
XMDR을 이용한 데이터웨어하우스 실시간 데이터 정제 시스템 설계

송홍율* · 정계동* · 최영근*

Design of Datawarehouse Real-Time Cleansing System using XMDR

Hong-Youl Song* · Kye-Dong Jung* · Young-keum Choi*

본 논문은 2010년도 광운대학교 교내 연구비 지원에 의해 연구되었음.

요 약

데이터웨어하우스는 기업에서 의사결정이나 기업의 정책을 결정하는데 사용하고 있다. 그러나 분산 환경에서 새로운 시스템이 추가되면 데이터 통합 측면에서 시스템간의 여러 가지 이질적인 특성으로 인해 많은 비용과 시간이 필요로 하게 된다. 따라서 이러한 이질적인 특성을 해결하기 위해 첫째, 데이터 구조의 이질성은 표준기관에서 제정한 표준스키마와 XMDR(eXtended Master Data Registry)를 이용하여 추상화된 쿼리를 생성하고, XMDR에 맞게 쿼리를 분리함으로써 구조적인 이질성을 해결한다. 둘째, 데이터 정의 및 표현의 이질성은 메타데이터에 대한 유사어와 데이터 값의 표현 방식을 정의한 메타데이터 사전을 이용함으로써 해결한다. 특히 본 논문에서는 XMDR을 이용하여 분산 시스템 통합시 로컬시스템의 영향을 최소화하고, 데이터웨어하우스의 정보를 실시간으로 생성하기 위해 분산된 환경에서 데이터 통합을 위한 표준화된 정보를 제공한다.

ABSTRACT

A datawarehouse is generally used in organizations for decision and policy making. And In a distribute environment when a new system is added, there needs considerable amount of time and cost due to the difference between the systems. Therefore, to solve this matter. Firstly, heterogeneous data structures can be handled by creating abstract queries according to the standard schema and by separating the queries using XMDR. Secondly, metadata dictionary which defines synonyms of metadata and methods for data expression is used to overcome difference of definition and expression of data. Especially, work presented in this thesis provides standardized information for data integration and minimizing the effects of integration on local systems in discrete environments using XMDR to create information of data warehouse in realtime.

키워드

XMDR, ETT, 온톨로지, 데이터웨어하우스

Key word

XMDR(eXtended Meta Data Registry), ETT, Ontology, Datawarehouse

I. 서론

최근 기업에서 의사결정이나 기업의 정책을 결정하는데 보다 적극적으로 활용하기 위한 방안으로 데이터웨어하우스가 제시되고 있다[1]. 데이터웨어하우스는 데이터를 정보로 활용하기 위해 로컬의 통합된 데이터로부터 시공간의 개념을 가진 새로운 정보를 생성하는 것으로 로컬 시스템의 데이터베이스 구조와 데이터웨어하우스의 데이터베이스 구조는 근본적으로 차이가 발생한다는 것을 고려해야 한다[2]. 본 논문에서는 XMDR을 이용한 데이터웨어하우스 실시간 정제 시스템을 제안한다. 이 시스템은 MD(Master Data) 시스템의 확장된 기능으로 XMDR을 이용하여, 데이터의 이질성을 해결한다. 또한, 일관된 기본 정보 관리로 각 로컬시스템으로부터 실시간 데이터를 추출 가공하여 데이터웨어하우스에서 최신 정보를 사용할 수 있는 방안을 제시한다. 제안된 시스템의 목적은 XMDR을 이용하여 분산 시스템 통합 시 로컬시스템의 영향을 최소화하고, 데이터웨어하우스의 정보를 실시간으로 생성하기 위해 분산된 환경에서 데이터 통합을 위한 표준화된 정보를 제공한다. 본 논문의 구성은 다음과 같다. 2장은 관련연구로 본 논문의 기반이 되는 데이터웨어하우스에 대하여 기술하고, 정제시스템의 근간이 되는 XMDR 구조에 대해 알아본다. 3장은 본 논문의 정제시스템에 대한 설계 방안을 기술한다. 4장에서는 시스템 적용 예와 비교 분석을 기술한다. 5장에서는 결론 및 향후연구과제에 대해 기술한다.

