)를 작성
진단기법	• 품질진단의 기법 종류 (예 - 메타정보, 프로파일링, 업무규칙)
업무규칙ID	• 업무규칙 생성시 업무규칙 정의서에 정의된 ID명

(# 유형 분류명 항목의 상세 설명은 “[서식-21] 데이터 분류체계 기준 정의서” 참고)

I

II

III

IV

V

부  
록

**서식-8** 품질진단 결과서

## 품질진단 결과서

2019.09

기관명

## 1 품질진단 수행 현황

작성방법 품질진단의 계획 대비 수행 현황을 기술한다.

저장영역	DBMS	주제 영역	전체		진단 대상(계획)		진단 대상(수행)	
			테이블수	컬럼수	테이블수	컬럼수	테이블수	컬럼수
수집	ORACLE	고객	115	800	55	480	54	470
수집	HBASE	고객	50	600	30	350	30	350
가공/분석	HBASE	고객	100	400	60	150	60	150
활용	MYSQL	고객	90	300	60	190	60	190
...	...	...	...		...	...	...	...

(# 상세 내역: [서식-2]품질진단대상목록(데이터셋-테이블), [서식-3] 품질진단 대상목록(항목-컬럼) )

## 2 품질진단 수행 결과내역

작성방법 품질진단 결과 내역을 요약 기술한다.

### ▶ 품질진단 오류 종합현황

#### ○ 정량진단(값) 오류율

데이터영역	완전성	유효성	일관성	정확성
수집	1.4%	1.5%	0.8%	0.3%
가공/분석	0.9%	0.3%	0.4%	0.2%
활용	0.1%	0.09%	0.001%	0.02%

#### ○ 정량진단(구조표준) 오류율

데이터영역	완전성(구조)	일관성(표준)
수집	0.3%	0.1%
가공/분석	...	...
활용	...	...

I

II

III

IV

V

부  
록

## ▶ 품질진단 오류 상세현황

### ○ 정량진단(값)

품질지표	품질상세지표	진단 내용	테이블수	컬럼수	전체건수	오류건수	오류율(%)	비고
유효성	여부	여부유효값진단	54	470	1000000000	350	0.000035	
	범위	금액유효범위진단	30	350	1200000000	1000	0.0001	
	날짜	...	60	150	2000000000	200	0.00001	
	코드	...	60	190	500000000	50	0.0001	
정확성	파생항목 정확성	...	...	...	...	...	...	
...	...	...	...	...	...	...	...	...
총계								

### ○ 정량진단(구조,표준)

품질지표	품질상세지표	진단내용	테이블수	컬럼수	오류건수	오류율(%)	비고
완전성	구조 완전성	테이블중복 진단	100	-	2	2	
		테이블 현행화 진단	100	-	1	1	
		컬럼 현행화진단	-	1000	4	0.4	
		식별자 누락 진단	-	1000	3	0.3	
일관성	표준 일관성	표준 도메인 진단	-	1000	10	1	
		엔티티명명규칙 진단	100	-	2	2	
		테이블 명명규칙 진단	100	-	2	2	
총계							

(# 상세 내역: [서식-7]품질진단 결과내역)

## 3 진단결과 총평

**작성방법** 품질진단 이후 진행되어야 할 사항에 대한 간단한 요약 및 일정을 기술한다.

예) 품질진단 결과의 전반적인 오류율을 비교하였을 때 유효성 지표에 대한 오류율이 높아 향후 유효성 지표 관련 품질진단 대상의 지속적인 관리가 필요해 보인다.

예) 품질진단 계획 대비 미수행 1개 테이블 10개 컬럼에 대한 품질진단은 2차 품질진단에 적용하여 수행 예정이다.

## 서식-9 오류원인 분석 내역

## ▶ 오류원인 분석 내역

분류		테이블				
저장영역	DBMS	주제영역	테이블명	테이블 한글명	전수/샘플	유형분류명
수집	ORACLE	숙박	T_DD_ACDT_RT_LT	일일숙박대여내역	전수	거래

컬럼							
컬럼명	컬럼 한글명	컬럼타입	컬럼길이	도메인명	PK여부	FK여부	NOT NULL
D_RT_CNT	일일대여건수	NUMBER	5				

품질진단 규칙 정의					
CTQ여부	품질지표(DQI)	품질상세지표	진단규칙내용	진단 기법	업무규칙ID
Y	정확성	논리관계 정확성		업무규칙	BR0001

품질진단 규칙 정의(업무규칙인 경우 SQL 생략)			품질진단 결과		
총건수	오류건수	오류내역	전체건수	오류건수	오류율(%)
SELECT COUNT(*)	SELECT COUNT(*)	SELECT 숙박대여일자	10,000,000	98	0.001

품질진단결과 분석(오류원인 분석)					
오류발생원인	값		구조	표준	관리체계
원천 데이터 오류	기 적재 데이터는 센터와 협의 후 직접 수정				

품질진단결과 분석		
관리체계	응용프로그램	개선 시기 (우선순위)
	일일숙박대여건수 수신 프로그램 체크 로직 추가	

I

II

III

IV

V

부  
록

▶ 주요 항목 설명

항목명	설명
CTQ여부	<ul style="list-style-type: none"> <li>CTQ 항목 여부 (예 - Y, N)</li> </ul>
품질지표(DQI)	<ul style="list-style-type: none"> <li>데이터 품질관리 지표(유효성, 정확성, 완전성, 일관성 등)</li> </ul>
품질상세지표	<ul style="list-style-type: none"> <li>데이터 품질관리 상세지표(참조무결성, 중복일관성, 범위유효성, 논리관계정확성, 단독완전성 등)</li> </ul>
진단규칙내용	<ul style="list-style-type: none"> <li>품질진단을 수행 할 내용을 기반으로 진단스크립트(sql)를 작성</li> </ul>
진단기법	<ul style="list-style-type: none"> <li>품질진단의 기법 종류 (예 - 메타정보, 프로파일링, 업무규칙)</li> </ul>
업무규칙ID	<ul style="list-style-type: none"> <li>업무규칙 생성시 업무규칙 정의서에 정의 된 ID명</li> </ul>
오류발생원인	<ul style="list-style-type: none"> <li>진단분석결과 오류가 발생한 원인(예 - 응용프로그램 검증 부재)</li> </ul>
값	<ul style="list-style-type: none"> <li>개선해야 할 내용이 값이면 개선 내용 기재</li> </ul>
구조	<ul style="list-style-type: none"> <li>개선해야 할 부분이 구조이면 개선 내용 기재</li> </ul>
표준	<ul style="list-style-type: none"> <li>개선해야 할 부분이 표준이면 개선 내용 기재</li> </ul>
관리체계	<ul style="list-style-type: none"> <li>개선해야 할 부분이 관리체계이면 개선 내용 기재</li> </ul>
응용프로그램	<ul style="list-style-type: none"> <li>개선해야 할 부분이 응용프로그램이면 개선 내용 기재</li> </ul>
개선 시기 (우선순위)	<ul style="list-style-type: none"> <li>개선해야 할 내용에 대해 업무중요도, 시스템영향도 등을 파악하여 개선시기 부여 (예시 - 최우선, 단기, 중장기)</li> </ul>

서식-10 품질진단 결과 분석서

# 품질진단 결과 분석서

2019.09

기관명

I

II

III

IV

V

부  
록



## 1 품질진단 결과내역

**작성방법** 품질진단 결과 내역을 기술한다.

### ▶ 지표 별 오류 현황

#### ○ 정량진단(값)

품질지표	품질상세 지표	진단 내용	테이블수	컬럼수	전체건수	오류건수	오류율(%)	비고
유효성	여부	여부 유효값진단	54	470	1000000000	350	0.000035	
	범위	금액 유효범위진단	30	350	1200000000	1000	0.0001	
	날짜	...	60	150	2000000000	200	0.00001	
	코드	...	60	190	50000000	50	0.0001	
정확성	파생항목 정확성	...	...	...	...	...	...	
...	...	...	...	...	...	...	...	...
총계								

#### ○ 정량진단(구조,표준)

품질지표	품질상세 지표	진단내용	테이블수	컬럼수	오류건수	오류율(%)	비고
완전성	구조 완전성	테이블중복 진단	100	-	2	2	
		테이블 현행화 진단	100	-	1	1	
		컬럼 현행화진단	-	1000	4	0.4	
		식별자 누락 진단	-	1000	3	0.3	
일관성	표준 일관성	표준 도메인 진단	-	1000	10	1	
		엔터티 명명규칙 진단	100	-	2	2	
		테이블 명명규칙 진단	100	-	2	2	
총계							

(# 상세 내역: [서식-7]품질진단 결과내역)

## 2 오류 원인 현황

**작성방법** 품질진단 결과 분석 내역의 오류발생 유형을 요약 기술한다.

### ▶ 지표 별 발생 유형 현황

저장영역	DBMS	오류발생유형 수				
		값 오류	구조 오류	표준 오류	관리체계 오류	응용프로그램 오류
수집						
가공/분석						
활용						

