

메타데이터 기반 데이터 웨어하우스 아키텍처*

- S병원 사례를 중심으로 -

김태훈**, 김종호**, 이희석**

A Metadata-oriented Data Warehouse Architecture

- S Hospital Case -

Kim, Taehun, Kim, Jongho, Lee, Heeseok

Data warehouse is an intelligent store of data. A metadata is critical for implementing data warehouse. Integrating data warehouse with its metadata helps create a more adaptive information system. This paper proposes a metadata-oriented data warehouse architecture that consists of seven components: legacy, extracting, operational data store, data warehouse, data mart, application, and metadata. A taxonomy for dataflow and metaflow is proposed for better understanding of the architecture. In addition, a metadata schema is built within the framework of the seven components. Lastly, the architecture with its metadata component is applied to a real-life data warehouse project for a large medical center in order to illustrate its practical usefulness.

* 본 연구는 산업자원부 산업기반기술개발사업과제(관리번호: A00-981-3302-09-1-2)로 수행되었음.

** 한국과학기술원 테크노경영대학원 기업정보시스템연구실

I. 서론

최근 들어 데이터 웨어하우스 (Data Warehouse: DW)의 기업 지식화에서의 중요성은 주지의 사실이다. DW 기술은 빠르게 발전되고 있으며 관련 신기술이 지속적으로 소개되어지고 있다. 그러나, DW에 관한 연구는 상대적으로 미비한 실정이다 [McFadden, 1996]. 그간의 연구는 주로 DW 이란 무엇인가에 초점을 두고 있으며, 어떻게 구축하는가라는 이슈는 최근에만 고려되고 있다 [Atre and Storer, 1995]. 또한 상업적보다 이론적인 면에 관점을 둘 필요성이 지속적으로 제기되고 있다.

DW의 아버지라는 Inmon (1994)은 DW를 “경영층의 의사결정지원을 위한, 통합적, 시계열적, 주제지향적, 비휘발성적인 데이터의 집합”으로 정의하고 있다. 즉, DW는 기업정보처리를 위한 기본적인 아키텍처 (Architecture)로 필요불가결한 요소인 것이다. 지속적으로 운영 데이터를 생산하는 운영 데이터베이스 (Operational Data-base) [Chaudhuri and Dayal, 1997]는 OLTP (On-line Transaction Processing) 어플리케이션에 이용된다. 반면에, DW는 이들 데이터를 통합하여 분석적인 정보를 산출한다는 점에서, OLAP (On-line Analytical Processing) 어플리케이션의 근간이 된다. OLAP은 OLTP을 위한 대안으로 최근에 대두된 용어이다 [Devlin, 1997].

DW 구축에는 다음의 두 가지 이슈가 중요하다. 비즈니스 사용자 (Business End-user)의 의사결정 지원에 근거해서 DW를 개발하는 것파이에 근거한 메타데이터 (Metadata)를 개발하는 것이다. 대부분의 DW 개발 방법론은 메타데이터를 개발하는 것을 고려하지 않고 있다. 기존 방법론은 다만 메타데이터를 소프트웨어로 간주하고 있다. 즉, DW와 메타데이터를 동시에 개발하려는 통합된 방법론이 필요한 것이다. 메타데이터가 데이터 웨어하우징 시스템 (Data Warehousing System) 의 핵심임에도 불구하고,

메타데이터 관리가 DW 개발 과정으로 분리되어져 왔기 때문에, 데이터 웨어하우징 프로젝트의 성공이 어려운 실정인 것이다 [Denzer and Guttler, 1996; Kutsche and Sunbul, 1999].

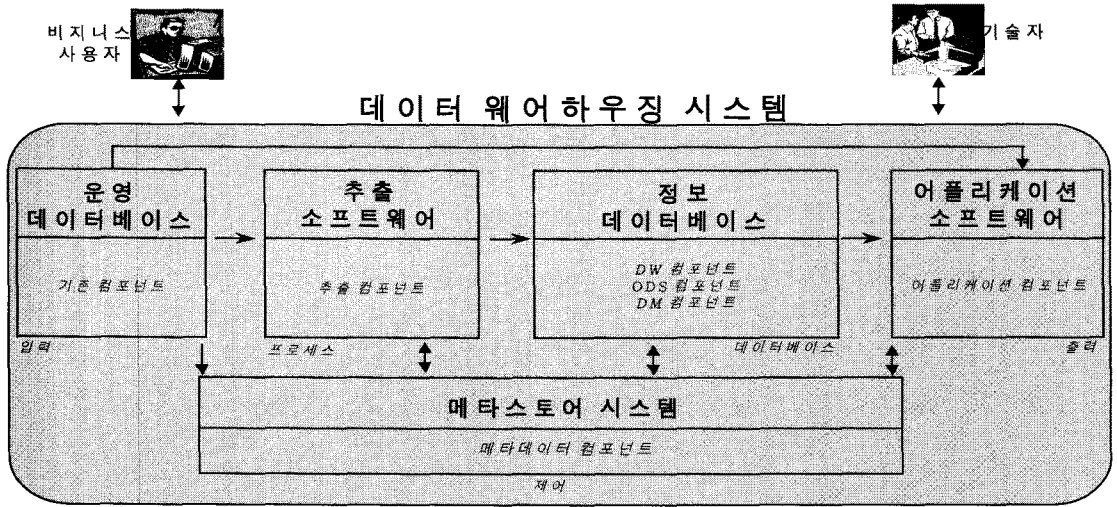
이러한 통합을 위해서, 본 연구는 데이터흐름 (Dataflow)과 메타흐름 (Metaflow)을 위한 분류 기준 (Taxonomy)을 제안한다. 이 기준에 근거하여, DW를 위한 아키텍처를 제시한다.

제안된 아키텍처는 7 컴포넌트 (Component)로 구성된다. 이 컴포넌트는 기존 (Legacy), 추출 (Extracting), 운영 데이터 저장소 (Operational Data Store: ODS) [Inmon et al., 1997], DW, 데이터 마트 (Data Mart: DM) [Hackney, 1997; Hoven, 1998], 어플리케이션 (Application) 및 메타데이터를 포함한다. 특히, 본 연구는 기술적 (Technical) 메타데이터와 비즈니스 메타데이터로 구성되어 있는 메타데이터 컴포넌트에 중점을 두고 DW를 위한 일반적인 (Generic) 메타데이터 스키마를 구현하였다. 데이터 웨어하우징 시스템이 주로 비즈니스 사용자를 위한 점을 고려하여 비즈니스 메타데이터에 초점이 주어졌다. 나아가서, 메타데이터 컴포넌트를 포함한 아키텍처의 실용적 유용성 검증을 위하여 실제 DW 프로젝트 사례분석을 하였다.

II. 데이터 웨어하우스: 시스템, 분류 및 아키텍처

2.1 데이터 웨어하우징 시스템

데이터 웨어하우징 시스템의 사용자는 비즈니스 사용자와 기술자 (Technical User)로 대별된다. 비즈니스 사용자는 임원, 임시 사용자 (Casual User), 비즈니스 분석가, 및 파워 사용자 (Power User)로 분류되어지며 [Poe, 1997], 기술자는 시스템 관리자, 어플리케이션 개발자, 운영자, 기술 보조자 및 설계자를 포함한다. <그림 1>은 데이터



<그림 1> 데이터 웨어하우징 시스템과 사용자

웨어하우징 시스템과 사용자의 구성을 나타낸다. 본 구성도는 Orr (1998)의 입력-프로세스-데이터베이스-출력 (Input-Process-Database-Output) 시스템 모형으로부터 확장되었다. 입력은 기존의 운영 데이터베이스에 해당된다. 프로세스는 추출 소프트웨어에 해당하며, 데이터베이스는 DW, ODS, DM으로 구성된 정보 데이터베이스 (Informational Database)에 해당된다. 출력은 정보 데이터베이스를 활용하는 어플리케이션 소프트웨어를 나타낸다. 다른 4가지 부분을 관리하는 제어 (Control)는 메타스토어 (Metastore) 시스템에 상응한다. 메타스토어는 메타데이터를 저장하는 데이터베이스이다.

용어 “데이터 웨어하우징”은 DW의 동적인

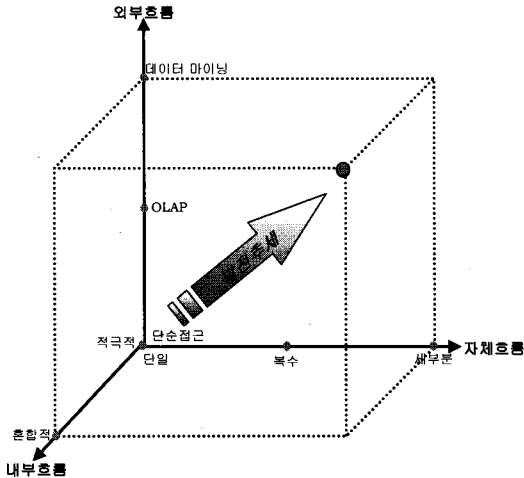
특성을 강조하기위해서 사용되어지며[Hackathorn, 1995], 이러한 특성은 다음에 기술하는 다섯 유형의 흐름 (Flow)에 의해 명확해질 것이다.

