



데이터 이해

Home > DB 구축 가이드 > DA 가이드 > 데이터 품질관리 이해 > 데이터 이해

데이터 품질 관리 프레임워크

표준 데이터

모델 데이터

관리 데이터

업무 데이터

+ 데이터 품질 관리 프레임워크

데이터에 관한 연구는 지난 30년 동안 IT 분야에 이론적으로나 산업적으로 엄청난 변화를 일으켜 왔다. 현재 데이터 품질 관리에 관한 연구는 경영 관리 사이클(Plan-Do-See) 관점에서 재해석되고 있으며, 많은 학계 및 산업계에서 데이터 품질에 대한 중요성을 인식하고 있다.

데이터 품질 관리란 조직 내외부의 지식 노동자와 최종 사용자의 기대를 만족시키기 위한 지속적 인 데이터 및 데이터 서비스 개선 활동을 말한다. 한국데이터베이스진흥원은 데이터 품질 관리 개선 과 관리를 위한 체계적인 계획을 수립하여 2003년에 데이터베이스 품질 관리 확장 모델을 개발한 후, 2005년에는 해당 실무에서 실제로 적용할 수 있는 데이터 베이스 품질 관리 지침과 해설서를 제 시함으로써 지속적인 개선을 통한 품질 관리를 지향하고 있다.

데이터 품질 요소에는 크게 데이터 값(Data Value), 데이터 서비스(Data Service), 데이터 구조 (Data Hierarchy), 데이터 관리 프로세스(Data Management Process) 등이 있다. 이러한 요소들 은 서로 연계되어 조직 데이터 품질에 영향을 주고 있어 종합적이고 체계적인 품질 관리 노력을 필요 로 한다. 그 결과 [표 6-1-1]과 같은 데이터 품질 관리 프레임워크를 개발하였다. 프레임워크를 구성 하고 있는 각 요소에 대한 설명은 본 과목에서 다루게 될 주요 내용으로, 2절 표준 데이터를 비롯해 본격적으로 데이터에 대한 이해를 하기 바란다.

[표 6-1-1] 데이터 품질 관리 프레임워크

대상 조직	데이터	데이터 구조	데이터 관리 프로세스
CIO / EDA (개괄적 관점)	데이터 관리 정책		
DA (개념적 관점)	표준 데이터	개념 데이터 모델 데이터 참조 모델	데이터 표준 관리 요구 사항 관리
Modeler (논리적 관점)	모델 데이터	논리 데이터 모델	데이터 모델 관리 데이터 흐름 관리
DBA (물리적 관점)	관리 데이터	물리 데이터 모델 데이터베이스	데이터베이스 관리
User (운용적 관점)	업무 데이터	사용자 뷰	데이터 활용 관리

※ [표6-1-1]의 'DA'는 Data Administrator를 의미함



데이터 품질 관리 프레임워크 표준 데이터 모델 데이터 관리 데이터 업무 데이터

+ 정의 및 관리 목적

표준 데이터란 정보시스템에서 사용되는 용어 및 도메인, 코드, 기타 데이터 관련 요소에 대해 공통된 형식과 내용으로 정의하여 사용하는 표준 관련 데이터를 의미한다. 표준 데이터는 정보시스템과 정보시스템 데이터의 품질 확보와 직결되는 요소로, 표준 데이터를 관리함으로써 기관이나 기업 전사 차원에서 단일화하고 표준화된 정보 시스템을 구현할 수 있다. 또한 데이터의 불일치나 데이터 오류를 방지하며, 표준화 되지 않은 데이터로 인해 야기되는 산출물 보정 작업 등을 최소화 함으로써 정보 시스템의 생산성을 향상시킬 수 있다. 표준 데이터를 관리함으로써 데이터에 대한 이해도를 높이고 의사소통을 원활하게 하며 데이터 통합을 효율적으로 수행할 수 있도록 한다.

현재 많은 기관 및 기업의 표준화 정도를 보면 대부분의 경우 단위 시스템별 표준화는 많은 부분에서 지켜지고 있지만 전사적인 표준화를 통해 통합 관리하는 경우는 매우 드물다. 그래서 전사 데이터 웨어하우스(EDW, Enterprise Data Warehouse)와 같은 통합 시스템을 구축할 경우 표준에 대한 재정비 및 재정비된 표준에 따른 기존 시스템에 대한 변경 작업을 위해 많은 인력 및 비용을 지불해야 한다. 표준 데이터는 관리 시스템 및 메타 관리 시스템을 도입하면 지속적이고 정량화된 관리가 가능하다.

+ 세부 관리 대상

▣ 표준 단어(Word) 사전

일반적으로 단어란 문법상 일정한 뜻과 구실을 가지는 말의 최소 단위를 의미하며, 정보 시스템에서 사용되는 표준 단어 사전이란 기업이나 기관에서 업무상 사용되며 일정한 의미를 갖고 있는 최소단위의 단어를 정의한 사전을 말한다. 표준 단어를 정의함으로써 업무상 편의나 관습에 따라 동일한 단어를 서로 다른 의미로 사용하는 경우(예, 자산 이관 시 '인수자', '이관자'를 주는 곳과 받는 곳에서 서로 상반되는 의미로 사용하는 경우가 있음), 혹은 하나의 단어에 다양한 의미를 부여(예, 처리자)하여 사용하는 등의 문제를 방지할 수 있다. 표준 단어 사전은 다음과 같은 기준에 따라 관리되어야 한다.

■ 표준성

표준 단어는 정보시스템 구축 대상 업무 범위에서 사용하고 있거나 일반적으로 사용되는 사전적 의미의 단어 가운데에서 추출해야 하며, 지나치게 업무에 의존적이거나 방언을 사용해서는 안되며 약어의 사용도 최소화해야 한다.

■ 참조 가능성

표준 단어는 기업이나 기관에서 새로운 업무를 정의할 때 참조할 수 있어야 한다.

■ 일반성

표준 단어는 일상적으로 사용하고 있는 사전적 의미의 단어와 의미상 크게 다르지 않아 일반인도 해당 단어의 의미를 이해할 수 있어야 한다.

■ 대표성

표준 단어는 동의어를 가질 수 있으나 표준 단어로 선언된 단어는 비슷한 의미의 동의어들을 대표할 수 있어야 한다.

표준 단어는 전사적으로 관리하고 있는 엔터티와 속성을 개별 단위로 하여 추출하며 추출된 단어는 동음이의어와 이음 동의어를 정비한 후 논리명(한글명)을 기준으로 물리명(영문명, 영문 약어명), 유사 용어까지 함께 정리하여 관리한다. 표준 단어 사전에는 개별 단어 외에도 동의어, 유의어, 반의어 등과 같은 단어 간의 구조도 함께 정의해야 한다.

▣ 표준 도메인(Domain) 사전

도메인이란 속성에 정의된 조건을 만족시키는 값의 범위를 의미하며, 표준 도메인은 전사적으로 사용되고 있는 데이터 중에서 논리적, 물리적으로 유사한 유형의 데이터를 그룹화하여 해당 그룹에 속하는 데이터의 유형과 길이를 정의한 것을 말한다. 도메인은 여러 개의 하위 도메인(복합 도메인)으로 구성되거나 하나의 도메인이 여러 개의 도메인에 중복적으로 사용될 수 있다. 표준 도메인은 다음과 같은 기준에 따라 관리되어야 한다.

■ 표준성

표준 도메인은 전사 차원에서 공통적으로 사용되는 속성을 대상으로 정의한다. 예를 들어 은행의 계좌번호는 은행 하위 업무나 상품에 따라 다르지 않으므로 표준 도메인을 정의하여 사용해야 한다.

■ 유일성

동일한 내용의 중복 도메인이 서로 다른 이름으로 선언되지 않도록 관리해야 한다.

■ 업무지향성

도메인은 지나치게 일반화하여 정의하기 보다는 업무의 특성을 충분히 반영할 수 있도록 선언하여 관리한다. 예를 들어 계좌번호의 도메인은 '-'가 없이 정의하는 것 보다 적절한 의미를 나타내도록 '-'를 이용하여 표현한다.

전사적으로 관리하고 있는 모든 데이터 속성 혹은 대표 속성 가운데에 DBMS(DatabaseManagement System)에 동일한 형태로 구현되는 속성들을 추출하여 그룹화한다. 모든 속성은 임의의 도메인에 할당되어야 하며, 하나 이상의 도메인에 복수로 할당되어서는 안 된다. 속성과 도메인은 상호 매핑하여 관리해야 하며 새로운 속성이 추가될 경우 해당 속성의 도메인을 선정, 등록할 것을 권장한다. 또한 도메인의 삭제는 해당 도메인을 사용하고 있는 속성이 없을 경우에만 가능하도록 해야 한다.

▣ 표준 용어(Terms) 사전

용어는 업무에서 자주 사용하는 단어의 조합을 의미하며, 표준 용어는 전사적으로 사용하는 엔터티와 속성을 대상으로 표준 단어 사전에 정의된 단어를 조합하여 정의한다. 단어는 개별적이나 용어는 업무와 조직의 성격에 따라 그 조합이 달라질 수 있다. 표준 용어를 정의함으로써 기업 내부에서 서로 상이한 업무 간에 의사소통이 필요한 경우 용어에 대한 이해 부족으로 유발되는 문제점을 최소화 할 수 있다. 표준 용어 사전은 다음과 같은 기준에 따라 관리되어야 한다.

■ 표준성

같은 기업 내부라도 업무별로 동일한 의미를 서로 다른 용어를 사용하여 표현하는 경우가 매우 많다. 따라서 표준 용어 사전은 용어의 표준화를 통해 용어 사용의 차이에 따라 발생하는 전사 차원의 혼란을 최소화 할 수 있어야 한다.

■ 일반성

용어가 지나치게 업무 관점에서만 정의되어 일반적으로 이해하기 힘들거나 의미상 혼란을 초래해서는 안 된다. 일반적 인 의미와 전혀 다르게 사용된 용어는 적절한 다른 용어로 대체하고, 새로운 용어 개발 또한 자제해야 한다.

■ 업무지향성

용어는 기업의 업무 범위 내에서 약어를 사용하거나 내부에서 별도로 정의하여 사용할 수 있다. 단 지나친 약어의 사용은 업무에 대한 이해도를 떨어뜨릴 수 있으므로 주의한다.

표준 용어는 전사적으로 보유하고 있는 엔터티와 속성을 대상으로 추출된 표준 단어를 조합하여 생성되며, 용어 사전은 엔터티 용어 사전과 속성 용어 사전으로 구분하여 정의 관리한다. 정의된 각각의 용어는 논리명(한글명)과 물리명(영문명)을 가지며, 용어 범위 및 자격 형식 등이 설명되어야 한다.

▣ 표준 코드

표준 코드에는 각 산업별로 법적, 제도적으로 부여하여 공통적으로 사용되는 코드뿐만 아니라 기관이나 기업 내부에서 정의하여 사용하는 코드가 대상이 된다. 표준 코드는 다음과 같은 기준에 따라 관리되어야 한다.

■ 재사용성

표준 코드는 기관이나 기업에서 자체적으로 정의하여 사용하는 것보다 표준화 기구나 정부, 공공 기관에서 정의한 코드를 재??도 관리를 용이하게 하는 데 더 효과적이다.

■ 일관성

코드는 업무 범위 내에서 가능한 한 유일하게 정의해야 한다. 동일한 내용의 코드를 사용 형태나 업무 범위에 따라 중복 정의하여 사용할 경우 전사 차원의 코드 데이터의 중복은 물론 코드 데이터 의 불일치(Inconsistency)라는 보다 심각한 문제를 야기할 수 있다.

■ 정보 분석성

가능한 범위의 데이터는 모두 코드화하여 관리한다. 즉 사용자가 텍스트로 직접 입력하는 값을 최소화하고 정의된 범위 안에서 선택하도록 함으로써 정보 분석 시에 데이터는 있으나 분석 가치가없는 데이터가 양산되지 않도록 한다.

전사적으로 사용하고 있는 코드를 추출하여 법·제도적으로 부여된 코드와 동일한지를 확인하고, 동일한 값을 가지는 코드를 통합하여 단일화 작업을 수행한다. 코드는 표준화 팀에서 엄격한 기준에 따라 관리해야 하며, 사용자 임의대로 코드 체계를 생성하거나 수정해서는 안 된다. 코드는 도메인과 밀접하게 연관되어 관리해야 하나 도메인에 값의 범위가 명확히 정의되어 있는 경우(예를 들어'여부'는'Y/N'으로 표기)에는 특별히 코드화하여 관리하지 않아도 된다.

▣ 데이터 표준 요소

데이터 표준 요소란 시스템을 설계하고 구축하는데 필요한 데이터 관련 요소의 표준이다. 데이터 관련 요소 표준 대상은 논리 데이터 모델의 주제 영역, 엔터티, 속성, 관계명을 포함하여 물리적 객체 대상인 Subject Areas, Relationships, Database & Instance, Indexes, Constraints, Sequences, 사용자 정의 Procedures & Functions, Synonyms, Views, Rollback Segments, Tablespace, File Names, Script Names 등의 명명 규칙을 포함한다.

시스템 운영에는 시스템 운영에만 필요한 본질적 요소와 시스템 운영자가 필요에 의해 생성한 요소들이 존재할 수 있다. 예를 들어 프로그램 수행 결과를 단순 적재하는 요소들은 문제 발생시 역추적에 필요하지만 시스템 운영의 필수 요소라고 할 수 없다. 데이터 관련 요소 중 관리 대상의 선별 기준은 시스템 운영에 필수적인 요소가 1차 대상이 될 수 있어야 한다.

데이터 표준 요소는 시스템 운영에 필요한 요소를 정확히 선별하여 관리해야 한다. 설계 및 구축에 필요한 요소를 추출하여 표준이 필요한 요소를 정의하고 그 요소에 대해 업무적 표준을 정의한다.

데이터 표준 요소는 다음과 같은 기준에 따라 관리되어야 한다.

■ 통합성

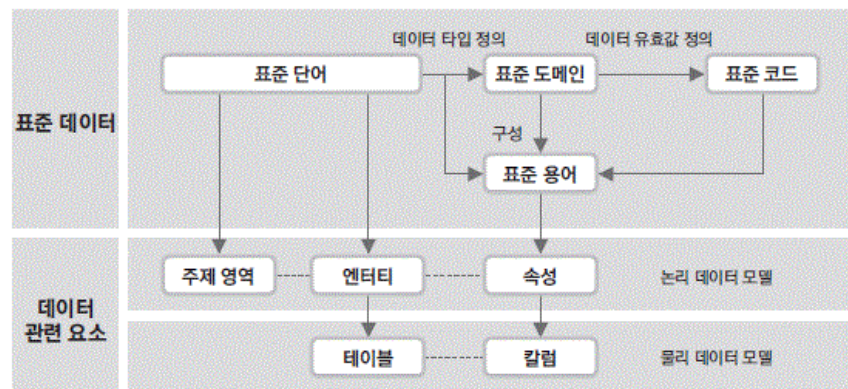
데이터 표준 요소의 각 요소는 전사적으로 통합하여 관리 및 적용해야 한다.

■ 일관성

정의된 표준 데이터가 데이터 모델 및 데이터베이스 스키마의 전 영역에 걸쳐 일관되게 적용되고 있는지 정기적으로 검토 확인한다.

+ 표준 데이터 상관도

표준 데이터 간의 상관 관계를 도식화하면 [그림 6-1-1]과 같다.



[그림 6-1-1] 표준 데이터 상관도

[데이터 품질 관리 프레임워크](#) [표준 데이터](#) [모델 데이터](#) [관리 데이터](#) [업무 데이터](#)

+ 정의 및 관리 목적

모델 데이터는 데이터 모델을 운용 관리하는데 필요한 데이터를 의미한다. 여기에는 데이터 참조 모델, 개념 데이터 모델, 논리 데이터 모델, 물리 데이터 모델에 대한 메타 데이터 및 DBMS 객체 정보가 포함된다. 데이터 모델에 대한 메타 데이터를 관리함으로써 데이터 구조에 대한 최신 정보를 유지하고 전사 차원에서 데이터 모델의 공유와 재사용성을 극대화하고 체계적인 데이터 모델의 변경관리를 가능하게 한다.

+ 세부 관리 대상

모델 데이터에서 다루는 세부 관리 대상은 데이터 참조 모델, 개념 데이터 모델, 논리 데이터 모델, 물리 데이터 모델에 대한 메타 데이터 및 DBMS 객체 정보 등이 있다. 이러한 모델 데이터는 다음과 같은 기준에 따라 관리되어야 한다.

■ 완전성

모델 데이터는 개념 데이터 모델, 논리 데이터 모델, 물리 데이터 모델, 데이터베이스와 같은 데이터 구조의 각 단계별 데이터 모델에 대한 모든 메타 데이터를 포함해야 한다.

■ 일관성

모델 데이터는 단어, 용어, 도메인 및 데이터 관련 요소 표준을 준수해 정의해야 한다.

■ 추적성

모델 데이터는 데이터 모델의 변경 이력에 대한 추적이 용이하고 과거 데이터 모델에 대한 활용 요구를 충족시켜야 한다.

■ 상호 연계성

모델 데이터는 데이터 구조를 입체적, 체계적으로 관리할 수 있도록 데이터 모델간의 상호 연관 관계를 표현해야 한다.

■ 최신성

모델 데이터는 데이터 구조의 각 단계별 데이터 모델과 업무 규칙은 물론, 실제 시스템에 구현된 물리 데이터와도 논리적으로 일치해야 한다.

■ 호환성

모델 데이터는 다른 종류의 관리 데이터와도 상호 호환이 가능해야 한다.

데이터 구조와 구조를 표현하는 모델 데이터는 별개로 관리한다. 데이터 모델에 변경 사항이 발생하면 변경 전과 변경 후의 데이터 모델과 이력은 물론, 데이터 모델 변경에 영향을 받은 응용 프로그램과 SQL의 변경 전과 변경 후의 내용도 함께 관리한다.



데이터 품질 관리 프레임워크	표준 데이터	모델 데이터	관리 데이터	업무 데이터
-----------------	--------	--------	---------------	--------

+ 정의 및 관리 목적

관리 데이터란 데이터베이스를 효과적으로 운영, 관리하기 위해 필요한 데이터를 의미한다. 여기에는 사용 관리 데이터, 장애 및 보안 관리 데이터, 성능 관리 데이터, 흐름 관리 데이터, 품질 관리 데이터 등이 포함된다. 데이터베이스는 크게 두 가지로 구분할 수 있다. 하나는 주로 기관이나 기업의 경영에 따라 OLTP에서 발생하는 운영계 시스템의 데이터베이스이고 다른 하나는 이러한 운영계 시스템으로부터 정보를 추출하여 기업의 의사 결정에 사용하는 분석계 시스템의 데이터베이스이다.

본 가이드의 세부 관리 대상 가운데 사용 관리 데이터, 장애 및 보안 관리 데이터, 성능 관리 데이터 등은 두 가지 모두 적용하되 주로 운영계 시스템에서 사용하는 데이터베이스에 초점을 두고 있다. 반면에 흐름 관리 데이터는 주로 분석계 시스템의 데이터베이스에 초점을 두고 있다.

+ 세부 관리 대상

▣ 사용 관리 데이터

사용 관리 데이터란 사용자가 데이터베이스를 효과적으로 사용할 수 있도록 지원하고 문제를 해결하는데 필요한 관리 데이터를 의미하며, 다음과 같은 기준에 따라 관리되어야 한다.

■ 데이터 활용도

주기적으로 데이터 사용 추세를 파악하여 저장 공간의 활용과 데이터로서의 활용 가치를 평가한다.

■ 사용자 만족도

사용자의 데이터베이스 관리에 대한 만족도는 제공되는 데이터에 대한 만족과 유지되는 데이터의 품질을 보증할 수 있다.

■ 문제 해결 소요기간

문제 발생에서 확인까지 소요되는 시간과 문제 확인 후 해결까지 소요되는 시간을 점검한다. 문제 해결 소요 기간은 데이터가 얼마나 체계적이고 구체적으로 관리되고 있는지를 가늠할 수 있는 잣대이다.

사용 관리 데이터의 관리 방법은 다음과 같이 요약할 수 있다.

- 일별, 주별, 월별로 데이터 변경 현황을 집계한다. 급격한 변화의 기준을 정하고 원인 및 추세 분석, 예상되는 문제점과 대책을 세운다.
- 월별로 데이터베이스 사용상의 문제점에 대한 개선 요구를 분석한다. 추세가 악화되는 원인을 파악한다.
- 문제 발견에 대한 경로를 다양하게 정의한다.
- 문제 원인을 유형별로 분류하고 처리 결과를 상세히 기록한다.(문제 정의, 관련 데이터베이스, 담당자와 관련자, 작업 진행 상황 등)

▣ 장애 및 보안 관리 데이터

장애 및 보안 관리 데이터란 데이터베이스의 정상적인 상태 유지나 효과적인 사용을 방해하는 사건을 사전에 예방하거나 사건 발생시에 신속한 복구가 이루어질 수 있도록 하는 데이터이며, 다음과 같이 기준에 따라 관리되어야 한다.

■ 주기적인 상태 기록

데이터베이스의 백업 주기, 백업 방법, 백업된 데이터의 안전한 보관과 백업된 데이터로부터의 정상적인 복구 여부의 관리하는 장애로부터 데이터의 안전성을 보장한다.

■ 복구 절차와 규칙

비상시 복구 절차와 적용되는 규칙의 완전성은 장애로부터 데이터의 안전성과 데이터 복구의 완전성을 보장한다.

■ 접근 통제

사용자 관리와 사용자 접근 권한의 관리는 내부 및 외부의 부적합한 사용자의 접근은 차단하고 권한 없는 자의 데이터베이스 접근을 차단하여 데이터의 안전성을 보장한다.

장애 및 보안 관리 데이터의 관리 방법은 다음과 같이 요약할 수 있다.

- 데이터베이스를 평가하여 중요도를 결정한다.
- 중요도에 따라 일별, 주별, 월별로 백업할 데이터를 분류한다.
- 백업 및 복구 절차를 확립하고 주기적으로 교육한다.
- 적용하는 규칙은 최대한 상세히 기술하되 중복이나 모순이 없는지를 확인한다.
- 백업 데이터의 보관 장소는 가급적 네트워크 및 서버가 다른 시스템과 분리되도록 하며 안전 장치를 설정한다.
- 데이터베이스에 대한 보안 규정을 수립하고 주기적으로 교육 및 홍보한다.
- 데이터베이스별로 사용자의 접근 권한을 명시하고 주기적으로 불법적인 접근을 검사하여 조치한다.

□ 성능 관리 데이터

성능 관리 데이터란 데이터베이스의 성능을 향상시키는데 필요한 관리 데이터를 의미하며, 다음과 같은 기준에 따라 관리되어야 한다.