II. 관련연구

2.1 ETT

ETT(Extract Transformation Transportation)는 데이터를 소스시스템에서 추출하여 데이터웨어하우스에 로드시킨 상태에서 정제작업까지 이르는 전 과정을 말하는 것으로 소스 데이터베이스인 운영시스템의 데이터를 변환, 정제하여 데이터웨어하우스에 적재하는 단계들이다[3]. 이 작업을 위해서는 우선 데이터웨어하우스와 소스데이터베이스의 데이터를 대응(mapping)한 매핑표가 필요하다. 매핑 표에는 소스데이터베이스의 데이터 구

조와 변환, 정제 알고리즘에 대한 정보가 기록된다[5]. 데이터웨어하우스에서 필요한 최종 테이블은 실적데이터를 가지고 있는 사실테이블과 공통적인 데이터를 관리하는 차원 테이블이 있다. ETT의 형태는 사실 테이블과 차원 테이블을 어떻게 설계하여 만들어 주느냐에 따라 시스템의 부하가 결정된다.

2.2 XMDR(eXtended MetaData Registry)

XMDR은 MDR을 확장한 개념으로 이전의 MDR에 전문 용어, 온톨로지, 시멘틱 웹(Semantic Web) 관리 기능을 확장한 것이다[6][7]. 본 논문에서의 XMDR의 각 구성 요소들은 메타 시멘틱 온톨로지(MSO:Meta Semantic Ontology)와 메타 로케이션(ML:Meta Location), 인스턴스 시멘틱 온톨로지(InSO:Instance Semantic Ontology)로 구성되어 있다[4].

III. XMDR기반의 실시간 정제시스템 설계

3.1 시스템 개요

본 시스템에서는 실시간 데이터 추출 및 정제작업을 위해 XMDR서버를 구성하였으며 XMDR 서버는 Query Manager, Transaction Manager, XMDR, Data Manager, Monitoring&Log Manager로 (그림 1)과 같이 구성하였다.

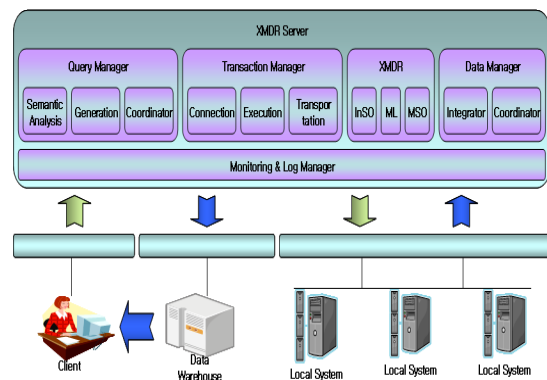


그림 1. XMDR기반의 정제시스템 모델
Fig. 1 Model for cleansing system based on XMDR

- **Query Manager:** 사용자가 요청한 질의를 분석하여 논리적인 메타쿼리로 변환한다. 변환시 의미적 이질성을 가진 조건은 Coordinator에 의해 확장된 조건으로 변환한다. 변환된 메타쿼리는 XMDR의 매핑맵을 통해 로컬쿼리로 다시 변환한다.
- **Transaction Manager:** 로컬시스템으로부터 추출한 자료를 Data Manager에게 제공하고, Data Manager에 의해 정제 및 통합된 데이터는 데이터웨어하우스로 전송한다.
- **Data Manager:** 각 로컬시스템으로부터 추출한 자료를 통합하고, 통합된 자료들 중 의미적 이질성, 표현 및 정의의 이질성을 가진 데이터는 Data Coordinator에 의해 일원화된 데이터로 조정된다.
- **Monitor&Log Manager:** 각 에이전트의 활동을 감시하며, 예외발생 시 로그를 생성하여, 관리자에게 통보하는 역할을 수행한다.