2.2 데이터흐름과 메타흐름을 위한 분류

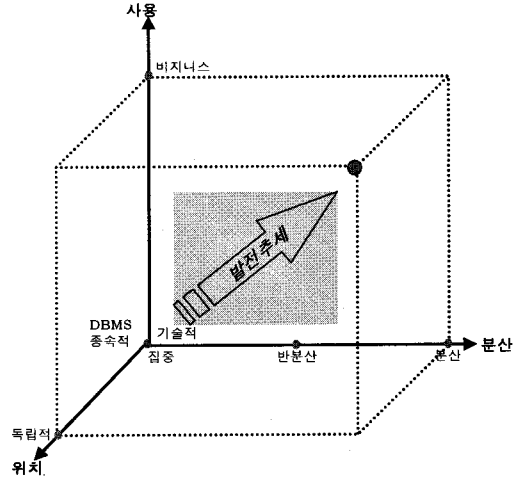
데이터흐름과 메타흐름은 데이터 웨어하우징 시스템의 모든 프로세스를 표현하므로 이에 대한 분류가 중요하다 [이희석 외, 1997]. Hackathorn (1995)은 독창적으로 이와 같은 흐름을 연구하였다. 내부흐름 (Inflow), 상향흐름 (Upflow), 하향흐름 (Downflow), 외부흐름 (Outflow), 및 메타흐름이 그것이다. 본 연구는 이 흐름을 채택하여 확장하였다 (<표 1> 참조).

<표 1> 데이터흐름과 메타흐름

흐름		설명	
데이터흐름 (Dataflow)	내부흐름 (Inflow)	운영 데이터베이스가 정제되어 정보 데이터베이스로 유입됨	
	자체흐름 (Ownflow)	상향흐름(Upflow)	고도로 자세한 데이터가 집약되어 요약되어짐
		하향흐름 (Downflow)	오래된 (Old) 데이터는 저장되거나 정화 (Purge) 기준에 따라 제거되어짐
	외부흐름 (Outflow)	사용자가 정형화 (Canned) 되거나 임시적 (Ad-hoc) 질의를 이용하여 정보 데이터베이스로부터 데이터를 얻음	
메타흐름 (Metaflow)		다양한 흐름과 관련한 메타데이터 흐름	



<그림 2> 데이터흐름을 위한 분류



<그림 3> 메타흐름의 분류

<그림 2>는 데이터흐름을 위한 분류도이다. 내부흐름 축은 기존 운영시스템에 접근하여 정보 데이터베이스를 구성하는 방법을 나타낸다. 내부흐름은 세가지 유형으로 대별된다 [Widom, 1995]. 우선, 사용자가 정보 데이터베이스없이 운영 데이터베이스에 직접 접근하는 경우이다. 이러한 접근방법을 가상 (Virtual) DW 접근방법이라고 부른다. OLAP 환경하에서의 의사결정지원은 이 방법을 채택하고 있으며, 나태한 (Lazy) 유형이다. 반대로, 사용자가 정보 데이터베이스의 도움으로 데이터를 접근하는 경우이다. 이러한 접근방법을 적극적 (Eager) 유형이라 부른다. 또한, 적극적 유형과 나태한 유형을 같이 사용하는 혼합적 (Hybrid) 유형이 있다. 다만, 나태한 유형은 실제로 DW 에서 다루지 않으므로, <그림 2>에서는 생략하였다.

두번째, 자체흐름 (Ownflow) 축은 데이터 분산 정도를 나타낸다. 단일 (Single) 유형은 하나의 데이터베이스를 사용한다. 복수 (Double) 유형은 두 개의 데이터베이스로 이루어지며, ODS와 DW, DW와 DM의 결합으로 이루어진다. DM은 DW와는 달리 ODS 정보를 담을 수가 없다. ODS는 오직 DW 만을 지원하기 위해 존재하며, DW 처럼 대량의 데이터를 가지고 있기 때문이다.

세부분 (Triple) 유형이 가장 복잡하나, 데이터 웨어하우징 시스템을 위해서는 이상적이다.

외부흐름은 단순접근 (Access), OLAP 및 데이터 마이닝 (Data Mining)으로 분류되어질 수 있다. 가장 단순한 유형인 단순접근은 사용자 주도의 질의를 통해서 단순 결과를 제공한다.

반대로, OLAP 유형은 사용자를 위한 데이터 큐브 (Data Cube)와 같은 다차원 시각 (Multi-dimensional View)을 제공한다. 마지막으로, 데이터 마이닝 유형은 하나의 질문을 묻기보다는 데이터간의 관계, 패턴, 규칙을 도출한다. 자체흐름과 외부흐름이 [부록 A]에 상세히 설명되었다.

<그림 3>은 메타흐름의 분류를 도식화한다. 본 분류는 3개의 축 -위치 (Situation), 분산 (Distribution) 및 사용 (Usage)-을 따라 구성된다. 데이터흐름처럼, 각 축은 데이터 웨어하우징 시스템의 특성에 영향을 미친다.

위치 축에 따르면, DBMS (Database Management System) 종속적 (Dependent) 또는 독립적 (Standalone)의 두가지 유형이 가능하다 [Duyn, 1982]. 독립적 시스템은 유연하며 이동가능한 반면, 종속적 시스템은 특정 DBMS에서 운영된다. 독립적 시스템은 인터페이스를 통해서 이질적인 DBMS 을 지원할 수 있으며, 다양한 DBMS 환경하에

서의 모든 데이터의 통합적 제어를 제공할 수 있다.

분산 축은 메타데이터 분산 정도를 나타내는 것이다. 집중은 한 장소 (Site)에 메타데이터를 저장하는 것이고, 분산은 여러 장소에 메타데이터가 저장된 것이며, 반분산은 집중과 분산의 결합 형태를 나타낸다. 데이터 분산과 관련된 많은 도구와 문제들이 데이터 웨어하우징 시스템에 적용되어질 수 있다 [Atre and Storer, 1995]. 메타데이터 역시 집중화된 혹은 분산화된 방식으로 유지되어 질 수 있다 [Ozsu and Valduriez, 1991].

사용 축을 따라서는, 비즈니스 메타데이터 (또는 어플리케이션 메타데이터)와 기술적 메타데이터 (혹은 시스템 메타데이터)로 대별된다. 기술적 메타데이터는 정보 데이터베이스에 관한 기술적인 정보를 포함하며, 주로 기술자를 위해서 사용되어진다. 비즈니스 메타데이터는 비즈니스 사용자에게 이해하기 쉬운 정보 데이터베이스의 비즈니스 관점을 제공한다 [White, 1995]. 메타흐름은 종속적, 집중적, 기술적에서 독

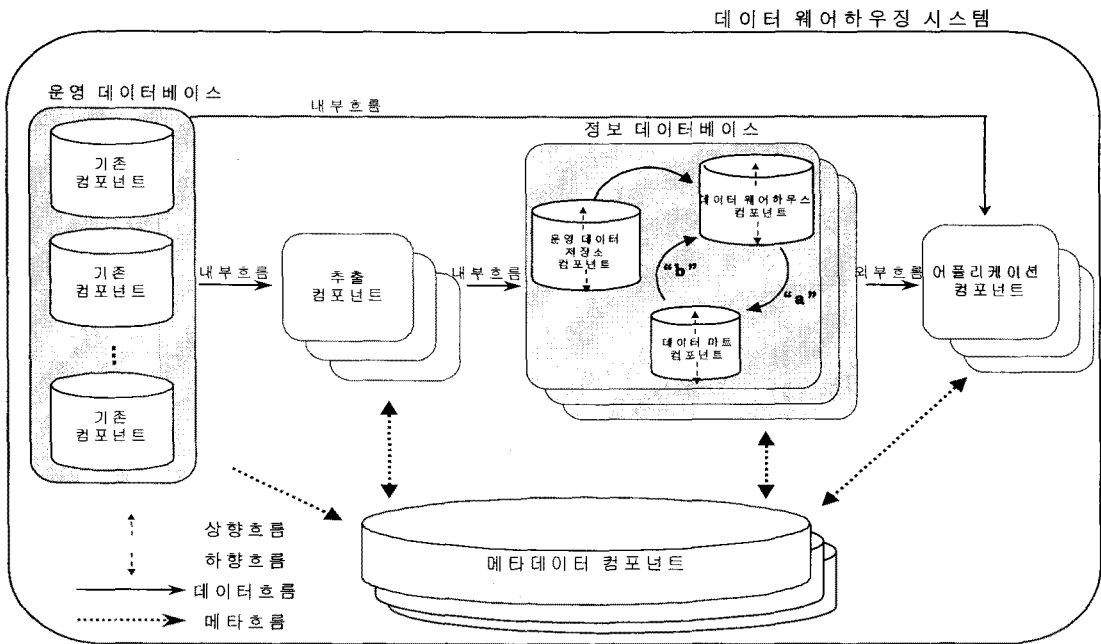
립적, 분산적, 비즈니스 유형으로 발전되고 있다.

데이터흐름을 지원하기 위해 아키텍처를 구성하는 컴포넌트는 상호 유기적인 관계성을 맺고 있으며, 아키텍처내의 데이터흐름 및 메타흐름은 DW 아키텍처를 이해하는 데 필 수적이다 [이희석 외, 1997]. 이러한 흐름들은 DW를 구축할 때 필요한 전략으로 활용된다.

2.3 데이터 웨어하우스 아키텍처

기존의 DW 아키텍처 [Chaudhuri and Dayal, 1997; Devlin, 1997; Gardner, 1998; Kosar, 1996; Murtaza, 1998; Noaman and Barker, 1997]는 한정된 컴포넌트로 구성되어 있다. Kosar (1996)의 아키텍처는 ODS 컴포넌트를 고려하지 않고 있으며, 메타흐름 분류를 제공하지 않고 있다.

Chaudhuri와 Dayal (1997)의 아키텍처는 메타데이터를 저장할 수 있는 집중화된 리퍼지토리 (Repository)를 고려했다는 점에서, Kosar의 아키텍처와 차별화된다.



<그림 4> 데이터 웨어하우스 아키텍처: 7 컴포넌트

인프라스트럭처 (Infrastructure)는 데이터 웨어하우징 시스템의 물리적 관점이며, 동일한 아키텍처로부터 수많은 인프라스트럭처가 도출되어질 수 있다 [Poe, 1997]. 본 연구는 데이터흐름과 메타흐름의 분류에 따라 아키텍처를 제안한다. 이 아키텍처는 <그림 4>에 도시된 것처럼 일곱가지의 컴포넌트로 이루어져 있다.