■ 주기적 성능 점검

데이터베이스의 성능 측정 기준과 측정 주기가 정립되어 있어야 하며 그에 대한 사용자의 만족도 역시 관리되어야 한다.

■ 성능 향상 수단

데이터베이스의 성능 향상을 위한 절차와 규칙을 정의하여 전반적인 데이터베이스 성능을 관리한다.

성능 관리 데이터의 관리 방법은 다음과 같이 요약할 수 있다.

- 성능 측정 기준을 정립한다. 기준은 모두 정량화한다.
- 일별, 주별, 월별로 성능을 측정하고 그 추세를 분석한다.
- 성능 향상을 위한 절차와 규칙을 정비한다. 질의어 최적화, 데이터베이스 구조 변경 등에 대한 절차와 규칙을 포함한다. 데이터베이스 관리에 따른 재구성 작업의 시기와 방법을 정의한다.
- 스토리지의 교체 및 확장 시기에 대한 규칙을 정립한다.

□ 흐름 관리 데이터

흐름 관리 데이터란 하나의 정보시스템 데이터를 다른 정보시스템으로 이동할 때 사용하는 소스 데이터와 타깃 데이터 간의 매핑 정보를 관리하는 데이터를 의미하며, 다음과 같은 기준에 따라 관리되어야 한다.

■ 안전성

데이터 이동이 필요한 모든 소스와 타깃을 정의하고 소스, 타깃 간의 매핑 규칙을 정의해야 한다.

■ 유효성

정의된 소스와 타깃의 매핑 규칙을 준수하고 이에 위배되는 데이터에 대한 클린징(Cleansing) 규칙이 정의되어 있어야 한다.

■ 데이터 정합성

소스와 타깃의 데이터가 매핑 규칙을 준수하여 데이터의 정합성이 보장되어야 한다.

흐름 관리 데이터의 관리 방법은 다음과 같이 요약할 수 있다.

- 소스 데이터와 타깃 데이터 간의 매핑 리스트를 작성하고, 타깃 시스템에서 필요로 하는 소스 데이터가 모두 포함되어 있는지 확인한다.
- 데이터 이동이 필요 없는 소스와 타깃의 매핑 여부를 검사한다.
- 삭제된 소스를 매핑 소스로 사용하고 있는지를 검사한다.
- 소스와 타깃의 데이터 구조가 동일한지 조사한다. 동일하지 않은 경우 변환 규칙을 적용하고 있는지 조사한다.
- 변환 규칙이 데이터 무결성 규칙을 준수하는지 검사한다. 그 결과가 데이터 정합을 보장하는지 검사한다.

□ 품질 관리 데이터

품질 관리 데이터란 데이터의 정합성을 확보하고 데이터 품질의 유지, 개선을 위한 데이터를 의미하며, 다음과 같은 기준에 따라 관리되어야 한다.

■ 품질 기준

시스템에서 관리하는 데이터의 품질 기준을 정의한다. 품질 기준은 데이터의 중요도에 따라 등급을 두어 관리할 수 있다.

■ 품질 점검 주기

데이터 품질 관리를 지속적, 정기적으로 수행하기 위해 데이터베이스 성능과 데이터 품질 등에 대한 측정 주기를 설정한다. 품질 점검 주기는 사용자의 요구 수준을 반영하여 결정한다.

■ 품질 검증 절차와 규칙

정의된 품질 기준을 적용하기 위한 데이터 품질 검증 절차와 규칙을 정의한다. 여기에서는 정의된 절차와 규칙을 따를

수 없는 예외 사항에 대한 조치 방안도 함께 고려되어야 한다.

■ 품질 개선 절차

측정된 품질 평가 결과를 반영하여 데이터의 품질을 향상시키고 고품질 데이터를 유지할 수 있는 절차와 방법을 정의한다.

데이터 품질 관리가 필요한 항목을 도출해야 하며, 여기에는 기본적으로 다음과 같은 항목들이 포함된다.

- 엔터티 무결성(Entity Integrity)
- 참조 무결성(Referential Integrity)
- 도메인 무결성(Domain Integrity)
- 속성, 칼럼의 비즈니스 규칙 적용
- 엔터티, 테이블(Table) 정의에 따른 데이터 생성, 변경, 삭제 규칙
- 트리거(Trieger) 등 사용자 정의 DBMS 객체의 작동 여부
- 데이터 복제 허용시 원본 데이터와 복제 데이터 간의 정합성

그 밖에 품질 기준에 어긋나는 부적합한 데이터에 대한 오류 수정 규칙을 정의한다.



데이터 품질 관리 프레임워크 표준 데이터 모델 데이터 관리 데이터 업무 데이터

+ 정의 및 관리 목적

업무 데이터란 기관이나 기업의 업무 및 비즈니스를 수행하는 데 필요한 데이터를 의미하며, 일반적으로 데이터 흐름에 따라 원천, 운영, 분석 데이터로 구분할 수 있다.

+ 세부 관리 대상

▣ 원천(Source) 데이터

원천 데이터란 운영 업무 데이터의 원천이 되는 현실 세계의 데이터로, 일반 문서, PC에 저장된 데이터 원천 파일, 이메일 및 팩스 등을 말하며, 통합적 시스템에 의한 관리보다는 원천 업무 데이터 소유주인 개인이나 단체에 의하여 관리되는 데이터를 의미한다. 원천 데이터는 다음과 같은 기준에 따라 관리되어야 한다.

■ 보안성

원천 데이터는 시스템이나 프로그램, 데이터베이스 객체에 의해 시스템적으로 관리되지 않아 허용되지 않은 사용자에게 노출될 위험성이 많으므로 중요 원천 데이터의 경우 보안에 각별히 유의해야 한다.

■ 안전성

원천 데이터는 재해 발생 시 데이터 손실률이 높고 손실된 원천 데이터의 복구가 매우 어려우므로 중요 원천 데이터의 경우 안전 관리의 수준이 높아야 한다.

■ 신뢰성

원천 데이터의 정확성과 신뢰성을 판단할 수 있도록 이와 관련된 근거를 정의하여 관리해야 한다.

데이터베이스 구축에 필요한 원천 데이터를 분류해 각 원천 데이터에 대한 접근 권한과 생성, 변경, 소멸 규칙을 정의한다. 원천 데이터의 검색은 일반적으로 시스템 내에 저장된 데이터를 검색하는 것보다 많은 시간이 소요될 수 있으므로 관리 체계를 명확히 정의해야 한다.

▣ 운영(Operation) 데이터

운영 데이터란 기업 및 기관의 목표 달성을 위해 데이터베이스에서 저장, 관리하여 활용하는 데이터로 단순한 입출력 작업 처리상 일시적으로 필요한 임시 데이터는 제외한다. 운영 데이터는 다음과 같은 기준에 따라 관리되어야 한다.

■ 정확성

실세계에 존재하는 원천 데이터와 동일한 데이터가 오류 없이 관리되어야 한다.

■ 일관성

데이터가 용어 정의, 규정, 표준, 속성 정의, 데이터 형식 등과 일치하여야 한다.

■ 최신성

제공 데이터가 가장 최근 형태로 갱신되어야 하고 데이터의 최신성 유지를 위하여 데이터 최신성 등급(매우 중요, 중요, 보통)을 둘 수 있다.

■ 완전성

정보 시스템 내의 저장된 데이터는 완전한 형태를 가지고 있어야 하며, 조직의 목표 달성을 위해 요구되는 데이터의 폭과 깊이의 관점에서 이를 제공할 수 있을 만큼의 데이터를 보유하고 있어야 한다.

■ 사용 용이성

정보시스템에서 제공하는 인터페이스, 도움말, 고객 지원 기능 등이 사용자가 데이터베이스를 이용하는 데 불편함이 없도록 제공되어야 한다.

■ 검색 용이성

정보 시스템에서 원하는 데이터를 추출하여 활용할 수 있도록 검색 관련 제반 기능과 검색 조건에 따른 검색 결과 및 출력 방식이 정확하며 적절하여야 한다.

데이터의 정확성, 일관성, 최신성, 완전성을 보장하기 위해 정의된 관리 기준과 관리 방법에 따라 주기적으로 데이터를 점검 관리한다. 사용 용이성과 검색 용이성은 성능 관리 데이터의 관리 기준과 관리 방법을 따를 수 있다.

▣ 분석(Analysis) 데이터

분석 데이터란, 운영 데이터의 추출(Extract), 변환(Transformation), 적재>Loading) 등의 과정을 통해 생성되는 데이터이다. 분석 데이터가 기관이나 조직의 업무나 제반 활동을 신속하게 지원할 수 있도록 하기 위해서는 최신성과 정확성을 갖춰야 하며, 다음과 같은 기준에 따라 관리되어야 한다.

■ 분석 주기

분석용 데이터의 원천이 되는 운영 데이터의 분석 및 변환 주기를 결정한다.

■ 마감 기한

운영 데이터를 분석용 데이터로 변환하기 위해 이용하는 운영 데이터의 특정 시점을 정의한다.

■ 요약 레벨

분석 데이터에 요구되는 요약 수준을 정의한다. 요약 수준은 운영 데이터의 범위와 깊이의 관점에서 고려되어야 한다.

■ 주제 지향성

분산되어 관리되는 운영 데이터를 통일된 주제 영역별로 분류할 수 있어야 한다.

■ 통합성

분석 데이터를 동일하고 일관된 표준('남/여', '1/0', ...)에 따라 분류할 수 있어야 한다.

■ 시계열성

일정 시간 동안 축적된 데이터를 다양한 시점별로 정의할 수 있어야 한다.

■ 비휘발성

데이터의 삭제, 갱신이 자주 일어나지 않고 검색 위주의 데이터로 구성되어야 한다.

운영 데이터를 분석 데이터로 추출, 변환, 적재하는 규칙을 정의한다. 또한 일반적으로 분석되는 데이터의 양이 매우 많을 수 있으므로 사용되는 데이터베이스의 특성에 맞는 관리 방법이 같이 병행 되어야 한다.

개념 데이터 모델

데이터 참조 모델

논리 데이터 모델

물리 데이터 모델

데이터베이스

사용자 뷰

정의 및 관리 목적

개념 데이터 모델이란 업무 요건을 충족하는 데이터의 주제 영역과 핵심 데이터 집합을 정의하고 관계를 정의한 모델을 의미한다. 기관이나 기업의 업무 특성에 적합한 주제 영역과 핵심 데이터 집합과의 관계를 정의하여 향후에 정의하게 될 상세 논리 데이터 모델과 물리 데이터 모델과의 데이터 구조적 정렬(Alignment)을 지원한다. 또한 주제 영역을 통해 전체 업무 범위와 업무 구성 요소를 확인 할 수 있다.

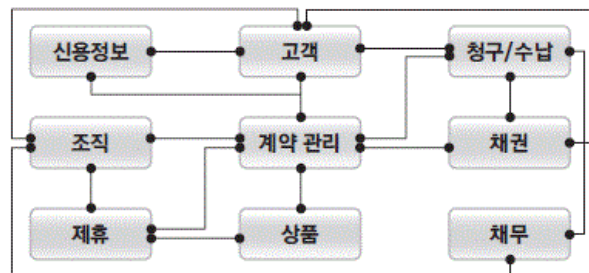
데이터 구조를 위한 데이터 모델은 건축물을 짓는 공법과 유사하다. 큰 건축물에서 방 하나의 인테 리어부터 시작하지 않고 건물의 용도, 건물의 전체적 구조, 쓰이는 자재의 분류부터 시작하듯이 데이터 구조에 대한 정의 역시 같은 방법으로 접근해야 한다.

개념 데이터 모델은 건축물의 조감도와 같이 구축하고자 하는 업무 모델의 핵심 데이터 구조를 그림으로써 전체 업무에 대한 큰 윤곽을 잡고 세부적인 단계로 나아갈 수 있게 한다. 개념 데이터 모델의 정의로부터 접근하는 것은 골조가 튼튼한 건물을 짓는 것과 같이 데이터 구조에 대한 기본 구조를 확실히 정의하고 업무 요건 변경에 취약하지 않는 데이터 구조를 세우는 것과 같다. 경우에 따라서는 개념 데이터 모델의 상위 모델인 개괄 데이터 모델을 둘 수 있다.

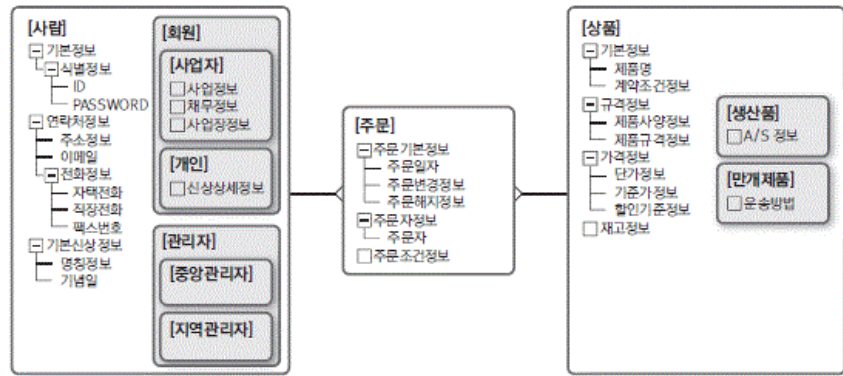
개괄 데이터 모델은 상위 수준의 전사 데이터 영역을 분류하여 표현한 것으로, 흔히 사용되는 주제 영역 관계도와 대응되는 부분이라고 할 수 있다. 반면에 개념 데이터 모델은 전사 수준에서 사용하는 데이터를 전체적으로 표현할 수 있는 틀로서, 단위 주제 영역별로 핵심 엔터티와 핵심 엔터티들 사이의 관계를 정의한 모델이라고 할 수 있다.

개괄 데이터 모델은 데이터 영역과 데이터 집합을 업무 영역에 국한하지 않고 전사적 관점에서 정의하는 것으로, 각 데이터 영역은 다른 데이터 영역과 관계를 가질 수 있다. 또한 기업의 이익 관점이 아닌 공익적인 관점에서 공통으로 사용되는 속성을 좀 더 원시화된 형태의 수준으로 정의할 수 있다. 데이터 구조에서의 세부 관리 대상은 개별적인 항목으로 관리하는 것이 아닌 개체-관계 다이어그램(ERD, Entity Relationship Diagram)으로 표현해 관리한다.

[그림 6-2-1], [그림 6-2-2]는 개괄 데이터 모델과 바커표기법을 기준으로 작성한 개념 데이터 모델의 예다. [그림 6-2-3]은 [그림 6-2-2]의 개념 데이터 모델을 IE표기법 기준으로 작성한 예다. 작성에 사용된 도구의 특성상 동일하게 표현할 수는 없으나 바커표기법을 기준으로 작성한 예에 대해 최대한 유사한 개념을 표현하도록 구성하였다.



[그림 6-2-1] 개괄 데이터 모델



[그림 6-2-2] 개념 데이터 모델

+ 세부 관리 대상

주제 영역(Subject Area)

주제 영역은 업무상 친밀도가 높은 데이터 집합을 하나의 주제 영역으로 정의한다. 일반적으로 업무를 명확히 구분하는 범위를 하나의 주제 영역으로 정의하기도 하나 주의할 점은 서로 다른 주제 영역간 공유하는 엔터티의 수가 가급적 적게 해야 한다. 다음과 같은 기준에 따라 관리되어야 한다.

원자성

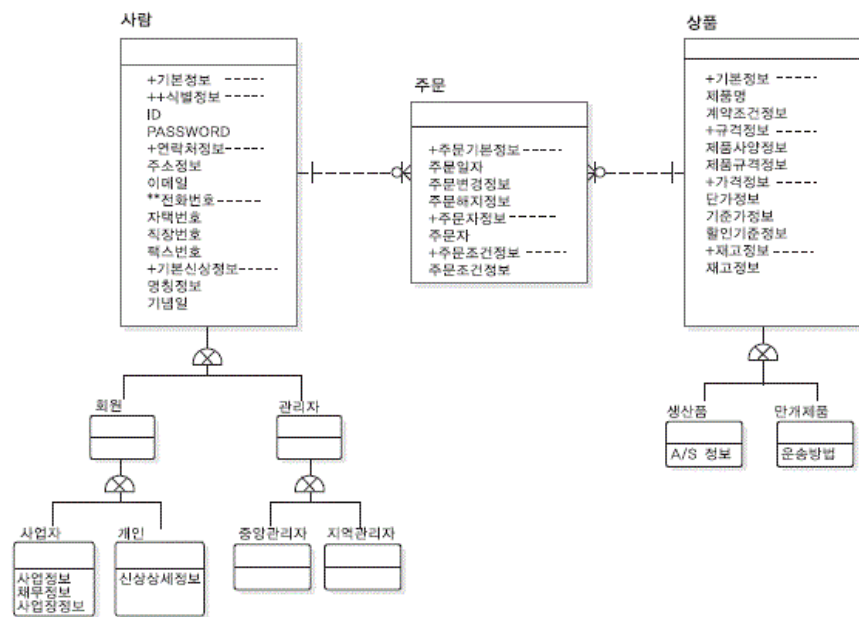
하나의 단위 주제 영역은 가급적 다른 주제 영역의 엔터티나 관계의 영향을 받지 않는 엔터티의 모임이어야 한다.

집중성

단위 주제 영역 내의 엔터티와 관계는 단위 주제 영역 내에 집중되어야 한다.

업무 지향성

주제 영역을 명명하는데 있어 업무적 명확성을 나타내는 단수 단위로 명명할 수 있어야 한다.



[그림 6-2-3] IE표기법에 따른 개념 데이터 모델 예

업무상 동일한 영역에서 다루는 것이 보다 효과적인 엔터티 집합들을 하나의 주제 영역으로 선언하며, 주제 영역은 업무의 다양성에 따라 여러 개로 나뉠 수 있다.

핵심 엔터티(Kernal Entity)

핵심 엔터티는 업무 영역 내에서 관리하고자 하는 데이터 집합으로 두 개 이상의 속성과 두 개 이상의 데이터 인스턴스를 가져야 하며 각각의 인스턴스는 개별적, 동질적, 독립적인 데이터 집합이며 영속적으로 존재하는 데이터 단위이다. 키(Key) 엔터티, 메인(Main) 엔터티인 핵심 엔터티는 업무의 근간이 되고 수많은 자식(업무, 트랜잭션) 엔터티를 만들 수 있는 상위 개념의 엔터티이다. 다음과 같은 기준에 따라 관리되어야 한다.

집합성

엔터티는 두 개 이상의 속성과 두 개 이상의 데이터 인스턴스를 갖는 데이터의 집합이어야 한다.

■ 식별성

엔터티는 하나 이상의 속성으로 엔터티의 각 데이터 인스턴스를 유일하게 구분할 수 있어야 한다.

■ 영속성

엔터티는 업무의 활동 주기에 따라 영속적으로 존재해야 하는 데이터 집합이다. 즉, 업무의 내용이 달라질 때 사라지거나 생성되어야 하는 데이터 집합은 엔터티로 선언하고 관리하기에 부적절하다.

■ 사용성

엔터티는 업무 범위 내에서 반드시 사용되어야 하는 데이터 집합이다. 선언은 되었으나 사용되지 않는다면 엔터티로서의 존재 가치가 없다.

■ 관계성

엔터티는 반드시 다른 엔터티와의 관계가 존재해야 한다. 관계가 없는 엔터티는 사용되지 않는 엔터티일 수 있으므로 사용성에 위배된다.

엔터티는 업무의 문서, 장표, 인터뷰, 관련 전문서적, 데이터 흐름도(Data Flow Diagram), 타시스템, 보고서, 현장 조사로부터 수집될 수 있다. 엔터티는 논리적인 단위로 정확히 분할하여 선언하되 하나의 엔터티가 의미상으로 다르게 보인다고 해서 중복되게 선언해서는 안된다.

이는 매우 중요 한 사항이다. 즉, 하나의 엔터티가 상태에 따라 다르게 보인다면 데이터의 동질성을 파악하여 서브타입으로 하나의 엔터티로 선언할 수 있는지 확인해야 한다.

▣ 핵심 관계

핵심 관계는 핵심 엔터티 간의 논리적인 관계를 나타낸 것으로 엔터티의 존재 형태나 상호 영향을 주는 비즈니스 규칙과 현재나 가까운 장래에 유용한 관계를 한정적으로 표현하며, 관계 명칭과 선택 사양과 관계 형태(Degree)를 갖는다. 관계형 데이터 모델에서 엔터티 간에는 반드시 관계가 존재해야 한다. 핵심 관계는 개념 데이터 모델 단계에서 M:M 관계를 그대로 유지할 수도 있고, M:M 관계가 해소된 엔터티를 포함할 수도 있다. 핵심 관계는 부모와 자식 간의 관계명을 반드시 정의하여야 하며, 다음과 같은 기준에 따라 관리되어야 한다.

■ 선택성

관계는 '필수'와 '선택'을 구별하여 표현할 수 있어야 한다.

■ 형태성

관계에는 1:1, 1:M, M:M의 형태가 정의되고 관리되어야 한다.

■ 업무 지향성

관계는 두 엔터티 간의 존재가 상호 어떤 영향을 미치는가를 명확히 표현할 수 있어야 한다. 즉, 자식 엔터티의 인스턴스 존재는 반드시 부모 인스턴스의 존재를 필요로 하지만 부모 인스턴스의 존재는 자식 인스턴스의 존재에 영향을 받지 않는다면 관계는 이와 같은 부모와 자식 엔터티 간의 존재 영향에 대하여 명확히 표현할 수 있어야 한다.

관계명은 구체적이어야 하며 엔터티 간의 주는 쪽(부모)과 받는 쪽(자식)의 관계가 명확해야 한다. 또한 반드시 관계를 갖는 데이터가 있어야 하는 경우와 관계를 갖는 대상 데이터가 없어도 되는 경우에 대한 선택성이 있어야 한다. 하나의 데이터와 하나 이상의 데이터에 대한 관계의 형태도 명확히 표현할 수 있어야 한다.