XMDR 서버는 사용자에게 실시간 분석자료를 제공하기 위해 요청자료에 대한 질의문에 포함되어 있는 시간과 공간의 조건을 분석하여 각 로컬 시스템에서 데이터 추출 시 해당 조건에 적합한 데이터만을 추출하여 Fact Table에 생성하도록 한다. XMDR를 이용한 실시간 정제 DATA추출작업은 다음과 같이 진행된다.

- Step 1. 사용자가 요구하는 Data Mart의 데이터를 추출할 분석/집합 쿼리를 분석한다.
- Step 2. 분석/집합 쿼리로부터 메타쿼리를 추출한다. 메타쿼리는 메타정보의 논리적 컬럼과 분석/집합 쿼리의 조건절로 구성되어 있고, 메타쿼리의 조건절은 XMDR의 InSO를 이용하여 분석/집합 쿼리의 조건절을 확장한 형태이다.
- Step 3. 메타쿼리를 컬럼 매핑 정보를 이용하여 로컬쿼리로 변환한다.
- Step 4. XMDR의 ML 정보를 이용하여 로컬시스템에 접속한다.
- Step 5. 로컬시스템에서 로컬쿼리를 실행하여 데이터를 추출한다.
- Step 6. 추출된 데이터를 XMDR의 MSO 정보를 이용하여 데이터를 정제한다.

Step 7. 추출 정제된 데이터를 Fact(Temp) Table에 생성한다.

Step 8. Fact(Temp) Table로부터 분석/집합 쿼리를 실행하여 Data Mart에 정보를 생성한다.

3.2 데이터추출

데이터 추출 속도는 일반적으로 데이터의 양과 비례한다. (그림1)의 Query Manager는 전체 데이터가 아닌 사용자가 요청하는 데이터만을 추출하여 추출속도를 향상시키기 위해 구문 분석기를 통해 사용자가 요청한 분석/집합쿼리를 형태대로 분리한 후에 메타정보, 예약어, 컬럼, 온톨로지 항목, 상수 등을 추출한다. 추출된 메타정보로부터 쿼리변환기는 메타정보의 매핑 컬럼을 조회하여 From절이 없는 논리적인 메타쿼리로 변환한다.

메타쿼리로 변환시 Coordinator는 XMDR의 InSO를 이용하여 조건절에서 추출한 정제 항목의 관계속성에 해당하는 메타자료조건 매핑맵으로부터 상수값과 동일한 의미를 가진 상수값들을 추출한다. 추출한 상수값들을 조건절에 포함시켜 사용자가 요청한 자료 추출시 의미적 이질성을 가진 전체 데이터를 조회할 수 있도록 하여 의미적 이질성을 극복하도록 한다. 관계속성은 Instance의미에 대한 Data층돌로 속성은 세가지로 분류하였다. 첫째, Has-a 관계로 Instance의 의미가 그룹에 속해있는 관계(가전제품C {TV,냉장고,세탁기})이다. 둘째, Is-a 관계로 Instance 의미가 세분화된 관계(TVC {LCD TV, LED TV, PDP TV})이다. 세 번째는 Equal 관계로 Instance의 의미는 동일하나 로컬스키마의 Instance값은 상이한 관계(서울=한양,텔레비전=TV)이다.

(표 1)는 본 논문에서 사용한 관계속성의 예제자료로 Has-a 관계는 분류그룹, Is-a 관계는 품목별분류, Equal관계는 동의어로 구분하였다. 메타자료명은 데이터 분석 방법의 분류에 따라 추가될 수 있으며, 지속적으로 관리가 필요한 대상이다. 메타쿼리의 컬럼 속성 설정시 온톨로지 항목인 경우에는 해당하는 관계속성의 메타자료명을 지정하여 쿼리변환기에 의해 분석/집합쿼리를 메타쿼리로 변환시 메타값에 해당하는 매핑값들을 조건절에 포함시킬 수 있도록 한다.

