
클라우드 환경에서 데이터 통합 관리를 위한 TMDM

문석재* · 신호영** · 정계동*

*광운대학교 **경북대학교

TMDM for Data Integration Management in Cloud Environment

Seok-jae Moon* · Hyo-young Shin* · Gye-dong Jeong**

*KwangWoon University, **KyungBok University

E-mail : msj8086@nate.com

요 약

클라우드 환경에서 기업들은 상호 연결되지 않은 여러 개의 시스템과 데이터베이스에 각각 마스터 정보를 분산 저장하여 사용하고 있다. 관리되지 않은 마스터 정보는 부정확하고 상호 불일치하기에 비즈니스 프로세스의 효율성을 저하시키고 최적의 의사결정을 할 수가 없게 된다. 효율적이고 오류 없는 비즈니스 프로세스 운용을 위해서는 고품질의 마스터 정보의 관리가 필요하다. 본 논문은 클라우드 환경에서 상호 연계되는 마스터 정보 간의 발생하는 이질적인 문제를 해결하고 비즈니스 프로세스를 효율적으로 운용하기 위한 방안으로 TMDM(Topic Maps Master Data Management)을 제안한다. TMDM는 데이터 간의 연관성을 고려한 Topic Maps를 이용하여 마스터 정보 간의 상호 불일치 문제를 해결하기 위해 제안된 지식 저장소이다. Topic Maps는 하나의 토픽을 통해 토픽이 표현하는 주제에 관련된 모든 지식 정보를 접근할 수 있도록 토픽간의 association을 통해 연결할 수 있다. 이러한 점은 클라우드 내에서 레거시 시스템 간 마스터 정보에도 적용할 수 있다.

ABSTRACT

In cloud environment, enterprises use a number of systems which are not interconnected and save master information in a distributed way in each of them. Master information which is not managed is incorrect and discord each other so that lowers efficiency of business process and disables optimum decision making. It is necessary to do high-qualified management of master information to operate efficient and errorless business process. This paper suggests TMDM as a solution to solve heterogeneous problems occurred between interconnected master information in cloud environment and to manage business process in an efficient method. TMDM is an information storage that is suggested to solve mutual discord problems between master information using Topic Maps that considers correlation between data. Topic Maps can be connected by association between topics in order to access through a topic to all related knowledge information which is described by the topic. This also can be applied to master information between legacy systems within cloud.

키워드

클라우드, 마스터 데이터, 토픽맵, XMDR

1. 서 론

클라우드 환경에서 기업이 관리하고 있는 마스터 정보는 대부분 여러 시스템에 분산 저장하여 관리되고 있다[1]. 마스터 정보의 관리는 비즈니스 프로세스의 효율적인 실행이 주된 목적이다. 그러나 관리되지 않은 마스터 정보는 상호 운용

상에서의 부정확한 정보를 제공하게 되어 비즈니스 프로세스의 운용이 불완전해지며 최적의 의사결정을 할 수 없게 된다. 따라서 클라우드 환경에서 기업들은 효율적이고 오류 없는 비즈니스 프로세스 실행을 위해 고품질의 마스터 정보 관리 및 연결은 필수 요건이다. 특히, ERP와 같은 업무코드의 새로운 체계, 프로세스 역할 정립이 필

요한 프로젝트 등에서 데이터 표준화(데이터 검증, 정제, 보강), 데이터 추출, 정제, 변환에서 마스터 정보 통합 방안이 필요하다.

최근 MDM(Matster Data Management)은 비즈니스의 수행과 분석의 기반이 되는 마스터 정보를 전사차원에서 통합적으로 관리함으로써 일관성(consistent) 있고 정확하게(accurate) 생성하고 유지하기 위한 원칙, 프로세스, 도구, 솔루션의 총합이다. 이 MDM은 기업의 합병이나 협업에 의한 새로운 시스템 구축 및 추가과정에서 마스터 정보의 상호운용성의 여러 가지 문제점들이 있다.