(# 상세 내역: [서식-9]오류원인 분석 내역)

## 3 개선 기회 도출

**작성방법** 오류 발생 원인에 따른 개선 기회에 대해 기술한다.

오류발생유형	오류발생원인	개선 기회
값	원천데이터 오류	원천데이터 개선
구조	<ul style="list-style-type: none"> <li>논리/물리 모델의 현행화 미흡</li> <li>데이터 모델관리 가이드 및 지침서 부재</li> </ul>	<ul style="list-style-type: none"> <li>논리/물리 모델 현행화 수행</li> <li>데이터 모델관리 가이드 및 지침 마련</li> </ul>
표준	<ul style="list-style-type: none"> <li>표준(용어, 도메인) 지침의 변경관리 미흡</li> </ul>	<ul style="list-style-type: none"> <li>표준(용어, 도메인) 정비 수행</li> </ul>
관리체계	<ul style="list-style-type: none"> <li>데이터 품질 관리 지침 가이드 부재</li> </ul>	<ul style="list-style-type: none"> <li>데이터 품질 관리 지침 마련</li> </ul>
응용프로그램	<ul style="list-style-type: none"> <li>연계/수집/생성 값의 형식 검증기능의 미흡/부재</li> </ul>	<ul style="list-style-type: none"> <li>형식 검증기능 보완 및 추가</li> </ul>

(# 상세 내역: [서식-9]오류원인 분석 내역)

I

II

III

IV

V

부  
록

**서식-11** 품질 개선 계획서

## 품질 개선 계획서

2019.09

기관명

## 1 품질개선 목적

**작성방법** 품질개선 수행의 목적을 기술한다.

예) 시스템 품질진단 정기 1차 점검에 따른 오류에 대한 품질 개선을 실시하고자 한다.

## 2 품질개선 대상

**작성방법** 품질오류 개선 대상에 대해 기술한다.

### ▶ 저장영역 별 오류 발생 유형 현황

저장영역	DBMS	오류발생유형 수				
		값 오류	구조 오류	표준 오류	관리체계 오류	응용프로그램 오류
수집	HBASE					
가공/분석	ORACLE					
활용	ORACLE					

(# 상세 내역: [서식-7]품질진단결과 분석결과 내역, [서식-10]품질진단 결과 분석서)

### ▶ 저장영역 별 오류 발생 현황

**작성방법** 오류 발생 유형 별 오류가 발생된 테이블 컬럼의 개수를 기술한다.

저장영역	DBMS	오류 발생 테이블 수	오류 발생 컬럼 수
수집	HBASE	10	18
가공/분석	ORACLE	5	9
활용	ORACLE	3	10

I

II

III

IV

V

부  
록

### 3 개선과제 정의

**작성방법** 개선 우선 순위 및 개선과제를 정의하여 기술한다.

#### ▶ 개선 과제 별 우선 순위 선정

우선순위	개선 수행방식	개선과제
최우선 개선 과제 (품질관리계획서의 일정)	유지보수 중 수행	기본적 오류 데이터 개선
	유지보수 중 수행	정제 된 데이터 일괄 재수신
	유지보수 중 수행	형식 검증 응용프로그램 개선
단기 개선 과제 (반기 단위)	유지보수 중 수행	데이터 모델 현행화
	유지보수 중 수행	표준(용어, 도메인) 현행화 수행
중장기 개선 과제 (년 단위)	품질개선 사업 발주	데이터 품질/모델 관리 지침 마련

#### ▶ 저장영역 별 우선순위 도출

**작성방법** 저장영역, DBMS 별 오류발생유형에 대한 개선 과제 별 우선순위를 기술한다.

예) 우선순위 유형 : ① 최우선 과제, ② 단기 과제, ③ 중장기 과제

저장영역	DBMS	오류발생유형				
		값	구조	표준	관리체계	응용프로그램
수집	HBASE	① 최우선 과제	① 최우선 과제	① 최우선 과제	② 단기 과제	① 최우선 과제
가공/분석	ORACLE	① 최우선 과제	② 단기 과제	② 단기 과제	② 단기 과제	② 단기 과제
활용	ORACLE	① 최우선 과제	③ 중장기 과제	③ 중장기 과제	③ 중장기 과제	③ 중장기 과제

## 4 품질개선 수행 일정

**작성방법** 품질개선 수행일정을 기술한다.

### ▶ 최우선 과제

수행TASK	수행일정	수행내용
1. 품질개선환경구성	2019/10/24 ~ 2019/10/25	1. 품질개선 수행 시스템 (프로그램, 데이터) 구성 2. 관련 이해관계자와의 협조체계 구성
2. 품질개선 수행 (최우선)	2019/10/26 ~ 2019/11/01	<ul style="list-style-type: none"> <li>개선 과제 별 품질개선 수행</li> </ul>
3. 품질개선 결과서 작성	2019/11/04 ~ 2019/11/05	<ul style="list-style-type: none"> <li>품질개선 결과 보고서 작성</li> </ul>

### ▶ 단기 및 중장기 과제

우선순위	과제명	수행일정	수행방법	수행예산
② 단기 과제	데이터 모델 현행화	2019/12/26 ~ 2019/12/31	운영인력 자체수행	해당없음
② 단기 과제	표준(용어, 도메인) 현행화 수행	2020/01/06 ~ 2020/01/10	별도 사업 추진	xxx만원
③ 중장기 과제	데이터 품질/모델 관리 지침 마련	...	...	...

I

II

III

IV

V

부  
록

## 서식-12 품질개선 결과내역

### ▶ 품질개선 결과내역

분류		테이블				
저장영역	DBMS	주제영역	테이블명	테이블 한글명	전수/샘플	유형분류명
수집	ORACLE	숙박	T_DD_ACDT_RT_LT	일일숙박대여내역	전수	거래

컬럼							
컬럼명	컬럼 한글명	컬럼타입	컬럼길이	도메인명	PK여부	FK여부	NOT NULL 여부
D_RT_CNT	일일대여건수	NUMBER	5				

품질진단 규칙 정의						
CTQ여부	진단대상선평여부	품질지표(DQI)	품질상세지표	진단내용	진단기법	업무규칙ID
Y		정확성	논리관계 정확성		업무규칙	BR0001

품질진단 규칙 정의			품질진단 결과		
총건수	오류건수	오류내역	전체건수	오류건수	오류율(%)
SELECT COUNT(*)	SELECT COUNT(*)	SELECT 숙박대여일자	10,000,000	98	0.001

품질진단결과 분석			
오류발생원인	값	구조	표준
원천 데이터 오류	기 적재 데이터는 센터와 협의 후 직접수정		

품질진단결과 분석		
관리체계	응용프로그램	개선 시기 (우선순위)
	일일숙박대여건수 수신 프로그램 체크로직 추가	

개선 수행							
값	구조	표준	관리 체계	응용프로그램	개선상태	최종개선일자	업무담당자
N				Y	개선중	2019-10-24	홍길동

## ▶ 주요 항목 설명

항목명	설명
CTQ여부	<ul style="list-style-type: none"> <li>CTQ 항목 여부 (예 - Y, N)</li> </ul>
품질지표(DQI)	<ul style="list-style-type: none"> <li>데이터 품질관리 지표(유효성, 정확성, 완전성, 일관성 등)</li> </ul>
품질 상세지표	<ul style="list-style-type: none"> <li>데이터 품질관리 상세지표( 참조무결성, 중복일관성, 범위유효성, 논리관계정확성, 단독완전성 등)</li> </ul>
진단규칙 내용	<ul style="list-style-type: none"> <li>품질진단을 수행할 내용을 기반으로 진단스크립트(sql)를 작성</li> </ul>
진단기법	<ul style="list-style-type: none"> <li>품질진단의 기법 종류 (예 - 메타정보, 프로파일링, 업무규칙)</li> </ul>
오류 발생 원인	<ul style="list-style-type: none"> <li>진단 분석결과 오류가 발생한 원인(예 - 응용프로그램 검증 부재)</li> </ul>
품질진단 결과 분석 (값/구조/표준/관리체계/응용)	<ul style="list-style-type: none"> <li>진단분석결과 개선내용을 해당 영역에 기재 (여러 영역에 개선이 필요하면 복수 기재)</li> </ul>
개선시기(우선순위)	<ul style="list-style-type: none"> <li>개선해야 할 내용에 대해 업무 중요도, 시스템 영향도 등을 파악하여 개선시기 부여 (예시 - 최우선 , 단기, 중장기)</li> </ul>
개선수행 (값/구조/표준/관리체계/응용)	<ul style="list-style-type: none"> <li>개선을 수행한 대상 영역 항목에 체크</li> </ul>
개선상태	<ul style="list-style-type: none"> <li>개선이 수행되는 과정(예 - 미개선, 개선 중, 개선 완료)</li> </ul>
최종 개선 일자	<ul style="list-style-type: none"> <li>최종 개선한 일자</li> </ul>

(# 유형 분류명 항목의 상세 설명은 “[서식-21] 데이터 분류체계 기준 정의서” 참고)