CATEGORY 컬럼이 온톨로지 항목일 때 상수가 '가전제품'인 경우 (표 1)의 예제자료에서 해당하는

매핑값을 추출하면 {'TV', '냉장고', '세탁기'}가 추출된다.

표 1. 관계속성의 예제자료
Table. 1 Sample Data Of Relation Attribute

메타자료명	관계속성	메타값	매핑값
분류그룹	Has-a	가전제품	TV
분류그룹	Has-a	가전제품	냉장고
분류그룹	Has-a	가전제품	세탁기
품목별분류	Is-a	TV	LCD TV
품목별분류	Is-a	TV	LED TV
품목별분류	Is-a	TV	PDP TV
동의어	Equal	컴퓨터	Computer
동의어	Equal	컴퓨터	PC
동의어	Equal	노트북	NoteBook
동의어	Equal	TV	텔레비전

추출된 값들의 품목별분류를 조회하면 {'LCD TV', 'LED TV', 'PDP TV'}가 추출된다. 추출된 값들의 동의어를 추출하면 '텔레비전'이 추출된다. 추출된 자료를 조건절에 포함시키면 'CATEGORY = '가전제품'인' 조건절은 CATEGORY IN ('TV', '냉장고', '세탁기', 'LCD TV', 'LED TV', 'PDP TV', '텔레비전')으로 확장된다. 관계속성에 의해 추출된 자료는 데이터 변형에서 메타값으로 변환하여 데이터의 이질성을 극복하고 일원화된 자료로 생성한 후 분석에 사용할 수 있도록 한다. 분석된 로컬쿼리를 이용하여 추출한 데이터를 Data Manager에게 제공하고, Data Manager로부터 정제된 데이터를 반환받아 Fact(Temporary) Table에 자료를 생성한다. Fact(Temporary) Table에서 Fact Table은 지속적으로 보관할 자료를 저장하고, Temporary Table은 실시간 분석용 자료를 Dynamic하게 생성하여 자료를 보관한다.

3.2 정제 및 변형

3.3.1 스키마 정제

메타쿼리의 구문형식은 "Select select-list Where conditions"의 형태를 지니고 있다. 메타쿼리에 From절을 명시하지 않는 이유는 구문분석기에서 추출한 메타정보와 매핑된 로컬시스템의 테이블정보를 로컬쿼리

의 From절로 명시하기 때문이다. 논리적인 메타정보와 물리적인 로컬시스템 테이블과의 관계를 정의하여 매핑맵을 생성함으로써 새로운 로컬시스템 추가시 로컬시스템의 변경없이 매핑관계의 정의만으로도 확장이 가능하게 한다. 메타쿼리는 로컬시스템으로부터 최하위 레벨의 소스 데이터를 추출하기 위한 논리적인 쿼리로 분석 및 집합 구문을 사용하지 않는다. 스키마 정제는 논리적인 메타쿼리를 물리적 로컬 스키마에서 질의를 수행할 수 있도록 로컬쿼리를 생성하는 작업이다. 이를 위해 스키마정제 테이블에 로컬 정보 및 로컬 테이블과 컬럼정보를 등록하고, 전사적 관점에서 통합할 메타정보를 등록한 후 메타정보의 컬럼과 로컬스키마 컬럼을 매핑하여 유연하게 메타쿼리를 로컬쿼리로 변경할 수 있도록 한다. 메타쿼리는 구문분석기를 통해 Select절, Where절로 구분하고, 메타컬럼, 예약어, 상수, 기호 등을 분리한다. 분석된 메타컬럼은 컬럼매핑정보를 통해 로컬컬럼으로 변환하고, 각종 예약어는 각 로컬스키마에서 사용하는 예약어로 변환한다. 변환된 로컬스키마별 조건절에 테이블매핑정보에 있는 로컬테이블의 기본 조건절을 포함시켜 조건절을 생성한다. 변환된 Select절과 Where절을 연결하여 로컬쿼리를 생성한다.