본 논문에서는 클라우드 환경에서 협업에 의한 시스템 통합하는 과정에서 비즈니스 프로세스를 위한 마스터 정보 간의 상호 연결할 때 발생하는 데이터 충돌 즉, 의미적, 논리적 구조 충돌을 해결할 수 있는 TMDM(Topic Master Data Management)을 제안한다. 클라우드 환경에서 분산된 마스터 정보들은 일반적으로 정보를 저장하기 위한 메타데이터 스키마와 실데이터(인스턴스 데이터)로 구성되는데, 이를 한 번에 상호 연결할 수는 없다. 그래서 메타데이터 스키마를 결합하여 통합 형태의 메타데이터 스키마를 만들고 이 통합 메타데이터에 실데이터(인스턴스 데이터)들 간의 충돌 이질성을 해결해서 연결시켜야 한다. 이렇게 상호 연결된 마스터 정보는 XMDR[1]과 같은 방법으로 글로벌 마스터 정보를 생성하여 로컬 마스터 정보와 매핑하여 글로벌 마스터 정보만으로 로컬 마스터 정보를 접근할 수 있도록 한다. 그리고 마스터 정보내의 실데이터들 간의 충돌은 Topic Maps에서 토픽과 연관관계(association)를 표현한 것과 같이 마스터 정보 내부에 메타데이터와 데이터를 표현하는 각 토픽 클래스와 토픽으로 등록하고, 메타데이터 사이의 관계, 데이터 사이의 관계 그리고 메타데이터와 실데이터 사이의 관계와 같은 연관관계를 제공한다. 이는 데이터의 의미 및 연관성을 기술할 수 있고, 이를 통해 다양한 형태의 의미충돌을 해결하여 비즈니스 프로세스의 수행과 분석의 기반이 되는 마스터 정보를 전사차원에서 통합적으로 관리할 수 있다. 본 논문의 구성은 2장에서는 Topic Maps Master Data에 대해서 기술하고, 3장에서는 TMDM이 적용된 Prototype 프레임워크를 기술한다. 그리고 마지막 5장에서 결론을 기술한다.

II. Topic Maps Master Data

TMDM은 Topic Maps와 MD를 결합한 것이다. Topic Maps는 온톨로지 역할을 수행하고, MD는 (Master Data Registry)MDR 기반으로 로컬 시스템 간의 마스터 정보 충돌을 해결하기 위한 매핑을 담당한다. 이 장은 Topic Maps과 MD를 적용하는 방법과 둘을 결합하는 방법에 대해 기술한다.

2.1 온톨로지 기반의 Topic Maps

Topic Maps[2]는 지식간의 관계성을 정의하여 데이터 간 충돌 문제를 해결 할 수 있다. 데이터 충돌은 다양한 유형이 있다. 충돌 유형에는 데이터 사이에서 발생하는 충돌, 데이터를 저장하기 위한 메타데이터 스키마에서 발생하는 충돌로 구분할 수 있다. 본 논문에서의 사용되는 Topic Maps는 마스터 정보간이 비즈니스 프로세스 수행 시 데이터 간의 충돌을 해결하는데 주목적을 둔다. 충돌 유형은 value, format, unit, precision, domain 이 있다[3]. Topic Maps의 토픽과 토픽들 간의 연관은 그림 1과 같이 표현된다. 예를 들어 상품 판매사에서 관리자가 구매자가 신청한 수량에 대한 재고 수량을 확인한다고 하자. 관리자는 검색 조건이 “키워드가 상품정보(상품명)을 포함하는 현재 구매 상품의 회사, 재고 수량을 검색”이라고 한다. Topic Maps에서 keyword가 Product인 토픽을 찾고, 이와 관련된 상품의 회사, 수량, 그리고 구매 신청자, 등과 같은 연관 정보를 찾는다. 이런 연관성을 검색의 조건에 포함시킴으로써 검색의 정확성을 향상시킬 수 있게 된다. 이 예제에 대한 Topic Maps 관계를 Protege를 이용하여 표현한 것이 그림 1과 같다.