데이터 마트는 기업의 매점 역할을 하는 것으로서, 부서별이나 기능별로 만들어지는 DW의 부분집합 (Subset) 이다 [Demarest, 1994]. <그림 4>의 "a"는 DW 컴포넌트가 DM 컴포넌트에 데이터를 제공하는 것을 의미하는 것이고 (하향식), "b"는 DM 컴포넌트가 DW 컴포넌트에 데이터를 제공하는 것을 의미한다 (상향식). 하향식에 의하면, DW가 개발되어진 다음 DM이 개발된 DW의 일부분으로 생성되어지는 것이다 [Hackney, 1997]. 상향식은 하향식의 반대로, DM을 DW의 블록으로 사용하는 방식이다. 추출과 어플리케이션 컴포넌트는 소프트웨어 프로그램인 반면에, DW, ODS, DM, 및 메타데이터 컴포넌트는 데이터베이스 형태를 취한다. 데이터베이스간의 상세한 비교가 [부록 B]에 정리되어 있다.

Ⅲ. 데이터 웨어하우스 메타데이터 컴포넌트

3.1 OLAP 환경하에서의 메타데이터

데이터 웨어하우징 시스템은 데이터 공급이 아니라, 정보의 사용이 주 목적이다. 데이터가 공급되었다고 해서 정보의 창조와 사용과 직결되지는 않는다. 메타데이터는 이점에서 중요하다. 메타데이터의 목적은 하나의 대상 (Object) (즉, 데이터) 에 하나의 이름 (즉, 의미)를 부여하는 것이다. OLTP 환경하에서의 메타데이터는 반복적 작업을 기반으로 해서, 데이터에게 의미를 준다. 그러나, OLAP에서의 중요성은 데이터 웨어하우징 시스템에 관한 정보로 확장되어 사용된다.

데이터 웨어하우징 시스템이 발전함에 따라, 메타데이터 또한 발전하고 있다. 메타데이터는 컴퓨터 시대가 도래한 이후로도 계속 존재하여 왔다. 메타데이터가 적용되고 있는 영역은 워크플로우 설계 및 관리, 문서 관리, CAD (Computer-aided Design) 및 어플리케이션 개발이다 [Bernstein and Edelstein, 1999; Mark and Roussopoulos, 1986].

<표 2> OLTP와 OLAP 메타데이터 비교

특징	메타데이터	OLTP	OLAP
버전 (Version) 관리		거의 없음 (갱신 위주)	다양한 버전 관리 (이력 정보)
사용자 수		적음 (기술자 > 비즈니스 사용자)	많음 (기술자 < 비즈니스 사용자)
메타데이터에 대한 관점		기능적 범위	전사적 범위
사용 환경		도구 위주	사용자 위주
메타데이터 사용의 목적		효율적	효과적
메타데이터 구조		비구조적 또는 구조적	구조적
메타데이터 생성		자동적	자동적 또는 수동적
통합성		비통합적	통합적
이해의 정도 (Legibility)		어려움	쉬움
변화가능성		없음	있음

OLAP 환경하에서는 기술자와 비즈니스 사용자 모두 메타데이터를 접근할 수 있다. 비즈니스 사용자는 의사결정지원을 위해서 메타데이터를 사용하는 반면에, 기술자는 데이터 웨어하우징 시스템 개발, 관리 혹은 유지보수를 위해서 사용한다. 메타데이터가 사용자의 의사결정 지원에 초점을 두고 있으므로, 메타데이터는 데이터 웨어하우징 시스템의 핵심이다. <표 2>는 OLTP와 OLAP 환경하의 메타데이터를 비교 요약한다. 본 연구의 중점은 OLAP 메타데이터에 있다.

3.2 핵심 메타데이터 스키마

현재, 데이터 웨어하우징 시스템을 위한 메타데이터 표준은 특정 벤더용이 대부분이다.

벤더 연합회 (Metadata Council)는 다양한 벤더 제품간의 메타데이터 교환을 위한 표준화 노력을 하고 있다 [Gill and Rao, 1996; Kelly, 1997]. 본 연합회는 시스템 리파지토리를 위해서 사용되

어질 수 있는 MIS (Metadata Interchange Standard)에 따른 메타데이터 스키마를 제시하고 있다 [Metadata Council, 1997].

본 연구에서 제안하는 메타데이터 스키마는 MIS 메타모델에 7 유형의 컴포넌트 관련 정보를 추가하였다. 기존 컴포넌트 정보 (MD_Legacy-Component), 추출 컴포넌트 정보 (MD_Extracting-Component), ODS, DW 및 DM 컴포넌트를 포함하는 정보 데이터베이스 컴포넌트 정보 (MD_InformationalDBComponent) 및 어플리케이션 컴포넌트 정보를 (MD_Application-Component)가 추가된 컴포넌트 관련 정보이다.

본 연구에서는 이 스키마를 핵심 메타데이터 스키마 (Core Metadata Schema)로 부르기로 한다. 스키마는 ER (Entity Relationship) 모델로 표현되었으며, 사이베이스의 PowerDesigner Data-Architect [Sybase Inc., 1999] 를 이용하여 도식화되었다. 총 28개의 개체로 구성된 스키마는 [부록 C]에 도식화되었으며, 개체에 대한 설명은

<표 3> 메타데이터 스키마 비교

연구 항목	Burton and Green (1996)	Inmon, Zachman, and Geiger (1997)	Barquin and Edelstein (1997)	Sachdeva (1998)	Katic et al. (1998)	Jarke et al. (1999)	핵심 메타데이터 스키마
분류 기준	위치	사용자, 위치 및 기능	기능	사용자	기능	기능	7 컴포넌트
분류	없음	Zachman 방식	없음	없음	없음	없음	데이터흐름과 메타흐름에 의한 분류
메타데이터 요소의 범주	테이블, 필드, 소스, 사용	동기, 엔터티, 행위, 위치, 사람, 시간	질의처리용, 구축용, 관리용	비즈니스 사용자용, 관리자용, 개발자용	DW 보안만을 위한 구조용, 접근용	DW 품질용	MIS 메타모델 정보, 7 컴포넌트 정보, 추가 메타데이터 정보
사용자 요구사항 반영 여부	반영할 수 없음 (고정 메타데이터 모델)	반영할 수 없음 (고정 메타데이터 모델)	반영할 수 없음 (고정 메타데이터 모델)	반영할 수 없음 (고정 메타데이터 모델)	반영할 수 없음 (고정 메타데이터 모델)	반영할 수 없음 (고정 메타데이터 모델)	반영됨 (유연한 메타데이터 모델)
모델 방법	ER (애트리뷰트 수준)	ER (엔터티 수준)	ER (엔터티 수준)	ER (애트리뷰트 수준)	플랫 파일 (Flat File) 형태	ER (애트리뷰트 수준)	ER (애트리뷰트 수준)
MIS 메타모델 포함 여부	미포함	일부 포함	미포함	미포함	미포함	미포함	완전히 포함

[부록 D]에 정리되었다.

핵심 메타데이터 스키마를 기존의 주요 메타데이터 스키마와 비교요약하여 <표 3>에 정리하였다. 대부분의 스키마는 ER 모델로 도식화되어 있다. 과거 연구의 스키마는 위치별 (Location) [Burton and Green, 1996; Inmon and Hackathorn, 1994], 사용자별 [Sachdeva, 1998; White, 1995] 또는 기능별 (Function) [Barquin and Edelstein, 1997; Devlin, 1997; Jarke et al., 1999; Katic et al., 1998; Moriarty and Mandracchia, 1996]로 구분되어져 있다. 그러나, 핵심 메타데이터 스키마는 7 컴포넌트에 의해 구성되어 상세한 설명이 용이하다. 핵심 메타데이터 스키마만이 MIS 메타모델을 완전히 포함하고 있음은 주목할만하다. 본 스키마는 데이터흐름과 메타흐름에 의한 분류에 의해서 나온 아키텍처에 근거한 컴포넌트의 스키마를 포함하고 있다.

Inmon et al. (1997)은 Zachman 방식에 의한 분류에 근거하고 있다. 메타데이터 요소의 범주에서, Katic et al. (1998)은 DW 보안만을 위한 메타데이터 요소를 제시하며, Jarke et al. (1999)은 DW 품질 (Quality)만을 위한 메타데이터 요소를 제시하고 있다. 대부분의 연구가 일반적인 요소들을 제시하고 있는데 반해, 본 연구는 표준안인 MIS 메타모델과 함께 추가적인 7개 컴포넌트 정보 및 메타데이터 요구사항을 포함한다. 따라서, 특정 프로젝트에는 추가된 요구사항이 핵심 메타데이터 스키마에 반영되어서 최종적 스키마가 도출된다.

IV. 사례: S병원

최근의 병원은 정보화로 인해서 상당한 양의 의료 데이터가 발생하며 이의 효과적인 이용에 관심을 가지고 있다. 그러나, 기존 병원정보시스템 (Hospital Information System: HIS)은 아직까지 일반관리와 원무관리 중심에서 벗어나지 못하고 있다. 환자 중심의 진료지원 및 임상

연구를 종합적으로 지원하기 위한 데이터 웨어하우스의 필요성이 대두되고 있다. 이에 본 연구는 제안된 데이터 웨어하우스 아키텍처를 실제 병원정보시스템에 적용하여 타당성을 검증하는데 주안점을 두었다. 1998년, S 병원은 병원 전역의 데이터 웨어하우징 시스템을 연구개발 프로젝트로 시작하였다. 프로젝트의 목적은 의료인과 비즈니스 사용자에게 다양한 소스들로부터 합병된, 비즈니스 데이터 및 임상 데이터를 제공하는 것이다.