I

II

III

IV

V

부  
록



**서식-13** 품질 개선 결과서

## 품질 개선 결과서

2019.09

기관명

## 1 품질개선 대상

**작성방법** 품질개선의 계획 대비 수행 현황을 기술한다.

## 2 우선 개선과제 정의

**작성방법** 품질개선의 우선 개선 과제에 대해 기술한다.

우선순위	개선 수행방식	개선과제
최우선 개선 과제 (품질관리계획서의 일정)	유지보수 중 수행	기본적 오류 데이터 개선
	유지보수 중 수행	정제된 데이터 일괄 재수신
	유지보수 중 수행	형식 검증 응용프로그램 개선

## 3 품질개선 수행 결과

**작성방법** 품질개선 결과를 기술한다.

### ▶ 오류유형 및 지표 별 개선 결과

#### ○ 영역 별 오류유형 개선 수행 현황

저장영역	DBMS	오류유형 수									
		값 오류		구조 오류		표준 오류		관리체계 오류		응용프로그램 오류	
		개선 대상	개선 완료	개선 대상	개선 완료	개선 대상	개선 완료	개선 대상	개선 완료	개선 대상	개선 완료
수집	HBASE										
가공/분석	ORACLE										
활용	ORACLE										

(# 상세 내역: [서식-10]품질진단 결과 분석서, [서식-12]품질개선 결과내역)

○ 지표 별 개선 전후 비교 현황

- 정량진단(값)

품질지표	품질상세지표	개선전			개선후	
		전체건수	오류건수	오류율(%)	개선건수	개선율 (%)
완전성	단독 완전성					
	조건 완전성					
	구조 완전성					
	공간객체 완전성					
유효성	범위 유효성					
	날짜 유효성					
	형식 유효성					
	코드 유효성					
	선후관계 유효성					
일관성	참조무결성					
	중복일관성					
	위상구조 일관성					
	개념적 일관성					
	연계일관성					
정확성	논리관계 정확성					
	파생항목 정확성					
	메타 정확성					
	그리드 정확성					
	위치 정확성					

- 정량진단(구조, 표준)

품질지표	품질상세지표	개선전			개선후	
		전체건수	오류건수	오류율(%)	개선건수	개선율 (%)
일관성	표준 일관성					
완전성	구조 완전성					

(# 상세 내역: [서식-7]품질개선 결과내역)

서식-14 품질 통제 계획서

# 품질 통제 계획서

2019.09

기관명

I

II

III

IV

V

부  
록

## 1 품질 통제 목적

**작성방법** 품질통제 수행의 목적을 기술한다.

예시) 데이터품질 정기진단 결과에 따른 지표별 오류율 개선 결과와 주기적인(반기·분기별) 지표별 성숙도 측정을 통해 품질개선 효과를 평가한다. 그리고 데이터 품질을 유지 또는 향상시키기 위한 품질목표 수준을 설정하고 지속적인 개선활동과 모니터링을 수행한다.

## 2 품질 수준 현황

**작성방법** 개선 전 후의 오류율을 비교하고 주기적인 성숙도 평가를 통해 현 수준의 데이터 품질수준을 평가한다.

### ▶ 지표별 오류율 개선 결과

평가지표		오류율(전체 건수 기준)	
품질지표	상세지표	개선 전	개선 후
일관성	식별자 정확성	1.5%	0%
	테이블간 일관성	2.0%	0%
	...	...	...
유효성	범위 유효성	1.2%	0%
	날짜 유효성	2.4%	0%
	형식 유효성	1.3%	0%
정확성	코드 정확성	0.5%	0.2%
	표준 정확성	0.5%	0.1%
완전성	...	0.5%	0%

### ▶ 지표별 성숙도 평가

품질지표	성숙도 평가			
	19-01	19-02	19-03	19-04
일관성				
유효성				
정확성				
완전성				
...				

### 3 품질 수준 목표설정

**작성방법** 현재 시점의 품질수준을 고려하여 상세지표별 목표 오류율과 품질지표별 목표 성숙도를 결정함으로써 정량적인 데이터 품질수준 목표를 설정한다.

평가지표		품질목표	
품질지표	상세지표	오류율	성숙도
일관성	식별자 정확성		
	테이블간 일관성		
	...		
유효성	범위 유효성		
	날짜 유효성		
	형식 유효성		
정확성	코드 정확성		
	표준 정확성		
완전성	...		

### 4 품질통제 활동 계획

**작성방법** 품질 수준 목표 달성을 위한 품질 활동 및 모니터링 계획을 수립한다.

예) 지속적인 품질개선을 위한 품질교육 및 문화정착, 시스템 품질점검 활동, 데이터 품질 이슈 및 품질수준 모니터링 활동, 품질오류 신고 접수 및 처리 활동 등

I

II

III

IV

V

부  
록

## 02

## 데이터 관리 산출물 서식



## 서식-21 데이터 분류체계 기준 정의서

## 1 데이터 분류체계 기준 정의

품질진단 대상 선정을 위해서는 “데이터 유형” 및 “데이터 구조”의 분류체계가 정의되어 있어야 하며, 전체 테이블 별로 제시된 분류체계를 참고하여 “데이터 유형” 및 “데이터 구조” 관점으로 테이블을 분류한다.

## 2 데이터 분류체계 기준

## ▶ 데이터 유형

데이터 유형	데이터 유형 설명	데이터 유형(예시)
메타 데이터 (Meta data)	<ul style="list-style-type: none"> <li>데이터에 대한 표준화 정보 및 데이터 특성을 설명하는 요약 정보로 정보명, 주제영역, 위치, 중요도, 품질수준, 다른 데이터와 연관성 등을 관리하는 정보</li> </ul>	<ul style="list-style-type: none"> <li>표준사전, 데이터셋의 특성정보(정보명, 위치, 주제영역, 출처, 중요도, 품질수준, 연관성 등)</li> </ul>
기준 데이터 (Master data)	<ul style="list-style-type: none"> <li>업무 프로세스의 중심이 되는 공통 개념의 데이터로 정보 값이 자주 변하지 않는 정보 및 참조정보(Reference data : 데이터의 값을 참조하는 코드성 정보)를 포함</li> </ul>	<ul style="list-style-type: none"> <li>제품명세, 고객정보, 사업자정보 등</li> <li>지역코드, 성별코드, 학력코드 등</li> </ul>
로그데이터 (Log data)	<ul style="list-style-type: none"> <li>시스템이 생성한 Log, 웹 상의 모든 흔적들의 Raw file 형식 데이터</li> </ul>	<ul style="list-style-type: none"> <li>로그기록, Web Log(Raw file) 등</li> </ul>

데이터 유형	데이터 유형 설명	데이터 유형(예시)
거래 데이터 (Transaction data)	<ul style="list-style-type: none"> <li>기업이나 기관의 고유한 업무활동 및 서비스에 의하여 생성되는 정보</li> </ul>	<ul style="list-style-type: none"> <li>신용/직불카드 및 금융 거래 자료, 유통업체 자료</li> </ul>
분석 데이터 (Analytics data)	<ul style="list-style-type: none"> <li>집계 데이터 및 분석을 통하여 결과로 생성된 정보</li> </ul>	<ul style="list-style-type: none"> <li>지역별/업종별 매출현황, 서울 인구 이동 분포, 종로 상권분석 등</li> </ul>
GIS 데이터 (GIS data)	<ul style="list-style-type: none"> <li>GIS 시스템에 의하여 생성한 공간 정보로 벡터정보와 속성정보로 구성된 정보</li> </ul>	<ul style="list-style-type: none"> <li>행정구역도, 지하매설물도, 농경지 지도, 산림도 등</li> </ul>
문서 데이터 (Document data)	<ul style="list-style-type: none"> <li>문서 작성기로 생성한 문서 정보(고유의 저장형식을 가짐)</li> </ul>	<ul style="list-style-type: none"> <li>일반문서, 논문, 보고서 등</li> </ul>
미디어 데이터 (Multimedia data)	<ul style="list-style-type: none"> <li>사진, 영상, 음성 등 미디어 정보</li> </ul>	<ul style="list-style-type: none"> <li>사진, 위성사진, 항공사진, 영상, 음성 등</li> </ul>
센서 데이터 (Sensor data)	<ul style="list-style-type: none"> <li>추적장치 정보(Tracking device data)를 포함한 각종 센서를 통하여 생성되는 정보</li> </ul>	<ul style="list-style-type: none"> <li>위치, 기상, 수질, 대기, IoT 등</li> </ul>
활동 데이터 (Online behavior data)	<ul style="list-style-type: none"> <li>의견정보(Opinion data), 웹 검색정보 등을 포함한 온라인 상에서 생성된 것으로 분석을 위하여 가공된 정보</li> </ul>	<ul style="list-style-type: none"> <li>인터넷 검색 및 페이지 뷰 인덱싱 정보(WebLog) 등</li> <li>이메일, 카카오톡, 트위터 등 공개 자료</li> </ul>



