

# 방대한 DB 길잡이 '메타데이터'

어떤 정보가 어디에 위치하고 그 내용이 어떤 것이며 어떤 기준으로 그러한 자료가 생성되었는가에 대한 정보들이 일관되고 통합되어 제공되지 않는다면 그 데이터 웨어하우스는 몇몇 사람만이 사용하는 전유물이 될 것이다. 데이터 웨어하우스에서 이런 것을 방지하기 위해서는 메타데이터의 역할을 정확히 이해하고 활용해야 한다. 이 글에서는 데이터 웨어하우스 구축시 메타데이터의 역할과 중요성에 대해 살펴보기로 한다.

홍승길/ 한국NCR 과장

우리가 도서관에서 필요한 자료를 찾으려 할 때 가장 많이 참조하는 것이 바로 로비에 비치된 카드 목록(card catalog)들이다. 카드 목록을 참조하면 필요로 하는 책이 도서관의 몇 층에 위치하고 있는지 그리고 그 자료가 몇 년도에 발간되었고 저자가 누구인지를 먼저 알 수 있게 된다.

만약 이러한 카드 목록 시스템이 존재하지 않는다면, 필요한 자료를 찾아내는 데 많은 시간이 소요될 것이고, 사용자들은 이 도서관을 이용하는데 매우 불편함을 느낄 것이다. 데이터 웨어하우스도 마찬가지이다.

아무리 유용하고 도움이 되는 정보들을 잘 꾸며 놓았다 하더라도, 어떤 정보가 어디에 위치하고 그 내용이 어떤 것이며 어떤 기준으로 그러한 자료가 생성되었는가에 대한 정보들이 일관되고 통합되어 제공되지 않는다면 그 데이터 웨어하우스는 몇몇 사람만이 사용하는 전유물이 될 것이다. 도서관에서의 카드목록의 역할을

수행하는 것이 바로 데이터 웨어하우스에서의 메타데이터이다.

이와 같이 메타데이터는 일반적으로 "데이터에 관한 데이터"로 정의된다. 관계형 데이터베이스에서의 메타데이터는 데이터베이스 내에 정의된 객체의 표현들이다. 다른 말로 표현하면, 테이블, 컬럼, 데이터베이스, 뷰 및 기타 객체에 대한 정의들이 바로 메타데이터이다. 데이터 웨어하우징과 관련되어 사용될 때에는, 메타데이터는 데이터 웨어하우스 객체를 정의하는 모든 것 - 테이블, 컬럼, 쿼리, 리포트, 비즈니스 규칙, 데이터 변형 알고리즘 등 - 을 지칭한다.

이러한 정의를 이해하는 것은 데이터 웨어하우스 구축의 전 과정에서 매우 중요하다. 메타데이터의 관리는 운영체 시스템으로부터 데이터를 추출하는 프로그램의 개발에서부터 데이터 웨어하우스로 데이터를 변형하여 적재하는 과정에 이르기까지 매우 견고하게 제어되어야 한다. 정보의 저장소를 활성화시키기 위해 변형

된 데이터가 비즈니스 질문에 정확하게 대답할 수 있어야만 데이터 웨어하우스는 경쟁적 이점을 획득하는데 유용하게 사용될 수 있기 때문이다.

## 메타데이터가 왜 중요한가?

메타데이터는 데이터에 대한 지도(map) 혹은 청사진이다. 도서관의 카드 목록이 도서관 내의 책들의 내용과 위치를 가리키듯이, 메타데이터는 데이터 웨어하우스 내의 다양한 정보를 지닌 객체의 위치와 의미를 가리킨다.

도서관과 마찬가지로, 데이터 웨어하우스는 자신이 관리하는 항목들의 목록을 반드시 관리해야 한다. 데이터 웨어하우스의 최종 사용자들은 마치 도서관의 사용자들과 유사하다 - 사용자들은 목록을 참조하여 작성한 선택 기준에 따라서 정보를 요구하게 된다. 그들의 요구를 만족시켜주는 프로세스는 데이터 웨어하우스 내의 어디에 정보가 저장되어 있는지 반드시 알고 있어야 한다. 그러므로 데이터

웨어하우스는 자신이 관리하는 정보에 대한 목록 기능을 충족시키는 요소 - 즉 메타데이터 - 를 반드시 지녀야 한다.

## 구축 과정에서 메타데이터의 중요성

데이터 웨어하우스 구축 과정은 엔지니어링 프로세스이다. 그러므로 이 과정은 발전적인 품질 향상을 위해서 그리고 재생산성을 제공하기 위해 명시적으로 반드시 문서화 되어야 한다.