3.3.2 데이터정제 및 변형

데이터 정제는 소스데이터의 형식, 내용을 검증하여 가치있는 데이터로 만드는 과정이다. 데이터는 로컬스키마 별로 데이터타입, 길이, 날짜형식, 수치에 대한 단위, 환율 등이 서로 상이할 수 있으므로 데이터웨어하우스에서 사용하기 위해서는 일원화된 정보로 정제하여야 한다. 정제속성은 병합, 일자, 통화, 단위 네가지로 구분하였다. 병합은 메타컬럼이 로컬스키마에서 여러 개의 컬럼으로 있는 경우 로컬쿼리 생성시 컬럼연결자를 사용하여 하나의 컬럼으로 인식할 수 있도록 하여 데이터 구조에서 오는 이질성을 해결한다. 일자는 메타컬럼의 타입과 일자표현형식이 로컬시스템과 상이한 경우 로컬시스템의 일자컬럼을 메타컬럼과 동일한 표현 방식으로 변환하여 데이터 정의 및 표현에서 오는 이질성을 해결한다. 통화는 여러 통화로 표현된 금액의 평가치를 하나의 표준통화로 변환하여 데이터 의미의 이질성을 해결한다. 단위는 여러 단위로 표현된 단위의 가치를 표준 단위로 변환하여 데이터 의미의 이질성을 해결

한다. 데이터 정제시 처리할 수 없는 데이터 구조의 이질성, 데이터 표현의 이질성, 데이터 정의의 이질성들은 Monitoring&Log Manager에 의해 로그에 기록하며 시스템 관리자가 확인할 수 있다. 데이터 변형은 일관성 있는 정보 활용을 위하여 의미적 이질성을 가진 데이터를 표준화된 형태인 메타값으로 변환하는 작업이다. 각 로컬 시스템으로부터 추출한 데이터는 'TV', '텔레비전', 'Television' 처럼 각각의 값은 틀리나 포함하는 의미는 동일하다. 데이터 분석시 이와 같은 자료는 동일한 자료로 인식하기 위해 Data Coordinator는 각각의 데이터를 XMDR서버의 MSO로부터 획득한 'TV'라는 메타값으로 변경한다.

3.4 작업흐름도

(그림 2)는 사용자가 데이터웨어하우스에 요청한 정보에 대하여 XMDR 서버가 로컬시스템으로부터 자료를 추출하여 결과를 전송하는 작업흐름도로 클라이언트, XMDR 서버 그리고 로컬 시스템부분들을 분할하여 수행하는 과정을 보였다. 각 과정의 작업들은 다음과 같다.

- Request():사용자가 데이터웨어하우스에 정보를 요청.
- SemAnalysis():데이터웨어하우스에 요청한 질의문 분석.
- QryConvert():분석한 질의문을 메타쿼리로 변환하고 메타쿼리를 로컬시스템의 질의문으로 변환
- QryCoordinator():질의문의 조건절에 포함된 의미적 이질성 자료의 검색조건을 InSO를 참조하여 확장된 조건절을 생성.
- GetAccessInfo():XMDR의 ML로부터 로컬시스템의 접속 정보를 조회
- LocalConn():로컬시스템에 접속
- ExecuteQry():로컬시스템으로부터 로컬쿼리를 실행하여 데이터를 조회
- Integrator():로컬시스템으로부터 추출한 데이터를 통합
- DataCoordinator():통합한 자료를 MSO를 참조하여 데이터 정제 및 변형작업 수행
- DataTrans():정제된 데이터를 데이터웨어하우스에 전송

- Response():사용자에게 요청한 정보의 결과를 전달
- Monitoring():각 과정별 작업 수행을 감시
- WriteLog():Monitoring으로부터 감지된 결과를 Log에 기록