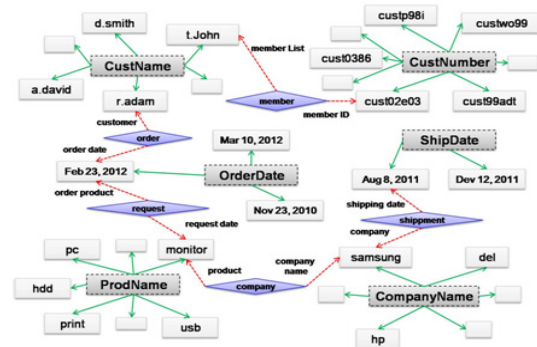


그림1. Topic Maps의 Topic과 Association

본 논문에서 제안하는 토픽맵에서의 토픽 클래스는 마스터 정보에서 정의된 글로벌 메타데이터 스키마를 토픽으로 사용한다. 그리고 인스턴스 토픽은 실데이터에 대한 토픽으로 원래 Topic Maps에서 제시하는 토픽으로 정의한다. 토픽들은 다양한 association이 존재한다. 고객명(CustName)과 상품명(ProdName)라는 토픽이 있을 때, 이 두 토픽 사이에는 고객이 주문한 상품이라는 order 연관성을 갖는다. Association의 구성은 역할(role)과 멤버(member)로 구성된다. 역할은 토픽들 간의 연관성을 나타낸 것이고, 멤버는 이 역할에 해당하는 실제 토픽이다. 예를 들어 “company”라는 연관성의 역할은 ‘product’와 ‘company name’이 있고, 멤버는 ‘product’ 역할에 해당하는 상품과 ‘company name’ 역할에 해당하는 samsung이라는 회사명이 된다. 여기서 실선은 토픽 간의 관계를 표현한 것으로 상위토픽(super topic)으로부터 하위토픽(sub topic)으로 상속 관계에 해당한다. 상속 관계는 두 가지 표현된다. 첫

째는 메타데이터 스키마 토픽간의 관계로 is-a, part-of, equal 관계로 구분할 수 있다. 둘째는 메타데이터 스키마 토픽과 인스턴스 토픽 사이의 관계를 표현한 것이다. 화살표 상에서 화살표의 시작은 상위토픽이고 화살표의 끝은 하위 토픽이 된다. 점선은 연관성(association)에 대한 멤버(member)를 지칭한다. 다시 말해 연관성에 관여하는 토픽들임을 알 수 있다. 또한 이러한 연관성 관계를 통해 토픽들 간의 계층적 접근뿐 아니라 계층 간의 수평 접근도 가능해진다.

본 논문의 Topic Maps을 구성하는 릴레이션은 다음과 같이 6가지 유형[2]으로 정리한다. 그리고 Topic Maps는 TM으로 표현한다.

TM = {T_s, T_i, S_a, I_a, R_a, O_i} - 정리

- T_s는 메타데이터 스키마 토픽을 관리하는 위한 릴레이션이다. 이것은 T_i를 표현하기 위한 개념토픽 역할로 상위 토픽클래스가 된다.
 - T_i는 인스턴스 토픽을 표현한 릴레이션이다 이는 인스턴스를 구성하는 값들의 부분집합이다
 - S_a는 로컬 시스템을 구성하는 메타데이터 스키마 토픽(T_s)간의 관계를 표현한 릴레이션이다 이 관계는 equal, is-a, part-of 관계로 구성되며, (α, β, γ) 형식으로 표현된다. {α, β} ∈ T_s이며 γ = {equal, is-a, part-of}이다. equal은 두 개의 T_s간의 동일한 관계를 의미한다 (X, Y, equal)는 X=Y로 표현한다. is-a는 T_s간의 상속관계를 의미한다 (X, Y, is-a)는 “X는 Y로 한정(specialization)된다.” 또는, “Y는 X로 일반화(generalization)된다.”는 것으로 X≧Y와 같이 표현된다. part-of는 T_s간의 포함관계를 의미한다 (X, Y, part-of)는 “Y는 X에 속한다(belong to)”는 의미를 가지고 X≧Y와 같이 집합기호로 표현된다.
 - I_a는 T_i간의 연관관계(association)를 표현하기 위한 릴레이션이다. 이는 연관관계의 구성요소인 역할(role) 타입과 멤버로 구성되고, 이를 통해 지식 계층의 수평적 지식검색이 가능해진다. 연관관계의 유형은 S_a의 관계유형과 동일하다. 수평적 지식 검색이란 다른 서브 트리에 있는 토픽을 일반적인 트리 운행 방식을 통해 접근하는 것이 아니라 연관관계를 통해 직접 접근하도록 한다는 것이다.
 - R_a는 T_s와 T_i간에서 발생할 수 있는 관계를 표현하는 릴레이션이다. 이는 조건에서 사용하는 관계가 된다. 질의문에서 사용되는 조건은 다음과 같다. field-name1 operation {field-name2 | values}에서 field-name1과 values를 이용할 경우 사용된다.
 - O_i는 토픽의 어커런스를 표현하는 릴레이션이다. 어커런스는 Topic Maps의 토픽과 실제 정보 자원을 연결하기 위한 정보를 관리하는 것으로 단순한 링크 정보라고 보기는 어렵다.
- 본 논문에서는 마스터 정보에서 로컬 시스템 접근과 관련된 정보를 관리하는 ML을 이용한다.