S 병원은 7개 지방병원과 대학병원으로 구성된 0 재단산하 의료센터의 하나이다. 이 중 S 병원이 가장 중추적 병원이며, 약 2,150 병상을 가지고 있다. 일일 평균 재원 환자는 2,000명이며 일일 평균 외래진료 환자는 약 4,708명이다 [송준식 외, 1998]. S 병원과 다른 병원과의 네트워크 구성도가 [부록 E]에 요약되어있다.

4.1 통합병원정보시스템과 의료용 데이터 웨어하우스

병원의 의료업무는 의료인과 환자의 상황뿐 아니라, 이것을 바탕으로 조직화된 의료행위의 실천을 필요로 한다 [서동희, 1997]. 즉, 병원 업무는 진료, 진료지원, 원무관리 및 일반관리의 네 유형으로 대별되며 [송운호 외, 1994], 상호간에 유기적인 협조관계를 통하여 환자진료와 병원경영이 수행된다.

의료용 데이터 웨어하우스는 사용자에게 따라, (i) 진료와 진료지원을 기반으로 하는 임상 데이터 웨어하우스 (Clinical Data Warehouse)와 (ii) 원무관리와 일반관리를 기반으로 하는 비임상 데이터 웨어하우스 (Non-clinical Data Warehouse)의 두 부분으로 나누어진다. 비임상 데이터 웨어하우스는 경영층의 의사결정을 위한 비즈니스 데이터 웨어하우스 (Business Data Warehouse)와 유사한 것이다. 임상 데이터 웨어하우스는 환자에 대한 정보를 저장하여 의료 서비스와

<표 4> 통합병원정보시스템

정보시스템 업무 유형	통합병원정보시스템				
	운영계 시스템			정보계 시스템	
진료	원격진료시스템 자동진단시스템 간호정보시스템	OCS EMR	CIS	영상 데이터 웨어하우징 시스템	의료용 데이터 웨어하우징 시스템
진료지원	PACS LIS RIS 약제처방시스템 약국정보시스템 의료정보검색시스템 교육관리시스템 의국관리시스템 대기전광판시스템 검사자동화시스템 급식관리시스템 종합건강관리시스템				
원무관리	환자 관리 시스템 수입 관리 시스템 보험 관리 및 청구 시스템 차트 저장 시스템	HMIS	비영상 데이터 웨어하우징 시스템		
일반관리	인사/급여시스템 회계/원가분석시스템 자산관리시스템 자재관리시스템 재고관리시스템				

의료 연구를 주목적으로 한다 [Pedersen and Jensen, 1998]. 병원정보와 DW에 대한 요약이 <표 4>에 정리되어있다.

의료용 데이터 웨어하우징에 관한 연구는 구축과 사례 연구들로 이루어져 있다. 퇴원정보에 대한 DM 구축 (Berndt and Hevner, 2000; 김유미 외, 2000), 핵의학과를 위한 영상 DM 구축 (범희승 외, 2000), 진료정보를 위한 DW 구축 (김선구 외, 2000) 등이 있다. Bates et al.

(1999)는 임상 데이터 웨어하우스의 정의와 아키텍처 일부를 제시하고 있다. 그러나 이 아키텍처는 질관리 측면만을 강조하고 있다. Nowack과 Niccolai (1997), Nussbaum (1998), Pedersen과 Jensen (1998) 및 Rardin (1996)은 전자의무기록 (Electronic Medical Record: EMR) 시스템

과 DW의 아키텍처의 중요성을 기술하고 있다. 특히, Pedersen과 Jensen (1998)은 임상 데이터 웨어하우스의 기술적 부분과 연구 부분에 대해서 심층적으로 제시하고 있다. 대부분의 관련 연구들이 의료용 데이터 웨어하우스의 청사진을 제시하고 있지 않으며, 구축 사례연구로써 끝을 맺고 있다. 본 연구는 이러한 기존 문헌을 바탕으로 통합병원정보시스템 (Integrated Hospital Information System: IHIS)의 분류와 그에 따른 아키텍처를 제시하였다.

통합병원정보시스템은 임상정보시스템 (Clinical Information System: CIS)과 병원경영정보시스템 (Hospital Management Information System: HMIS)으로 구성된다. IHIS 사용자는 크게 의사, 간호사, 연구자 등을 포함하는 의료인과 원무관리

자, 경영자 등의 비의료인으로 나누어진다. 의료인은 임상정보시스템과 임상 데이터 웨어하우스에 중요성을 두고 있으며, 비의료인들은 병원 사무와 병원경영관점에서 수익증대에 관심을 두고 있다. 이들은 각자 다른 데이터와 분석 목적을 요구한다.

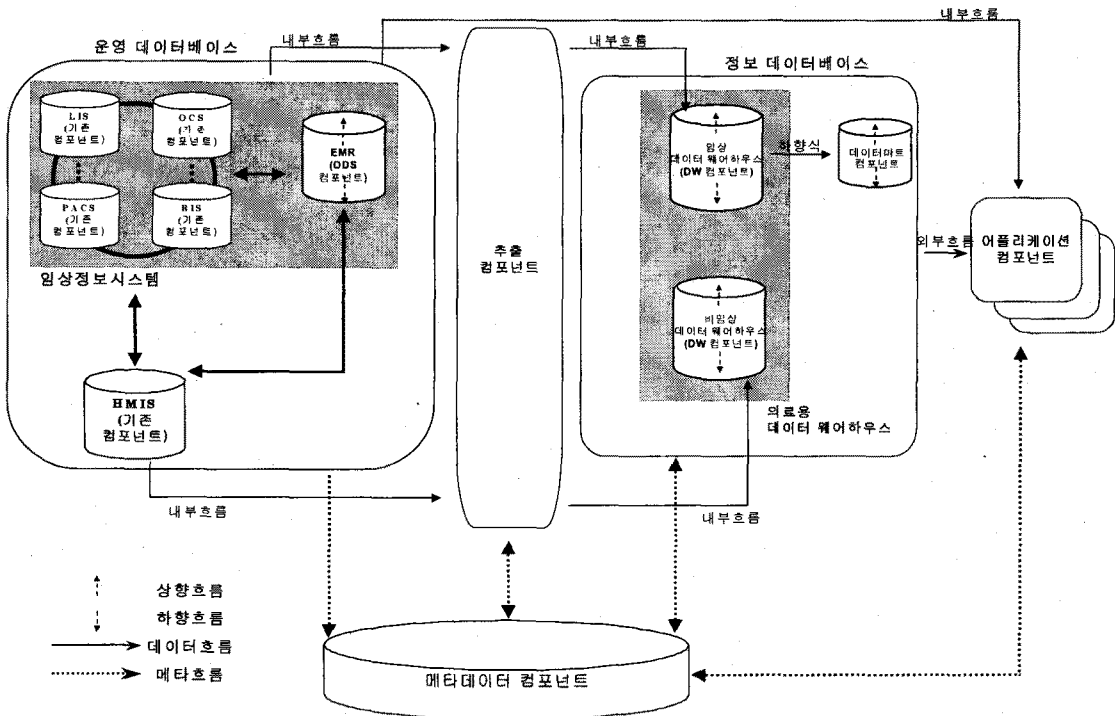
CIS 에는 처방전달시스템 (Order Communication System: OCS), 의료영상저장전송시스템 (Picture Archiving and Communication System: PACS), 전자의무기록 시스템이 근본이 된다. 1994년부터 사 병원은 OCS를 운영하고 있다. OCS 트랜잭션은 입원전 (Pre-admission) 기록, 입원환자 (Inpatient) 및 외래환자 (Outpatient) 등록, 신생아 정보, 환자 정보, 입원/퇴원/이송 (Admission/Discharge/Transfer: ADT) 정보를 포함하고 있다. PACS는 1997년부터 운영되고 있으며, HMIS, OCS 및 방사선정보시스템 (Radiological Information System: RIS)과 통합운영하고 있다.

EMR 시스템은 임상 데이터 웨어하우스의 핵심

요소이다 [Nowack and Nicolai, 1997; Nussbaum, 1998; Pedersen and Jensen, 1998; Rardin, 1996]. 현재는 정형외과에서만 시험적으로 운영되고 있다. 국내 대부분의 병원에는 EMR 시스템이 설치되어 있지 않은 실정이다. 국내 병원 전산화 추세는 OCS에서 EMR 시스템으로 확대되어가고 있다 [조훈, 1998]. 다른 중요한 CIS 시스템에는 RIS 와 임상병리정보시스템이 있다.