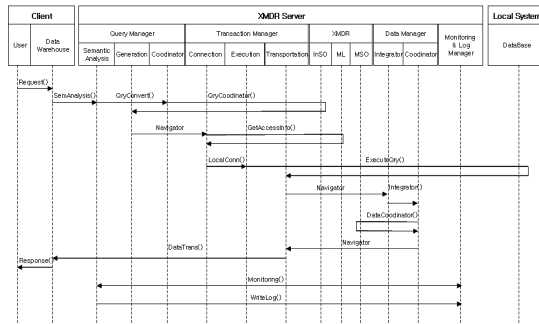


그림 2. 시스템 작업 흐름도
Fig. 2 System Work Flow

IV. 시스템적용 및 비교분석

4.1 적용 예

본 논문에서 제안된 시스템 적용을 위한 환경으로는 운영 체제는 Windows XP, 데이터베이스는 Oracle, Tool은 Developer2000을 사용하였다.

컬럼명	데이터형	길이	기본값	외래키	비고
PSTONE	VARCHAR2	30			이름
PSTONE_ZIP	VARCHAR2	30			우편번호
PSTONE_ZONE	VARCHAR2	30			지역명
PBOOK_NAME	VARCHAR2	30			책 제목
PBOOK_CODE	VARCHAR2	30			책 코드
PBOOK_YEAR	VARCHAR2	30			출판 연도
PAUTHOR	VARCHAR2	30			작가
PLANGUAGE	VARCHAR2	30			언어
PBOOK_SIZE	NUMBER	30			페이지
PBOOK_YEAR	NUMBER	30			출판 연도
PBOOK_YEAR	NUMBER	30			출판 연도
PBOOK_YEAR	NUMBER	30			출판 연도
PPRICE_IN	NUMBER	30			입장권
PPRICE_OUT	NUMBER	30			출장권
PQUANTITY_IN	NUMBER	30			입장권 수
PQUANTITY_OUT	NUMBER	30			출장권 수
PCURRENCY_OUT	NUMBER	30			출장권 단가
PCURRENCY_IN	NUMBER	30			출장권 단가
PCURRENCY_OUT	NUMBER	30			출장권 단가
PCURRENCY_IN	NUMBER	30			출장권 단가

그림 3. 메타 정보 및 컬럼등록
Fig. 3 Registration Meta Information & Column

(그림 3)은 메타정보 및 메타컬럼정보를 등록하는 화면이다. 메타정보 및 메타컬럼정보는 논리적인 메타쿼리를 구성하는 항목으로 사용자는 메타정보 및 메타컬

럼으로 질의를 수행 할 메타쿼리를 작성한다. 등록항목 중 자료매핑명은 XMDR의 InSO를 위한 정보로서 메타 쿼리 변환시 메타자료매핑 정보로부터 데이터를 조회하여 조건절을 확장하는데 사용한다. 확장된 조건절에는 의미적으로 연관된 데이터를 포함하여 질의를 수행할 수 있도록 하여 검색의 효율성을 증가시켰다.

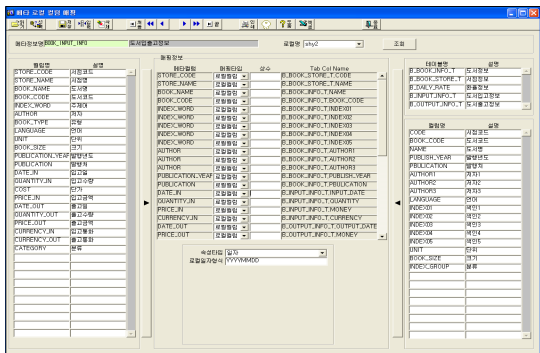


그림 4. 메타 로컬 컬럼 매핑
Fig. 4 Mapping Meta & Local Column

(그림 4)는 메타정보의 컬럼과 로컬시스템의 컬럼정보를 매핑하는 화면이다. 매핑타입은 로컬컬럼과 상수 두가지로 구분하였다. 매핑타입이 로컬컬럼인 경우에는 메타쿼리를 로컬쿼리로 변환시 메타컬럼과 매핑된 로컬컬럼으로 컬럼을 변환하고, 상수인 경우에는 상수값에 등록된 값으로 메타컬럼을 변환한다. 새로운 분산 시스템 통합시 메타컬럼과 로컬컬럼을 매핑하는 작업만으로 로컬시스템의 변경없이 통합이 가능해지므로 시스템의 확장이 용이해진다.