## 서식-22 핵심 품질 항목(CTQ) 정의서

### ▶ 핵심품질항목(CTQ) 정의서

분류		테이블			
저장영역	DBMS	주제영역	테이블명	테이블 한글명	유형분류명
수집	ORACLE	숙박	T_DD_ACDT_RT_LT	일일숙박대여내역	거래

컬럼				CTQ 항목					비고
컬럼명	컬럼 한글명	도메인명	데이터 타입	활용도수준	점수	순위	선정 여부	우선대상여부	
D_RT_CNT	일일대여건수		NUMBER	중	90	1	Y	Y	

### ▶ 주요 항목 설명

항목명	설 명
저장영역	<ul style="list-style-type: none"> <li>데이터 저장영역 명 (예 - 수집/저장 , 가공/융합, 분석/활용 영역)</li> </ul>
유형 분류명	<ul style="list-style-type: none"> <li>데이터 유형 분류체계에 따른 데이터셋(테이블)의 유형분류 명 (예 - 메타, 기준, 로그, 거래, 분석, GIS, 문서, 미디어, 센서, 활동 등)</li> </ul>
도메인명	<ul style="list-style-type: none"> <li>도메인 정의서에 정의된 도메인 명</li> </ul>
활용도 수준	<ul style="list-style-type: none"> <li>해당 대상이 업무적으로 미치는 영향도( 예 - 상, 중, 하)</li> </ul>
선정 여부	<ul style="list-style-type: none"> <li>CTQ 선정 및 제외 여부 (예 - Y,N)</li> </ul>
점수	<ul style="list-style-type: none"> <li>CTQ 평가항목(영향도, 가중치) 합산 점수 (예- 0~100)</li> </ul>
순위	<ul style="list-style-type: none"> <li>총점 기준 순위 (예 - 1~3순위)</li> </ul>
선정 여부	<ul style="list-style-type: none"> <li>CTQ 항목 선정 여부</li> </ul>
우선 대상 여부	<ul style="list-style-type: none"> <li>CTQ 평가항목(영향도, 가중치) 중 상위 우선 여부(예 - Y,N)</li> </ul>

(# 유형 분류명 항목의 상세 설명은 "[서식-21] 데이터 분류체계 기준 정의서" 참고)

## 서식-23 업무규칙 정의서

## ▶ 업무규칙 정의서

분류		테이블			
저장영역	DBMS	주제영역	대상 테이블명	대상 테이블 한글명	유형분류명
수집	ORACLE	숙박	T_DD_ACDT_RT_LT	일일숙박대여내역	거래

컬럼								
컬럼명	컬럼 한글명	데이터 타입	컬럼길이	도메인명	PK 여부	FK 여부	NOT NULL 여부	CTQ 여부
D_RT_CNT	일일대여건수	NUMBER	6					Y

품질 지표 및 업무규칙						
품질지표 (DQI)	품질상세지표	업무규칙ID	업무규칙 내용	총건수	오류건수	오류내역
정확성	논리관계 정확성	BI_00001	일일숙박 대여건수는...	SELECT COUNT(*)	SELECT COUNT(*)	SELECT 숙박 대여일자

담당자		비고
업무규칙 작성일자	업무규칙 담당자	
20190930	홍길동	

I

II

III

IV

V

부  
록

## ▶ 주요 항목 설명

항목명	설명
저장영역	<ul style="list-style-type: none"> <li>데이터 저장영역 명 (예 - 수집, 가공/분석, 활용)</li> </ul>
DBMS	<ul style="list-style-type: none"> <li>DBMS(Oracle, MySql, DB2 등), NoSQL(MongGo DB, Neo4j, Redis) 등</li> </ul>
주제영역	<ul style="list-style-type: none"> <li>테이블의 주제영역</li> </ul>
유형분류명	<ul style="list-style-type: none"> <li>데이터 유형 분류체계에 따른 데이터셋(테이블)의 유형 분류명 (예 - 메타, 기준, 거래, 로그, 분석, GIS, 문서, 미디어, 센서, 활동)</li> </ul>
CTQ여부	<ul style="list-style-type: none"> <li>CTQ 항목 여부 (예 - Y, N)</li> </ul>
품질지표(DQI)	<ul style="list-style-type: none"> <li>데이터 품질관리 지표(유효성, 정확성, 완전성, 일관성 등)</li> </ul>
품질상세지표	<ul style="list-style-type: none"> <li>데이터 품질관리 상세지표( 참조무결성, 중복일관성, 범위유효성, 논리관계정확성, 단독완전성 등)</li> </ul>
업무규칙내용	<ul style="list-style-type: none"> <li>품질진단을 수행할 내용을 기반으로 진단 스크립트(sql)를 작성</li> </ul>
총건수	<ul style="list-style-type: none"> <li>총 건수를 구할 수 있는 스크립트(sql문)</li> </ul>
오류건수	<ul style="list-style-type: none"> <li>오류 건수를 구할 수 있는 스크립트(sql문)</li> </ul>
오류내역	<ul style="list-style-type: none"> <li>업무규칙 내용에 해당하는 오류 내역을 구하는 스크립트(sql문)</li> </ul>
업무규칙 작성일자	<ul style="list-style-type: none"> <li>업무규칙을 작성한 일자</li> </ul>
업무규칙 담당자	<ul style="list-style-type: none"> <li>업무규칙을 도출한 담당자</li> </ul>

(# 유형 분류명 항목의 상세 설명은 “[서식-21] 데이터 분류체계 기준 정의서” 참고)

## 서식-24 품질 이슈 목록

### > 품질 이슈 목록

이슈 번호	이슈 유형	이슈명	이슈 내용	시스템명	요청 부서	요청자	요청 일자	검토 부서	검토 담당자

### > 주요 항목 설명

항목명	설명
이슈번호	• 품질이슈의 일련번호
이슈유형	• 품질이슈의 유형(예시- 데이터 값, 프로세스 등)
이슈명	• 품질이슈의 명칭
이슈내용	• 품질이슈의 상세내용
시스템명	• 품질이슈가 발생하는 시스템 명
요청부서	• 품질이슈 요청부서
요청자	• 품질이슈 요청자 명
요청일자	• 품질이슈 요청일자
검토부서	• 품질이슈 검토부서
검토담당자	• 품질이슈 검토담당자 명

I

II

III

IV

V

부  
록

# 01

## 개요



- [1] 진단 대상 데이터 모델 예시
- [2] 진단예시 > 단독 완전성
- [3] 진단예시 > 조건 완전성
- [4] 진단예시 > 구조 완전성
- [5] 진단예시 > 코드 유효성
- [6] 진단예시 > 형식 유효성
- [7] 진단예시 > 여부 유효성
- [8] 진단예시 > 날짜 유효성
- [9] 진단예시 > 범위 유효성
- [10] 진단예시 > 선후관계 유효성
- [11] 진단예시 > 참조 무결성
- [12] 진단예시 > 중복 일관성
- [13] 진단예시 > 표준 일관성
- [14] 진단예시 > 논리관계 정확성
- [15] 진단예시 > 파생항목 정확성

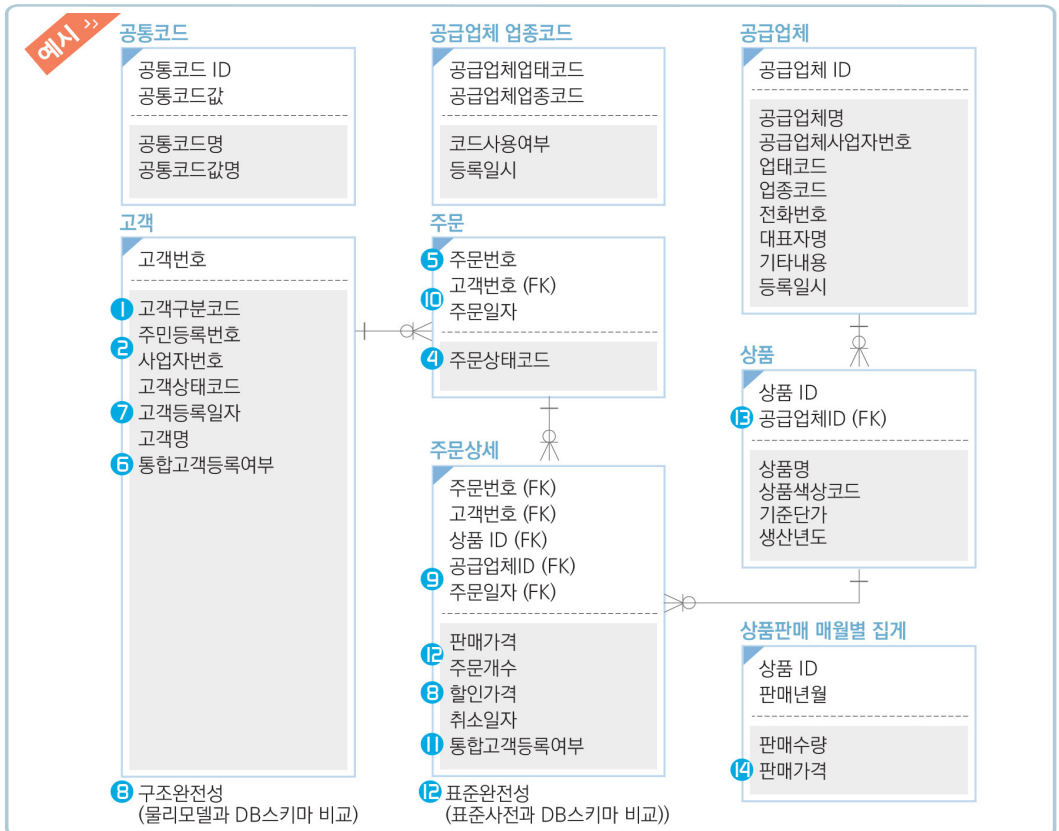
## 02

## 진단 규칙 예시

## 1 진단 대상 논리 데이터모델 예시

상품판매 DB 데이터 모델을 사례로 데이터 품질관리 지표별 정량적 진단규칙을 도출하는 예시와 상세한 도출 방법을 설명하고자 한다.