데이터구조 이해

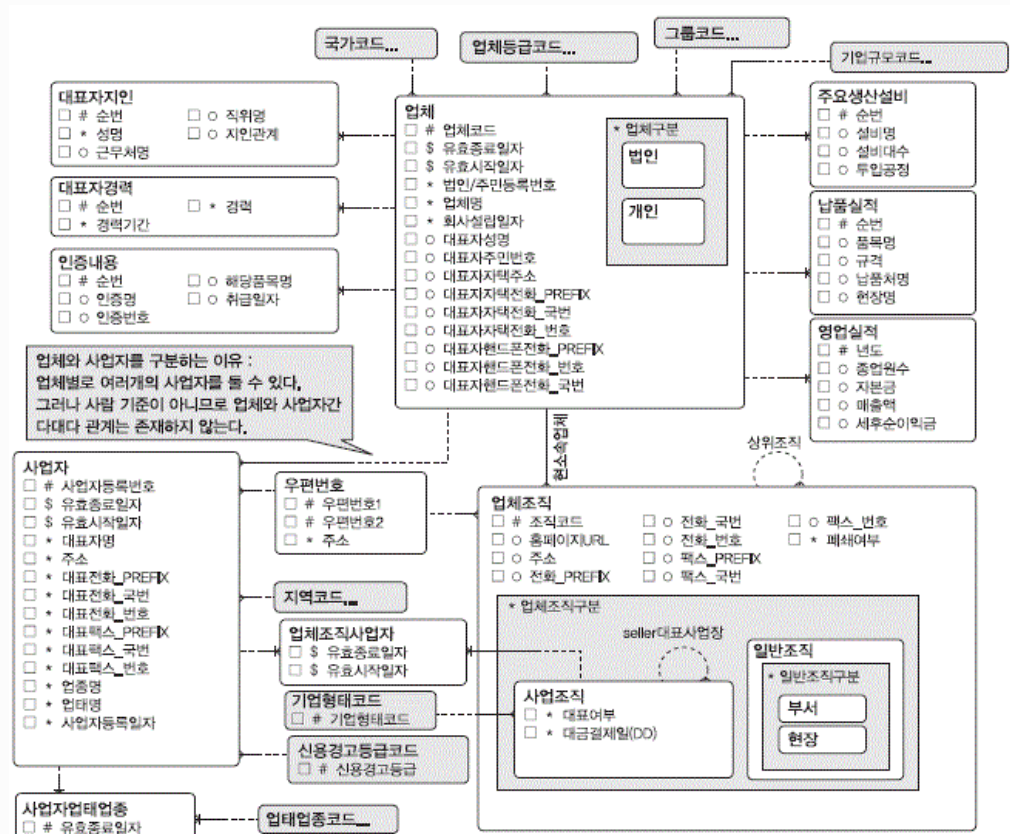
Home > DB 구축 가이드 > DA 가이드 > 데이터 품질관리 이해 > 데이터구조이해

개념 데이터 모델 **데이터 참조 모델** 논리 데이터 모델 물리 데이터 모델 데이터베이스 사용자 뷰

+ 정의 및 관리 목적

데이터 참조 모델(DRM, Data Reference Model)이란 업무 영역별, 주제 영역별 표준 데이터 집합, 관리 항목들이 표기되어 재사용이 가능한 데이터 모델을 말한다. 전사적 업무 혹은 범용적 데이터 모델 정의시 기존에 검증된 데이터 모델을 참조함으로써 데이터 모델의 정확성과 재사용성을 높이고 새로운 데이터 모델 정의에 따른 시간과 비용을 절감할 수 있다. 또한 새로운 데이터 모델링시 참조 모델을 활용함으로써 정보의 누락을 예방할 수 있으며, 기존에 검증된 데이터 참조 모델을 이용하여 자사 데이터 모델의 오류를 확인하거나 보완할 수 있다.

중앙 기관이나 부서에서 양질의 데이터 참조 모델을 채택, 활용함으로써 단기간에 하부 기관이나 부서 데이터 모델의 품질을 개선할 수 있다. 하지만 기업을 대상으로 한 데이터 참조 모델과 공공 기관을 대상으로 한 데이터 참조 모델은 관리 및 활용 측면에서 서로 상이할 수 있다. 예컨대 기업 데이터 참조 모델은 개별 기업의 정보 유출 및 보안 문제로 활성화되기 어려울 수 있다. 그러나 공공 기관의 경우 상위 기관이 하위 기관에게 데이터 참조 모델의 사용을 제도화할 수 있으며, 이에 따라 데이터 참조 모델을 활용한 공공 기관 데이터 모델의 품질 표준화 및 데이터 품질 개선을 신속하게 추진할 수 있다. 데이터 참조 모델에 대한 예로 [그림 6-2-4]를 들 수 있다. [그림 6-2-5]는 [그림 6-2-4]의 데이터 참조 모델을 IE표기법을 기준으로 작성한 예이다.



[그림 6-2-4] 데이터 참조 모델 예

■ 분류성

데이터 참조 모델은 업무 영역과 업종은 물론 데이터 구조 각 단계와 데이터 참조 모델의 범위 내 에서도 분류될 수 있어야 한다.

데이터 참조 모델은 중앙 기관이나 상부 조직에서 정의하고 하부 기관이나 조직에서 공유, 활용 가 능하도록 관리해야 한다.



개념 데이터 모델

데이터 참조 모델

논리 데이터 모델

물리 데이터 모델

데이터베이스

사용자 뷰

+ 정의 및 관리 목적

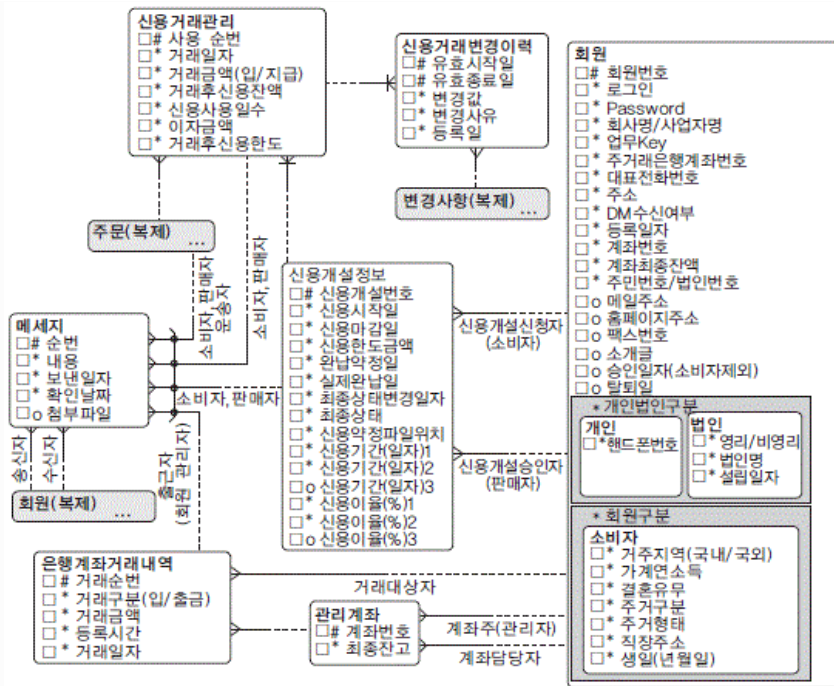
논리 데이터 모델이란 개념 데이터 모델을 상세화 하여 논리적인 데이터 집합, 관리 항목, 관계를 정의한 모델을 말한다.

논리 데이터 모델은 전체 데이터 구조에서 가장 핵심을 이루는 모델로서 전체 업무 범위와 업무 구성요소를 확인할 수 있다.

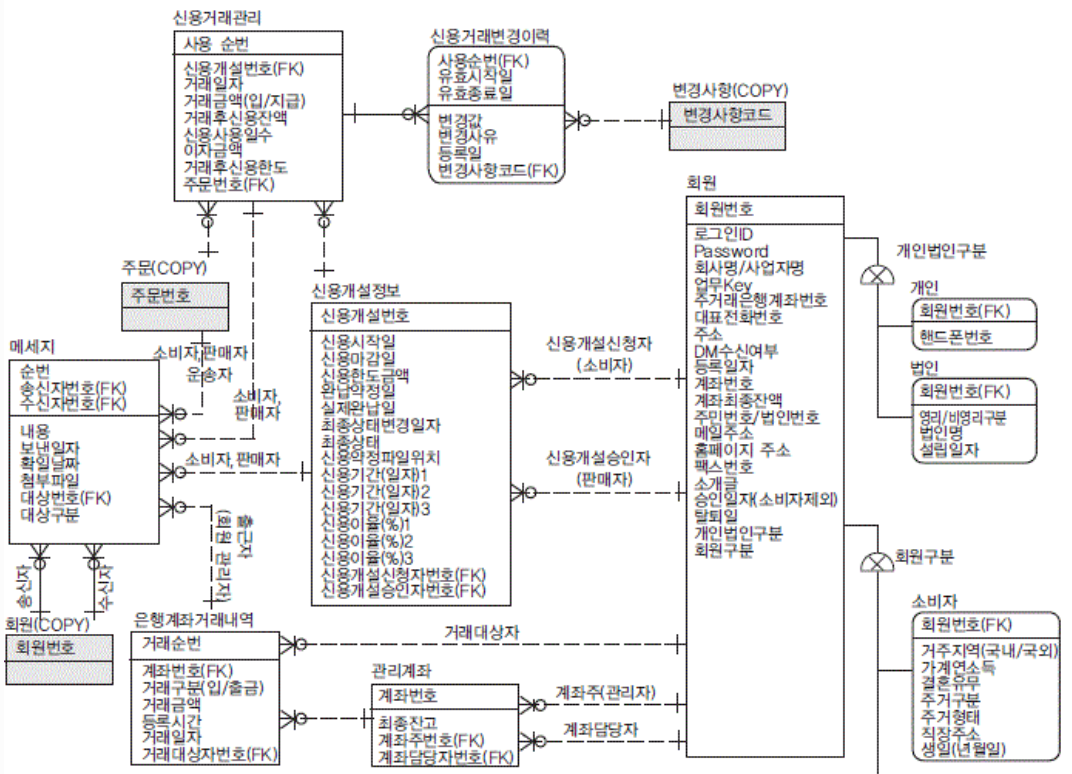
따라서 모든 업무의 데이터 구조를 구체적으로 정의하고 최신의 내용으로 관리될 수 있도록 해야 한다. 논리 데이터 모델은 데이터 구조 정의시 상세하게 정의될 수 있는 모든 정보를 포함해야 하며, 논리 데이터 모델이 구체적이고 상세할 수록 업무에서 관리하는 모든 데이터 구조는 상세하게 관리될 수 있다.

데이터가 논리 데이터 모델 단계에서 상세하게 관리되면 불필요한 데이터 중복을 방지할 수 있다. 데이터의 중복은 결과적으로 중복된 데이터로 인해 발생하는 데이터의 불일치(Inconsistency)를 방지하므로 궁극적으로 목적하는 데이터의 품질을 보장할 수 있다. 즉, 구조적이고 상세화된 논리 데이터 모델의 관리와 현재의 논리 데이터 모델이 현재의 업무를 반영하는 주의 깊은 논리 데이터 모델의 관리는 데이터의 품질 보증을 위한 핵심 사항이다.

논리 데이터 모델의 관리는 무엇보다 해당 논리 데이터 모델이 현재의 업무 상태를 그대로 반영해 야 한다. 물리 데이터 모델이나 데이터베이스의 객체는 변경하였으나 논리 데이터 모델에는 변경 사항을 반영하지 않는다면 논리 데이터 모델의 정확성이 떨어지고 결과적으로 관리하는 데이터의 품질 에도 영향을 주게 된다. 가능하다면 논리 데이터 모델의 변경 사항 이력도 관리되는 것을 권장한다. 이력이 관리되면 사안에 따라 과거 일정 시점의 논리 데이터 모델로 복귀할 수 있다. 또한 논리 데이터 모델에서 엔터티나 속성, 관계명을 표현하는 용어는 표준 데이터에서 명시한 표준 단어와 표준 용어 내에서 정의할 것을 권장한다. 논리 데이터 모델에 대한 예로 [그림 6-2-6]을 들 수 있다. [그림 6-2-6]에 나타난 모델은 식별자 상속모드를 사용하지 않고 작성하여 부모 엔터티의 식별자가 자식 엔터티에 관계 속성으로 나타나지 않았음을 감안하여야 한다. [그림 6-2-7]은 [그림 6-2-6]의 논리 데이터 모델을 IE표기법을 기준으로 작성한 예이다.



[그림 6-2-6] 논리 데이터 모델 예



[그림 6-2-7] IE표기법을 기준으로 작성한 논리 데이터 모델 예

세부 관리 대상

주제 영역 (Subject Area)

주제 영역은 업무상 친밀도가 높은 데이터 집합을 하나의 주제 영역에서 선언하여 관리한다. 논리 데이터 모델의 주제 영역 관리 기준은 개념 데이터 모델의 주제 영역 관리 기준을 따른다. 주제 영역을 통해 전체 업무 범위 결정과 업무 구성 요소를 확인할 수 있다.

엔터티 (Entity)

엔터티는 개념 데이터 모델의 정의를 포함하고 이력 관리와 동질성, 독립성 정보가 보다 더 상세히 파악된 서브타입 정보가 추가될 수 있으며, 다음과 같은 기준에 따라 관리되어야 한다.

■ 완전성

엔터티는 개별적인 데이터 집합으로 두 개 이상의 속성과 두 개 이상의 인스턴스를 유지해야 한다.

■ 영속성

현재 관리하고 있는 데이터 집합이거나 앞으로도 관리할 데이터 집합이다.

■ 식별성

엔터티의 인스턴스를 개별적으로 구별할 수 있는 하나 이상의 속성이 존재해야 한다.

■ 동질성

하나의 데이터 집합인 엔터티에는 동질의 데이터가 모인 데이터 집합이어야 한다.

■ 정규화

엔터티는 일반적으로 3차 정규화까지 정규화 되는 것을 권장한다.

개념 데이터 모델의 관리 방법을 따르면서 정보의 상세화에 따라 엔터티 정의, 데이터 발생 규칙 등의 세부 정보를 추가 관리한다.

▣ 관계(Relationship)

관계는 개념 데이터 모델의 정의를 포함하고 상세 논리 데이터 모델 단계에서 모든 M:M 관계는 해소되어야 한다. 다음과 같은 기준에 따라 관리되어야 한다.

■ 선택성

관계는 필수와 선택으로 나누어질 수 있다. 필수 관계는 해당 관계를 갖는 인스턴스가 반드시 엔터티에 존재해야 함을 의미한다.

■ 관계 형태

관계는 1:1 혹은 1:M, M:M의 관계를 가질 수 있다.

■ 관계 명칭

관계는 관계명이 명확한 경우 표현에 있어 생략 가능할 수 있으나 일반적으로 반드시 엔터티와 엔터티 간의 관계 설정시 관계명을 갖는다.

개념 데이터 모델의 관리 방법을 따르면서 엔터티의 통합, 관계의 통합에 따라 관계는 순환 관계, 배타 관계가 추가되어 보다 복잡해 질 수 있다. 또한, 관계에서 내포하는 비즈니스들을 상세하게 정의하여 관리한다.

▣ 속성(Attribute)

속성은 엔터티 내에서 관리하고자 하는 정보의 항목들을 말하며, 다음과 같은 기준에 따라 관리되어야 한다.

■ 원자성

의미 있는 최소 단위까지 분할되어야 하며, 하나의 속성은 동시에 여러 상태의 정보를 포함할 수 없다.

■ 일관성

하나의 속성은 하나의 데이터 유형을 가리키며 하나의 데이터만 관리한다.

■ 무결성

참조되는 속성의 데이터는 해당 속성을 참조하는 속성의 데이터와 일치해야 한다.

■ 정보성

업무와 관련해 의미 있는 범위 내에서 상세화의 수준이 결정되어야 한다. 속성은 엔터티의 관리 항목 범위 내에서 초기에 결정된 후 사용자의 요구에 따라 무분별하게 증가될 우려가 있다. 속성은 기존의 정보에서 추출이 가능하지 않을 때 새로 추가될 수 있고 속성의 상세화에 따라 엔터티가 추가될 수 있다.

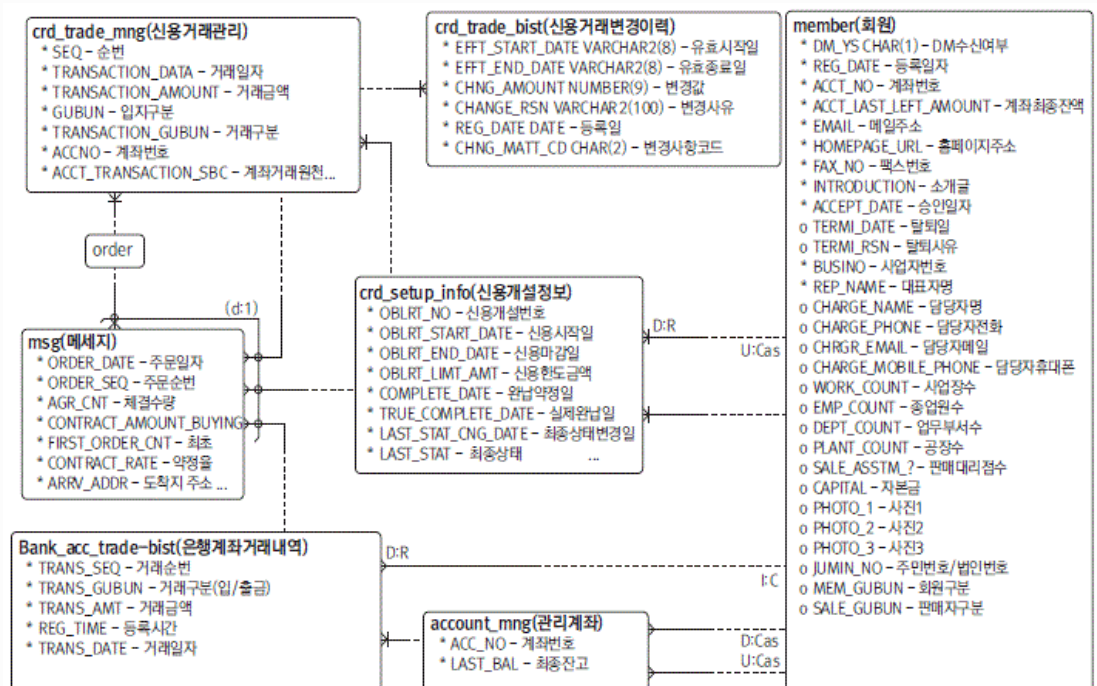
개념 데이터 모델 데이터 참조 모델 논리 데이터 모델 **물리 데이터 모델** 데이터베이스 사용자 뷰

+ 정의 및 관리 목적

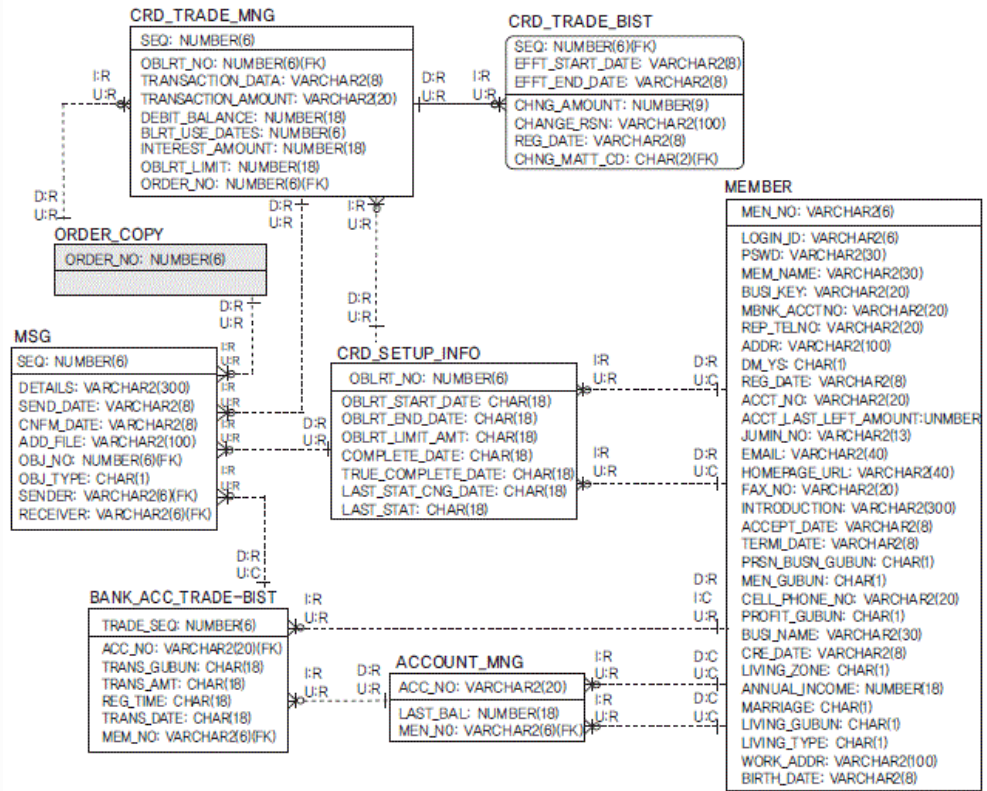
물리 데이터 모델이란 논리 데이터 모델을 DBMS의 특성 및 성능을 고려하여 구체화시킨 모델을 말한다. 물리 데이터 모델은 DBMS 선정 이후에 해당 DBMS 상에서 최상의 성능을 보장하도록 논리 데이터 모델에서 저장하는 데이터의 물리적 특성을 최대한 반영하여 설계하고 이를 관리한다. 논리 데이터 모델이 1:1로 데이터베이스의 객체로 대응되어 생성되지 않으므로 DBMS의 성능을 최대한 살릴 수 있고 저장되는 데이터의 특성을 충분히 반영할 수 있다.

물리 데이터 모델로의 설계 단계에서 샘플 데이터를 이용하여 논리 데이터 모델의 적합성을 재 검증할 수 있다. 물리 데이터 모델의 설계를 위해서는 업무 요건과 필요에 따라 사용자 화면이 완성되어야 하므로 사용자 애플리케이션과 상호 검증 하에 설계될 수 있다. 물리 데이터 모델의 설계 시점은 애플리케이션의 설계나 업무 요건이 명확해지는 단계이므로 업무 요건을 반영한 물리 데이터 모델을 설계할 것을 권장한다.

물리 데이터 모델의 테이블명, 관계명, 칼럼명 등은 표준 데이터에서 명시한 표준 단어와 표준 용어 규칙에 따른 물리명을 선언하고 이를 기준으로 하여 생성할 것을 권장한다. 물리 데이터 모델에서는 무엇보다 도메인의 선언이 중요하며 도메인 규칙에 대한 충실한 준수는 물리 데이터 모델 내에서 유지하는 데이터를 고품질로 유지할 수 있는 필수 조건이 될 것이다. 물리 데이터 모델에 대한 예로 [그림 6-2-8]을 들 수 있다. [그림 6-2-9]는 IE표기법을 기준으로 작성한 예이다.