2.2 XMDR을 이용한 Master Data

XMDR은 데이터의 상호운용성을 확보하기 위한 목적으로 제안되었다[4]. XMDR은 메타데이터 등록과 인증과정으로 표준화된 메타데이터를 유지 관리하며, 메타데이터의 명세와 의미 공유를 이용하여 메타데이터 요소들 간의 접근성을 유지시킨다. 또한 이 기종 데이터베이스 시스템간의 공유되는 개념을 온톨로지로 해결한다. 이 온톨로지는 여러 로컬 시스템이 운용하고 있는 메타데이터 스키마들 사이에 발생하는 충돌들을 해결하는데 쓰인다.

표 1. 메타데이터 스키마 충돌 유형

종류	설명
이름 충돌	스키마(엔티티, 속성, 관계)의 동음이의어, 동의어에 따른 문제
식별자 충돌	동일 데이터 집합을 가지는 개체에서 서로 다른 식별자(key) 할당
구조적 충돌	동일 데이터 집합을 가지는 개체에서 서로 다른 구조를 가지는 문제
모델링 충돌	동일 데이터에 대한 데이터 모델의 차이에서 발생하는 문제
제약조건 충돌	동일 속성이 가지는 제약조건에서 발생하는 문제
데이터 타입 충돌	동일 속성에 대한 데이터 타입 간의 문제

XMDR은 표1과 같은 충돌 문제를 해결하기 위한 것이며, MSO(Meta Semantic Ontology)와 MLO(Meta Location Ontology)으로 구성된다[5]. MSO는 글로벌 메타데이터 스키마와 로컬 메타데이터 스키마 사이에 매핑정보를 정의한 온톨로지이고 ML은 로컬 시스템의 접근 정보를 정의한 것이다.

● **Meta-Semantic Ontology Area**

MSO는 메타데이터 스키마의 구조적 의미적 충돌을 해결한 온톨로지이다. MSO는 글로벌 메타데이터 스키마를 구성하는 표준 항목을 결정한다. MSO는 표준 항목과 로컬 시스템의 메타데이터 스키마 정보 간의 매핑정보, 메타데이터 스키마 간의 이질적 관계에 대한 솔루션 정보 데이터타입 간의 변환정보를 세부적으로 관리한다. 여기서 표준 항목이란 로컬 시스템의 로컬 메타데이터 스키마의 대표 스키마로 실제로 존재하는 스키마가 아니라 통합을 위한 가상의 메타데이터 스키마이다. 이것은 사용자에게 제공되는 인터페이스를 구성하기 위한 항목이며 로컬 시스템 사이에 데이터를 이동하기 위한 기준이다

● **Meta-Location Ontology Area**

로컬 시스템에 접근하여 검색하기 위한 위치정보, 권한정보, 각 로컬 시스템에 적합한 질의로 변환에 사용하게 될 데이터베이스 정보와 개체 정보 그리고 검색을 위한 우선순위 정보를 관리하고 제공하는 역할을 수행한다. 이것은 사용자가 실제 데이터의 위치를 몰라도 접근할 수 위치투

명성을 제공한다. Meta Location Ontology의 표현은 관계속성과 지역속성으로 구분한다. 관계속성은 표준 온톨로지와 메타데이터 스키마와 지역속성 연결을 위한 것이다. 지역속성은 로컬 시스템에 접근하기 위해서 사용되는 권한 이름에 대한 속성, 그리고 위치 정보 속성을 정의한 것이다. 이처럼 정의된 정의된 속성들에 의해서 표준 온톨로지에 표현된 데이터의 표준과 관계성과 매핑을 통해 각 로컬 시스템의 MDR에 접근할 수 있는 방법을 제공하게 된다.

III. 적용 프레임워크

본 논문의 적용 프레임워크는 이질성에 의해 발생하는 여러 문제 때문에 실시간 통합이 불가능한 이종 데이터베이스를 통합하고 의미적인 상호 운용성을 제공할 수 있어야 한다. 발생하는 문제점은 데이터의 값, 입자성(granularity), 표현 방식, 형식, 포맷, 단위 등과 같은 값의 표현에 따른 차이와 스키마의 차이 때문에 발생한다. 이 프레임워크는 시맨틱 에이전트를 통해 여러 분산 시스템의 데이터 간에 존재하는 의미상의 충돌을 실시간 탐지하고 해결을 제공함으로써 분산된 데이터, 이질적인 데이터에 대한 지식이 없는 사용자에게 분산된 시스템에 접근을 용이하게 하고 wrapper 에이전트를 통해 로컬 시스템에 접근한다. 이에 대한 전반적인 시스템 흐름은 그림과 같다.