## ▶ 진단 대상 논리 데이터모델



▶ 데이터 품질 지표별 진단 예시

품질 지표	품질상세지표	진단대상	비고
완전성	단독 완전성	고객구분코드	진단 예시 1
	조건 완전성	주민등록/사업자번호	진단 예시 2
	구조 완전성	물리모델 <-> DB 스키마	진단 예시 3
유효성	코드 유효성	주문상태코드	진단 예시 4
	형식 유효성	주문번호	진단 예시 5
	여부 유효성	통합고객등록여부	진단 예시 6
	날짜 유효성	고객등록일자	진단 예시 7
	범위 유효성	할인가격	진단 예시 8
	선후관계 유효성	주문일자 <-> 배송일자	진단 예시 9
일관성	참조 무결성	고객번호	진단 예시 10
	중복 일관성	통합고객등록여부	진단 예시 11
	표준 일관성	표준사전 <-> DB 스키마	진단 예시 12
정확성	논리관계 정확성	상품(공급업체ID, 색상코드)	진단 예시 13
	파생항목 정확성	판매가격	진단 예시 14

2 진단 예시 1 > 단독 완전성

단독 완전성 품질지표를 기준으로 ‘고객구분코드’ 항목의 업무규칙 진단기법에 의한 진단규칙 도출 예시

1. 진단대상 분석(업무규칙)

1.1 ‘고객구분코드’ 항목의 NOT NULL 제약조건(구조) 확인

- 단독완전성 항목이 NOT NULL 구조인 경우 품질진단 대상에서 제외하며 NULL 값이 허용 되는 경우 품질진단 대상이 될 수 있다.

테이블명	컬럼명	NULL여부
고객	고객번호	N
	고객구분코드	Y
	주민등록번호	Y

(조회 SQL - Oracle 11g 기준)

```
SELECT  TABLE_NAME      AS 테이블명
        , COLUMN_NAME     AS 컬럼명
        , NULLABLE        AS NULL여부
FROM    ALL_TAB_COLUMNS
WHERE   TABLE_NAME = '고객';
```

## 1.2 고객 테이블 데이터 구성 확인

고객번호	고객구분코드	주민등록번호	사업자번호
E0001	C02		10-181-39750
E0002		76****-*****	
E0003	C02		60-381-13200

(조회 SQL - Oracle 11g 기준)

```
SELECT  고객번호
        , 고객구분코드
        , 주민등록번호
        , 사업자번호
FROM    고객 ;
```

## 1.3 고객구분코드 업무규칙 정보 확인

- '고객구분코드' 항목의 값은 물리적으로는 NULL을 허용하는 구조이나 업무적으로 NOT NULL 이어야 한다.

## 2. 업무규칙 대상 분석 결과 및 진단규칙 도출

- 업무규칙 정보를 참고하여 '고객구분코드' 항목에 NULL 값이 있는 경우 오류로 도출하는 진단 규칙을 적용한다.

(진단규칙 SQL - Oracle 11g 기준)

```
SELECT 고객번호
        , 고객구분코드
FROM    고객
WHERE   고객구분코드 IS NULL ;
```

I

II

III

IV

V

부  
록



### 3 진단 예시 2 > 조건 완전성

조건 완전성 품질지표를 기준으로 ‘고객구분코드’, ‘주민등록번호’/‘사업자번호’ 항목의 업무규칙 진단기법에 의한 진단규칙 도출 예시

#### 1. 진단대상 분석(업무규칙)

##### 1.1 ‘주민등록번호’, ‘사업자번호’ 항목의 NOT NULL 제약조건(구조) 확인

- 조건완전성 진단 대상(주민등록번호, 사업자번호) 항목이 NOT NULL 구조인지 확인 필요하며 조건의 대상이 되는 항목의 구조가 NULL 값이 허용 되는 경우 품질진단 대상이 될 수 있다.

테이블명	컬럼명	NULL여부
고객	고객번호	N
	고객구분코드	Y
	주민등록번호	Y
	사업자번호	Y

(조회 SQL - Oracle 11g 기준)

```
SELECT  TABLE_NAME      AS 테이블명
        , COLUMN_NAME     AS 컬럼명
        , NULLABLE        AS NULL여부
FROM    ALL_TAB_COLUMNS
WHERE   TABLE_NAME = '고객';
```

##### 1.2 고객 테이블 데이터 구성 확인

고객번호	고객구분코드	주민등록번호	사업자번호
E0006	C01	76****-*****	
E0007	CO1		
E0008	CO1	80****-*****	
E0009	CO2		
E0010	CO2		20-782-07314

(조회 SQL - Oracle 11g 기준)

```
SELECT  고객번호
        , 고객구분코드
        , 주민등록번호
        , 사업자번호
FROM    고객 ;
```

※ 고객구분코드 값 : CO1 - 개인 , CO2- 사업자

## 1.3 고객구분코드 관련 업무규칙 정보 확인

- '고객구분코드' 항목의 값이 '개인'일 경우 '주민등록번호' 항목의 값만 존재해야 하며, '고객구분코드' 항목의 값이 '사업자'일 경우에는 '사업자번호' 항목의 값만 채워져야 하는 규칙에 위배된 경우 오류로 도출할 수 있다.

## 2. 진단대상 분석 결과 및 진단규칙 도출(업무규칙)

- 업무규칙 정보를 확인한 결과 다른 항목(고객구분코드) 조건에 의해 NULL 값이 물리적으로 허용되는 주민등록번호, 사업자번호 항목은 조건항목(고객구분코드) 값에 따라 NULL 인 데이터가 존재할 수 없는 경우에 위배된 경우 오류로 도출할 수 있다.

(진단규칙 SQL - Oracle 11g 기준)

```
SELECT  고객번호, 고객구분코드, 주민등록번호, 사업자번호
FROM    고객
WHERE   (고객구분코드 = 'C01' AND 주민등록번호 IS NULL)
OR      (고객구분코드 <> 'C01' AND 주민등록번호 IS NOT NULL)
OR      (고객구분코드 = 'C02' AND 사업자번호 IS NULL)
OR      (고객구분코드 <> 'C02' AND 사업자번호 IS NOT NULL) ;
```

I

II

III

IV

V

부  
록

## 4 진단 예시 3 > 구조 완전성

구조 완전성 품질지표를 기준으로 ‘물리모델’, ‘DBMS 물리 스키마’ 간 업무규칙 진단기법에 의한 진단규칙 도출 예시

### 1. 진단대상 분석(업무규칙)

#### 1.1 물리모델 정보 확인

- 물리모델과 DBMS물리스키마 정보를 비교 하여 불일치 하는 항목이 존재하는지 확인 필요(고객 테이블 물리모델 예시)

CUSTOMER (고객) 테이블					
컬럼명	컬럼한글명	타입	PK 여부	길이	NOT NULL 여부
CUST_NO	고객번호	VARCHAR2	Y	10	Y
CUST_DVCD	고객구분코드	VARCHAR2	N	5	N
JUMIN_RGST_NO	주민등록번호	VARCHAR2	N	15	N
BS_RGST_NO	사업자번호	VARCHAR2	N	15	N
CUST_STCD	고객상태코드	VARCHAR2	N	5	N
CUST_RGST_DT	고객등록일자	VARCHAR2	N	8	N
CUST_NM	고객명	VARCHAR2	N	10	N
COMB_RES_YN	통합고객등록여부	VARCHAR2	N	1	N

#### 1.2 DBMS 물리스키마 정보 확인

- 물리모델과 DBMS 물리스키마 정보를 비교하여 불일치하는 항목이 존재하는지 확인 필요

컬럼명	컬럼한글명	타입	PK 여부	길이	NOT NULL 여부
CUST_ID	고객ID	VARCHAR2	Y	10	Y
CUST_DVCD	고객구분코드	VARCHAR2		5	
JUMIN_RGST_NO	주민등록번호	VARCHAR2		15	
BS_RGST_NO	사업자번호	VARCHAR2		10	

### 2. 진단대상 분석 결과 및 진단규칙 도출(업무규칙)

- ‘물리모델’, ‘DBMS 물리스키마’를 근거로 물리모델로 설계된 내용(타입, 사이즈 등)이 실제 DBMS에 구현되어 있는지 진단한다.