4.2 의료용 데이터 웨어하우스 아키텍처 및 구현

프로젝트 팀은 <그림 4>에 제안된 DW 아키텍처에 근간하여 의료용 데이터 웨어하우스 아키텍처를 <그림 5>와 같이 도출하였다. ODS 컴포넌트와 유사한 EMR 시스템은 DW를 위한 토대로써 사용되어진다 [Pedersen and Jensen, 1998]. EMR 시스템내의 모든 환자 정보는 구조화된 형태로 존



<그림 5> 의료용 데이터 웨어하우스 아키텍처

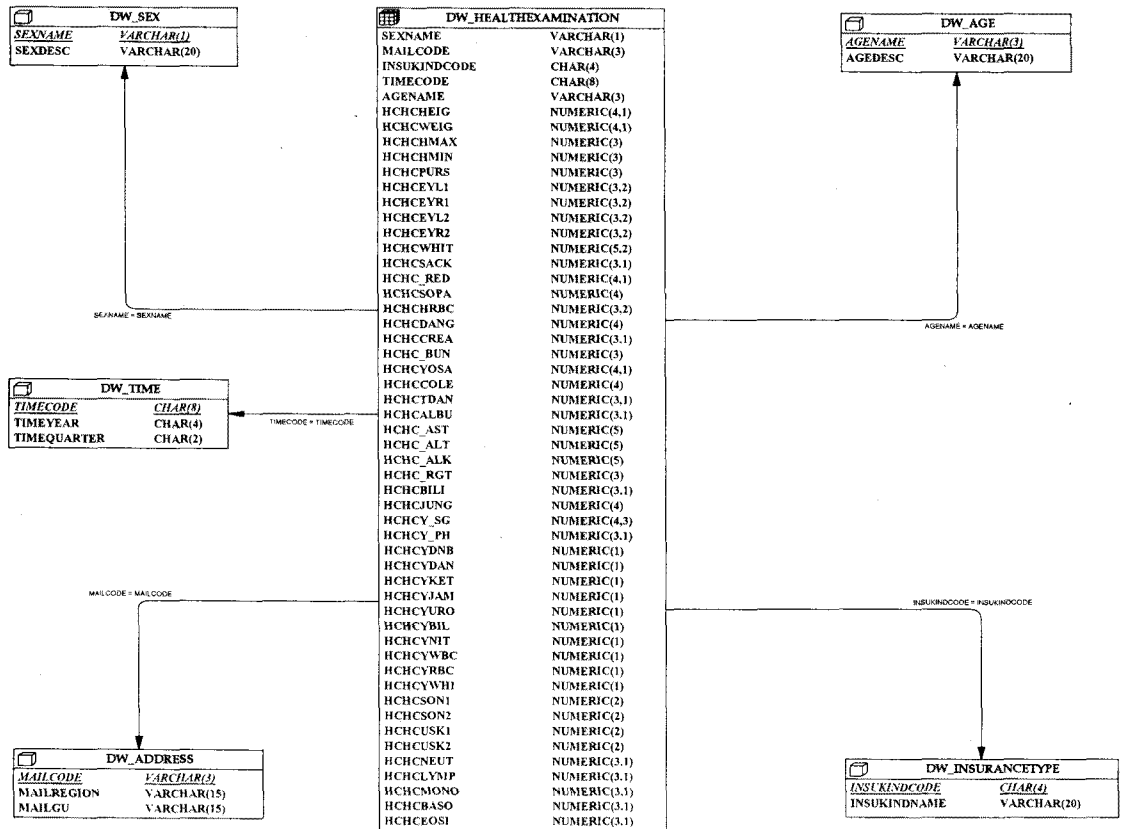
재한다. 즉, EMR 데이터는 어떤 환자의 건강 상태를 파악하기 위하여, 환자에 대한 데이터를 환자중심으로 저장한다 [Pedersen and Jensen, 1998].

DM 은 각 과의 특정한 관점을 위해서 필요하다. <그림 5>에서, 내부흐름은 임상정보시스템과 병원경영정보시스템으로부터 임상 및 비임상 DW 로의 흐름을 가지고 있다. 외부흐름은 정보 데이터베이스로부터의 흐름을 가진다. 운영시스템에서 EMR 시스템과 OCS, RIS, PACS, LIS의 통합, 더 나아가서 H MIS 와의 통합은 어려운 작업이다. 그러나, 이러한 운영시스템은 의료인과 비의료인을 지원하기 위해서 단단하게 (Tightly) 통합되어야한다. 현재, S 병원에서 이러한 통합 작업을 진행중이다.

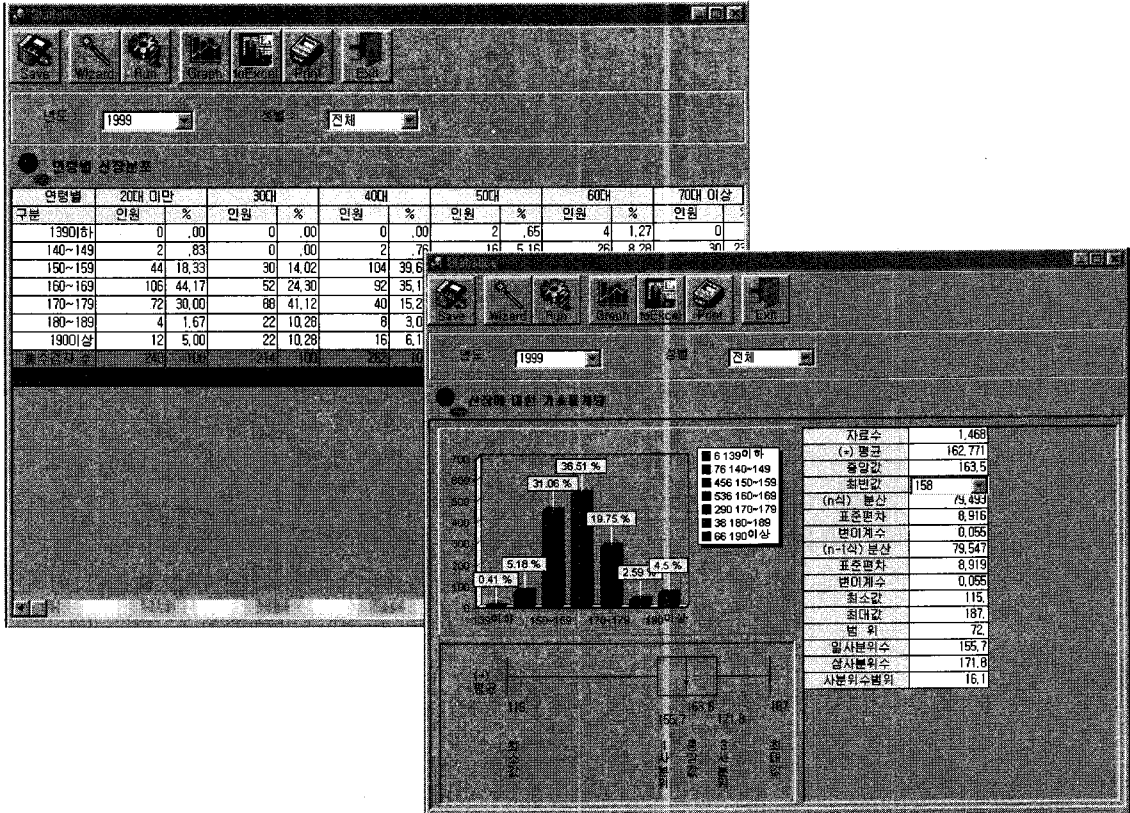
S 병원 의료용 데이터 웨어하우스 아키텍처

는 EMR 시스템이 개발되어 있지 않기 때문에, ODS 컴포넌트를 포함하지 않았다. 데이터흐름을 위한 분류에 따라, 적극적 (내부흐름), 단일 (자체흐름) 및 OLAP (외부흐름) 유형을 채택하였다. 메타흐름에 의한 분류에 따라서는, 독립적 (위치), 비즈니스 (사용) 및 집중화된 (분산) 유형을 채택하였다. 임상 데이터 웨어하우스의 소스는 S 병원의 임상정보시스템이며, 비임상의 소스는 병원경영정보시스템이다.

의료용 데이터 웨어하우스는 싸이베이스 IQ 12 DBMS 엔진에서 운영되고 있으며, 800 기가바이트 용량을 가진 썬 울트라 엔터프라이즈 450 상에서 작동되고 있다. 임상 데이터 웨어하우스는 종합검진 등을 포함하는 다양한 주제를 포함하고 있다. <그림 6>은 임상 데이터 웨어하



<그림 6> 임상 데이터 웨어하우스를 위한 스타 스키마 일부



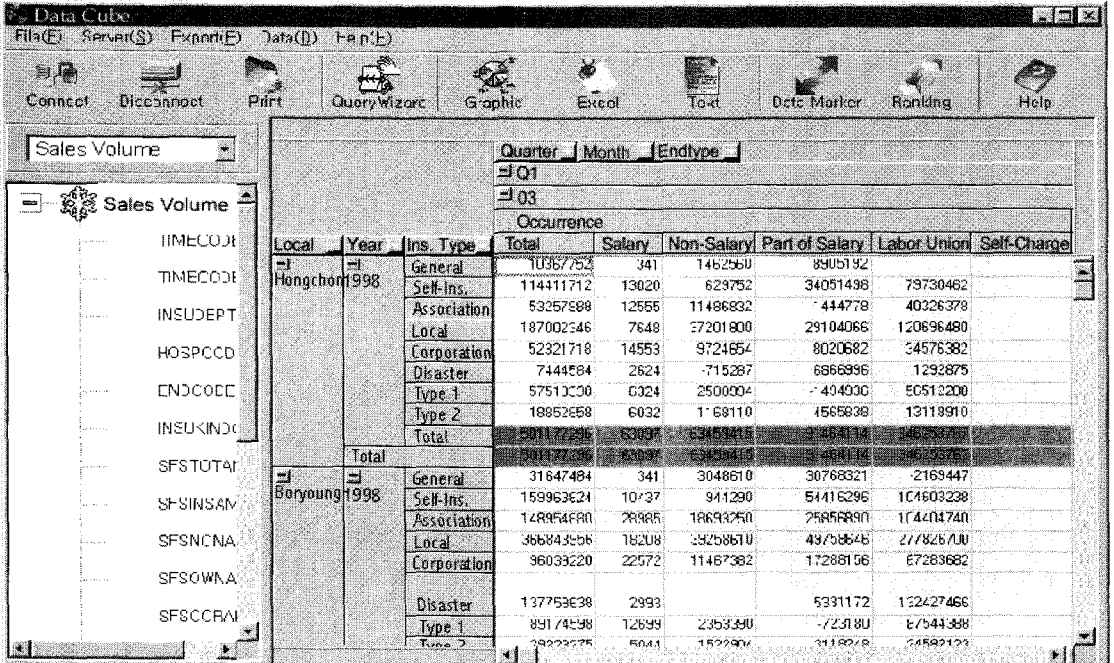
<그림 7> 종합검진을 위한 분할표와 기초통계량 결과

우스에서 종합검진에 대한 스타 (Star-shaped) 스키마의 일부를 보여주고 있다. 스키마의 시간 입도 (Granularity)는 분기-기반 (Quarter-based) 이다. 비임상 데이터 웨어하우스 스타 스키마 일부는 [부록 F]에 요약되어 있다.