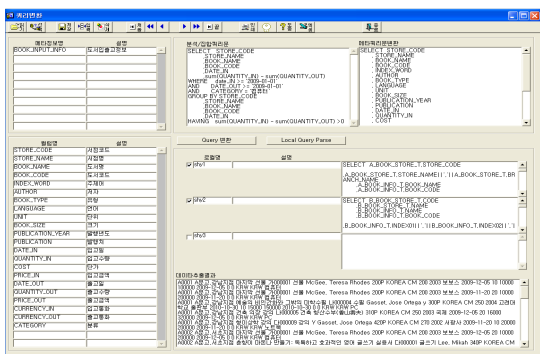


그림 5. 쿼리 변환 및 데이터 추출
Fig. 5 Transformation SQL & Execute SQL

(그림5)는 XMDR 정보를 이용 사용자가 메타정보를 참조하여 작성한 SQL문을 분석하여 Fact Table에 자료를 생성할 메타쿼리문으로 변환하고, 변환된 메타쿼리문을 로컬 쿼리문으로 변환하여 실행한 결과를 보여주는 화면이다. 분산된 로컬시스템의 데이터를 통합하기 위해 XMDR의 ML에 등록된 데이터베이스 정보를 이용하여 각 로컬시스템에 직접 접속하고 InSO를 이용하여 조건절을 확장함으로써 검색의 효율성을 증가시켰으며, MSO를 통해 일관성있는 데이터를 제공함으로써 정합성을 보장하도록 하였다.

4.2 비교 분석

Oracle ODS와 본시스템은 새로운 분산 시스템의 통합시 데이터 표준화를 통한 데이터의 이질성을 극복하고, 로컬 어플리케이션의 영향을 최소화하여 통합 작업을 할 수 있도록 하였다. (표 2)는 Oracle ODS와 본 시스템의 비교표이다.

표 2. 비교 분석
Table. 2 Comprise of Systems

비교항목	Oracle ODS	본 시스템
데이터 통합방식	Data HUB를 통한 데이터 통합	로컬시스템에 직접 접속하여 데이터를 추출
확장성	수집정보 일원화체계로 확장성 용이	XMR Map을 통한 확장
이질성극복	통합을 고려한 데이터 표준화	XMR을 이용하여 데이터의 구분, 구조, 의미의 이질성 극복
일관성	Data HUB를 변경하여 일관성 유지	XMR Setup을 통한 일관성 유지
정합성	데이터 표준화를 통한 정합성 확립	XMR 데이터 정제 및 병렬을 통한 정합성 보장
검색효율성	Data HUB를 이용한 검색 질의 수행	의미적으로 연관된 데이터를 포함한 검색 질의 수행
실시간정보제공	실시간 ETL 가능 제공	실시간 정제 정보 제공
DBMS의 용량	다양한 분석함수 활용	다양한 분석함수 활용

데이터 통합방식에서 Oracle ODS는 Data HUB를 이용하여 데이터를 중앙에서 집중하는 방식으로 로컬 시스템이 증가할수록 데이터 허브에 대한 스키마와 다른 로컬 시스템과의 연계를 새로이 구축해야 하는 확장성의 단점이 있다. 본 시스템은 XMDR의 ML정보를 통하여 로컬 시스템에 직접 접속, 데이터를 추출하는 방식을 사용하였으며, 스키마의 이질성은 논리적인 메타데이터 스키마와 물리적인 로컬 시스템 스키마를 매핑하는 방식을 사용하여 새로운 로컬 시스템의 확장시 로컬 시스템의 변경을 최소화하여 유연하게 통합할 수 있도록 하였다. 그리고 로컬 시스템의 변경에서 발생하는 데이터의 일관성은 Oracle ODS는 Data HUB의 변경을 통해

일관성을 유지하였고, 본 시스템에서는 XMDR의 MSO 정보를 통해 일관성을 유지할 수 있도록 하였다.