[그림 6-2-8] 물리 데이터 모델 예



[그림 6-2-9] IE표기법을 기준으로 작성한 물리 데이터 모델 예

+ 세부 관리 대상

▣ 주제 영역(Subject Area)

주제 영역은 분산 DBMS의 고려나 업무 영역에 따라 다른 스키마의 설계로 대응을 고려할 수 있다. 논리 데이터 모델에서 정의한 주제 영역은 물리 데이터 모델에서 스키마나 서버로 분산될 수도 있으나 경우에 따라서는 하나의 서버에 하나의 스키마 내에서 테이블의 명명 관례(Naming Convention)에 의하여 물리적 주제 영역을 구분하여 관리할 수도 있다. 물리 데이터 모델의 주제 영역 관리 기준은 개념 데이터 모델과 논리 데이터 모델의 관리 기준을 따른다. 논리적인 주제 영역은 분산 DBMS의 인스턴스나 스키마와의 대응 관계에 따라서 관리되어야 한다.

▣ 테이블

테이블은 데이터의 물리적 특성과 DBMS의 특성에 따라 하나의 물리적 저장 장소인 테이블 혹은 서브타입이나 업무적 특성에 따라 하나 이상의 물리적 테이블로 분할될 수 있으며, 다음과 같은 기준에 따라 관리되어야 한다.

■ 연속성

테이블의 데이터는 현재 관리하고 있는 데이터이며 앞으로도 관리해야 할 필요가 있다.

■ 식별성

테이블 내의 레코드들은 하나 이상의 칼럼 데이터에 의해 구별 가능해야 한다.

테이블 내에 저장되는 데이터의 생명주기와 일정 기간 유지되어야 하는 데이터의 양이 반영된 설계 정보를 관리한다.

▣ 관계(Relationship)

관계는 부모 테이블과 자식 테이블 간의 데이터 생성, 삭제, 변경 규칙을 정의할 수 있으며, 다음과 같은 기준에 따라 관리되어야 한다.

■ 생성 규칙

자식 테이블의 데이터 생성시 부모 테이블에 참조되는 데이터가 반드시 존재해야 한다.

■ 변경 규칙

부모 테이블의 키 데이터가 변경되면 참조하는 자식 테이블의 참조 데이터는 같이 변경되거나 혹은 자식 데이터가 존재하면 부모 테이블의 키 데이터는 변경되지 못한다.

■ 삭제 규칙

부모 테이블의 데이터가 삭제되면 해당 데이터를 참조하는 자식 테이블의 데이터가 함께 삭제되거나 혹은 자식 데이터가 존재하면 부모 테이블의 데이터는 삭제될 수 없다.

관계는 업무 규칙이므로 DBMS 수준에서 관리할 것인지 혹은 애플리케이션 수준에서 관리할 것인지 먼저 결정되어야 하

며, 트리거에 의한 자동 변경은 DBMS 오류시 추적이 어려우므로 가능한 최소화 한다.

▣ 칼럼

데이터아키텍처 전문가 가이드 칼럼은 표준화된 도메인 내에서 업무 규칙이 반영된 데이터를 저장하도록 정의한다. 하나의 칼럼 데이터는 같은 데이터 유형(Type)을 갖는다. 비슷한 데이터 유형과 표현을 갖는 칼럼의 물리적 속성 을 도메인으로 정의하여 관리할 수 있다. 칼럼은 물리적 성질이 다양하나 표준화된 도메인 내에서 물리적 특성이 선택되도록 한다.

즉, 저장하는 데이터의 개별 단위가 가능하면 모두 표준화되어 도메인으로 미리 선언되고 칼럼의 물리적 특성은 미리 선언된 도메인 내에서 선택될 것을 권장한다. 이러한 칼럼 관리는 부하가 따를 수 있으나 데이터의 품질을 지속적으로 관리할 수 있는 최선의 방안이 될 것이다.



개념 데이터 모델 데이터 참조 모델 논리 데이터 모델 물리 데이터 모델 **데이터베이스** 사용자 뷰

+ 정의 및 관리 목적

데이터베이스는 물리 데이터 모델을 구현한 결과물이며 구축된 실제 데이터가 저장되는 데이터 저장소를 의미한다. 물리 데이터 모델이 구현된 데이터베이스 저장소인 테이블과 속도를 위한 인덱스, 비즈니스 규칙이 반영된 제약 사항 및 데이터베이스를 효과적으로 운영하기 위한 객체를 정의하고 관리한다.

데이터베이스 내에서 생성된 모든 객체는 DBMS 자체의 데이터 사전에 의해 관리된다. 데이터 사전에서는 항상 현재의 상태와 DBMS 객체 자체의 정보만 관리되므로 논리 데이터 모델과의 정렬이 나 각 DBMS 객체에 대한 논리적 정보가 데이터 품질 관리 관점에서 함께 관리된다면 데이터 아키텍처 관점에서 더 유효할 것이다. 참고로 데이터 사전(Data Dictionary) 정보와 DBMS 객체 관리 를 위한 정보는 구분되어야 한다. DBMS의 데이터 사전은 DBMS의 객체를 운영하기 위한 정보이므로 데이터 구조상의 논리적 정보를 같이 관리하지는 않는다. 이는 논리 데이터 모델과 DBMS 객체 정보가 분리되어서 운영되고 상호의 형상 유지가 따로 관리되는 구조적 차이이기도 하나 이로 인한 데이터 아키텍처 관점에서 구조 정렬이 유지되지 않는 문제로 확대된다. 따라서 논리 데이터 모델과 DBMS 객체가 데이터 아키텍처 관점에서 논리적 연계성을 유지하는 방안이 데이터 품질 관리 방안 에서 무엇보다 중요한 사항이라 할 수 있다.

+ 세부 관리 대상

▣ 저장 공간

저장 공간은 데이터베이스에서 데이터를 저장할 공간을 필요로 하는 테이블과 인덱스를 정의하는 영역(Tablespace, Data File)이다. 데이터베이스에서 운영하는 모든 정보는 대표적으로 테이블과 인덱스로 구분되어 저장되며, 데이터의 관리 는 기본적으로 저장 공간의 관리로부터 시작될 수 있다. 다음과 같은 기준에 따라 관리되어야 한다.

■ 안전성

저장 공간을 위한 시스템의 디스크 영역은 시스템의 다른 프로그램 수행 영역으로부터 분리되어 다른 프로그램으로부터 안전하게 보호되어야 한다. 이를 위하여 보통의 업무 시스템은 DBMS 전용 서버를 운영하기도 한다. 또한 데이터가 물리적으로 저장되는 실제 공간이므로 외부의 위협이나 재해로부터 저장공간의 존재가 보호되어야 한다.

■ 보안성

데이터가 물리적으로 저장되는 실제 공간이므로 허가 받지 않은 프로그램이나 사용자에게 대하여 안전하게 접근 제어가 이루어져야 한다.

■ 확장성

데이터는 지속적으로 증가할 수 있으므로 저장 공간의 확장과 물리적 디스크 영역의 할당이 충분하고 편리하게 수행되어야 한다.

■ 성능 보장

대용량의 데이터가 DBMS 운영 중 수시로 호출되고 저장되므로 저장 공간을 할당한 물리적 디스크는 빠른 성능을 유지할 수 있는 제품과 구조적 배치가 이루어져야 한다.

데이터베이스 관리자(DBA, Database Administrator)는 관리 기준에 따라 성능과 보안을 고려 하여 시스템의 저장 공간을 수시로 확인하여 데이터의 최종 보안과 안전성을 제일의 목적으로 관리 해야 한다. 또한 개발자나 사용자의 편의에 따라 무분별하게 데이터가 확장되는 것을 제어하고 적절한 수준의 저장 공간에 대한 백업도 병행되어야 한다.

▣ 테이블

테이블은 엔터티와 속성으로 정의한 것이다. 데이터의 특성에 따라 클러스터, 파티션 등의 다양한 방법이 적용될 수 있다. 데이터의 증가 추이도 관찰되고, 이에 따른 물리적 성질 변화도 필요할 수 있다. 다음과 같은 기준에 따라 관리되어야 한다.

■ 주기성

테이블 내의 데이터는 일정한 주기에 따라 백업되거나 성능을 위하여 재생성 될 수 있다.

■ 다양성

테이블 내의 데이터는 성능을 위하여 적절한 분산 전략과 테이블 저장 공간 정의 방식에 따라 파티션(Partition), 클러스터(Cluster), 인덱스 구성 테이블(IOT, Index Organized Table)등 여러 형태로 정의될 수 있다.

■ 보안성

테이블은 권한과 사용에 따라 제한된 범위의 사용자에게 테이블 단위, 칼럼 단위로 접근, 생성, 변경, 삭제 규칙이 정의되어야 한다.

■ 논리성

테이블의 추가와 칼럼의 추가는 반드시 논리 데이터 모델을 참조하여 반영되어야 한다. 논리 데이터 모델을 근거로 하지 않은 DBMS상의 테이블과 칼럼의 추가는 무분별한 중복 데이터를 양산하게 되어 결과적으로 데이터의 품질을 보장할 수 없다.

데이터베이스 관리자는 관리 기준에 따라 성능과 보안을 고려하여 테이블의 데이터를 관리해야 하고 개발자나 사용자의 편의에 따라 무분별하게 테이블이 생성되는 것을 적절하게 제한하여야 한다. 테이블 생성에 대한 예로 [그림 6-2-10]을 들 수 있다.

```
SQL> CREATE TABLE s_emp (
    Id                VARCHAR2(7),
    last_name         VARCHAR2(25)
    CONSTRAINT s_emp_last_name_nn NOT NULL,
    first_name        VARCHAR2(25),
    userid            VARCHAR2(8),
    start_date        DATE,
    comments          VARCHAR2(255),
    manager_id        NUMBER(7),
    title             VARCHAR2(25),
    dept_id           NUMBER(7),
    salary            NUMBER(11,2),
    commission_pct    NUMBER(4,2),
    CONSTRAINT s_emp_id_pk PRIMARY KEY (id),
    CONSTRAINT s_emp_userid_uk UNIQUE (userid),
    CONSTRAINT s_emp-commission_pct_ck
    CHECK (commission_pct IN (10, 12.5, 15, 17.5, 20)));
```

[그림 6-2-10] 테이블 생성 DDL 예

□ 제약 조건

제약 조건은 NOT NULL, DEFAULT, FOREIGN KEY CONSTRAINT, CHECK 조건 등의 비즈니스 규칙은 칼럼에 정의할 것을 권장하나 테이블 간의 관계 적용 제약 규칙은 애플리케이션과 병행하여 적용할 수 있으며, 다음과 같은 기준에 따라 관리되어야 한다.

■ NOT NULL

테이블에 데이터 생성시 데이터가 반드시 존재해야 하는 칼럼을 정의한다.

■ DEFAULT

데이터가 반드시 존재해야 하는 칼럼에 확정 값을 정의할 수 없을 때 기본 데이터를 정의할 수 있다.

■ FOREIGN KEY

물리 데이터 모델에서 정의한 관계의 입력, 삭제, 생성 규칙을 정의하여 관리한다.

■ CHECK

특정 칼럼에는 미리 정의한 데이터 종류 혹은 범위 내의 데이터만 존재하도록 정의한다.

칼럼에 대한 제약 조건의 반영 역시 논리 데이터 모델의 속성 정의와 맞추어져야 한다. 관계에 대한 규칙은 애플리케이션에 의해 유지될 수 있다.

□ 인덱스

인덱스는 논리 데이터 모델에는 반영되어 있지 않으나 데이터의 접근 속도를 빠르게 하기 위한 데이터 저장소의 하나이다. 인덱스는 업무 요건에 따라 다양하게 정의될 수 있다. 그러나 구성하는 칼럼의 중복도가 높을수록 저장 공간의 낭비와 데이터 입력, 삭제, 갱신시에 오히려 속도에 악영향을 줄 수 있다. 인덱스는 사용하는 상용 RDBMS의 종류에 따라 다

양한 종류가 존재할 수 있다. 일반적으로 B*Tree 형태로 유지된다. 인덱스는 저장 공간의 재사용이 거의 없으므로 주기적으로 인덱스를 재생성 하는 것을 권장한다.

▣ 트리거

트리거는 테이블과 연계되어서 미리 규정된 함수를 수행한다. 트리거의 생성 시 이전(Before) 키 를 사용하여 Tuple(Row, Record)에 어떤 이벤트가 발생하기 전에 기동될 수 있도록 규정할 수 있으며, 반대로 이후(After) 키수도 있다. 트리거는 실행될 때 다른 트리거를 연쇄적으로 기동할 수 있다. 따라서 트리거의 생성과 사용은 신중하게 정의되어야 하며 잘못된 트리거의 사용으로 원하지 않는 결과를 얻을 수도 있다. 또한 동일한 테이블에 동 일한 이벤트를 지정하는 하나 이상의 트리거를 정의할 수 있으나, 이는 트리거의 기동 순서를 예측할 수 없게 하므로 보다 주의 깊은 관리가 필요하다.

▣ DB 링크

DB 링크(Link)는 원격지에 있는 데이터베이스를 연결하여 한 곳의 서버에서 다른 서버에 있는데 이터를 하나의 SQL문 내에서 다룰 수 있다. 분산 서버 환경에서 하나의 서버에서 다른 서버 혹은 다른 데이터베이스 인스턴스에 위치하는 테이블의 데이터를 손쉽게 호출하고자 할 때 정의한다. DB 링크가 제대로 생성되었으나 실제 질의시 연결에 실패하는 경우가 자주 발생하므로 실제 사용에서 작동 여부에 주의가 필요하다. 또한 DB 링크의 남용은 SQL 수행 속도의 저하를 가져올 수 있다.

▣ 프로시저

함수(Function)와 프로시저(Procedure)를 사용자가 정의하여 사용할 수 있다. 함수와 프로시저는 프로그램 SQL 문으로 작성되며 사용자 함수는 SQL 문과 더불어서, 프로시저는 해당 프로시저의 이름을 호출함으로써 사용할 수 있다. 테이블의 데이터는 SQL 문에 의해서 입력, 수정, 삭제가 수행 되나 복잡한 업무를 수행할 때 혹은 같은 유형의 SQL이 반복 사용될 때 사용자 업무에 맞게 개선된 프로그램 유형의 SQL을 함수나 프로시저로 정의한다. (함수와 프로시저가 공용성이 보다 강조되면 동의어를 선언하여 다른 스키마에서 정의한 함수와 프로시저의 재사용률을 높이도록 한다.) 함수나 프로시저 내에 잘못 사용한 SQL 문장은 전체적인 수행 속도를 크게 저하시킬 수 있음을 주의한다.

▣ 뷰

뷰(View)는 사용자 관점의 데이터를 보기 위하여 생성한 객체로 실제 물리적인 저장 공간을 필요로 하는 것이 아닌 사용자가 정의한 SQL문의 수행 결과를 보여주는 가상 데이터 영역이다. 상용 RDBMS의 종류에 따라 실제 물리적인 저장 공간을 갖는 뷰도 존재할 수 있다. 중요한 데이터에 대한 접근 제한과 데이터베이스 복잡 완화, 복잡한 데이터베이스 디자인 숨김, 이질 데이터에 대한 분산 질의를 포함한 복잡한 질의 단순화, 사용자 접근 관리 단순화 등을 위하여 생성한다. 뷰의 장점은 뷰의 단점이 될 수 있다. 복잡한 뷰는 사용자의 의도를 제대로 파악할 수 없고 전체적으로 수행되는 SQL의 업무를 파악하기 힘들게 하며 속도에도 영향을 줄 수 있으므로 중첩 뷰의 사용과 너무 복잡한 뷰의 사용은 자제해야 할 것이다.

▣ 동의어

동의어(Synonym)는 테이블에 대한 별명을 의미한다. 일반 동의어는 모든 유저가 만들 수 있고, 소 유권은 만든 자신만이 사용할 수 있다. 공용 동의어는 PUBLIC을 사용한 동의어로, 데이터베이스 관리자만이 만들고 삭제할 수 있는 동의어이다. 주제 영역을 스키마로 정의했을 때 다른 스키마에 정의 했지만 업무상 자주 상호 참조해야 하는 테이블 데이터에 대하여 동의어를 정의할 수 있다. 동의어는 하나의 객체를 여러 스키마에서 공용으로 사용하고자 할 때 생성하는 것을 권장한다. 다른 스키마에서 생성된 객체를 읽기 권한이 있어도 객체에 대한 접근이 번거로울 때 동의어를 사용하여 간편화할 수 있다.

▣ 롤

롤(Role)은 데이터베이스 객체에 대하여 생성, 삭제, 읽기 변경 권한에 대한 그룹(Group)을 만든다. 롤을 부여하고 제거할 수 있는 사람은 데이터베이스 관리자 뿐이다. 데이터베이스 객체를 관리할 때는 사용의 편리함도 중요하나 테이블 내의 데이터의 보안과 관리도 중요하다. 권한 그룹을 생성하여 데이터베이스를 사용하는 사용자의 권한을 적절히 제한하여 데이터베이스 객체를 보호하고 객체 내의 데이터를 보호하기 위해 롤을 정의할 수 있다. 상용 RDBMS에서는 자주 사용되는 것과 중요도가 높은 권한들을 묶어서 몇 개의 기본적인 롤을 제공한다. 롤의 사용은 기본적으로 상용 RDBMS에서 제공하는 것을 사용하는 것에서 시작하되 시스템 내의 보안 규칙에 따라 보다 다양하게 정의하여 사용할 수 있다.



개념 데이터 모델

데이터 참조 모델

논리 데이터 모델

물리 데이터 모델

데이터베이스

사용자 뷰

+ 정의 및 관리 목적

사용자 뷰는 데이터를 제공하는 정보시스템상의 화면이나 출력물을 의미한다. 데이터 품질 관리 활동의 결과물인 데이터는 화면, 출력물과 같은 사용자 뷰를 통해 제공된다. 따라서 데이터에 대한 만족도를 극대화하기 위해서는 데이터 제공 매개체인 사용자 뷰도 관리되어야 한다. 사용자 뷰에 대한 관리는 시스템 개발 이후 유지, 보수 차원에서 좀더 의미가 있을 수 있다. 화면, 출력물과 시스템의 구조적 관계를 정의해 관리하면, 사용자 뷰를 개선 관리하기 위해 수행하는 데이터 모델이나 SQL 등에 대한 일련의 변경 작업을 신속하고 정확하게 수행할 수 있다. 사용자 화면과 출력물에 대한 관리가 필요한 이유는 다음과 같다.

- 사용자 뷰는 사용자 요구 사항과 업무의 정확성 및 단순성을 반영한 최종 결과이다.
- 사용자 화면과 출력물은 사용자가 요구하는 수준에 따라 개발되어야 하므로 시스템 개발 의뢰자와 시스템 개발자의 업무적 약속으로 정의할 수 있다.
- 사용자 출력물은 데이터 품질 관리의 전반에 걸쳐 수행한 작업의 결과물로서 사용자에게 제공되는 최종 산출물이라고 할 수 있다.

+ 세부 관리 대상

▣ 화면

화면은 정보시스템이 생성한 최종 산출물의 제공 인터페이스로 최종 사용자 화면과 시스템 관리자용 화면이 있으며, 다음과 같은 기준에 따라 관리되어야 한다.

■ 편의성

사용자 화면을 통해 처리되는 모든 작업 절차는 직관적이고 편리해야 한다.

■ 검색성

사용자는 화면을 통해 원하는 정보를 신속하고 정확하게 검색할 수 있어야 한다.

■ 지원성

화면을 처음 사용하는 사용자가 사용상 도움을 원할 때 적절한 수준으로 제공해야 한다.

■ 시스템 성능

화면을 통해 처리되는 모든 작업은 적절한 속도와 성능(예: 3초 이내의 응답성)을 유지해야 한다.

화면의 용도와 제약 사항에 대한 정보는 물론 해당 화면에 대한 사용자 요구 사항과 도움말 등을 함께 관리해야 한다.

▣ 출력물

출력물은 정보시스템을 통해 생성되는 산출물을 의미한다. 여기에는 보고서, 장표, 전표 등과 같은 산출물은 물론 해당 출력물을 생성하는 응용 프로그램까지 포함된다. 일반적으로 출력물은 최종 사용자에게 제공되거나 정보시스템 내부에서 분석용으로 활용된다. 정보시스템을 통해 생성되는 모든 출력물을 관리할 수는 없으며 대개의 경우에 출력물은 사용자 화면을 통해 제공되므로, 해당 화면을 관리하는 것으로 출력물의 품질 관리를 갈음한다. 그러나 사용자 화면이 없는 경우에는 출력물 생성에 관련된 업무와 출력물 예제(스캔 받은 출력물)를 시스템 관리 툴을 통해 관리한다.

데이터 관리 정책

데이터 표준 관리

요구 사항 관리

데이터 모델 관리

데이터 흐름 관리

데이터베이스 관리

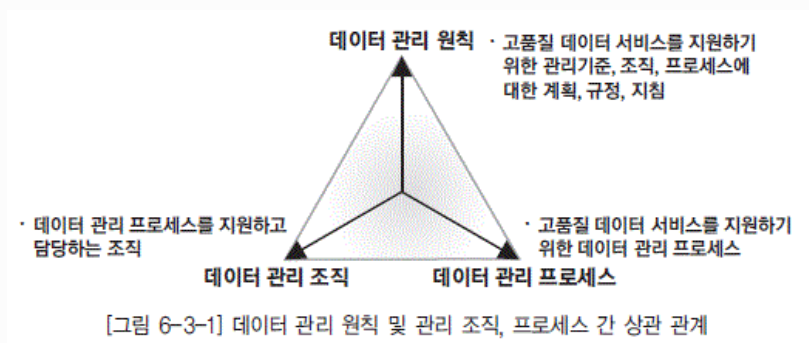
데이터 활용 관리

+ 정의 및 관리 목적

데이터 관리 정책은 기업의 비전과 목표 달성에 필요한 데이터의 확보 계획과 확보된 데이터의 효과적인 운영 관리 체계 및 계획을 정의하는 작업을 말한다. 기업은 데이터 관리 정책을 수립함으로써 기관의 비전과 목표에 맞는 데이터를 확보하고, 확보된 데이터를 사용자가 원하는 시간에 원하는 형태로 안정적으로 서비스 할 수 있는 기본 원칙과 관리체계를 구축할 수 있다. 또한 데이터에 대한 의사결정의 기초로 활용함으로써 체계적이고 일관성 있는 데이터 관리가 가능해진다. 데이터 관리 정책 자체는 일상적이고 반복적인 활동이 아니므로 프로세스를 관리하는 것보다는 정책에 의해 관리 활동이 적절히 이루어지는지를 점검하는 것이 필요하다. 즉 정기적/비정기적인 데이터 품질 평가 활동을 통한 검토 작업이 필요하다.