이 과정의 결과로 발생된 메타데이터는 그러므로 변경 관리의 목적을 위해서 버전관리가 되어야 한다. 예를 들면, 데이터 웨어하우스 목록의 모든 운영체 스키마, 변형 등을 위해 사용된 비즈니스 규칙과 프로세스 규칙들은 반드시 포착되고 버전관리가 되어야 한다.

데이터 웨어하우스를 위한 메타데이터의 설계는 운영체 시스템을 지원하기 위한 데이터베이스의 분석 및 설계와는 크게 다르다. 운영체 데이터베이스 설계 과정의 초점은 atomic 데이터로 속성이 주어진 정규화 된 데이터 모델을 생성하는 것이다.

지금까지의 가장 주요한 관심은 정규화를 사용한 데이터 중복성의 제거였다. 데이터 중복성의 제거의 목적은 데이터 업데이트의 문제를 방지하고 데이터의 일치성을 유지하기 위함이었다. 데이터 웨어하우스를 위한 메타데이터의 설계상의 초점은 종종 상당한 중복성과 함께 수 많은 관계를 사용

자에게 제시하는 것이다.

정보 분석가에게 있어서 또 다른 변화는 운영체 시스템이 지니는 현재의 메타데이터에 대한 중요성이다. 대부분의 운영체 시스템은 데이터베이스 및 데이터의 현재의 구조에 대해서 운영된다. 오래된 데이터는 오래된 데이터의 구조와 함께 보존된다. 데이터 웨어하우스 내에서는 기존(historical) 데이터를 추출하기 위해서 메타데이터가 사용되어야 한다. 그러므로 운영체 데이터베이스의 메타데이터가 현재부터 제대로 관리되어야 한다.

메타데이터의 중요한 측면은 추출, 정제 및 리엔지니어링 과정을 통해서 소스

로부터 데이터 웨어하우스로의 모든 방향으로 매핑을 유지해야 하는 필요성이다. 이러한 매핑은 다음의 목적을 위해 유지되어야 한다:

- ▲ 데이터 품질의 확인
- ▲ 동기화 및 갱신
- ▲ 통합

데이터 웨어하우스 구축과정 중에서 OLAP을 적용하거나 마이닝 과정을 거치면서도 메타데이터는 생성된다. 그러나 데이터 웨어하우스의 구축과정 중에서 데이터의 이동 및 변환이 가장 많이 발생하

## 메타데이터란 무엇인가?

메타데이터는 다음과 같은 객체의 유형들로 구성된다.

- 데이터 웨어하우스 서버, 데이터베이스, 테이블 등의 위치와 설명
- 상품, 시장 및 회계전표와 같은 business dimension hierarchy에 걸쳐있는 규칙 및 그것들에 대해 자동적인 drill-up/drill-down을 가능케 만드는 규칙
- 사실(Fact)에 대해서 기술적인 측면에서 명명된 데이터의 명칭과, 최종 사용자의 측면에서 정의한 이름 혹은 alias
- 최종 사용자가 직접 정의한 연산에 대한 규칙
- (최종 사용자가) 직접 정의한 aggregation, 연산 및 최종 사용자가 작성한 분석을 조회하고 변경하며 분배하기 위한 개인, 워크그룹 및 전사적인 차원의 보안
- 원본 소스와 변형에 대한 설명
- 데이터 웨어하우스의 테이블과 속성에 대한 논리적인 정의
- 테이블과 컬럼에 대한 물리적 정의 및 그들의 성격

- 복수의 운영체 소스로부터의 선택 논리
- 데이터 웨어하우스 테이블 상호간의 매핑
- 추출 이력
- Alias 정보
- Aggregation 알고리즘
- 주제 영역의 위치
- Relationship의 이력
- 소유권한/관리권한
- 접근 유형
- 참조되는 테이블 및 암호화 된 데이터
- (데이터 등의) 노후화 및 삭제의 근거
- 데이터 품질 표시자
- 보안
- 측정의 단위
- 운영체 소스의 신분
- 속성 대 속성의 간단한 매핑
- 물리적인 성격의 변환
- 암호화 및 참조 테이블의 변환
- 이름의 변경
- Key의 변경
- 사용되는 기본설정값 및 기본설정의 근거

는 과정이 바로 데이터 트랜스포메이션이다. 그러므로 데이터 트랜스포메이션의 과정 별로 메타데이터의 필요성과 역할을 살펴보면 다음과 같다.