V. 결론 및 향후연구

본 논문에서 XMDR을 이용한 데이터웨어하우스 실시간 데이터 정제 시스템을 제안하였다. 본 시스템의 장점은 분산된 환경에서 데이터를 통합하기 위한 표준화된 정보를 제공하여 일관성있는 정보를 생성할 수 있도록 하고, 정보관리의 일원화를 통해 정보의 품질을 보장하였다. 또한 새로운 분산 환경 통합 시 로컬 시스템의 변경없이 XMDR에 정보를 추가함으로써 데이터웨어하우스를 위한 데이터 통합을 할 수 있도록 하여, 데이터 통합 작업 시 비용과 시간이 절감되는 효과를 얻게 하였다.

제안된 시스템에서는 실시간처리를 위해 Data Mart에 데이터를 생성하기 위한 쿼리를 분석하여 추출하는 데이터의 범위를 최소화 하도록 하였다. 그러나 실제 Data Mart의 기본이 되는 Fact Table을 생성하기 위한 추출 대상 데이터가 대용량이라면 실시간 정제시스템의 응답시간은 그만큼 늦어질 수밖에 없다. 따라서 대용량 기반의 데이터 추출 및 정제 시간을 단축하기 위한 방법에 관한 연구가 필요하다.

참고문헌

- [1] Immon, W.H. and R.D. Hackathorn. "Using the Data Warehouse", New York: John Wiley & Sons, 1992
- [2] Inmon, W.H and R.d. Hackathorn, "Using the Data Warehouse", New York: Jonh Wiley & Sons, 1994
- [3] M. Asif Naem, Gillian Dobbie, Gerald Weber, "An Event-Based Near Real-Time Data Integration Architecture", Department of Computer Science, The University of Auckland
- [4] Kevin D. Keck and John L. McCarthy, "XMDR: Proposed Prototype Architecture Version 1.01", <http://www.xmdr.org>, February 3, 2005
- [5] Andrew White, David Newman, Debra Logan, John Radcliffe, "Mastering Master Data Management", Garther, 2006

- [6] 정계동, 황치곤, 최영근, "분산 환경에서 XMDR을 이용한 예약 정보 시스템", 한국해양정보통신학회 논문지 Vol.11 No.10 pp.1957-1967, 2007
- [7] 국윤규, "하이브리드 에이전트에 의한 효율적인 데이터 그리드 시스템", 광운대학교 박사학위 논문, 2006.

저자소개

송홍율(Hong-Youl Song)



2008년 학점은행 전자계산학 이학사
2010년 광운대학교 유비쿼터스
컴퓨팅학과 석사
2005년~ 현재 빌트윈 컨설턴트

※관심분야: XMDR, 데이터웨어하우스, 비즈니스인텔리전스

정계동(Gye-Dong Jung)



1985년 광운대학교 전자계산학졸업
1992년 광운대학교 산업정보학석사
2000년 광운대학교 컴퓨터과학박사
1993년~2004년 광운대학교
정보과학원 교수

2005년~ 현재 광운대학교 교양학부 교수

※관심분야: XML 분산시스템, 분산 컴퓨팅기술, 이동에이전트

최영근(Young-Keun Choi)



1980년 서울대학교 수학교육과
이학사
1982년 서울대학교 계산통계학과
이학석사

1989년 서울대학교 계산통계학과 이학박사
1983년~ 현재 광운대학교 컴퓨터과학과 교수
1992년~2000년 광운대학교 전산정보원 원장
2002년~2005년 광운대학교 교무연구처장
※관심분야: 객체지향설계, 분산시스템, 이동에이전트, 상호운용성