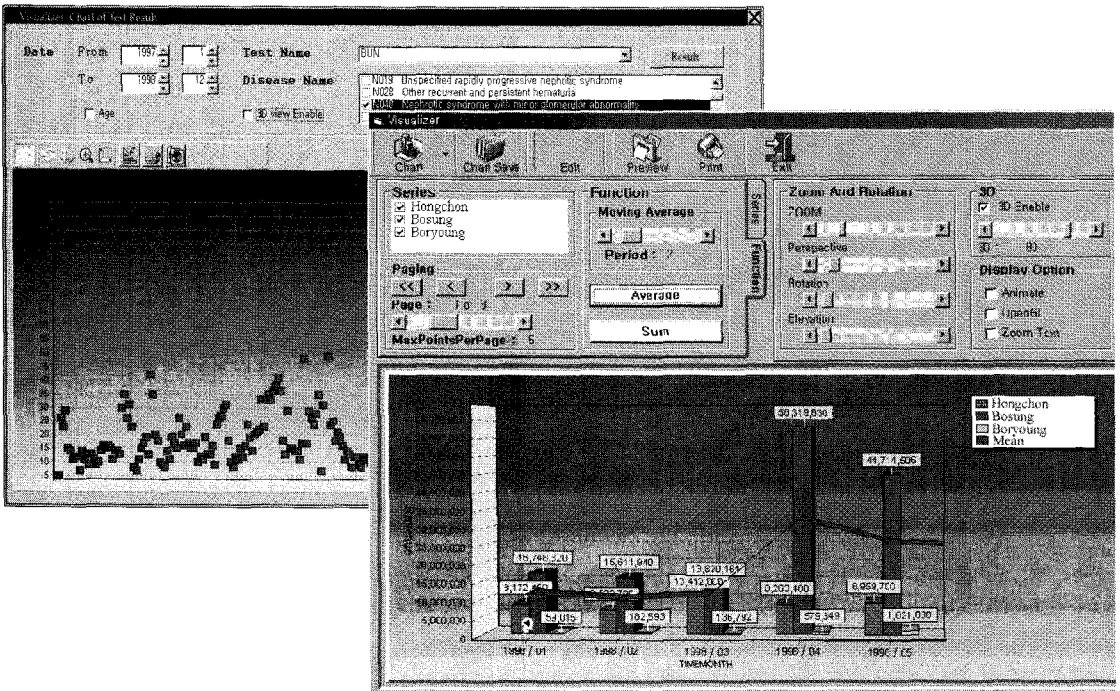
<그림 6>에서 차원 테이블 (Dimension Table)의 키에는 밑줄이 그어져 있다. 사실 테이블 (Fact Table)의 키는 차원테이블들의 키들로 구성되어진다. 이 스키마는 한 개의 사실 테이블과 다섯 개의 차원테이블로 구성되어있다 (DW_SEX, DW_TIME, DW_ADDRESS, DW_AGE, 및 DW_INSURANCETYPE). DW_HEALTHEXAMINATION 사실 테이블은 HCHCCOLE (총 콜레스테롤), HCHCEYL1 (첫번째 왼쪽 시력검사 결과), HCHCYBIL (빌리루빈 검사 결과), HCHCYURO (요

빌리루빈 검사 결과), HCHCYKET (케톤 검사 결과)의 차원 테이블에 의해서 요약되어진다.

<그림 7>은 MS 비주얼 베이직 6.0으로 개발된 어플리케이션 컴포넌트의 일부를 나타낸다. 메타데이터에 근간을 둔 마법사 기능 (Wizard Function)을 통해, 2X2 분할표가 구성가능하다. 각 셀 (Cell)은 성, 시간, 주소, 연령, 및 보험유형과 같은 차원에 의한 요약데이터를 나타낸다. OLAP 도구는 드릴다운 (Drill-down), 드릴업 (Drill-up), 드릴어크로스 (Drill-across), 및 피벗팅 (Pivoting)의 기능을 가지고 있다. 이 도구는 메타데이터를 기반으로 한다. 즉, 정보 (예를 들면, 사용자, 요소, 레코드, 차원)를 메타스토어로부터 추출해서 사용한다. <그림 7>에서 연령별 신장분포가 분할표로 나타나있다.



<그림 8> 데이터 큐브 도구의 일부



<그림 9> 시각화 도구의 일부

어플리케이션 컴포넌트는 통계 (Statistics), 증례 (Case Management), 데이터 큐브 (Data Cube) 및 시각화 (Visualizer) 도구로 구성된다. 통계 도구는 기초 또는 고급 기법 (예를 들면, 회귀 분석이나 분산분석 기법) 을 사용한다 (예를 들면, <그림 7>). 증례 도구는 환자의 상태가 괜찮은지, 비용 효과적으로 환자가 돌봐지고 있는지를 점검하는 데 사용한다 [Murder and Brick, 1997]. 데이터 큐브 도구는 사용자에게 다차원 데이터를 제공하는 것이 주목적이다 (예를 들면, <그림 8>). 시각화 도구는 의료인에게 환자 데이터에 대해서 도식적 관점을 제공하는 기능을 가지고 있다 (예를 들면, <그림 9>). <그림 8>과 <그림 9>는 마법사 기능을 통해서 수행되며, 메타데이터를 근간으로 개발되었다.

4.3 메타데이터 컴포넌트와 프로토타입

도메인에 고유한 요구사항과 추가적인 요구사항이 의료인, 비의료인 및 기술자와의 인터뷰를 통해서 조사되었다. 우선, 보안에 대한 강화가 요구되었다 (예를 들면, 그룹에 대한 권한). 데이터 웨어하우징 시스템의 보안 요구사항은 다른 분산 컴퓨팅 시스템의 경우와 유사하지만, DW 개발과정에서 보안에 대한 고려는 잘 이루어지지 않고 있다 [Finne, 1997; Katic et al., 1998; Warigon, 1997]. 본 연구에서는 사용자, 보안 객체, 보안 유형 및 권한으로 이루어진 임의의 (Discretionary) 보안 모형 [Essmayr et al., 1998; Kirkgoze et al., 1997]을 확장하여 채택하였다. 보안 데이터도 일종의 메타데이터인 것이다 [Katic et al., 1998; Nussbaum, 1998]. 또한, 사용자가 좋아하는 질의, 리포트, 차트 및 다른 객체의 트리 구조를 어플리케이션 화면에서 가지길 원하는 점이 주목되었다. 이 기능은 인터넷 브라우저에서 책갈피 (Bookmark) 기능과 유사한 것이다. 이러한 정보를 '서브젝트 (Subject)' 라고 명명한 메타데이터 엔터티로 만들었다. 나아가

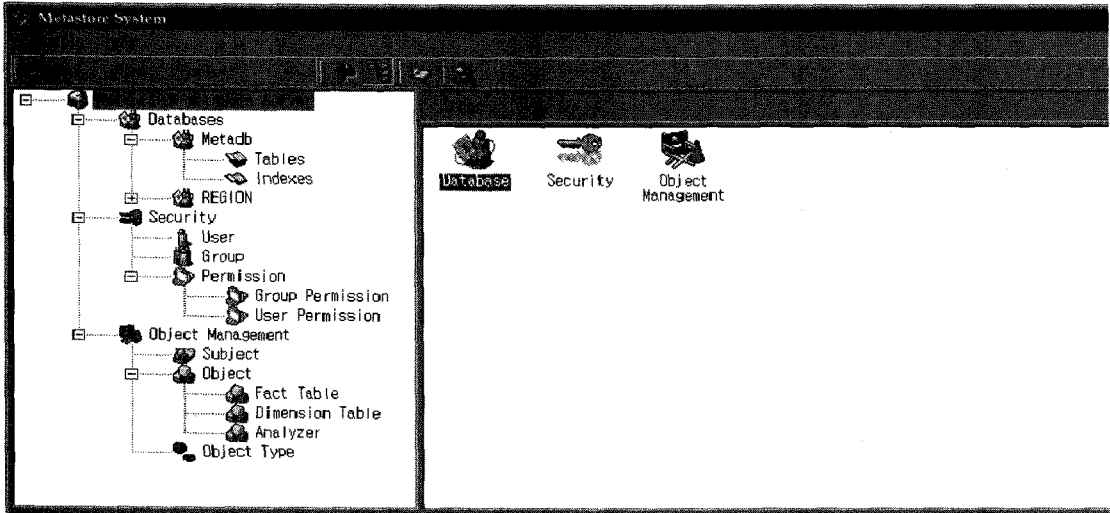
서, 메타스토어 시스템을 쉽게 유지하기 위해서, 기술자들은 레코드, 차원, 요소 엔터티를 '객체 (Object)' 엔터티로 이동시키길 원했다. 의료인들은 데이터 마이닝을 원하였으나, 실제로는 메타데이터를 근간으로 한 OLAP 도구가 채택되었다. 현재, 통계적 데이터 마이닝 모듈은 개발되어 있다. 이상에서 언급한, 추가적 요구사항을 핵심 메타데이터 스키마에 추가한 후, 최종 메타데이터 스키마가 도출되었다.

<그림 10>은 메타데이터를 관리하고 탐색하는 메타스토어 프로토타입 시스템을 나타낸다. 이것은 MS 비주얼 베이직 6.0을 이용해서 개발되었으며, 메타스토어는 의료용 데이터 웨어하우스와 같은 하드웨어의 싸이베이스 IQ 12 DBMS 엔진에 탑재되어 있다.

시스템의 주 메뉴는 'Database,' 'Security,' 및 'Object Management' 로 구성되어 있다.