■ Data Source 추출

메타데이터는 복수개의 소스로부터 데이터 웨어하우스로 운영체 데이터를 통합하는 작업에 있어서 핵심 요소이다. 메타데이터는 다음과 같은 데이터 추출과정에서 필요로 된다.

- ▲ 소스 필드를 확인
- ▲ 데이터의 이력에서 데이터 구조의 변경을 추적
- ▲ 고의로 혹은 실수로 입력되지 않은 데이터 필드에 대해 기본설정값을 자동적으로 적용
- ▲ 속성간의 매핑
- ▲ 속성 변환
- ▲ 소스 및 타겟 스키마에 대한 완전한 설명
- ▲ 항목별 소스에서 타겟으로의 매핑
- ▲ 트랜스포메이션 과정 중에 정의된 변형 변수
- ▲ 데이터 변형 규칙
- ▲ 데이터 추출 옵션
- ▲ 예외 처리 옵션
- ▲ 변형 단계 및 리포트

■ Data 정제 및 리엔지니어링

데이터 트랜스포메이션에서 데이터 정제 및 리엔지니어링 단계는 다양한 소스 시스템으로부터의 데이터를 깨끗이 해야

하는 책임을 진다. 이 단계에서의 메타데이터는 다음의 행위들을 추적해야 한다.

- ▲ 통합 및 분배
- ▲ Aggregation
- ▲ 사전연산 및 도출
- ▲ 변형 및 재매핑

■ 메타데이터의 끊임없는 저장과 관리

데이터 웨어하우스가 구축되고 난 후, 최종 사용자에게 메타데이터는 지도의 역할을 수행한다. 실제로 데이터 웨어하우스 개발 과정에서 발생되고 포착된 모든 메타데이터가 데이터 웨어하우스 사용자에게 유용하거나 관심거리가 되는 것은 아니다.

메타데이터 중에서 어떤 정보가 가용하고 접근 가능한지를 정의한 메타데이터만이 주요 관심거리가 된다. 사용자들은 또한 질의(query)의 예상 실행 시간을 받아보는 것에도 관심을 가진다. 그러므로 데이터 웨어하우스 구축 과정에서 발생한 메타데이터는 적절하게 저장되고 편성되어야 한다. 이러한 메타데이터의 저장과 관리 방법에는 크게 두 가지가 존재한다.

■ 데이터 웨어하우스 전용 메타데이터 디렉토리

이러한 유형은 오직 데이터 웨어하우스 어플리케이션만이 사용할 수 있는 메타데이터를 관리한다.

■ 전사 메타데이터 디렉토리

전사 메타데이터 디렉토리는 메타데이터를 저장, 분류 및 관리하는데 일반적으로 사용되는 범용의 리포지토리이다. 이 리포지토리는 데이터 웨어하우스를 위한 메타데이터 뿐만 아니라 그 외에 모든 메타데이터를 위해 사용된다.

이러한 메타데이터의 관리는 제품을 활용하거나 혹은 자체적으로 개발할 수도 있다. 메타데이터 관리의 중요성이 데이터 웨어하우스의 개발 및 구축에서 주요한 역할을 수행하므로 각 업체에서 메타데이터 관리의 주요한 측면을 지원하는 통합된 정보 디렉토리 서비스를 제공하고 있다. 그러나 현재 이러한 메타데이터의 관리에 대한 표준화는 아직 정립되어 있지 않은 실정이다. 그리하여 톨간의 메타데이터의 교환 및 관리에 아직은 어려움이 따른다.

이렇게 데이터 웨어하우스에서 메타데이터 관리의 역할에 대한 관심이 증가하면서 메타데이터 관리 제품이 상호간에 메타데이터 정보를 교환하도록 허용하는 메타데이터 표준이 필요하게 되었다. 메타데이터표준화에 있어서 다음의 영역에서 주요한 도전에 직면하고 있다.

■ 메타데이터 관리

메타데이터의 정의를 표준화하는 것은 이해하기 쉽고 의사소통이 잘 되는 메타데이터 요소들에 대한 명명, 데이터 type 및 길이의 표준화 그리고 의미가 풍부하고 설명적인 glossary를 유지하기 위한 프로세스를 구축하는 것이다.