+ 세부 관리 대상

데이터 관리 정책은 데이터 관리 원칙, 데이터 관리 조직, 데이터 관리 프로세스에 대한 체계 및 계획을 수립하는 것을 의미한다. [그림 6-3-1]을 통해 각각의 세부 관리 대상 간의 상관 관계를 알 수 있다.



▣ 데이터 관리 원칙

데이터 관리 정책은 데이터 관리 원칙, 데이터 관리 조직, 데이터 관리 프로세스에 대한 체계 및 계획을 수립하는 것을 의미한다. [그림 6-3-1]을 통해 각각의 세부 관리 대상 간의 상관 관계를 알 수 있다.

■ 준수성

데이터 관리 원칙은 기업의 비전과 목표에 맞는 데이터를 확보하고 데이터 관리 목적을 달성할 수 있도록 정의하여야 한다.

■ 불가변성

수립된 데이터 관리 원칙은 아키텍처 원칙 변화에 의한 불가피한 경우를 제외하고는 쉽게 바뀌지 않도록 정의한다.

■ 이해성

데이터 관리 원칙은 쉽게 이해할 수 있어야 하며, 의미가 불분명하여 발생하는 혼란을 최소화해야 한다.

■ 완전성

데이터 관리 원칙은 정책 수립에 필요한 모든 사항을 정의해야 한다.

■

■ 일관성

데이터 관리 원칙은 원칙간의 충돌이 없도록 정의하고, 충돌이 발생한 경우에는 분명한 의사결정을 할 수 있도록 명시해야 한다.

데이터 관리 원칙은 문서화(Statement, Rationale, Implication)하여 관리하며, 변경은 데이터 관리와 관련된 담당자 및 사용자에게 의해 이루어지도록 한다.

▣ 데이터 관리 프로세스

데이터 관리 프로세스는 고품질의 데이터를 지속적이고 안정적으로 서비스하기 위해 각 기관의 특성에 맞게 정의한 프로세스 간의 연관 관계를 정의한 프로세스를 의미하며, 다음과 같은 기준에 따라 관리되어야 한다.

■ 준수성

데이터 관리 메인 프로세스는 데이터 관리 원칙에 맞게 정의되어야 한다.

■ 완전성

데이터 관리 메인 프로세스는 각 기관의 기존 프로세스에 대한 특성을 고려하여 정의하고 정의된 메인 프로세스는 데이터와 관련된 모든 요소가 빠짐없이 관리될 수 있도록 정의되어야 한다.

■ 상호 운용성

데이터 관리 메인 프로세스에는 기존의 다른 프로세스(변화 관리, 프로젝트 관리 등)와 상호 연관 관계가 명확하게 정의되어 있어 적용함에 문제가 없어야 한다.

데이터 관리 원칙에 준하여 데이터 관리 프로세스 목록을 도출하고 도출된 프로세스 간의 상호 운용성을 고려하여 메인 프로세스를 정의한다.

▣ 데이터 관리 조직

데이터 관리 조직 관리는 각 기관에서 정의한 데이터 관리 프로세스를 지원하고 담당할 담당자와 조직을 정의하는 것을 말하며, 다음과 같은 기준에 따라 관리되어야 한다.

■ 명확성

데이터 관리를 담당할 담당자가 선정되어 있고 담당자별로 수행해야 할 역할이 명확하게 정의되어야 한다.

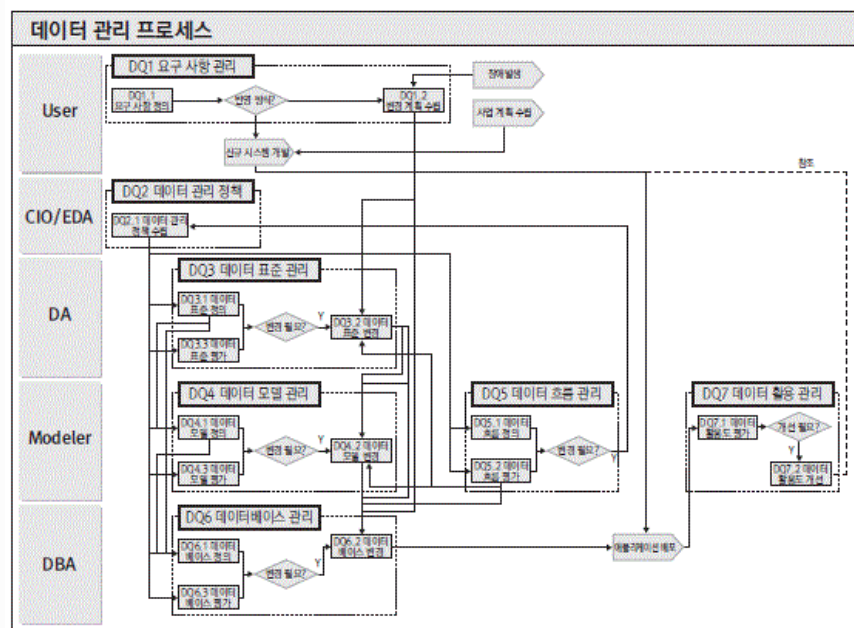
■ 운영성

데이터 관리 조직 구성원은 해당 역할 및 업무를 수행하는데 필요한 능력을 갖추고 있거나 능력에 필요한 교육 프로그램 등의 운영을 통해 관리 프로세스에 맞게 원활한 수행이 가능해야 한다.

각 기업에서는 체계적으로 정의한 데이터 관리 메인 프로세스를 수행하기 위해 필요한 역할을 정의하고 정의된 역할을 수행할 수 있는 담당자를 선정해야 한다.

+ 데이터 관리 프로세스

▣ 데이터 관리 프로세스



[그림 6-3-2] 데이터 관리 프로세스

※ [그림 6-3-2]의 'DA'는 Data Administrator를 의미함

■ 사업 계획 수립

[그림 6-3-2]의 사업 계획 수립은 사업에 대한 기회 분석, 사업 전략 및 계획 수립, 실행, 평가로 구성된다. 사업 기회 분석은 선진 사례 분석 및 기관의 환경 분석 결과를 바탕으로 새로운 사업 기회를 도출하는 작업이다. 사업 전략 및 계획 수립은 도출된 사업별로 전략을 수립하고 중장기 계획을 수립하는 작업으로, 이해 관계자에 대한 검증을 통해 사업 계획을 확정하는 작업이 반드시 필요하다. 이렇게 확정된 사업 계획을 토대로 실행하고 실행된 결과를 평가하는 작업을 반복적으로 수행함으로써 해당 기관의 운영 연속성을 보장할 뿐만 아니라 고품질의 서비스를 사용자에게 제공할 수 있다. 사업 계획 수립 프로세스는 데이터 관리 정책을 수립하는 기본 자료로 활용된다.

■ 신규사업

[그림 6-3-2]의 신규 시스템 개발은 신규 사업을 수행하는 데 필요한 시스템을 개발하고자 할 경우와 사용자의 수정/보완 요청 사항 중 대규모/장기적인 작업이 필요한 요건인 경우에는 신규 시스템 개발 작업을 수행하며, 소규모/단기적인 요건은 변경 작업을 수행한다. 신규 시스템의 개발은 기업별로 보유하고 있는 표준 개발 방법론에 정의된 절차에 의해 진행되어야 하며, 데이터 표준 및 데이터 참조 모델을 통해 전사적인 데이터베이스의 품질을 고도화하는 작업을 수행해야 한다.

■ 장애 발생

[그림 6-3-2]의 장애 발생은 비정상적인 애플리케이션 동작 및 데이터 오류, 시스템 오류 등으로 인해 고객에 대한 서비스의 품질을 떨어뜨리는 사건을 말한다. 우리가 일반적으로 이야기하는 장애 관리는 이러한 문제를 빠르게 해결함으로써 기업의 피해를 최소화하는 작업을 말한다. 긴급 장애(사후에 표준 및 설계서의 변경 처리를 수행함)를 제외한 장애 발생 시에도 사용자의 요구에 따른 변경 관리 절차와 동일한 변경 관리 프로세스를 수행한다.

■ 애플리케이션 배포

[그림 6-3-2]의 애플리케이션 배포는 테스트가 완료된 애플리케이션 및 데이터베이스를 운영 환경에 이관하는 작업뿐만 아니라 사용자가 안정적인 작업을 포함한다.

■ DQ1.1 요구 사항 정의

[그림 6-3-2]의 DQ1.1 요구 사항 정의는 비즈니스의 연속성 및 장애에 따른 위험성을 사전에 제거, 최소화하기 위해 사용자의 요구 사항을 수집·분석하는 작업이다. 요구 사항을 기준으로 데이터베이스의 변경에 따른 영향도를 분석하고 분석 결과를 토대로 적용 우선순위를 정의한다. 요구 사항의 영향도 및 중요도 분석 후에는 규모와 적용 시점을 고려하여 신규 시스템 개발로 해결을 할지 또는 기존 시스템의 변경으로 해결할지를 결정해야 한다.

■ DQ1.2 변경 계획 수립

[그림 6-3-2]의 DQ1.2 변경 계획 수립은 기존 시스템의 변경이 필요한 변경 사항인지, 표준 변경 요소인지, 모델 변경 요소인지를 판단하고 해당 작업을 수행하기 위한 작업자 배정 및 일정 계획을 수립하는 작업이다. 변경 계획 수립시에는 데이터 관련 변경 계획뿐만 아니라 애플리케이션과 기술에 대한 변경 계획도 포함시켜 종합적인 변경 계획이 수립될 수 있도록 작업을 수행해야 한다.

■ DQ2.1 데이터 관리 정책 수립

[그림 6-3-2]의 DQ2.1 데이터 관리 정책 수립은 사업 계획에 기반을 둔 기업의 비전과 목표를 달성하기 위해 필요한 데이터 확보 계획과 확보된 데이터를 효과적으로 관리, 유지하기 위한 체계 및 계획을 정의하는 작업을 말한다. 세부적인 작업 내역으로는 데이터베이스 품질과 관련된 프로세스를 정의하고, 정의된 프로세스를 수행하는 작업 주체를 선정하고, 선정된 작업 주체가 해당 작업을 원활하게 수행할 수 있는 능력을 배양할 수 있는 교육 체계를 수립하는 것이다.

■ DQ3.1 데이터 표준 정의

[그림 6-3-2]의 DQ3.1 데이터 표준 정의는 해당 기관에서 사용되는 용어 및 도메인, 코드, 데이터 관련 요소에 대한 표준을 전사적으로 정의하는 작업으로 표준에 따른 원칙을 정의하고 사용자의 표준화 요건을 수렴한 후 각 표준화 요소에 대한 전사 표준을 정의한다.

■ DQ3.2 데이터 표준 변경

[그림 6-3-2]의 DQ3.2 데이터 표준 변경은 정의된 데이터 표준(단어 표준, 도메인 표준, 코드 표준, 데이터 관련 요소 표준)에 대한 신규 및 추가 요청을 반영하는 변경 관리 작업이다. 변경이 요청된 표준을 수정하고 표준 변경에 따라 조정이 필요한 모델 변경 사항을 분석하여 모델 변경을 요청함으로써 표준화된 데이터 모델을 유지할 수 있도록 한다.

■ DQ3.3 데이터 표준 평가

[그림 6-3-2]의 DQ3.3 데이터 표준 평가는 해당 기관에서 전사적으로 정의한 용어, 도메인 및 코드 표준의 준수 현황을 평가하는 작업으로 정의된 표준과 데이터 모델과의 매핑을 통해 표준 준수 여부를 체크하고 미 준수 데이터에 대해서는 원인 및 변경 영향도 분석 결과를 반영하여 개선 작업을 수행한다.

■ DQ4.1 데이터 모델 정의

[그림 6-3-2]의 DQ4.1 데이터 모델 정의는 신규 시스템 개발시에 데이터 모델링 작업을 통해 설계된 개념 데이터 모델, 데이터 참조 모델, 논리 데이터 모델, 물리 데이터 모델을 전사적으로 생성, 유지하기 위해 필요한 작업을 말한다. 만약, 기존 시스템의 데이터 모델이 생성되어 관리되지 못하고 있다면 별도의 작업 계획을 수립하여 현재 운영 중인 데이터베이스의 스키마와 동일한 데이터 모델을 정의해야 한다.

■ DQ4.2 데이터 모델 변경

[그림 6-3-2]의 DQ4.2 데이터 모델 변경은 사용자 요구 사항에 적합한 서비스를 제공하기 위해 데이터 표준 및 참조

모델을 토대로 데이터 모델을 변경하는 작업이다. 변경 작업 수행 시에는 개념 데이터 모델과 논리 데이터 모델, 물리 데이터 모델이 상호 연관 관계를 유지할 수 있도록 변경 관리가 동시에 이루어져야 한다. 모델 변경 시에는 다른 영역에서 정의된 요소를 중복 요청한 것인지, 데이터의 정합성에 맞게 변경 처리하였는지를 고려해 처리해야 한다.

■ DQ4.3 데이터 모델 평가

[그림 6-3-2]의 DQ4.3 데이터 모델 평가는 해당 기관에서 전사적으로 관리하고 있는 데이터 모델을 평가하는 작업으로 개념 데이터 모델-논리 데이터 모델간, 논리 데이터 모델-물리 데이터 모델간, 물리 데이터 모델-DB간 매핑 작업과 연결정보(Alignment) 분석 작업을 실시하여 발생된 오류에 대한 데이터 모델 개선 작업을 수행하고, 영향도 분석을 거쳐 DBMS에 대한 개선 작업을 수행한다.

■ DQ5.1 데이터 흐름 정의

[그림 6-3-2]의 DQ5.1 데이터 흐름 정의는 원천 데이터(문서, Text, DB 등)를 수기로 생성하거나 추출, 변환, 적재, 가공을 통해 목표 데이터베이스에 저장하는 데이터의 라이프사이클을 통제, 관리하는 작업으로 정기적/비정기적인 배치 작업 및 정형/비정형 데이터의 배치 작업을 포함한다.

■ DQ5.2 데이터 흐름 평가

[그림 6-3-2]의 DQ5.2 데이터 흐름 평가는 소스 데이터를 생성하여 타겟 데이터로 저장/관리되는 데이터의 정합성을 평가하는 작업이다. 데이터 흐름 점검 기준과 지표를 설정하고 데이터의 정합성을 체크하여 오류 데이터를 분석하고 영향도 분석 결과를 반영하여 개선 작업을 수행한다.

■ DQ6.1 데이터베이스 정의

[그림 6-3-2]의 DQ6.1 데이터베이스 정의는 데이터베이스를 안정적으로 운영, 유지하는데 필요한 정기적/비정기적 작업을 말한다. 여기에는 데이터 모델에 적합한 데이터베이스 구성 및 백업, 보안, 복구, 성능 관리 등이 있다.

■ DQ6.2 데이터베이스 변경

[그림 6-3-2]의 DQ6.2 데이터베이스 변경은 요구 사항에 따라 변경된 데이터 모델을 토대로 데이터베이스를 변경하는 작업을 말한다.

■ DQ6.3 데이터베이스 평가

[그림 6-3-2]의 DQ6.3 데이터베이스 평가는 현재 설정된 데이터베이스의 객체에 지정한 제약 조건과 객체 유형을 확인하여 해당 규칙이 최적의 성능을 보장하고 데이터의 오류를 방지하기에 적합한지를 평가하는 것이다.

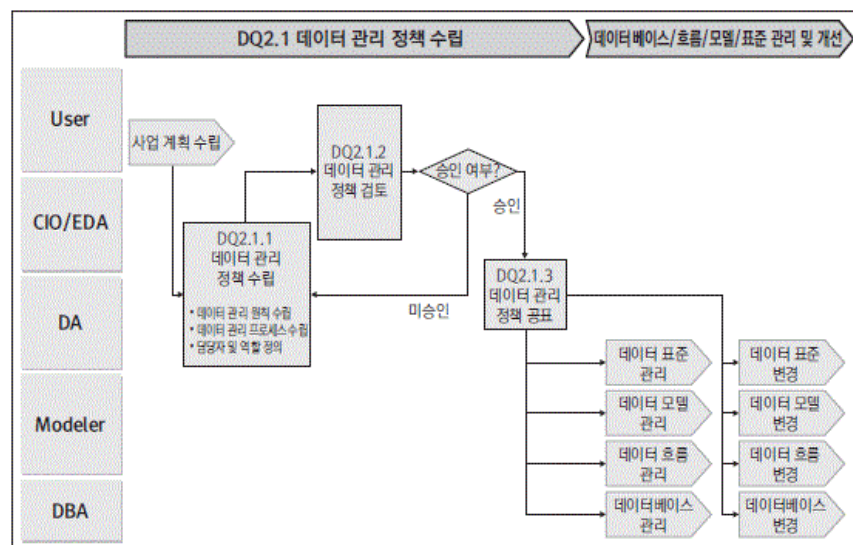
■ DQ7.1 데이터 활용도 평가

[그림 6-3-2]의 DQ7.1 데이터 활용도 평가는 데이터의 활용도를 높이기 위해 핵심 데이터를 수집한 후 이를 대상으로 활용도 측정 기준을 마련하여 데이터의 활용도를 측정하는 작업이다.

■ DQ7.2 데이터 활용도 개선

[그림 6-3-2]의 DQ7.2 데이터 활용도 개선은 데이터 활용도의 저하 원인과 데이터 품질이 충족되지 못한 원인을 분석하여 개선 방안을 마련하고, 데이터 품질 개선 활동을 통해 데이터 활용도를 높이기 위한 작업이다.

▣ 데이터 관리 정책 수립 프로세스



[그림 6-3-3] 데이터 관리 정책 수립 프로세스

■ DQ2.1.1 데이터 관리 정책 수립

[그림 6-3-3]의 DQ2.1.1 데이터 관리 정책 수립은 비즈니스나 IT의 환경 변화에 따라 데이터 관리 정책의 수립 및 변경이 필요한 경우, 필요한 관련 자료를 수집하여 정책 자료를 작성한다. 정책 작성 시에는 데이터 관리 원칙에 대한 수립과 데이터 관련 프로세스의 정의, 관련 담당자의 역할 정의 등의 내용이 포함되어야 한다. 정책 수립 작업은 전사 데이터 관리자(EDA, Enterprise Data Administrator)가 작업하는 것이 일반적이나 데이터 관리자(DA, Data Administrator)가

작업을 수행한 후 EDA가 검토하여 최종안을 확정하는 경우도 있다.

■ DQ2.1.2 데이터 관리 정책 검토

[그림 6-3-3]의 DQ2.1.2 데이터 관리 정책 검토는 수립된 정책(안)을 토대로 CIO/EDA 및 관련 사용자, 관련 데이터 관리자 등이 참석하여 정책에 대한 완전성 및 일관성, 실현 가능성 등을 검토 하여 승인 처리한다.

■ DQ2.1.3 데이터 관리 정책 공표

[그림 6-3-3]의 DQ2.1.3 데이터 관리 정책 공표는 확정된 데이터 관리 정책을 선포하고, 정책 변 경에 따른 데이터 관리 프로세스의 정의 및 수정이 필요한 경우 이를 수행토록 한다. 전문성이 요 구되는 활동인 경우 담당자의 교육 훈련이 이루어지도록 한다.

데이터 관리 정책

데이터 표준 관리

요구 사항 관리

데이터 모델 관리

데이터 흐름 관리

데이터베이스 관리

데이터 활용 관리

+ 정의 및 관리 목적

데이터 표준 관리는 데이터 표준화 원칙에 따라 정의된 표준 단어 사전 및 도메인 사전, 표준 용어 사전, 표준 코드, 데이터 관련 요소 표준 등을 기관에 적합한 형태로 정의하고 관리하는 작업을 말하며, 데이터베이스 설계와 개발을 지원하고 전사적인 데이터 표준의 사용 및 재사용을 통해 시스템간 상호 운용성, 데이터 공유, 시스템 통합, 비즈니스 프로세스 개선 등을 지원한다. 데이터 표준 관리는 전사적으로 공통된 표준을 사용하게 함으로써 데이터의 일관성과 정확성을 유지할 수 있다.

또한 데이터 표준 관리는 지속적인 표준화에 대한 교육과 개선/모니터링 활동으로 표준이 조직과 관련 담당자에게 체화 되도록 한다. 데이터 표준은 현업의 의견이 반영되어야 하겠지만, 관습적으로 잘못 사용되어 온 용어를 모두 수용할 수는 없으므로 조정이 필요하다. 또한 표준의 적용은 신규 개발 시점에서 이루어지고 기존 시스템과의 중복 표준이 허용될 수 있다. 표준 관리 대상 및 적용 대상이 많을 경우 표준화 도구 등을 활용한 자동화를 고려할 수 있다.

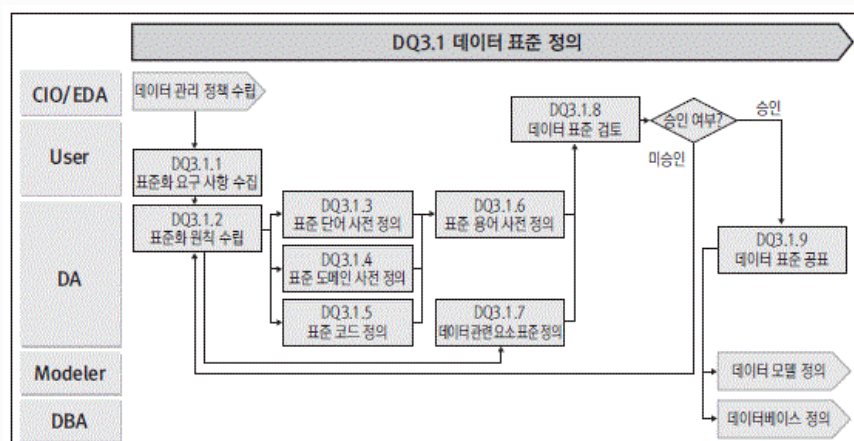
+ 세부 관리 대상

세부 관리 대상으로 표준 데이터가 있으며, 해당 데이터와 관련된 내용은 앞서 1장 데이터 이해와 2장 데이터 구조 이해에서 다루었으므로 여기서는 생략하기로 한다.

+ 데이터 표준 관리 프로세스

▣ 데이터 표준 관리 프로세스

[그림 6-3-4]와 [그림 6-3-5]는 데이터 표준 정의 및 데이터 표준 변경에 대한 데이터 표준 관리 프로세스를 보여준다



[그림 6-3-4] 데이터 표준 정의 프로세스

■ DQ3.1.1 표준화 요구 사항 수집

[그림 6-3-4]의 DQ3.1.1 표준화 요구 사항 수집은 현업, 모델러, 및 사용자 뷰 운영자를 대상으로 데이터 표준과 관련된 요구 사항을 인터뷰와 설문지 조사를 통해 수집하고, 전사 데이터 표준 대상 후보를 추출하여 개선 방안을 도출한다.