'Database'는 메타스토어에 저장된 메타데이터를 관리한다. 'Security'는 사용자와 그룹 정보를 나타낸다. DW 관리자만이 새로운 사용자와 그룹을 등록시킬 수 있으며, 기존 사용자와 그룹을 삭제시킬 수 있다. 'Permission'는 객체를 사용자와 그룹을 대응시킨다. 이 시스템은 개인 사용자가 어플리케이션 컴포넌트에 로그인할 때, 사용자 또는 그룹으로 로그인 하는지를 점검한다. 사용자로 로그인하면 어플리케이션은 사용자 프로필, 서브젝트, 객체, 및 권한 메타데이터를 메타스토어로부터 가지고 와서, 사용자의 어플리케이션 화면에 나타난다. 각 사용자 또는 그룹은 기능 메뉴와 객체의 서브젝트로 구성된 스크린 화면을 가진다. 'Object Management'는 테이블, 애트리뷰트, 및 분석과 같은 객체를 관리한다. 'Analyzer'는 통계, 증례, 데이터 큐브 및 시각화 기능에 관한 메타데이터를 최신의 상태로 유지한다.

핵심 메타데이터 스키마 내의 질의 및 리포트 엔터티는 통계 기능관련 엔터티로 통합하였다. 'Fact Table'과 'Dimension Table' 내의 테이블



<그림 10> 메타스토어 프로토타입 시스템

과 애틀리뷰트는 사용자에게 할당되어진다.

게다가, 신경망 알고리즘과 같은 새로운 객체는, 언제든지 'Analyzer' 에게로 도입될 수 있다.

메타데이터 컴포넌트는 사 병원 사용자의 요구에 따라 지속하여 보장될 것이다. 현재, 메타데이터에 근간한 데이터 마이닝 어플리케이션 컴포넌트의 개발은 계획중에 있으며, 회귀분석과 분산분석과 같은 통계적 데이터 마이닝 모듈은 개발중이다.

4.4 토론

본 사례를 통하여 데이터 웨어하우징 시스템을 구축하려는 계획을 가질 때, 병원 경영자들이 고려할 주요사항은 다음과 같이 파악되었다. 첫째, 임상 데이터 웨어하우스의 성공적인 구축을 위해서는 EMR 시스템이 선행되어 개발될 필요가 있다. 운영데이터가 나쁜 품질의 데이터 라면, 대응하는 EMR 시스템의 품질 또한 좋지 않을 것이다. 사 병원이 전사적 EMR 시스템을 도입하려는 계획을 가지고 있는 것은 다행스러운 일이다.

둘째, 명명 (Naming)과 코드 표준화가 중요하

다. 본 사례에서는, 많은 운영데이터들이 코드화 되어 있지 않았고, 자유로운 메모형식으로 저장 되어 있었다. 의료인들은 그들 자신의 분류체계를 가지길 원하기 때문에 현행의 질병 분류 표준, ICD-10 (International Classification Disease 10th Revision),을 사용하지 않으려는 경향이 있다. ICD-10 으로는 새로운 분류 (예를 들면, 새로 개발된 외과 수술법 등) 기준을 수용하기가 어렵다. 그러므로, 진보된 분류 체계가 임상 부문에서 개발되어질 필요가 있다 [Pedersen and Jensen, 1998].

DW는 응용측면과 기술측면에서 병원 경영에 영향을 미친다. 우선, 응용측면에서는 의약품의 남용과 오용이 의료인들의 처방비교에 의해서 줄어들 수 있을 것이다. 최근 WHO (World Health Organization)의 보고에 의하면 국내의 항생제 남용이 세계 최고 수준이라고 밝혔다. 1997년 기준으로 서태평양지역 12개 국가에서 페렴 구균에 대한 페니실린 내성률을 조사한 결과 한국이 84%로 세계 최고이다. 서태평양지역 15개 국 가운데 임균에 대한 페니실린 내성률은 한국이 91%로 필리핀 95%에 이어 두번째이며 일본은 4%였다. 이 경우처럼 항생제 사용 적절성

비교와 같은 임상의학 분야의 질 (Quality) 판단지표는 현 시스템에서 통합된 관점을 제공하지 못하기 때문에 데이터 웨어하우스를 통해서 지속적으로 측정되고 평가되어 질 수 있다. 게다가, DW는 단면적 연구, 증례 연구 및 코호트 연구를 위해서 사용되는 연구 데이터를 의료인들에게 제공할 수 있다. 이러한 연구들로부터 나온 결과들은 환자들에 대한 서비스를 향상시키고 학술적 성취를 이루게끔 한다. 2000년에 시행되는 의약분업화는 많은 병원들에게 재정적 불안정을 야기시킬 수 있다. DW에서 가능한 지표들은 병원 경영 비용 (예를 들면, 인건비)을 감소시키는 것을 도울 수 있을 것이다.

기술적 측면에서는, 메타데이터 기반 아키텍처는 메타데이터 관리 및 어플리케이션을 위한, 값비싸고 상용의 도구를 필요로 하지 않는다. DW 프로젝트 팀은 메타데이터 기반의 어플리케이션 컴포넌트를 쉽게 개발할 수 있다. 데이터 마이닝을 위한 메타데이터로 확장시킬 수 있으며, 이는 유지비용을 절감시킬 수 있다. 본 메타데이터 기반의 아키텍처는 다른 기업의 유사 프로젝트에 적용할 수 있을 것이다.

V. 결 론

오늘날 정보 처리분야에서, 데이터 웨어하우스는 기업 정보 통합의 필요불가결한 수단이다. 특히, 메타데이터 컴포넌트는 데이터 웨어하우스 시스템을 성공적으로 구현하고, 이용하는 데 중요하다. 본 연구는 메타데이터를 포괄하는 DW

구축용 아키텍처와 이를 위한 메타데이터 스키마를 제안하였다.

본 연구의 주요 공헌은 다음과 같다. 첫째, DW 아키텍처의 제안이다. 본 아키텍처는 7개의 컴포넌트로 구성되어 있으며, 데이터흐름과 메타흐름의 분류에 근간하고 있다. 둘째, 데이터 웨어하우스 시스템을 위한 메타데이터 모형의 제안이다. 모형은 메타데이터 표준, 컴포넌트 정보 및 추가적인 요구사항으로 구성되었다. 셋째, 제안된 아키텍처를 의료용 데이터 웨어하우스 프로젝트에 적용하였다. 메타데이터에 근거한 어플리케이션 컴포넌트와 메타스토어 프로토타입 시스템이 개발되어 실용성이 입증되었다.

향후의 몇 가지 연구분야는 다음과 같다; 우선, 웹 (Web)을 통한 메타데이터 컴포넌트 개발이다 [Denzer and Guttler, 1996; Ivy, 1997; Kurz and Tjoa, 1997]. 웹 브라우저 (Browser)는 사용하기 쉽고, 저렴한 가격 때문에 경쟁력이 있다. 둘째, 의료용 DW 에서의 연구 이슈 (예를 들면, 복잡한 데이터 모델링, 고급의 분류기법, 임상 프로토콜을 위한 지원, 표준화 작업) 이다 [Pedersen and Jensen, 1998]. 현재의 DW는 이러한 이슈의 대부분을 지원하지 않고 있다. 셋째, 본 연구의 아키텍처와 메타데이터를 다른 사례에 적용하여 그 확장성을 검증하는 것이다. 예를 들면, 행정 [Weilbach and Viktor, 1999], 전자 [Berra et al., 1998], 생산 [Eicker et al., 1997; Gorla and Krehiel, 1999], 마케팅 [Albrecht et al., 1997] 또는 통신 [Conine, 1998; Shin et al., 1998] 분야에 적용이 가능할 것이다.

〈참 고 문 헌〉

- [1] 김선구, 김종득, 이은주, 백철현, 이해석, 최진욱, 김성권, 진료, 연구를 위한 서울대학교병원 데이터 웨어하우스의 구축 사례, 제16차 대한의료정보학회 춘계학술대회, 2000.
- [2] 김유미, 강성홍, 김원중, 퇴원요약DB를 활용한 병원데이터베이스 마케팅, 제16차 대한의료정보학회 춘계학술대회, 2000.
- [3] 범희승, 박성희, 신찬호, 정환정, 송호천, 핵의학 의료영상 관리를 위한 미니 데이터웨어하우징, 제16차 대한의료정보학회 춘계학술대회, 2000.
- [4] 서동희, 처방전달체계를 이용한 병동 간호업무 정보체계, 대한의료정보학회지 3(1), 1997, pp. 49-65.
- [5] 송근식, 신명진, 이수현, 이주희, 오용호, 서울중앙병원의 PACS 구축 및 운영 현황, 대한PACS 학회지, 4, 1998, pp. 5-12.
- [6] 송운호, 조용구, 강석중, 국방 의료체계 구축방안 연구, 정보화저널 1(2), 1994.
- [7] 이재욱, 곽연식, 민원기, 최진욱, 조한익, 데이터 웨어하우징 개념을 이용한 검진자료의 분석: 데이터 웨어하우징과 검진자료, 대한의료정보학회지, 4(2), 1998, pp. 111-114.
- [8] 이희석, 홍의기, 김태훈, 데이터 웨어하우스를 이용한 임원정보시스템 구축용 아키텍처, 경영정보학연구, 7(1), 1997, pp. 7-24.
- [9] 조훈, 최근 병원정보시스템의 경향: 의료정보의 관점에서, 병원협회지, 1998년 12월.
- [10] Albrecht, J., Lehner, W., Teschke, M., and Kirsche, T., Building a real data warehouse for market research, *Proc. of the 8th IEEE Intl Workshop on Database and Expert Systems Applications*, 1997, pp. 651-656.
- [11] Atre, S. and Storer, P., Data distribution and warehousing, *DBMS*, 10, 1995, pp. 54-62.
- [12] Barquin, R.C. and Edelstein, H.A., *Planning and Designing the Data Warehouse*, Prentice Hall, 1997.
- [13] Bates, D.W., Pappius, E., Kuperman, G., Sittig D., Burstin, H., Fairchild, D., Brennan, T.A., and Teich, J.M., Using information systems to measure and improve quality, *International Journal of Medical Informatics*, 53, 1999, pp. 115-124.
- [14] Berndt, D.J. and Hevner, A.R., Hospital discharge transactions: a data warehouse component, *Proc. of the 33rd IEEE Annual HICSS*, 2000.
- [15] Bernstein, P.A. and Bergstraesser, T., Metadata support for data transformations using Microsoft Repository, *IEEE Bulletin of the Technical Committee on Data Engineering*, 22 (1), 1999, pp. 9-14.
- [16] Berra, P.B., Mitkas, P.A., and Liuzzi, R.A., Dynamic data mining using an electro-optical data warehouse, *IEEE Information Technology Conference*, 1998, pp. 83-86.
- [17] Blobel, B. and Holena, M., Comparing middleware concepts for advanced healthcare system architecture, *International Journal of Medical Informatics*, 46(1), 1997, pp. 69-85.
- [18] Burton, P. and Green, S., Metadata: the key to data warehouse design, *Working Paper*, The Institute for Systems Research, The University of Maryland, 1996.
- [19] Chaudhuri, S. and Dayal, D., An overview of data warehousing and OLAP technology, *SIGMOD Record*, 26(1), 1997, pp. 65-69.
- [20] Conine, R., The data warehouse in the telecommunications industry, *Proc. of the 1998 IEEE Network Operations and Manage-*