(진단규칙 SQL - Oracle 11g 기준)

```
SELECT  A.TABLE_NAME      AS 테이블명
        , A.COLUMN_NAME   AS 컬럼명
        , B.COMMENT       AS 컬럼한글명
        , A.DATA_TYPE      AS 타입
        , A.DATA_LENGTH    AS 길이
```

```

, A.NULLABLE          AS   NULL여부
FROM   ALL_TAB_COLUMNS A
       LEFT OUTER JOIN ALL_COL_COMMENTS(모델정보) B
       ON (A.TABLE_NAME = B.TABLE_NAME
AND       A.COLUMN_NAME = B.COLUMN_NAME)
WHERE   B.COLUMN_NAME IS NULL
OR       A.DATA_TYPE <> B.DATA_TYPE
OR       A.DATA_LENGTH <> B.DATA_LENGTH ;

```

I

II

III

IV

V

부  
록

## 5 진단 예시 4 > 코드 유효성

코드 유효성 품질지표를 기준으로 '주문상태코드' 항목의 프로파일 진단기법을 활용한 진단규칙 도출 예시

### 1. 진단대상 분석

#### 1.1 주문 테이블 주문상태코드 항목 프로파일 분석

- '고객구분코드' 항목 프로파일(DISTINCT , COUNT) 수행

테이블명	DISTINCT	COUNT
주문	A	50
	B	1200
	FA	2

(조회 SQL - Oracle 11g 기준)

```
SELECT      주문상태코드
            , COUNT(*) AS CNT
FROM        주문
GROUP BY    주문상태코드 ;
```

#### 1.2 공통 코드 데이터 구성 확인

- 공통코드 테이블에 정의 된 '주문상태코드' 항목의 값은 'A','B'로 정의 되어 다.

공통코드ID	공통코드명	코드값	코드값명
1001	주문상태코드	A	주문
1001	주문상태코드	B	주문취소

(조회 SQL - Oracle 11g 기준)

```
SELECT      공통코드ID
            , 공통코드명
            , 공통코드값
            , 공통코드값명
FROM        공통코드
WHERE       공통코드명 = '주문상태코드' ;
```

### 2. 진단대상 분석 결과 및 진단규칙 도출

- 공통코드 테이블에 정의 되어 있지 않은 '주문상태코드' 항목 값은 오류로 도출한다.

(진단규칙 SQL - Oracle 11g 기준)

```
SELECT      주문.주문상태코드 AS 주문상태코드
            , COUNT(*)        AS 코드건수
```

```
FROM      주문 LEFT OUTER JOIN 공통코드
          ON ('1001' = 공통코드.분류코드
              AND 주문.주문상태코드 = 공통코드.코드값)
WHERE      공통코드.코드값 IS NULL
GROUP BY  주문.주문상태코드 ;
```

I

II

III

IV

V

부  
록

## 6 진단 예시 5 > 형식 유효성

형식 유효성 품질지표를 기준으로 '주문번호' 항목의 프로파일 진단기법을 활용한 진단규칙 도출 예시

### 1. 진단대상 분석

#### 1.1 주문 테이블 주문번호 항목 프로파일 분석

- '주문번호' 항목 프로파일(Pattern) 수행

테이블명	컬럼명	PATTERN	길이	COUNT
주문	주문번호	99999XXXXX	10	1500
		9999XXXXXX	10	7

(조회 SQL - Oracle 11g 기준)

```
SELECT   LENGTH(주문번호)      AS 길이
        , TRANSLATE (주문번호, '0123456789ABCDE...abcde...'
                     , '9999999999XXXXX...XXXXX...' ) as PATTERN
        , COUNT (*) AS CNT
FROM     주문
GROUP BY LENGTH(주문번호)
        , TRANSLATE (주문번호, '0123456789ABCDE...abcde...'
                     , '9999999999XXXXX...XXXXX...' )
```

#### 1.2 주문 테이블 데이터 구성 확인

- 주문 테이블의 '주문번호' 항목 값은 아래와 같이 구성 되어 있다.

주문번호	고객번호	주문일자	주문상태코드
00015DAAAL	E0006	20191101	02
0267AAQBG	E0008	20191214	01

(조회 SQL - Oracle 11g 기준)

```
SELECT  주문번호
        , 고객번호
        , 주문일자
        , 주문상태코드
FROM    주문 ;
```

#### 1.3 주문번호 생성 규칙 확인

- ERD, 업무규칙 정의서, 업무담당자 등 다양한 경로의 정보를 참고하여 주문번호 생성 규칙을 확인 할 수 있다.

항목명	데이터 발생 규칙
주문번호	영문5자리+숫자5자리(예시: 99999ZZZZZ)

## 2. 진단대상 분석 결과 및 진단규칙 도출

- 주문번호 항목의 프로파일 및 생성 규칙 확인 결과 주문번호는 '영문5자리 + 숫자5자리'로 구성되어야 한다. 이에 규칙에 위배되는 값은 오류로 도출한다.

(진단규칙 SQL - Oracle 11g 기준)

```
SELECT  주문번호
        , 고객번호
        , 주문일자
        , 주문상태코드
FROM    주문
WHERE   REGEXP_REPLACE(주문번호,'[0-9]{5}[a-zA-Z]{5}',") > '0' ;
```

I

II

III

IV

V

부  
록



## 7 진단 예시 6 > 여부 유효성

여부 유효성 품질지표를 기준으로 '통합고객등록여부' 항목의 프로파일 진단기법을 활용한 진단 규칙 도출 예시

### 1. 진단대상 분석

#### 1.1 고객 테이블 통합고객등록여부 항목 프로파일 분석

- '통합고객등록여부' 항목 프로파일(DISTINCT,COUNT) 수행

테이블명	통합고객등록여부	CNT
고객	Y	650
	N	400
	1	13

(조회 SQL - Oracle 11g 기준)

```
SELECT   통합고객등록여부
          , COUNT(*) AS CNT
FROM     고객
GROUP BY 통합고객등록여부 ;
```

#### 1.2 여부 도메인 생성 규칙 확인

- ERD, 도메인정의서 등 다양한 경로의 정보를 참고하여 통합고객등록여부 항목의 생성 규칙을 확인 할 수 있다.

도메인명	TYPE	길이	정의
여부	VARCHAR2	1	유효값은 'Y' , 'N' 으로 정의한다.

### 2. 진단대상 분석 결과 및 진단규칙 도출

- 통합고객등록여부 항목의 프로파일 및 생성 규칙 확인 결과 'Y','N'으로 구성되어야 한다. 규칙에 위배되는 값은 오류로 도출한다.

(진단규칙 SQL - Oracle 11g 기준)

```
SELECT   고객번호
          , 통합고객등록여부
FROM     고객
WHERE    통합고객등록여부 NOT IN ('Y','N') ;
```

## 8 진단 예시 7 > 날짜 유효성

날짜 유효성 품질지표를 기준으로 '고객등록일자' 항목의 프로파일 진단기법을 활용한 진단규칙 도출 예시

### 1. 진단대상 분석

#### 1.1 고객 테이블 고객등록일자 항목 프로파일 분석

- '고객등록일자' 항목 프로파일(Pattern) 수행

테이블명	컬럼명	PATTERN	COUNT
고객	고객등록일자	YYYY-MM-DD	3
		YYYYMMDD	950

(조회 SQL - Oracle 11g 기준)

```
SELECT    TRANSLATE (고객등록일자, '0123-45-67'
                    , '9999-99-99' ) as PATTERN
          , COUNT (*) AS CNT
FROM      고객
GROUP BY  TRANSLATE (고객등록일자, '0123-45-67'
                    , '9999-99-99' ) ;
```

#### 1.2 고객 테이블 데이터 구성 확인

- 고객 테이블의 '고객등록일자' 항목의 값은 아래와 같다.

고객번호	고객구분코드	고객등록일자
E0004	C01	2019-01-03
E0019	C01	20191205
E0250	C01	20190101

(조회 SQL - Oracle 11g 기준)

```
SELECT  고객번호
        , 고객구분코드
        , 고객등록일자
FROM    고객 ;
```

#### 1.3 등록일자 생성 규칙 확인

- ERD, 업무규칙 정의서, 업무담당자 등 다양한 경로의 정보를 참고하여 등록일자 생성 규칙을 확인 할 수 있다.

항목명	데이터 발생 규칙
등록일자	YYYYMMDD 형식을 준수한다.

### 2. 진단대상 분석 결과 및 진단규칙 도출

- '고객등록일자' 항목의 프로파일 결과 및 생성 규칙 확인 결과 고객등록일자 항목의 값 중

I

II

III

IV

V

부  
록

YYYYMMDD 형식에 위배되는 데이터는 오류로 도출한다.