■ DQ3.1.2 표준화 원칙 수립

[그림 6-3-4]의 DQ3.1.2 표준화 원칙 수립은 현행 정보시스템에서 적용하고 있는 데이터 표준 원칙과 모델 데이터, 업무 데이터를 수집하여 현행 데이터 표준에 대한 개선 방안을 토대로 향후에 적용할 전사 데이터 표준화 원칙을 수립한다. 지속적으로 표준 데이터를 관리하고 개선할 수 있도록 데이터 표준 지침을 작성한다.

■ DQ3.1.3 표준 단어 사전 정의

[그림 6-3-4]의 DQ3.1.3 표준 단어 사전 정의는 기존 모델 데이터 및 용어집을 통해 해당 기관에서 사용되고 있는 모든 단어를 추출하여 정의된 표준화 원칙에 따라 한글명, 영문명, 영문 약어명 등을 정의하고 단어의 종류와 유형을 분류한다.

■ DQ3.1.4 표준 도메인 사전 정의

[그림 6-3-4]의 DQ3.1.4 표준 도메인 사전 정의는 데이터의 물리적 데이터 특성과 사용 빈도, 업무적 특성을 고려하여 정의된 표준화 원칙에 따라 도메인을 분류하고 도메인별 데이터 타입을 정의한다.

■ DQ3.1.5 표준 코드 정의

[그림 6-3-4]의 DQ3.1.5 표준 코드 정의는 기존 모델 데이터를 통해 코드를 선별하여 현 코드집에 누락된 코드 정보를 수집하고, 수집된 코드 정보와 표준화 원칙을 바탕으로 표준 코드를 정의한다. 통합 요구 사항과 통합 필요성의 제기 시에는 코드 통합 대상을 추출하고 해당 코드를 활용하는 사용자를 대상으로 인터뷰 및 설문 등을 실시하여 표준 코드를 정의한다. 향후 지속적인 관리를 위해 코드별 오너십(Ownership)을 부여한다.

■ DQ3.1.6 표준 용어 사전 정의

[그림 6-3-4]의 DQ3.1.6 표준 용어 사전 정의는 표준 단어, 표준 도메인, 표준 코드를 조합하여 정의된 표준화 원칙에 따라 표준 용어를 정의하고 용어의 설명을 수집한다.

■ DQ3.1.7 데이터 관련 요소 표준 정의

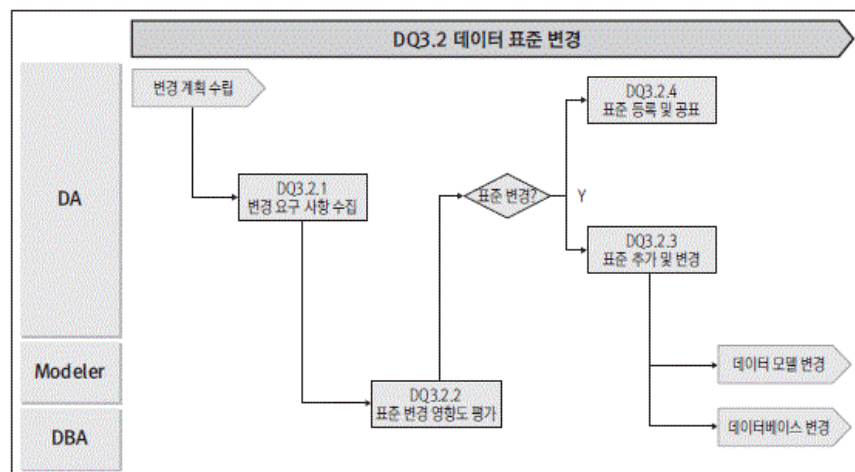
[그림 6-3-4]의 DQ3.1.7 데이터 관련 요소 표준 정의는 정의된 표준 데이터와 표준화 원칙을 바탕으로 업무적 용도와 물리적 특성을 고려하여 데이터 관련 요소 표준을 정의한다.

■ DQ3.1.8 데이터 표준 검토

[그림 6-3-4]의 DQ3.1.8 데이터 표준 검토는 데이터 관리자가 정의한 표준 데이터가 업무적 용도와 물리적 특성을 고려하여 표준화 원칙에 위배됨이 없이 정확하게 정의되었는지를 확인하고 표준 예외 사항은 표준화 원칙에 피드백하여 처리한다.

■ DQ3.1.9 데이터 표준 공표

[그림 6-3-4]의 DQ3.1.9 데이터 표준 공표는 정의된 표준화 프로세스에 따라 전사 시스템에 표준화 원칙이 적용 가능하도록 확정된 데이터 표준을 배포하고 표준 데이터 관리에 대한 이해 및 적용을 위한 교육을 실시한다.



[그림 6-3-5] 데이터 표준 변경 프로세스

■ DQ3.2.1 변경 요구 사항 검토

[그림 6-3-5]의 DQ3.2.1 변경 요구 사항 검토는 요청된 표준 변경 요구 사항이 기존에 정의된 데이터 표준을 사용해서도 처리 가능한 요건인지를 먼저 검토하고, 추가 및 변경이 필요하다고 판단 되는 경우에만 추가/변경 작업을 요청한다. 만약 기존 표준만으로도 처리 가능한 요건이라면 데이터 모델 변경 작업이나 변경 취소 처리를 한다.

■ DQ3.2.2 표준 변경 영향도 평가

[그림 6-3-5]의 DQ3.2.2 표준 변경 영향도 평가는 표준의 변경 시에 기존 테이블이나 칼럼에 영향을 미치므로 해당 표준의 변경으로 인해 변경이 필요한 테이블 및 속성, 기타 요소들을 파악하고 해당 모델러(Modeler)에게 해당 작업을 요청한다. 변경 영향도 평가 작업 시 누락된 영향 요소가 없는지 철저히 파악하도록 한다.

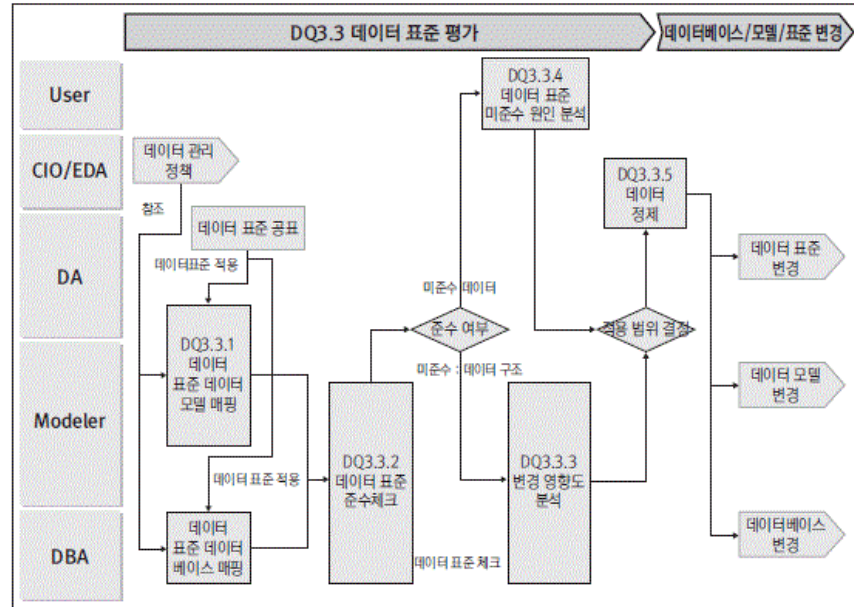
■ DQ3.2.3 표준 추가 및 변경

[그림 6-3-5]의 DQ3.2.3 표준 추가 및 변경은 표준 변경 요소에 대한 내역을 데이터 표준화 원칙에 맞게 추가 및 변경한다. 변경 작업이 완료되면 변경된 사항을 토대로 영향도 평가 작업 및 공표 작업을 요청한다.

■ DQ3.2.4 표준 등록 및 공표

[그림 6-3-5]의 DQ3.2.4 표준 등록 및 공표는 표준 추가 및 변경 작업을 통해 변경된 데이터 표준 내역을 공표하여 향후 모델링 작업 및 데이터베이스 관리 작업 시에 활용하도록 한다. 표준 변경 내역에 대한 올바른 적용을 위해 교육을 고려할 수도 있다.

□ 데이터 표준 개선 프로세스



[그림 6-3-6] 데이터 표준 개선 프로세스

■ DQ3.3.1 데이터 표준-데이터 모델 매핑

[그림 6-3-6]의 DQ3.3.1 데이터 표준 - 데이터 모델 매핑은 용어 표준, 도메인 표준, 명명 규칙 표준을 데이터 모델(개념, 논리, 물리)에 반영하는 작업을 말한다. 예를 들어 데이터 표준화 생성 프로세스에서 생성된 용어 표준(속성 표준, 칼럼 표준 등)을 데이터 모델에 반영하는 작업이다. 통상적으로는 데이터 모델을 생성하는 작업을 수행하면서 데이터 표준에 따라 모델을 생성하게 된다. 하지만 이미 만들어진 데이터 모델에 대해서는 각 데이터 모델에 용어 표준을 적용하여야 한다.

■ DQ3.3.2 데이터 표준 준수 체크

[그림 6-3-6]의 DQ3.3.2 데이터 표준 준수 체크는 데이터 표준과 데이터 객체(데이터 모델, 데이터베이스 객체) 간에 데이터 표준을 준수하고 있는지를 체크하는 과정이다. 각 기관이 제정한 데이터 표준(표준 용어, 표준 도메인, 명명 규칙 등)에 대해 각각의 데이터 모델 객체, 데이터베이스 객체들이 표준을 준수하고 있는지 체크한다. 이러한 단계들은 통상적으로는 주기적으로 수행된다.

■ DQ3.3.3 변경 영향도 분석

[그림 6-3-6]의 DQ3.3.3 변경 영향도 분석은 앞선 체크 과정에서 데이터 표준 미준수 부분에 대한 영향을 분석하는 과정이다. 구체적으로는 데이터 표준을 변경할 때의 영향, 데이터 모델을 변경할 때의 영향, 데이터베이스 객체를 변경했을 때의 영향 등으로, 데이터 표준을 준수하지 않아 발생할 수 있는 영향들을 분석하는 과정이다. 이러한 분석 과정을 통하여 데이터 표준 변경, 데이터 모델 변경, 데이터베이스 변경 가운데 하나의 프로세스를 분기하게 된다.

■ DQ3.3.4 데이터 표준 미준수 원인 분석

[그림 6-3-6]의 DQ3.3.4 데이터 표준 미준수 원인 분석은 실 데이터 값에 대해서 데이터 표준을 지키고 있는지를 체크하여 표준 미준수의 원인을 분석하는 과정이다. 실제로 데이터 표준에서 정의한 코드 도메인 또는 표준 코드 값들을 실제 데이터들이 준수하고 있는지를 체크하는 과정을 들 수 있다.

■ DQ3.3.5 데이터 정제

[그림 6-3-6]의 DQ3.3.5 데이터 정제는 앞선 데이터 표준을 준수하지 않은 데이터에 대해서 여러 분석 작업을 통하여 데이터를 수정하는 과정을 말한다.

데이터 관리 정책

데이터 표준 관리

요구 사항 관리

데이터 모델 관리

데이터 흐름 관리

데이터베이스 관리

데이터 활용 관리

+ 정의 및 관리 목적

요구 사항 관리란 데이터를 비롯하여 관련 애플리케이션 및 시스템 전반에 걸친 사용자의 요구를 수집하고 분류하여 반영하는 작업을 말한다. 사용자의 정보 요구 사항을 종합적으로 검토, 확인하여 요건에 적합하도록 시스템을 개선, 반영함으로써 사용자의 만족도를 높이고 고품질의 서비스를 가능하게 한다.

사용자의 요구 사항 관리는 데이터와 관련된 요구 사항만을 따로 정의하는 것보다는 애플리케이션 과 관련된 요구 사항과 비즈니스에 대한 요구 사항 등을 통합하고 함께 정의하여 관리해야 한다. 이번 절에서는 대부분의 경우 요구 사항 관리 프로세스가 애플리케이션에 대한 요구 사항만을 관리함으로써 데이터에 대한 요구 사항이 원활하게 유지, 관리되지 못하는 점을 고려하여 데이터와 관련된 요구 사항에 대한 관리 방법을 제시하고자 한다. 따라서 각 기관에서 수행 중인 요구 사항 프로세스에 데이터 요구 사항 관리를 포함시킬 수 있는 참고 자료로 활용할 수 있다.

+ 세부 관리 대상

▣ 외부 인터페이스 요건

외부 인터페이스 요건이란 외부 기관이나 외부 시스템과의 사이에서 발생하는 모든 입/출력에 대한 요건으로 외부 기관이 추가 변경되거나 제도 및 기준 등이 변화하여 인터페이스 형식이 바뀌었을 경우가 이에 해당된다. 외부 인터페이스 요건은 다음과 같은 기준에 따라 관리되어야 한다.

■ 중복성

동일한 형태의 인터페이스가 중복되어 정의되지 않도록 정의한다.

■ 표준 준수도

관련 인터페이스와 관련된 국제 표준 및 국가 표준이 존재할 경우 그에 적합한 형태로 제공되어야 한다.

또한, 외부 인터페이스 요건에는 항목 이름, 목적 설명, 입력의 원천 및 출력의 방향, 유효 범위, 시간, 다른 입/출력과의 관계, 데이터 포맷, 최종 메시지 등이 포함되어 관리되어야 한다.

▣ 기능 개선 요건

기능 개선 요건이란 애플리케이션에서 입력 발생하는 출력에 대한 요건을 말하며, 다음과 같은 기준에 따라 관리되어야 한다.

■ 불가변성

기능 개선 요건이 향후에 재변경되지 않도록 근본적인 개선 방안을 요청해야 한다.

■ 범용성

많은 사용자가 편리하게 사용할 수 있는 요건을 우선적으로 요청해야 한다.

또한 기능 개선 요건에는 입력에 대한 유효 체크, 정확한 처리 순서, 비정상 상태에 대한 반응(오버플로우, 통신 장비, 에러 처리), 매개변수의 기능, 입출력 관계, 입출력 순서, 입력을 출력으로 변환하는 공식 등이 포함되어 관리되어야 한다.

▣ 성능 개선 요건

성능 개선 요건이란 해당 기관의 사용자가 필요로 하는 성능 개선 사항으로 정적인 수치 요구 사항은 지원하는 터미널 수, 지원하는 동시 사용자 수, 처리하는 정보의 양과 종류 등이 있으며 동적인 수치 요구 사항은 일정한 기간 내에 처리하는 트랜잭션이나 작업의 수 등이 있다. 성능 개선 요건은 다음과 같은 기준에 따라 관리되어야 한다.

■ 실현 가능성

해당 성능 개선 요구 사항이 현행 기술 수준과 서비스 특성을 고려할 때 구현 가능한 요건인지를 확인한 후 제시해야 한다.

■ 측정 가능성

측정이 불가능한 모호한 형태로 요건이 제시되면 안된다. 예를 들어 ‘좀 더 빠르게 제공되었으면 좋겠다’는 표현보다는 ‘현재 15초 정도 소요되는 것을 5초 이내로 개선했으면 좋겠다’와 같이 구체적으로 제시해야 한다.

또한, 성능 개선 요건은 각 기관의 서비스 특성을 고려하여 정적, 동적 기준을 마련하고 해당 기준에 맞게 서비스되고 있는지를 모니터링 작업을 통해 항시적으로 관리해야 한다.

□ 보안 개선 요건

보안 개선 요건이란 중요 데이터에 대한 훼손, 변조, 도난, 유출에 대한 물리적 접근 통제(제한 구역, 동제 구역 등) 및 사용 통제(인증, 암호화, 방화벽 등)에 대한 요건을 말하며, 다음과 같은 기준에 따라 관리되어야 한다.

■ 불변성

보안 개선 요건이 향후에 재변경되지 않도록 근본적인 개선 방안을 요청해야 한다.

■ 실현 가능성

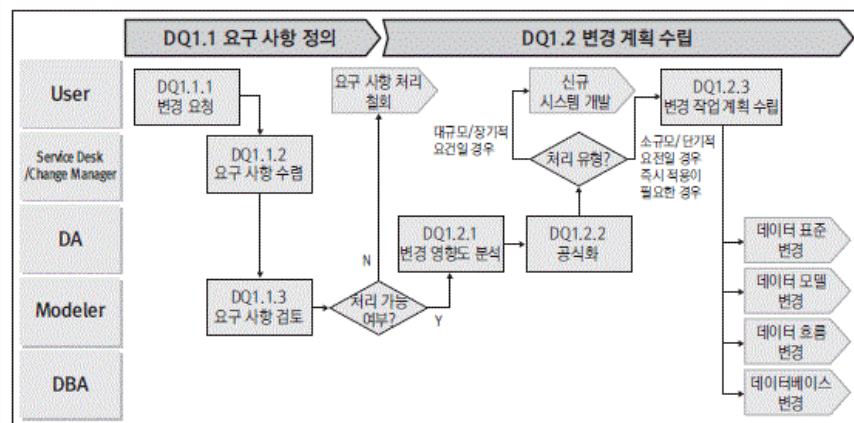
해당 보안 개선 요구 사항이 현행 기술 수준과 서비스 특성을 고려할 때 구현 가능한 요건인지를 확인한 후 제시해야 한다.

보안 개선 요건에 대한 관리를 위해서는 먼저 보안 관리가 필요한 정보에 대한 등급을 설정하고, 해당 등급별로 접근 가능한 이용자의 등급을 관리해야 한다. 또한 접근 방식에 있어서의 접근 통제 기준 및 사용 통제 기준이 제시되어야 한다.

해당 기준에 따라 모니터 작업을 통해 안정적인 서비스가 제공될 수 있도록 관리해야 한다.

✚ 요구 사항 관리 프로세스

여기서 말하는 요구 사항 관리 프로세스는 ‘데이터 요건 분석’과목에서 다루는 내용과는 차이가 있다. ‘데이터 요건 분석’과목에서의 요건 분석은 사전 요구 사항 분석을 의미하며, 여기서는 사후 요구 사항 변경 관리를 의미하므로 그 차이점에 대해서 참고하기 바란다.



[그림 6-3-7] 요구 사항 관리 프로세스

■ DQ1.1.1 변경 요청

[그림 6-3-7]의 DQ1.1.1 변경 요청은 사용자가 해당 기관의 시스템을 활용하면서 발생하는 외부 인터페이스 및 기능, 성능, 보안 등의 요건을 요구 사항 변경 신청서를 통해 변경 요청한다. 요구사항 변경 신청서는 [그림 6-3-8]과 같은 내용으로 구성될 수 있다.

요구사항 변경 신청서							
번호				요구사항 명			
요청자 정보	부서명		성명		연락처		요청일
반영자 정보	부서명		성명		연락처		반영일
요청 내역							
상세 요구사항				반영 방법		확인	
						(반영자)	(사용자)
반영효과							

[그림 6-3-8] 요구 사항 변경 신청서 예

■ DQ1.1.2 요구 사항 수렴

[그림 6-3-7]의 DQ1.1.2 요구 사항 수렴은 사용자로부터 요청된 변경 요청서를 수집하여 변경 신청서 작성 규칙에 맞게 정확하게 정의했는지를 확인하고 해당 요건을 검토할 처리담당자 (Modeler)를 지정한다.

■ DQ1.1.3 요구 사항 검토

[그림 6-3-7]의 DQ1.1.3 요구 사항 검토는 요청된 요구 사항과 관련된 자료 및 기준, 시스템 등을 확인하여 처리 가능 여부를 판단하고 처리 가능한 경우에 데이터 관리자를 통해 공식화를 요청 한다.

■ DQ1.2.1 변경 영향도 분석

[그림 6-3-7]의 DQ1.2.1 변경 영향도 분석은 변경 요청된 내역을 토대로 변경에 따른 영향이 미치는 설계서 및 애플리케이션, 데이터베이스 등을 도출하고, 그 결과로 변경 영향도 분석서를 작성 한다.

■ DQ1.2.2 공식화

[그림 6-3-7]의 DQ1.2.2 공식화는 영향도 분석을 통해 변경 처리가 요구되는 관련 담당자를 수집하여 공식화를 하고 해당 담당자들과의 협의를 통해 승인 방식은 규모 및 기간, 시급성에 따라 결정되며, 처리 방법은 신규 시스템 개발 방식과 기존 시스템 변경 방식이 있다. 공식화 단계는 요구 사항 승인서를 통해 진행된다.

■ DQ1.2.3 변경 작업 계획 수립

[그림 6-3-7]의 DQ1.2.3 변경 작업 계획 수립은 영향도 평가서를 통해 관련된 업무 영역 및 관련 시스템 내역을 토대로 작업 일정 계획을 수립한다. 작업 일정 계획에는 표준과 설계서 변경, 데이터베이스 및 애플리케이션 수정, 테스트, 이관 등의 작업이 명시되어야 하며, 작업별로 작업 담당자 및 일정이 정의된 변경 작업 계획서를 작성한다.

데이터 관리 정책

데이터 표준 관리

요구 사항 관리

데이터 모델 관리

데이터 흐름 관리

데이터베이스 관리

데이터 활용 관리

+ 정의 및 관리 목적

데이터 모델 관리란 데이터 요구 사항 관리에 의해 변경되는 데이터 구조를 모델에 반영하는 작업 절차와 데이터베이스 시스템 구조와 동일하게 데이터 모델을 유지하도록 하는 작업 절차를 말한다. 데이터 모델은 기관의 비즈니스 목적에 맞는 최적화된 데이터 서비스를 제공하고 데이터베이스를 구성하고 유지하기 위해 체계적으로 관리되어야 한다.