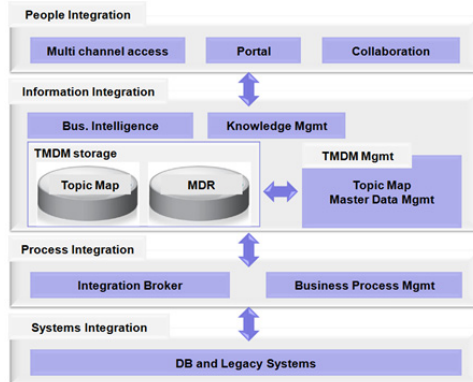


그림 2. 적용 프레임워크

- People Integration: 사용자들은 이 계층을 통해 TMDM에 의해서 통합된 마스터 정보들을 접근하거나 데이터 조회를 위한 Portal 인터페이스를 제공하고, 협업에 의해 전사적 데이터를 통합 할 수 있다.
- Information Integration: 이 계층은 비즈니스 프로세스를 효율적으로 운용할 수 있도록 하기 위한 것으로, TMDM은 데이터 간의 연관성을 고려한 Topic Maps를 이용하여 마스터 정보간의 발생하는 상호 불일치 문제를 해결하기 위해 제안된 지식 저장소이다.
- Process Integration: 이 계층은 비즈니스 프로세

스 관리 하는 계층으로 프로세스 간의 데이터 교환을 연결해주는 계층이다.

- Systems Integration: 이 계층은 가장 마지막 하위 계층으로 데이터 소스 즉, 클라우드 환경에서 마스터 정보에 의해 운용되고 있는 데이터베이스와 레거시 시스템들이다

그림 3은 사용자가 통합 검색을 하기 위한 인터페이스로 웹 기반이다

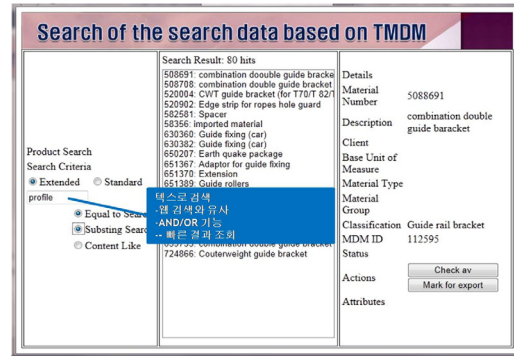


그림 3. TMDM 기반의 데이터 검색 인터페이스

IV. 결론

본 논문에서는 클라우드 환경에서 기업간 정보 연계되는 마스터 정보에서 발생하는 이질적인 문제를 해결하였고, 비즈니스 프로세스를 효율적으로 운용하기 위한 방안으로 TMDM을 제안하였다. TMDM은 데이터 간의 연관성을 고려한 Topic Maps를 이용하여 마스터 정보 간의 상호 불일치 문제를 해결하기 위해 제안된 지식 정보를 제공하였다. TMDM을 이용한 데이터 통합 관리 프레임워크는 중앙의 TMD 저장소와 로컬 비즈니스 어플리케이션 간 정교 교환 지원 원격 시스템으로부터 데이터 추출 및 동기화 지원 마스터 데이터 접근을 위한 API 및 웹 서비스를 제공한다. 그리고 마스터데이터 처리를 위한 관리 도구 마스터데이터 규칙 및 처리를 위한 기능을 제공한다

참고문헌

- [1] 이미영, 클라우드 기반 대규모 데이터 처리 및 관리 기술, 전자통신동향분석, 제24권, 제4호, 2009.
- [2] <http://www.isoTopicMaps.org/> ISO/IEC 13250 Topic Maps (Second Edition) 19 May 2002.
- [3] KyeDong Jung, ChiGon Hwang, Byungkook Jeon, YounKenu Choi, "TMDR-Based Semantic Agent System for Distributed Data Integration", CCA 2012.
- [4] ISO/IEC 11179 eXtended Metadata Registry(XMDR), part 1-6.
- [5] SeokJae Moon, GyeDong Jung, YoungKeun Choi, "XMDR-DAI Based on GQBP and LQBP for Business Process", AST2010, CCIS 74, 2010.