(진단규칙 SQL - Oracle 11g 기준)

```
SELECT   고객번호
        , 고객등록일자
FROM     고객
WHERE    (NOT REGEXP_LIKE(고객등록일자,'^[0-9]
        OR NOT REGEXP_LIKE(고객등록일자,'^
        (([0-9][0-9][0-9][0-9][0-9][1-9][0-2][0-9])|([1][0-2][3][0-1])$'))
        OR REGEXP_LIKE(고객등록일자, '[0230|0231|0431|0631|0931|1131]$');
```

※ 본 예시는 윤달에 의한 조건은 예외로 작성

## 9 진단 예시 8 > 범위 유효성

범위 유효성 품질지표를 기준으로 '할인가격' 항목의 프로파일 진단기법을 활용한 진단규칙 도출 예시

### 1. 진단대상 분석

#### 1.1 주문상세 테이블 '할인가격' 항목 프로파일 분석

- '통합고객등록여부' 항목 프로파일(DISTINCT,COUNT) 수행

테이블명	컬럼명	MIN(할인가격)	MAX(할인가격)
주문상세	할인가격	-12000	8770

(조회 SQL - Oracle 11g 기준)

```
SELECT      MIN(할인가격) AS MIN할인가격
            ,MAX(할인가격) AS MAX할인가격
FROM        주문상세 ;
```

#### 1.2 할인가격 생성 규칙 확인

- ERD, 업무규칙정의서 등 다양한 경로의 정보를 참고하여 할인가격 항목의 생성 규칙을 확인 할 수 있다.

항목명	데이터 발생 규칙
할인가격	할인가격의 범위는 0~150,000 로 정의한다.

#### 1.3 주문상세 테이블 데이터 구성 확인

- 주문상세 테이블의 '할인가격' 항목 값은 아래와 같다.

주문번호	고객번호	판매가격	할인가격
11276AB	E0046	93850	-12000
21987AK	E0070	50000	1000
11387AZ	E0124	45000	450

### 2. 진단대상 분석 결과 및 진단규칙 도출

- 할인가격 항목의 프로파일 및 데이터 생성 규칙 확인 결과 범위 값 0~150,000 으로 구성되어야 하며, 규칙에 위배되는 값은 오류로 도출한다.

(진단규칙 SQL - Oracle 11g 기준)

```
SELECT      주문번호
            , 고객번호
            , 판매가격
            , 할인가격
FROM        주문상세
WHERE 할인가격 NOT BETWEEN 0 AND 150000 ;
```

I

II

III

IV

V

부  
록

## 10 진단 예시 9 > 선후관계 유효성

선후관계 유효성 품질지표를 기준으로 '주문일자', '배송일자' 항목의 프로파일 진단기법을 활용한 진단규칙 도출 예시)

### 1. 진단대상 분석

#### 1.1 하나의 테이블에 날짜' 도메인유형 항목이 2개 이상인 대상 확인

테이블명	컬럼명	데이터타입	데이터길이
주문상세	주문일자	VARCHAR2	14
	배송일자	VARCHAR2	8
	취소일자	VARCHAR2	8

(조회 SQL - Oracle 11g 기준)

```
SELECT A.TABLE_NAME, COLUMN_NAME, DATA_TYPE, DATA_LENGTH
FROM ALL_TAB_COLUMNS A
      INNER JOIN (SELECT OWNER, TABLE_NAME, COUNT(*) AS CNT
                  FROM ALL_TAB_COLUMNS
                  WHERE OWNER='TEST_SVR'
                  AND REGEXP_LIKE(COLUMN_NAME,
                                  '_DT|DTM|_DAY|_YM|_MM$')
                  GROUP BY OWNER, TABLE_NAME
                ) B
ON (A.OWNER = B.OWNER
    AND A.TABLE_NAME = B.TABLE_NAME)
WHERE B.CNT > 1
AND REGEXP_LIKE(COLUMN_NAME, '_DT|DTM|_DAY|_YM|_MM$');
```

#### 1.2 주문상세 테이블 데이터 구성 확인

- 배송일자는 주문일자 항목보다 앞서는 값이 존재 할 수 없다.

주문번호	고객번호	주문일자	배송일자
00005DAAAB	E0004	20190421	20190101
00011AAABD	E0145	20191210	20191212
000125AOADB	E0009	20191011	20191118

(조회 SQL - Oracle 11g 기준)

```
SELECT      주문번호
            , 고객번호
            , 주문일자
            , 배송일자
FROM        주문상세 ;
```

## 2. 진단대상 분석 결과 및 진단규칙 도출

- 주문상세 테이블의 '주문일자', '배송일자' 항목의 데이터를 조회한 결과 '배송일자' 항목은 '주문일자' 항목보다 앞서는 값이 존재하는 경우 오류로 도출한다.

(진단규칙 SQL - Oracle 11g 기준)

```
SELECT  주문번호
        , 고객번호
        , 주문일자
        , 배송일자
FROM    주문상세
WHERE   배송일자 IS NOT NULL
AND     주문일자 > 배송일자 ;
```

I

II

III

IV

V

부  
록

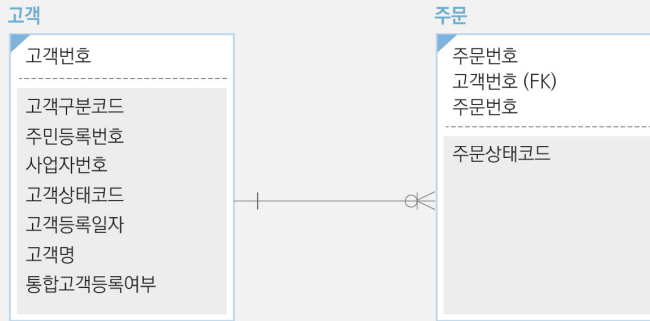
## 11 진단 예시 10 > 참조 무결성

참조 무결성 품질지표를 기준으로 '고객번호(FK)'항목의 업무규칙 진단기법을 활용한 진단규칙 도출 예시

### 1. 진단대상 분석(업무규칙)

#### 1.1 ERD 모델 및 데이터 규칙정의/업무규칙 정보 확인

- ERD 모델 관계선 및 업무적으로 발생 되는 데이터발생 규칙 등을 참고하여 진단규칙을 도출 할 수 있다.



#### 1.2 주문, 고객 테이블 데이터 확인

주문테이블			고객테이블
주문번호	고객번호	주문일자	고객번호
00005DAAAB	00005	20190503	00005
00006DAAAC	00007	20190803	00007
00003GGGGB	00008	20191123	

(조회 SQL - Oracle 11g 기준)

```
SELECT  주문.주문번호
        , 주문.주문일자
        , 주문.주문고객번호
        , 고객.고객번호
FROM    주문 LEFT OUTER JOIN 고객
        ON (주문.주문고객번호 = 고객.고객번호)
```

### 2. 진단대상 분석 결과 및 진단규칙 도출

Foreign Key 제약조건이 없는 항목을 대상으로 하여 ERD 논리 모델을 분석하여 진단 규칙을 도출한다.

(진단규칙 SQL - Oracle 11g 기준)

```
SELECT  고객.고객번호
```

```
        , 고객.고객등록일자  
        , 고객.고객명  
        , 주문.주문번호  
        , 주문.주문일자  
        , 주문.주문고객번호  
FROM      주문 LEFT OUTER JOIN 고객  
        ON (주문.주문고객번호 = 고객.고객번호)  
WHERE      고객.고객ID IS NULL ;
```

I

II

III

IV

V

부  
록



12

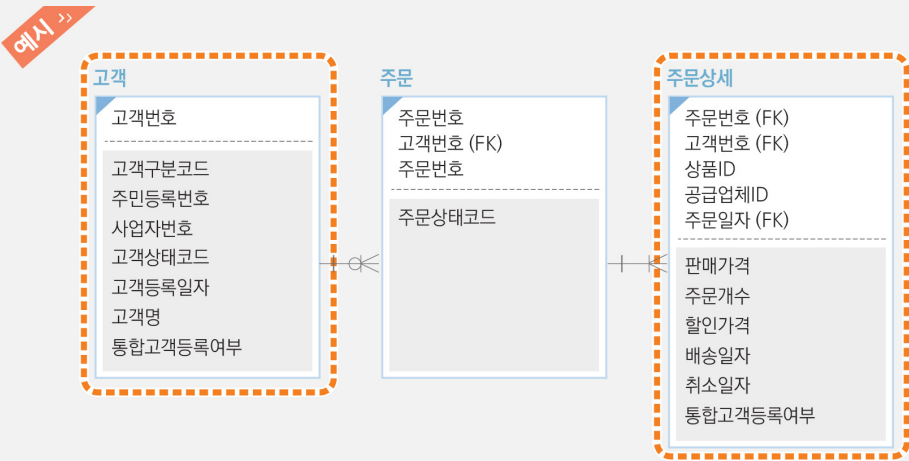
진단 예시 11 > 중복 일관성

중복 일관성 품질지표를 기준으로 ‘통합고객등록여부’ 항목의 업무규칙 진단기법에 의한 진단규칙 도출 예시

1. 진단대상 분석(업무규칙)

1.1 ERD 모델 및 정의서/업무규칙 정보 확인

- ERD & 속성(컬럼) 정의서에 정의 된 반정규화 관련 내용 및 업무담당자, 운영담당자 등 관련자들의 정보를 참고하여 진단규칙을 도출할 수 있다.
- 업무규칙 정보를 확인한 결과 주문상세 테이블의 ‘고객번호’, ‘통합고객등록여부’ 항목의 원천 데이터는 고객 테이블의 ‘고객번호’, ‘통합고객등록여부’ 항목임이 확인되었다.