■ 메타데이터 표현 및 분류

표현 및 구분 스키마의 주 목적은 메타

(표) 메타데이터 관련 표준화 업체 및 기관

메타데이터 모델		메타데이터 교환 방법	
표준	주도 기관	표준	주도 기관
MOF	OMG	XMI	OMG
MSR/OIM	마이크로소프트	MDIS	MDC

데이터를 기술에 기반한 class로 분류하는 것이다. 기업에서 수 많은 저장 기술들이 사용되므로, 이 분류는 매우 다양할 수 있다. 분류 스키마의 일정 부분은 메타데이터의 상이한 class들간의 관계를 표현하고 관리하는데 필요로 된다.

## 메타데이터의 표준화 동향

메타데이터와 관련된 표준화 작업들은 현재 다양한 형태로 진행중이다. 메타데이터를 위해 사용되는 기술들의 표준화는 MDC(MDIS), OMG(MOF, UML, XMI), W3C(XML), EIA(CDIF)등의 기관들이 주도하고 있고, 업체로는 마이크로소프트, IBM, 오라클, 유니시스 등이 활발하게 활동하고 있다. 데이터 웨어하우스와 관련된 메타데이터의 표준화는 메타데이터 모델 부분과 메타데이터 교환 방법(Exchange Method)부분으로 구분된다

### ■ MOF(Meta Object Facility)

MOF는 OMG가 채택한 메타 모델을 정의하기 위한 사양이다. 현재는 적은 수의 제품만이 지원하고 있으나, 점차 영향력을 확장하는 추세이다. MOF는 분산 환경 하에서 MOF에 호환되는 모든 메타 모델을 위한 OMG의 메타데이터를 관리하기 위한 리포지토리 API이다. 이 API는 CORBA의 IDL을 사용하며 MOF는 CORBA, COM, 자바 컴포넌트를 지원한다. MOF에서도 데이터 웨어하우스를 위한 모델 작업을 진행하고 있다.

### ■ MRS/OIM(Microsoft Repository/Open Interface Model)

마이크로소프트 리포지토리를 위한 메

타 모델이며 마이크로소프트를 위시한 여러 업체의 의해 작성되었다. COM 인터페이스와의 연속성을 위해 UML을 사용하며 데이터 웨어하우징 및 어플리케이션을 위한 도메인을 포함한다. 마이크로소프트는 OIM에 대한 제어를 MDIS에 일임하였으며 향후 MDIS와의 긴밀한 통합이 예상된다.

### ■ XMI(XML Metadata Interchange)

1998년도에 OMG로 제안되었으며 객체에 대한 정보의 교환을 위해 XML을 사용한다. 메타데이터 모델을 위해서 MOF를 사용하며 IBM, 오라클, 유니시스 등에 의해 주도된다. 1999년도 초반에 OMG에 의해 수용될 것으로 보인다.

### ■ MDIS(Metadata Interchange Specification)

MDIS는 메타데이터를 ASCII 형태와 배치 타입으로 교환이 가능하도록 지원한다. 이런 단순한 구조를 채택한 이유는 업체들이 단기간 내에 용이하게 구현하도록 지원하기 위해서이다.

MDC에는 ETI, 하이퍼리온 솔루션즈(Hyperion Solutions), IBM, 플래티늄(Platinum), 코그노스(Cognos), SAS, 사이베이스, NCR 등 50여개 업체가 참여하고 있고 마이크로소프트가 1998년 12월 정식 회원으로 가입함으로써 메타데이터 교환 방법 분야에서 표준화를 보다 역동적으로 추진하고 있다.

현재 어퍼투스/칼레튼(Apertus/Carleton), ETI, IBM, 원미닝(One-Meaning), 플래티늄/로직웍스(Platinum/LogicWorks), 그리고 사이베이스/인텔텍스(Sybase/Intellidex) 등이 제품에 MDIS 표준을 구현하였다. MDIS는 컨버전 언어를 E/R에 기반을 두고 있기 때문에 객체에 대한 지원이 미약하며 업체마다 구현에 약간의 불일치가 존재한다.

또한 기존에 치중했던 기술적인 메타데이터를 넘어서서 비즈니스 메타데이터(Business Rule)의 서술 방법에 대한 표준도 정의하고 있다. 그래서 MDIS는 XML 기반의 'Tag Language'를 지원하는 방향으로 움직이고 있다. ☞

**사무실 이전 안내**

# 다우기술

- 이전일자: 1999년 3월 7일(일요일)
- 이전장소: 서울특별시 강남구 대치동 1002번지  
코스모타워 14~15층
- 전화 및 팩스번호: 종전과 동일
- 대표전화: 02)3450-4500 FAX: 02)552-7022