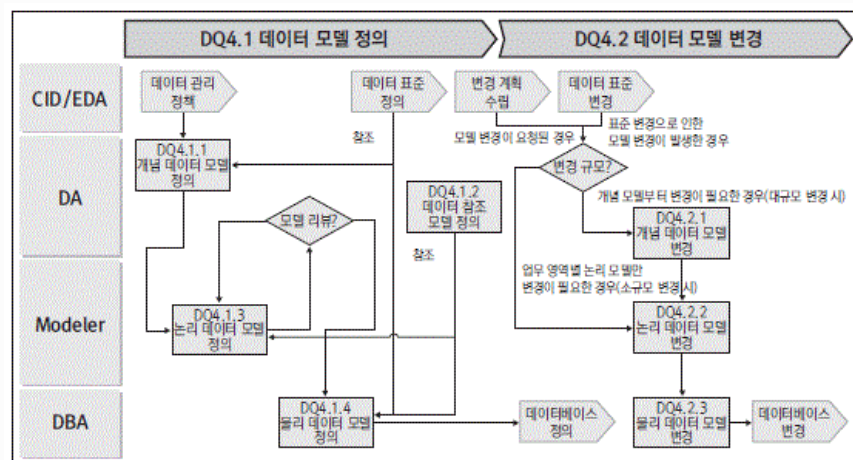
개념 데이터 모델을 관리함으로써 논리, 물리 데이터 모델의 연관 관계 분석을 통한 전사 데이터 구조에 대한 파악이 가능하다. 또한 물리 데이터 모델과 데이터베이스 간의 상관 관계 분석을 통해 현재 운영 중인 데이터베이스와 동일한 모델 확보를 통해 유지 보수 및 체계적인 전사 데이터베이스 관리가 가능하다. 이외에도 데이터 참조 모델을 활용함으로써 일정 수준 이상의 데이터 모델 및 고품질의 데이터 서비스가 가능하다. 개념 데이터 모델, 논리 데이터 모델, 물리 데이터 모델을 관리하기 위해서는 모델링 도구를 활용하는 것이 효과적이다. 또한 업무 영역별 개별 관리 방식보다는 데이터의 중복 및 불필요한 인터페이스를 확인하고 조정이 가능한 전사적 통합 관리 방식을 권장한다.

+ 세부 관리 대상

세부 관리 대상으로 개념 데이터 모델, 데이터 참조 모델, 논리 데이터 모델, 물리 데이터 모델이 있으며, 해당 데이터 모델과 관련된 내용은 앞서 2장 데이터 구조 이해에서 다루었으므로 여기서는 생략한다.

+ 데이터 모델 관리 프로세스

□ 데이터 모델 관리 프로세스



[그림 6-3-9] 데이터 모델 관리 프로세스

■ DQ4.1.1 개념 데이터 모델 정의

[그림 6-3-9]의 DQ4.1.1 개념 데이터 모델 정의는 각 기관의 비전을 수립하는 데 필요한 데이터 주제 영역을 정의하고, 세부적인 내역보다는 전사 정보를 중복되지 않고 확장성 있게 설계하는 데 초점을 맞춘다. 데이터의 주제 영역과 핵심

데이터 집합 및 데이터 집합 간의 관계를 정의하여 항 후에 정의하게 될 상세 논리 데이터 모델과 물리 데이터 모델과의 데이터 구조적 정렬을 지원한다.

■ DQ4.1.2 데이터 참조 모델 정의

[그림 6-3-9]의 DQ4.1.2 데이터 참조 모델 정의는 업무 영역별, 주제 영역별 표준 데이터 집합, 관리 항목들이 표기되어 재사용이 가능한 데이터 모델을 정의하는 작업이다. 이미 검증된 데이터 모델을 참조함으로써 데이터 모델의 정확성과 재사용률을 높이고 일정 수준 이상의 설계 품질을 보장할 수 있다.

■ DQ4.1.3 논리 데이터 모델 정의

[그림 6-3-9]의 DQ4.1.3 논리 데이터 모델 정의는 비즈니스 규칙을 토대로 업무의 모든 데이터 구조를 상세하고 구체적으로 정의한 모델로, 데이터 참조 모델 및 데이터 표준을 참고하여 설계 작업을 수행한다. 개념 데이터 모델 정의 시에는 다른 주제 영역 간의 인터페이스 추출 작업에 초점을 맞춘다면, 논리 데이터 모델 정의 작업 시에는 개념 데이터 모델의 인터페이스를 토대로 주제 영역 내의 연관 관계를 중심으로 설계 작업을 수행한다. 논리 데이터 모델 정의 작업이 완료되면 데이터 관리자 및 사용자 등과 함께 재검토 작업을 수행하여 해당 비즈니스 요건에 적합한 형태로 설계되었는지를 검토한다.

■ DQ4.1.4 물리 데이터 모델 정의

[그림 6-3-9]의 DQ4.1.4 물리 데이터 모델 정의는 논리 데이터 모델 및 데이터 표준을 기준으로 대상 데이터베이스의 물리 특성을 고려하여 최적의 성능이 발휘될 수 있도록 상세한 설계 작업을 수행한다.

■ DQ4.2.1 개념 데이터 모델 변경

[그림 6-3-9]의 DQ4.2.1 개념 데이터 모델 변경은 사용자 요구 사항의 특성에 따라 모델 변경 요청 및 표준에 대한 변경 요청으로 분리되는데, 이 중에서 변경 규모가 클 경우(다른 주제 영역 간의 인터페이스 조정 및 핵심 엔터티 타입의 변경, 핵심 엔터티 타입 간의 관계 변경) 개념 데이터 모델의 변경 작업이 발생된다. 개념 데이터 모델의 변경 시에는 반드시 논리 데이터 모델 및 물리 데이터 모델의 변경이 발생된다.

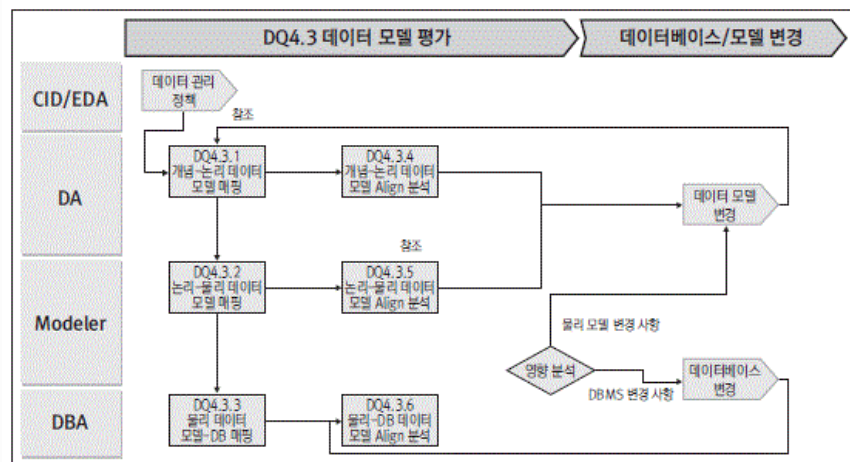
■ DQ4.2.2 논리 데이터 모델 변경

[그림 6-3-9]의 DQ4.2.2 논리 데이터 모델 변경은 개념 데이터 모델이 변경되거나 개념 데이터 모델의 변경이 없는 작은 규모의 변경(주제 영역 내의 인터페이스 조정 및 엔터티 타입의 변경, 엔터티 타입 간의 관계 변경, 속성 변경)이 요청된 경우, 또는 데이터 표준이 변경된 경우 논리 데이터 모델의 변경 작업이 수행된다. 논리 데이터 모델 변경 시에는 다른 주제 영역에 동일한 형태의 데이터 집합이 존재하는지 중복성을 검토하고, 데이터 표준 및 데이터 참조 모델 등을 참조하여 표준화된 모델을 유지할 수 있도록 한다.

■ DQ4.2.3 물리 데이터 모델 변경

[그림 6-3-9]의 DQ4.2.3 물리 데이터 모델 변경은 변경 요청된 내역을 논리 데이터 모델 및 데이터 표준, 데이터베이스의 물리 특성 등을 참고하여 최적의 성능을 발휘할 수 있도록 물리 데이터 모델 변경 작업을 수행한다.

□ 데이터 모델 개선 프로세스



[그림 6-3-10] 데이터 모델 개선 프로세스

■ DQ4.3.1 개념-논리 데이터 모델 매핑

[그림 6-3-10]의 DQ4.3.1 개념-논리 데이터 모델 매핑은 개념적으로 생성된 데이터 집합 또는 관리 항목과 논리 데이터 모델 사이의 구조적 정렬 정보를 생성하는 작업으로, 데이터 아키텍처 관점에서 개념 데이터 모델의 각 객체와 논리 데이터 모델의 각 객체 간 연결 정보를 설정한다.

■ DQ4.3.2 논리-물리 데이터 모델 매핑

[그림 6-3-10]의 DQ4.3.2 논리-물리 데이터 모델 매핑은 비즈니스 규칙을 토대로 업무의 모델 데이터 구조와 이를 바탕으로 데이터베이스의 물리적인 특성을 고려하여 정의한 물리 데이터 모델 간의 구조적 연결 정보를 설정한다. 예를 들어 하나의 엔터티가 여러 개의 테이블이 되는 경우도 존재한다.

■ DQ4.3.3 물리 데이터 모델-DB 매핑

[그림 6-3-10]의 DQ4.3.3 물리 데이터 모델-DB 매핑은 물리 데이터 모델(최종 설계 도면)과 DBMS 카탈로그(건축물) 정보와의 구조적 연결 정보를 설정한다. 대부분의 모델링 툴은 물리 데이터 모델과 DB 간의 연결 정보를 자동으로 생성한다.

■ DQ4.3.4 개념-논리 데이터 모델 얼라인(Align) 분석

[그림 6-3-10]의 DQ4.3.4 개념-논리 데이터 모델 얼라인 분석은 개념 데이터 모델에 정의된 모델이 실제 논리 데이터 모델에 구체적으로 정의되지 않은 모델이 존재하는지를 체크하는 등의 차이 분석 작업을 말한다. 이렇게 분석된 결과에 의해 개념 데이터 모델의 변경 또는 논리 데이터 모델 변경의 프로세스를 수행하게 된다.

■ DQ4.3.5 논리-물리 데이터 모델 얼라인 분석

[그림 6-3-10]의 DQ4.3.5 논리-물리 데이터 모델 얼라인 분석은 논리 데이터 모델과 물리 데이터 모델 사이의 차이를 분석한다. 모델러 관점에서 변경 사항을 분석하여 해당 데이터 모델에 대한 변경을 수행하게 된다.

■ DQ4.3.6 물리 데이터 모델-DB 얼라인 분석

[그림 6-3-10]의 DQ4.3.6 물리 데이터 모델-DB 얼라인 분석은 물리 데이터 모델과 실제 DB와의 차이를 분석한다. 이렇게 함으로써 데이터 모델에 표현되지 않는 DB 객체가 있는지를 분석할 수 있다. 물론 정상적인 데이터 모델 변경 프로세스 또는 데이터베이스 변경 프로세스를 수행했다면 이러한 차이는 발생하지 않을 것이다.



데이터 관리 프로세스 이해

Home > DB 구축 가이드 > DA 가이드 > 데이터 품질관리 이해 > 데이터 관리 프로세스 이해

데이터 관리 정책

데이터 표준 관리

요구 사항 관리

데이터 모델 관리

데이터 흐름 관리

데이터베이스 관리

데이터 활용 관리

+ 정의 및 관리 목적

데이터 흐름 관리란 소스 데이터(문서, Text, DB 등)를 수기로 생성하거나 추출, 변환, 적재를 통해 생성하여 타깃 데이터베이스에 저장·가공하는 것을 관리하는 절차를 말한다. 각 기관이 관리하고 있는 데이터가 생성, 변경되고 활용되는 생명주기를 관리함으로써 전사 데이터에 대한 현황 파악 및 최적화된 형태로 활용되고 있는지를 확인할 수 있다.

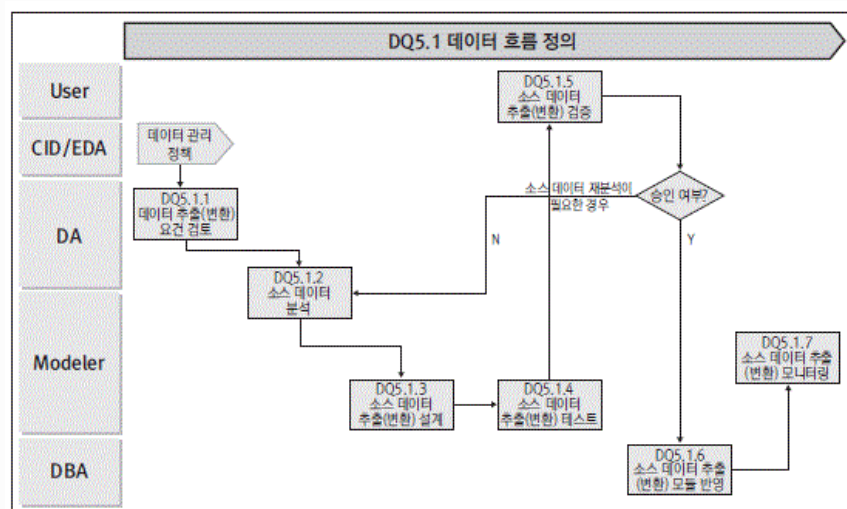
개념 데이터 모델을 관리함으로써 논리, 물리 데이터 모델의 연관 관계 분석을 통한 전사 데이터 구조에 대한 파악이 가능하다. 또한 물리 데이터 모델과 데이터베이스 간의 상관 관계 분석을 통해 현재 운영 중인 데이터베이스와 동일한 모델 확보를 통해 유지 보수 및 체계적인 전사 데이터베이스 관리가 가능하다. 이외에도 데이터 참조 모델을 활용함으로써 일정 수준 이상의 데이터 모델 및 고품질의 데이터 서비스가 가능하다. 개념 데이터 모델, 논리 데이터 모델, 물리 데이터 모델을 관리하기 위해서는 모델링 도구를 활용하는 것이 효과적이다. 또한 업무 영역별 개별 관리 방식보다는 데이터의 중복 및 불필요한 인터페이스를 확인하고 조정이 가능한 전사적 통합 관리 방식을 권장한다.

+ 세부 관리 대상

세부 관리 대상으로 개념 데이터 모델, 데이터 참조 모델, 논리 데이터 모델, 물리 데이터 모델이 있으며, 해당 데이터 모델과 관련된 내용은 앞서 2장 데이터 구조 이해에서 다루었으므로 여기서는 생략한다.

+ 데이터 흐름 관리 프로세스

▣ 데이터 흐름 관리 프로세스



[그림 6-3-11] 데이터 흐름 관리 프로세스

■ DQ5.1.1 데이터 추출(변환) 요건 검토

[그림 6-3-11]의 DQ5.1.1 데이터 추출(변환) 요건 검토는 현업 업무를 위해 사용자로부터 접수한 요구 사항 중에서 데

이터를 추출(변환)하여 해당 데이터베이스에 적재해야 하는 요건을 검토한다. 데이터베이스 관리자는 해당 요건 검토 시 전사아키텍처 뷰에서 데이터 정책/표준을 기준으로 요건 반영 여부 및 방법에 대한 내용을 검토한다.

■ DQ5.1.2 소스 데이터 분석

[그림 6-3-11]의 DQ5.1.2 소스 데이터 분석을 위해 모델러는 소스 데이터를 추출(변환)하여 해당 데이터베이스에 적재하기로 결정된 요건에 대해 소스 데이터 관점에서 해당 테이블 및 컬럼에 대한 내용을 분석한다. 컬럼 분석 시에는 컬럼에 대한 속성(데이터 타입 및 건수 등을) 중심으로 분석하여 해당 요건을 지원할 수 있는 소스 데이터를 식별해 소스 데이터 분석서를 작성한다. [그림 6-3-12]와 같은 형식으로 작성할 수 있다.

소스 데이터 분석서					
부서명		담당영역		담당자	
작성일		요구사항 번호			
상세 요구사항	소스 테이블 및 컬럼 정보				비고
	소스 테이블명	소스 컬럼명	소스 데이터 타입	데이터 건수	

[그림 6-3-12] 소스 데이터 분석서 예

■ DQ5.1.3 소스 데이터 추출(변환) 설계

[그림 6-3-11]의 DQ5.1.3 소스 데이터 추출(변환) 설계를 위해 모델러는 소스 데이터의 변환 로직 및 적재 로직을 설계한다. 설계 시 고려할 사항은 타깃 데이터베이스의 테이블 및 컬럼을 식별하여 매핑 로직 분석서를 작성하며, [그림 6-3-13]과 같은 형식으로 작성할 수 있다.

매핑 로직 분석서						
부서명		담당영역		담당자		
작성일		요구사항 번호				
소스			타깃			변환 로직
테이블명	컬럼명	데이터타입	테이블명	컬럼명	데이터타입	

[그림 6-3-13] 매핑 로직 분석서 예

■ DQ5.1.4 소스 데이터 추출(변환) 테스트

[그림 6-3-11]의 DQ5.1.4 소스 데이터 추출(변환) 테스트를 위해 모델러는 추출(변환) 설계에 따라 소스 데이터를 테스트 형식으로 타깃 데이터베이스에 적재한다. 테스트를 통해 적재(가공)된 데이터의 정확성을 사전 정의된 소스 데이터 대상 검증 항목을 기준으로 대상 검증 분석서를 작성한다. [그림 6-3-14]와 같은 형식으로 작성할 수 있다.

소스 대 타깃 대상 검증 분석서							
부서명		담당영역		담당자			
요구사항명	대사 기준	검증 항목	엔터타입명	검증여부	검증내용	검증 항목	사용자 확인

[그림 6-3-14] 소스 대 타깃 대상 검증 분석서 예

■ DQ5.1.5 소스 데이터 추출(변환) 검증

[그림 6-3-11]의 DQ5.1.5 소스 데이터 추출(변환) 검증을 위해 사용자는 소스 데이터 추출(변환) 테스트에서 작성된 대상 내용을 바탕으로 해당 요건이 타깃 데이터베이스에 정확하게 반영되어 데이터가 적재되었는지를 확인한다. 해당 요건에 따라 타깃 데이터베이스로 데이터가 정확하게 적재되지 않았을 경우에는 소스 데이터 분석을 다시 하도록 한다. 소스 데이터는 정확하게 타깃 데이터 베이스로 적재(가공)되었지만 테스트를 통해 확인된 내용이 해당 업무를 지원하지 못할 경우에는 사용자 요구 사항을 재정의한다.

■ DQ5.1.6 소스 데이터 추출(변환) 모듈 반영

[그림 6-3-11]의 DQ5.1.6 소스 데이터 추출(변환) 모듈 반영을 위해 데이터베이스 관리자는 사용자의 검증이 완료된 소스 데이터 추출(변환) 변화를 운영 환경으로 적용한다. 해당 모듈을 운영 환경으로 적용하면서 데이터베이스 관리자는 다른 시스템 및 프로그램에 영향을 주거나 성능을 저해하는 사항이 없는지를 확인한 후 반영한다.

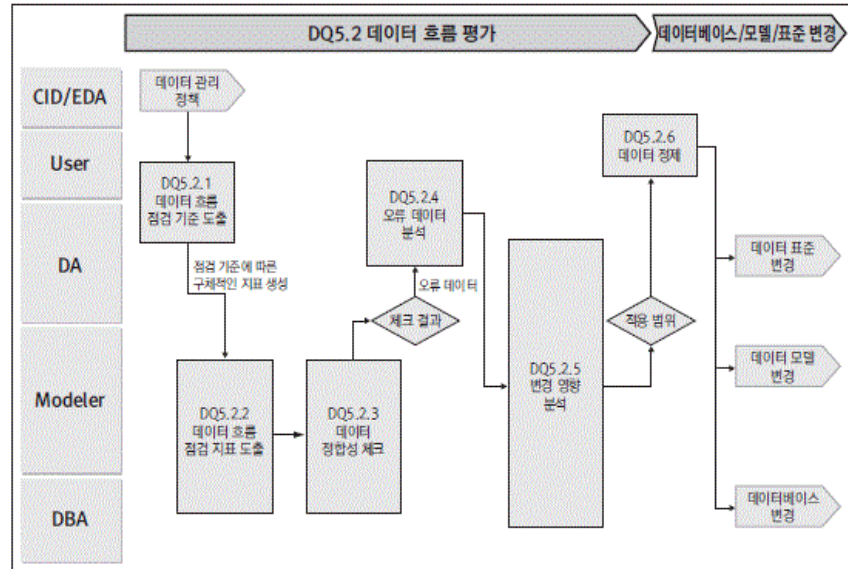
■ DQ5.1.7 소스 데이터 추출(변환) 모니터링

[그림 6-3-11]의 DQ5.1.7 소스 데이터 추출(변환) 모니터링을 위해 모델러는 운영 환경에 적용된 소스 데이터 추출(변환) 모듈을 정해진 규칙에 따라 주기적으로 모니터링하여 그 결과를 데이터 관리자에게 보고한다. [그림 6-3-15]와 같은 형식으로 작성할 수 있다.

소스 데이터 추출(변환) 모니터링 결과 보고서						
부서명		담당영역			담당자	
모듈 번호	모듈명	모듈 실행 상태	모듈 실행 속도 상태	소스 데이터 현황	타깃 데이터 적재 현황	조치 사항

[그림 6-3-15] 소스 데이터 추출(변환) 모니터링 결과 보고서 예

▣ 데이터 흐름 개선 프로세스



[그림 6-3-16] 데이터 흐름 개선 프로세스

■ DQ5.2.1 데이터 흐름 점검 기준 도출

[그림 6-3-16]의 DQ5.2.1 데이터 흐름 점검 기준 도출은 데이터 오류를 최소화하기 위해 지속적으로 품질 점검을 통해 관리되어야 할 기준을 도출하는 과정이다. 예로는 데이터 구조적 측면의 흐름 점검 기준, 데이터 정합성 유지 방안 등을 들 수 있다. 사용자가 설정한 품질 지표별로 데이터 이동에 대한 규칙들을 구체적인 기준으로 생성하여 관리할 수 있다.

■ DQ5.2.2 데이터 흐름 점검 지표 도출

[그림 6-3-16]의 DQ5.2.2 데이터 흐름 점검 지표 생성은 앞선 데이터 흐름 점검 기준으로 구체적인 데이터 흐름의 정합성을 체크할 수 있는 지표들을 도출한다. 각각의 지표들은 데이터 흐름에 대한 정합성을 체크할 수 있는 구체적인 데이터 이동으로 체크될 수 있는 형태로 도출되어야 한다.

■ DQ5.2.3 데이터 정합성 체크

[그림 6-3-16]의 DQ5.2.3 데이터 정합성 체크는 앞선 지표에 따른 구체적인 체크 모듈을 실행하여 정합성을 체크하는 과정이다.

■ DQ5.2.4 오류 데이터 분석

[그림 6-3-16]의 DQ5.2.4 오류 데이터 분석은 데이터 정합성 검증을 통하여 추출된 오류 데이터에 대한 분석을 수행한다. 오류의 원인 분석에 따라 데이터 관리의 각 요소에 적절한 조치를 수행하게 된다.

■ DQ5.2.5 변경 영향 분석

[그림 6-3-16]의 DQ5.2.5 변경 영향 분석은 앞선 체크 과정에서 오류 데이터의 원인에 대한 분석을 통하여 구체적으로는 데이터 표준을 변경할 때의 영향, 데이터 모델을 변경할 때의 영향, 데이터베이스 객체를 변경했을 때의 영향 등을 분석하는 과정이다.