1.2 주문상세 및 고객 테이블 데이터 확인

주문상세 테이블			고객 테이블	
주문번호	고객번호	통합고객등록여부	고객번호	통합고객등록여부
00001AAAAA	00001	Y	00001	
00002CCCCA	00002	Y	00002	Y
00003GGGGB	00003	Y	00003	Y
11236SDFGH	00005	Y	00005	N

(조회 SQL - Oracle 11g 기준)

```
SELECT  주문상세.주문번호
        , 주문상세.주문고객번호
        , 주문상세.통합고객등록여부
        , 고객.고객번호
        , 고객.통합고객등록여부
FROM    주문상세 INNER JOIN 고객
ON      (주문상세.주문고객번호 = 고객.고객번호)
```

## 2. 진단대상 분석 결과 및 진단규칙 도출

ERD & 속성(컬럼) 정의서에 정의된 반정규화 관련 내용 및 업무담당자, 운영담당자 등 관련자들의 정보 바탕으로 주문 테이블의 '주문고객번호', '주문통합고객등록여부' 항목의 원천은 고객 테이블의 '고객번호', '통합고객등록여부' 항목으로 동일한 시점에 원천데이터와 중복데이터가 같은 값을 가지고 있어야 함을 진단 규칙으로 도출할 수 있다.

(진단규칙 SQL - Oracle 11g 기준)

```
SELECT  주문고객번호
        , 주문_통합고객등록여부
        , 고객_통합고객등록여부
FROM    주문상세 INNER JOIN  고객
        ON (주문상세.주문고객번호 = 고객.고객번호)
WHERE   주문상세.통합고객등록여부 <> 고객.통합고객등록여부 ;
```

I

II

III

IV

V

부  
록

## 13 진단 예시 12 > 표준 일관성

표준 일관성 품질지표를 기준으로 '표준관리', 'DBMS 테이블'간 프로파일 진단기법을 활용한 진단규칙 도출 예시

### 1. 진단대상 분석

#### 1.1 표준관리(표준용어사전) 문서 확인

표준용어	표준도메인	데이터 타입	데이터 길이
휴대폰번호	전화번호	VARCHAR2	20
집전화번호	전화번호	VARCHAR2	20

#### 1.2 DBMS 테이블 정보

컬럼명	데이터 타입	데이터 길이
핸드폰전화번호	VARCHAR2	20
집전화번호	VARCHAR2	20

### 2. 진단대상 분석 결과 및 진단규칙 도출(업무규칙)

- 표준관리(표준용어사전) 문서와 DBMS 테이블 정보 분석 결과 표준용어사전에 등록되어 있지 않은 항목을 컬럼명으로 사용하고 있음이 확인되어 오류로 도출한다.

(진단규칙 SQL - Oracle 11g 기준)

```
SELECT B.테이블명
      , B.컬럼명
```

```
FROM   표준용어사전 A, DBMS테이블정보 B
```

```
WHERE  A.표준용어 != B.컬럼명 ;
```

※ 본 예시는 표준관리 문서가 DB에 적재 된 경우를 가정

## 14 진단 예시 13 &gt; 논리관계 정확성

논리관계 일관성 품질지표를 기준으로 ‘상품명’, ‘공급업체명’ 항목의 업무규칙 진단기법에 의한 진단규칙 도출 예시

## 1. 진단대상 분석(업무규칙)

## 1.1 ERD 모델 및 데이터 규칙정의/업무규칙 정보 확인

- ERD, 속성(컬럼) 정의서에 정의 된 테이블/컬럼 내용 및 업무담당자, 운영담당자 등 관련자들의 정보를 참고하여 진단규칙을 도출 할 수 있다.
- “「상품」테이블의 「상품명」이 본체인 상품 중 화이트(WH) 색상인 상품은 MS사에서만 공급한다”라는 업무규칙이 확인되었다.

## 1.2 ‘공급업체’ 테이블 정보 확인

공급업체ID	공급업체명
15476	기린전자
18555	MS

(조회 SQL - Oracle 11g 기준)

```
SELECT   공급업체ID
        , 공급업체명
FROM     공급업체 ;
```

## 1.3 ‘상품’ 테이블 정보 확인

상품ID	공급업체ID	상품명	상품색상코드
A549	15476	키보드	BL
A550	15476	본체	WH
A551	18555	본체	WH

(조회 SQL - Oracle 11g 기준)

```
SELECT   상품ID
        , 공급업체ID
        , 상품명
        , 상품색상코드
FROM     상품 ;
```

## 2. 진단대상 분석 결과 및 진단규칙 도출

‘상품’테이블의 정보를 확인한 결과 ‘상품명’이 본체인 상품 중 화이트(WH) 색상인 상품에서 ‘MS’사가 아닌 ‘기린전자’로 등록 된 값이 존재함을 확인하였다. 따라서 아래와 같은 진단규칙을 도출할 수 있다.

(진단규칙 SQL - Oracle 11g 기준)

```
SELECT   상품ID
        , 공급업체ID
```

I

II

III

IV

V

부  
록

```
        , 상품명  
        , 상품색상코드  
        , 기준단가  
        , 생산년도  
FROM   상품 A  
        LEFT OUTER JOIN 공급업체 B  
        ON (A.공급업체ID = B.공급업체ID)  
WHERE  A.상품명 = '본체'  
AND    A.상품색상코드 = 'WH'  
WHERE  B.공급업체명 <> 'MS';
```

## 15 진단 예시 14 &gt; 파생항목 정확성

파생항목 정확성 품질지표를 기준으로 '판매가격' 항목의 업무규칙 진단기법에 의한 진단규칙 도출 예시

## 1. 진단대상 분석(업무규칙)

## 1.1 ERD 모델 및 데이터 규칙정의/업무규칙 정보 확인

- 수치(수,울,량) 도메인 컬럼 중 연산에 의해 데이터가 발생하는 컬럼을 대상으로 업무담당자, 운영담당자 등 관련자들의 정보를 참고하여 진단규칙을 도출할 수 있다.
- 주문상세 테이블의 '판매가격' 항목의 연산규칙은 아래 예시와 같다.  
(판매가격 연산규칙 예시 - 기준단가 X 주문개수 - 할인가격 = 판매가격)

## 1.2 주문상세 및 상품 테이블 정보 확인

주문상세 테이블					상품 테이블
주문번호	상품ID	판매가격	주문개수	할인가격	기준단가
11279JIIH	A547	141230	50	8770	3000
11265SDFGH	A547	50200	17	800	3000
11260SDFGH	A547	96000	32	300	3000
11266SDFGH	A785	1200000	14	900	85000
11281JIIH	B732	52766	54	1234	1000

(조희 SQL - Oracle 11g 기준)

```
SELECT    A.주문번호
          , A.상품ID
          , A.판매가격
          , A.주문개수
          , A.할인가격
          , B.기준단가
FROM      주문상세 A LEFT OUTER JOIN 상품 B
          ON (A.상품ID = B.상품ID
AND A.공급업체ID = B.공급업체ID) ;
```

## 2. 진단대상 분석 결과 및 진단규칙 도출

- 주문상세 테이블의 '판매가격' 항목의 데이터 생성 규칙(연산규칙)을 기반으로 아래와 같은 진단규칙을 도출할 수 있다.

(진단규칙 SQL - Oracle 11g 기준)

```
SELECT    A.주문번호
          , A.상품ID
          , A.판매가격
          , B.기준단가 * A.주문개수 - A.할인가격 AS 연산결과
FROM      주문상세 A LEFT OUTER JOIN 상품 B
```

```
ON (A.상품ID = B.상품ID
AND    A.공급업체ID = B.공급업체ID)
WHERE  A.판매가격 < B.기준단가 * A.주문개수 - A.할인가격 ;
```

# 빅데이터 플랫폼 및 센터 데이터 품질관리 가이드

2020년 2월 초판 인쇄

2020년 2월 초판 발행

발행인 문 용 식

발행처 한국정보화진흥원

대구광역시 동구 첨단로 53

본 가이드는 2019년 한국정보화진흥원의 「빅데이터 플랫폼 및 센터 데이터품질관리」 용역 사업의 결과 산출물입니다.

본 가이드 내용의 무단 전재 및 재배포를 금하며, 인용 시에는 반드시 자료의 출처인 한국정보화진흥원 “빅데이터 플랫폼 및 센터 데이터 품질관리 가이드” 라고 밝혀주시기 바랍니다.

본 가이드의 내용과 관련한 문의는 한국정보화진흥원 지능데이터본부(지능데이터기반팀)로 연락주시기 바랍니다.