■ DQ5.2.6 데이터 정제

[그림 6-3-16]의 DQ5.2.6 데이터 정제는 데이터 정합성을 지키지 않은 오류 데이터의 원인 분석 결과에 따라 데이터 관리의 각 요소에 적절한 조치를 수행하고, 데이터 값을 수정하는 과정을 말한다.



데이터 관리 프로세스 이해

Home > DB 구축 가이드 > DA 가이드 > 데이터 품질관리 이해 > 데이터 관리 프로세스 이해

데이터 관리 정책

데이터 표준 관리

요구 사항 관리

데이터 모델 관리

데이터 흐름 관리

데이터베이스 관리

데이터 활용 관리

+ 정의 및 관리 목적

데이터베이스 관리란 원활한 데이터 서비스를 위해 필요한 데이터베이스를 안정적으로 운영, 관리 하는 데 필요한 작업을 체계화하는 것으로 백업, 보안, 튜닝, 모니터링 등의 작업이 포함된다. 데이터베이스 관리 작업은 데이터베이스와 데이터베이스에 저장된 데이터가 오류 및 훼손 없이 안정적으로 서비스될 수 있도록 데이터베이스에 대한 생성 및 변경, 보안, 성능 개선, 백업 관리를 지속적으로 수행할 수 있도록 체계화하는 작업이라고 할 수 있다.

데이터베이스 관리 체계화를 통해 데이터의 오류 및 훼손 없이 사용자가 원하는 데이터를 원하는 시간에 원하는 형태로 정확하고 안정적으로 서비스함으로써 안정적이고 지속적인 업무 활동의 기반을 마련할 수 있다.

데이터베이스 관리 프로세스 중 보안 관리는 데이터베이스에 대한 보안뿐만 아니라 애플리케이션, PC, 문서 등 각 기관이 보유하고 있는 모든 자원에 대한 외부 침입으로부터의 보호이기 때문에 별도의 상세화된 지침을 통해 정의하는 것이 효과적이다. 또한 보안에 대한 전반적인 관리는 별도의 보안 담당자를 선임하는 것이 좋으며, 데이터베이스 관리자의 경우 데이터베이스 및 데이터베이스에 저장된 데이터에 대한 보안에 책임을 갖도록 한다.

+ 세부 관리 대상

세부 관리 대상으로 표준 데이터, 모델 데이터, 관리 데이터, 업무 데이터, 데이터베이스가 있다. 해당 데이터와 관련된 내용은 앞서 1장 데이터 이해와 2장 데이터 구조 이해에서 다루었으므로 여기 서는 생략한다.

+ 데이터베이스 관리 프로세스

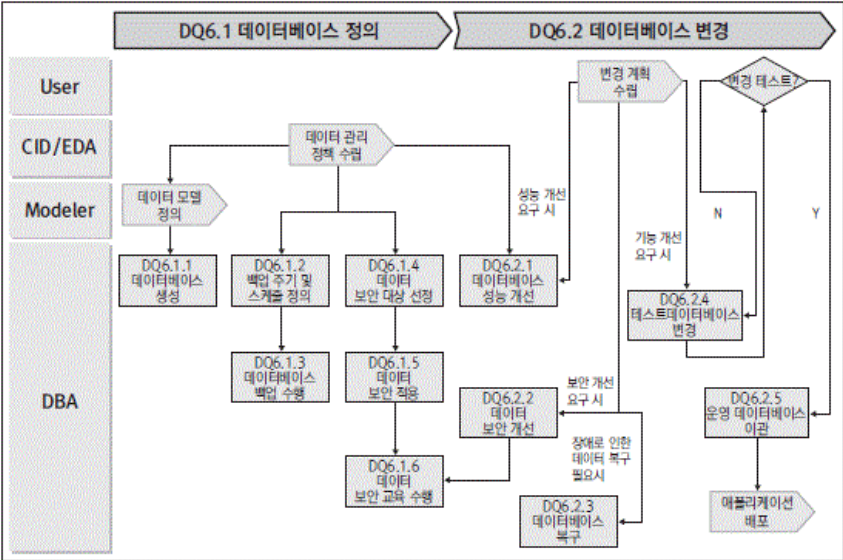
▣ 데이터베이스 관리 프로세스

■ DQ6.1.1 데이터베이스 생성

[그림 6-3-17]의 DQ6.1.1 데이터베이스 생성은 비즈니스 요건에 맞게 설계된 데이터 모델을 토대로 작성된 DDL문을 가지고 데이터베이스의 물리 특성을 고려하여 데이터베이스를 구성한다.

■ DQ6.1.2 백업 주기 및 스케줄 정의

[그림 6-3-17]의 DQ6.1.2 백업 주기 및 스케줄 정의는 어떠한 장애가 발생되더라도 사용 중인 데이터의 완전 복구가 가능하도록 백업 주기 및 스케줄이 정의되어야 한다. 따라서 일 단위, 주 단위, 월 단위, 년 단위의 백업 주기별 백업 내용을 정의해 관리해야 한다. 백업 주기별 스케줄 표는 [그림 6-3-18]과 같은 형식으로 작성할 수 있다.



[그림 6-3-17] 데이터베이스 관리 프로세스

백업 스케줄 표			
주기	시각	백업내용	백업수단
수시	AM 4:00	전일 Data Export	Export
매일			

[그림 6-3-18] 백업 주기별 스케줄 예

■ DQ6.1.3 데이터베이스 백업 수행

[그림 6-3-17]의 DQ6.1.3 데이터베이스 백업 수행은 백업 주기별 스케줄 표를 참고로 하여 백업 을 수행한다. 백업 수행 절차는 데이터베이스 가동 상태가 정상인지를 확인하고 배치 작업이 없는 시간을 배정하여 백업을 수행한다. 백업 수행이 완료되면 로그 파일 및 루트 메일(Root Mail)을 통해 결과를 확인한다. 백업 결과에 대해서는 일지를 반드시 기록해야 하며, 실패한 경우에도 작성 하도록 한다. 백업 실패 시 재수행을 하며, 재수행이 불가능할 경우 각 기관에서 정의한 장애 관리 프로세스에 의해 처리하도록 한다. 백업 수행과 관련되는 백업 일지 및 백업 관리 대장은 [그림 6- 3-19], [그림 6-3-20]과 같은 형식으로 작성할 수 있다.

백업일지						
백업번호		작업일		작업자		승인자
백업목적						
구분	대상 데이터	시작시간	완료시간	백업수단	완료 여부	신청자
작업결과						

[그림 6-3-19] 백업 일지 예

백업관리 대장								
백업번호	대상데이터	백업수단	보관주기	보관장소	보관 안료일	폐기일	보관 담당자	폐기 담당자

[그림 6-3-20] 백업 관리 대장 예

■ DQ6.1.4 데이터 보안 대상 선정

[그림 6-3-17]의 DQ6.1.4 데이터 보안 대상 선정은 보호되어야 할 자산의 파악 및 가치에 대한 평가 작업을 수행하고, 시스템에 존재하는 취약점 및 위험 요인에 대한 분석 작업을 수행한다. 분석된 결과를 토대로 보안 대상을 선정하고 선정된 대상에 보안 등급을 적용하여 중요도에 따른 차 등 보안 적용을 수행한다.

■ DQ6.1.5 데이터 보안 적용

[그림 6-3-17]의 DQ6.1.5 데이터 보안 적용은 보안 관리 대상별 중요도에 따른 보안을 적용하는 작업으로, 물리적 접근 보안 및 네트워크 보안, 서버 및 운영체제 보안, 데이터베이스 보안, 응용 시스템 보안, PC 보안 등 종합적인 보안 적용이 필요하다. 그 중에서도 데이터베이스 보안은 데이터베이스 사용자에게 대한 보안 관리와 데이터베이스 자체에 대한 보안, 데이터에 대한 보안으로 나눌 수 있으며, 데이터의 중요도에 따른 체계적인 적용이 필요하다.

■ DQ6.1.6 데이터 보안 교육 수행

[그림 6-3-17]의 DQ6.1.6 데이터 보안 교육 수행은 기관별로 수립된 데이터 보안 정책을 년 1회 이상 전 구성원을 대상으로 실시해야 하며, 교육 평가 작업 등을 통한 고품질의 교육이 될 수 있도록 체계화한다. 보안 교육의 예로는 컴퓨터 및 인터넷에 대한 구조 및 해킹과 관련된 일반 내용을 담은 정보 보호 입문부터 인터넷 서버 보안 교육, 네트워크 보안 교육, 해킹 및 대응 등과 같은 세 부적인 교육까지 체계적으로 수립·시행되어야 한다.

■ DQ6.2.1 데이터베이스 성능 개선

[그림 6-3-17]의 DQ6.2.1 데이터베이스 성능 개선은 해당 기관의 사용자가 필요로 하는 성능 개선 사항으로, 정적인 수치 요구와 동적인 수치 요구를 처리하는 작업인 데이터베이스 성능 개선을 들 수 있다. 성능에 대한 작업은 사용자의 요구에 의해 개선되는 경우도 있지만 자체적인 성능개선 대상을 선정하고 선정된 대상에 대해 일별, 주별, 월별 모니터링 작업을 통해 일정 정도 성능 기준에 미달했을 경우 조정하는 작업도 있다. 이러한 주기적인 성능 개선 작업은 안정적인 서비스를 위한 근본적인 해결 방법이라 하겠다. 이때 작성해야 하는 것으로 [그림 6-3-21]과 같은 성능 개선 관리 대장이 있다.

성능 개선 관리 대장							
번호	성능 개선명	원인	조치내역	신청일	처리일	신청자	처리 담당자

[그림 6-3-21] 성능 개선 관리 대장 예

■ DQ6.2.2 데이터 보안 개선

[그림 6-3-17]의 DQ6.2.2 데이터 보안 개선은 중요 데이터에 대한 훼손, 변조, 도난, 유출에 대한 물리적 접근 통제(제한 구역, 통제 구역 등) 및 사용 통제(인증, 암호화, 방화벽 등)에 대한 요건이 발생되었을 경우 보안 장치를 개선하는 작업으로, 데이터 보안 개선을 들 수 있다. 보안 작업 또한 성능 개선 작업과 마찬가지로 보안 대상에 대해 항시적인 모니터링 작업을 통한 주기적인 개선 작업이 안정적인 데이터 관리를 위한 근본적 해결 방법이라 하겠다. 보안 개선 관리 대장은 [그림 6-3-22]와 같은 형식으로 작성할 수 있다.

보안 개선 관리 대장							
번호	보안 개선명	원인	조치내역	신청일	처리일	신청자	처리 담당자

[그림 6-3-22] 보안 개선 관리 대장 예

■ DQ6.2.3 데이터베이스 복구

[그림 6-3-17]의 DQ6.2.3 데이터베이스 복구는 장애 등으로 인해 데이터에 대한 전반적인 훼손 및 에러로 인해 기존 백업된 데이터로의 복구 작업이다. 안정적인 복구 작업을 위해서는 복구와 관련된 작업 절차를 사전에 정의하고 정의된 절차에 따라 사전 테스트 작업을 수행해야 한다. 복구 관리 대장은 [그림 6-3-23]과 같은 형식으로 작성할 수 있다.

복구 관리 대장								
번호	대상 데이터	원인	복구수단	관련 백업번호	작업 결과	복구수행 시작시간	복구수행 완료시간	복구 담당자

[그림 6-3-23] 복구 관리 대장 예

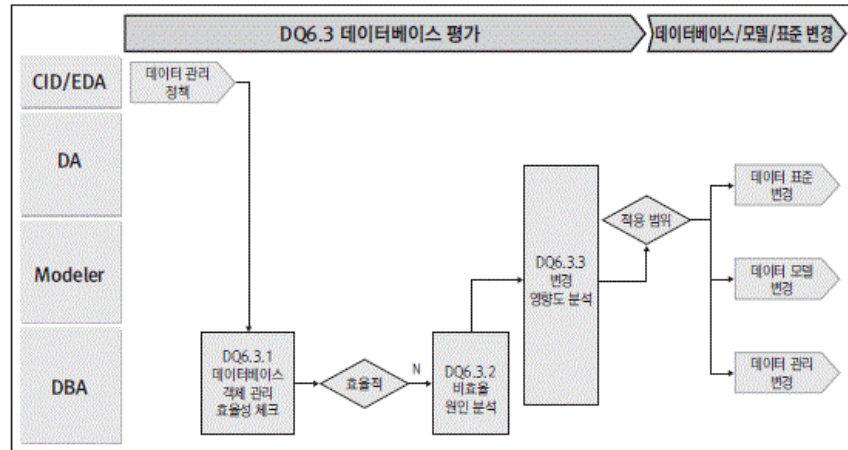
■ DQ6.2.4 테스트 데이터베이스 변경

[그림 6-3-17]의 DQ6.2.4 테스트 데이터베이스 변경은 변경 요청에 의해 제시된 요건에 따라 변경된 데이터 모델을 토대로 작성된 DDL문을 가지고 데이터베이스의 물리 특성을 고려한 테스트 데이터베이스를 변경된 데이터 모델과 동일한 형태로 변경하는 작업이다.

■ DQ6.2.5 운영 데이터베이스 이관

[그림 6-3-17]의 DQ6.2.5 운영 데이터베이스 이관은 테스트 데이터베이스에 변경된 내역을 토대로 해당 애플리케이션에 대한 문제점을 확인하는 단위 테스트와 타 애플리케이션과의 인터페이스를 테스트하는 통합 테스트, 사용자의 만족도를 확인하는 사용자 테스트 등을 수행한 후 안정성 및 정확성이 확보되면 운영 데이터베이스에 해당 변경 내역을 반영한다.

▣ 데이터베이스 개선 프로세스



[그림 6-3-24] 데이터베이스 개선 프로세스

■ DQ6.3.1 데이터베이스 객체 관리 효율성 체크

[그림 6-3-24]의 DQ6.3.1 데이터베이스 객체 관리 효율성 체크는 현재 설정된 데이터베이스의 객체에 지정한 제약 조건과 객체 유형을 확인하여 최적의 성능을 보장하고 데이터의 오류를 방지하기 위한 객체 관리 규칙들인지 평가하는 것이다.

■ DQ6.3.2 비효율 원인 분석

[그림 6-3-24]의 DQ6.3.2 비효율 원인 분석은 현재 설정한 객체 관리 유형이나 객체 유형이 비효율적 성능을 보인다면 해당 원인을 분석하는 것이다. 예를 들어 불필요한 제약 규칙의 남발이나 저장 공간 지정의 부적절, 데이터 객체에 데이터 분할 저장 방법의 문제, 인덱스 구성 칼럼의 순서 등이 원인이 될 수 있다.

■ DQ6.3.3 변경 영향도 분석

[그림 6-3-24]의 DQ6.3.3 변경 영향도 분석은 비효율을 개선하기 위하여 데이터베이스 내에서 제약 조건이나 객체 유형을 변경할 수도 있으나 테이블의 통합/분리의 변경이 요구된다면 물리 데이터 모델의 변경이 요구될 수도 있다. 또한 객체의 도메인이 변경될 수 있다면 데이터 표준의 변경이 역으로 요구될 수도 있다.

데이터 관리 정책

데이터 표준 관리

요구 사항 관리

데이터 모델 관리

데이터 흐름 관리

데이터베이스 관리

데이터 활용 관리

+ 정의 및 관리 목적

데이터 활용 관리란 데이터의 활용 여부를 점검하거나 활용도를 높이기 위해 측정 대상 데이터와 품질 지표를 선정하여 품질을 측정하고 분석하여 품질을 충족시키지 못하는 경우, 원인을 분석하여 담당자로 하여금 조치하도록 하는 작업을 말한다. 애플리케이션에서 활용되지 않는 데이터를 점검하여 데이터베이스의 사용 환경을 개선하고 업무적 중요도가 높은 데이터에 대한 품질의 평가와 개선으로 데이터의 활용도를 높인다. 데이터 활용 관리를 통해 데이터의 정확성을 저하시키는 원인을 분석하고 개선함으로써 지속적인 데이터의 품질을 높이고 활용성을 높일 수 있다.

품질 저하를 개선하기 위해서는 조직 간의 원활한 협조 체계와 지속적인 활동이 필요하다. 미활용 테이블/칼럼에 대해 정리를 할 경우, 관련 항목을 사용하는 다른 애플리케이션이 있는지를 철저히 검토해야 한다. 데이터 품질 평가는 한 번의 수행을 통해 이루어지는 것이 아니라 반복적이고 지속적으로 진행되었을 때 고품질의 데이터를 유지할 수 있다.

+ 세부 관리 대상

▣ 핵심 데이터

핵심 데이터는 회사의 고객, 프로세스, 시장 환경, 재무 정보 등에 직접적으로 영향을 미치는 중요성이 높은 데이터를 말하며, 다음과 같은 기준에 따라 관리되어야 한다.

■ 완전성

데이터의 모든 값은 의미 있게 채워져 있어야 한다.

■ 일관성

데이터의 값은 동일하게 관리되어야 한다.

■ 최신성

데이터의 값은 실제 세계의 객체들이 가지고 있는 값과 같아야 한다.

■ 유효성

데이터의 값은 업무 규칙을 준수해야 한다.

■ 유일성

데이터의 값은 동일 테이블에서 중복 관리되어서는 안 된다.

■ 명확성

데이터의 의미가 혼동되지 않도록 분명하게 관리되어야 한다.

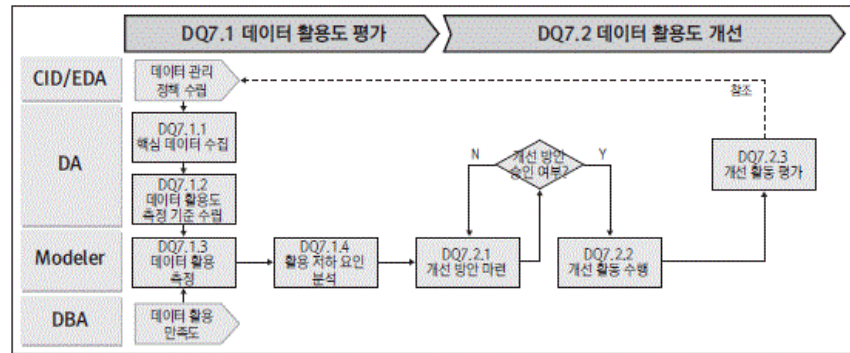
핵심 데이터는 업무 프로세스상의 중요성, 재무적 관점에서 관리의 필요성, 최종적인 사용자의 활용성 등을 기준으로 도출해 해당 테이블의 칼럼 수준으로 관리한다.

▣ 측정 방법

데이터의 업무적인 규칙 및 물리적 특성(도메인, 유효성 등)을 반영한 데이터 품질 측정 기준을 말한다. 업무적인 규칙과 물리적인 특성이 정확하게 반영되도록 하며, 핵심 데이터별로 측정 방법을 관리한다.

+ 데이터 활용 관리 프로세스

▣ 데이터 활용 관리 프로세스



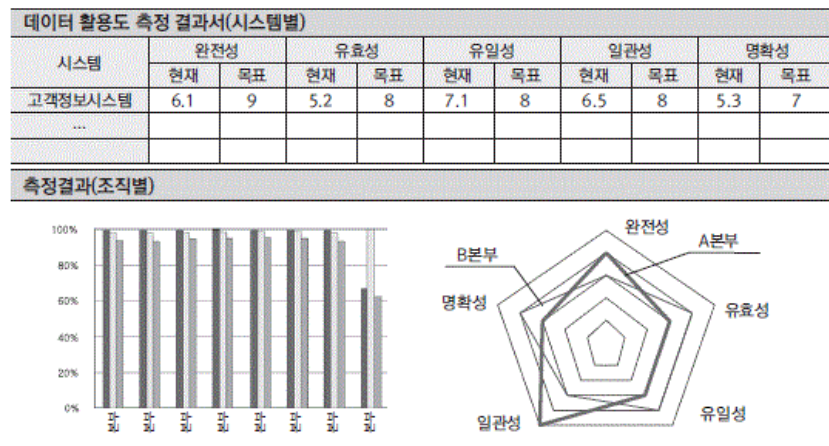
[그림 6-3-25] 데이터 활용 관리 프로세스

■ DQ7.1.1 핵심 데이터 수집

[그림 6-3-25]의 DQ7.1.1 핵심 데이터 수집은 개선 대상이 되는 데이터를 선정 기준에 따라 선정 하고, 업무부하 및 시스템 부하를 고려하여 측정 데이터 량을 조정한다.

■ DQ7.1.2 데이터 활용도 측정 기준 수립

[그림 6-3-25]의 DQ7.1.2 데이터 활용도 측정 기준 수립은 데이터별 활용도 측정 기준을 정량적으로 마련하고 데이터 활용 개선 목표치를 설정하여 향후 개선 작업에 대한 평가 작업 수행 시 활용한다.



[그림 6-3-26] 데이터 활용도 측정 결과서 예

■ DQ7.1.3 데이터 활용 측정

[그림 6-3-25]의 DQ7.1.3 데이터 활용 측정은 데이터 활용도 측정 기준에 따른 활용도 평가 작업을 수행하고, 데이터 활용도 측정 결과서를 작성한다. [그림 6-3-26]과 같은 형식으로 작성할 수 있다.

■ DQ7.1.4 활용 저하 요인 분석

[그림 6-3-25]의 DQ7.1.4 활용 저하 요인 분석은 데이터 활용의 저하를 유발한 비즈니스적, IT 적 원인을 데이터의 생성, 갱신, 변환, 활용 관점에서 도출하고, 데이터 활용 저하 원인 분석서를 작성한다. [그림 6-3-27]과 같은 형식으로 작성할 수 있다.

구분		원인	내용
비즈니스	생성	미입력	의미를 정확히 알 수 없어 입력하지 않음
	...		
	변환	변환오류	변환 규칙의 오류로 잘못된 데이터 생성
IT	...		

[그림 6-3-27] 데이터 활용 저하 원인 분석서 예

■ DQ7.2.1 개선 방안 마련

[그림 6-3-25]의 DQ7.2.1 개선 방안 마련은 활용 저하 원인별로 개선 방안을 마련한다. 개선 방안은 주로 프로세스 개선, 표준화, 클린징의 범주로 분류될 수 있다.

■ DQ7.2.2 개선 활동 수행

[그림 6-3-25]의 DQ7.2.2 개선 활동 수행은 승인된 개선 방안과 원인별로 도출된 개선 방안의 활동 계획에 따라서 개선 활동을 추진한다.

■ DQ7.2.3 개선 활동 평가

[그림 6-3-25]의 DQ7.2.3 개선 활동 평가는 개선 활동을 평가하는 과정으로, 측정 목표치를 초과한 데이터에 대해서는 개선 항목에서 제외시키거나 목표치를 조정한다. 종합적인 수행 결과를 정리하여 향후 활동에 활용할 수 있도록 한다