- ment Symposium, 1998, pp. 205-209.
- [21] Demarest, M., Building the data mart, *DBMS*, 7(8), 1994, pp.44-50
- [22] Denzer, R. and Guttler, R., Requirements of the meta information models for the environmental domain, *Journal of Computing and Information*, 2(1), 1996, pp. 1328-1335.
- [23] Devlin, B., *Data Warehouse from Architecture to Implementation*, Addison Wesley, 1997.
- [24] Duyn, J.V., *Developing a Data Dictionary System*, Prentice Hall, 1982.
- [25] Eicker, S., Jung, R., Nietsch, M., and Winter, R., Development of data warehouses for production controlling systems, *PICMET 97*, 1997, pp. 725-728.
- [26] Essmayr, W., Kapsammer, E., and Tjoa, A.M., Meta-data for enterprise-wide security administration, *Proc. of the 3rd IEEE Metadata Conf.*, 1999.
- [27] Finne, T., What are the information security risks in decision support systems and data warehousing? *Computer & Security*, 16(3), 1997, pp. 197-204.
- [28] Gardner, S.R., Building the data warehouse, *Communications of the ACM*, 41(9), 1998, pp. 52-60.
- [29] Gill, H.S. and Rao, P.C., *The Official Guide to Data Warehousing*, Que, 1996.
- [30] Gorla, N. and Krehiel, S., Summers Rubber Company designs its data warehouse, *Interfaces*, 29(2), 1999, pp. 104-117.
- [31] Hackathorn, R.D., Data warehousing energizes your enterprise, *Datamation*, 1995, pp. 38-42.
- [32] Hackney, D., *Understanding and Implementing successful Data Marts*, Addison Wesley, 1997.
- [33] Hoven, J.V.D., Data marts: plan big, build small, *Information Systems Management*, 1998, pp. 71-73.
- [34] Inmon, W.H. and Hackathorn, R.D., *Using the Data Warehouse*, John Wiley, 1994.
- [35] Inmon, W.H., Zachman, J.A., and Geiger, J.G., *Data Stores, Data Warehousing, and the Zachman Framework*, McGraw-Hill, 1997.
- [36] Ivy, W.C., Advances in DSS metadata delivery via the WWW, *CAUSE Annual Conf.*, 1997.
- [37] Jarke, M., Jeusfeld, M.A., Quix, C., and Vassiliadis, P., Architecture and quality in data warehouse- an extended repository approach, *Information Systems*, 24(3), 1999, pp. 229-253.
- [38] Katic, N., Quirchmay, G., Schiefer, J., Stolba, M., and Tjoa, A.M., A prototype model for data warehouse security based on metadata, *Proc. of the 9th IEEE Intl Workshop on Database and Expert Systems Applications*, 1998, pp. 300-308.
- [39] Kelly, S., *Data Warehousing in Action*, John Wiley, 1997.
- [40] Kimura, M., Ohe, K., Yoshihara, H., Ando, Y., Kawamata, F., Tsuchiya, F., Furukawa, H., Horiguchi, S., Sakusabe, T., Tani, S., and Akiyama, M., MERIT-9: a patient information exchange guideline using MML, HL7 and DICOM, *International Journal of Medical Informatics*, 51(1), 1998, pp. 59-68.
- [41] Kirkgöze, R., Katic, N., Stolba, M., and Tjoa, A.M., A security concept for OLAP, *Proc. of the 8th IEEE Intl Workshop on Database and Expert Systems Applications*, 1997, pp. 619-626.
- [42] Kosar, D., Architectural framework for building a warehouse, *DB2 Magazine*, 1996, pp. 10-15.
- [43] Kurz, A. and Tjoa, A.M., Data warehousing

- within intranet: prototype of a web-based executive information system, *Proc. of the 8th IEEE Intl Workshop on Database and Expert Systems Applications*, 1997, pp. 627-632.
- [44] Kutsche, R. and Sünbül, A., A meta-data based development strategy for heterogeneous, distributed information systems, *Proc. of the 3rd IEEE Metadata Conf.*, 1999.
- [45] Mark, L. and Roussopoulos, N., Metadata management, *IEEE Computer*, 19(12), 1986, pp. 26-36.
- [46] McFadden, F.R., Data warehouse for EIS: some issues and impacts, *Proc. of the 29th IEEE Annual HICSS*, 1996, pp. 120-129.
- [47] Metadata Council, *Version 1.1 Metadata Interchange Specification (MDIS)*, Standard Report, 1997.
- [48] Moriarty, T. and Mandracchia, C., Heart of the warehouse, *Database Programming and Design*, 1996, pp. 70-74.
- [49] Murder, C.G. and Brick, L.L., *The Case Management Sourcebook: A Guide to Designing and Implementing a Centralized Case Management System*, McGraw-Hill, 1997.
- [50] Murtaza, A.H., A framework for developing enterprise data warehouses, *Information Systems Management*, 1998, pp. 21-26.
- [51] Noaman, A.Y. and Barker, K., Distributed data warehouse architectures, *Journal of Data Warehousing*, 2(2), 1997, pp. 37-50.
- [52] Nowack, W.J. and Niccolai, M.J., The electronic medical record and the data warehouse, *Proc. of the 16th Southern Biomedical Engineering Conference*, 1997, pp. 178-180.
- [53] Nussbaum, G.M., The best little data warehouse, *Journal of Healthcare Information Management*, 12(4), 1998, pp. 79-93.
- [54] Orr, K., Data quality and system theory, *Communications of the ACM*, 41(2), 1998, pp. 66-71.
- [55] Özsu, M.T. and Valduriez, P., *Principles of Distributed Database Systems*, Prentice Hall, 1991.
- [56] Pedersen, T.B. and Jensen, C.S., Research issues in clinical data warehousing, *Proc. of the 10th IEEE Intl Conference on Scientific and Statistical Database Management*, 1998, pp. 43-52.
- [57] Poe, V., *Building a Data Warehouse for Decision Support*, Second Edition, Prentice Hall, 1997.
- [58] Rardin, K.D., Information warehousing within the CPR, *Imaging World Magazine*, 1996, S5-S6.
- [59] Sachdeva, S., Meta data architecture for data warehousing, *Data Management Review*, (3), 1998.
- [60] Shin, S., Park, G., Lee, W., and Lee, S., Case study: how to make telecom pricing strategy using data warehouse approach, *Proc. of the 31st IEEE Annual HICSS*, 1998.
- [61] Sybase Inc., *PowerDesigner DataArchitect - User's Guide Version 6.1.3*, Document Number: AA04631, 1999.
- [62] Warigon, S., Data warehouse control and security, *Association of College and University Auditors LEDGER*, 41(2), 1997, pp. 3-7.
- [63] Weilbach, J.F.J. and Viktor, H.L., A data warehouse for policy making: a case study, *Proc. of the 32nd IEEE Annual HICSS*, 1999.
- [64] White, C., The key to a data warehouse, *Database Programming and Design*, (2), 1995, pp. 23-25.
- [65] Widom, J., Research problems in data warehousing, *Proc. of the 4th Intl Conf. on Information and Knowledge Management*, 1995, pp. 25-30.

[부록 A] 자체흐름 및 외부흐름 비교

<표-1> 자체흐름

정보 데이터 베이스 유형	ODS	DW	DM	설명	복잡성
단일 (Single)	✓			단기간의 의사결정 지원에 적합하며 주로 운영 데이터에 이용	저 ↓ 고
		✓		장기간의 의사결정 지원에 적합하며 전사적 의사결정에 필요한 모든 데이터를 포괄	
			✓	부서별 혹은 기능별로 장기간의 의사결정 지원에 적합하며 독립적 DM 으로서도 불림	
복수 (Double)	✓	✓		운영 데이터와 의사결정 지원 데이터 모두를 제공하며 DW 부하 (Workload) 를 경감하고 장단기의 의사결정 지원에 적합	
			✓	분산 데이터베이스 환경에 적합하며 DW, DM 의 결합 순서는 상하향식 또는 하상향식에 따라 다르게 구성됨 기업 전체의 관점 혹은 부서적 관점에 대한 DW 부하를 경감	
세부분 (Triple)	✓	✓	✓	장단기의 의사결정 지원을 위한 운영 데이터 및 분산된 의사결정 지원 데이터	

<표-2> 외부흐름

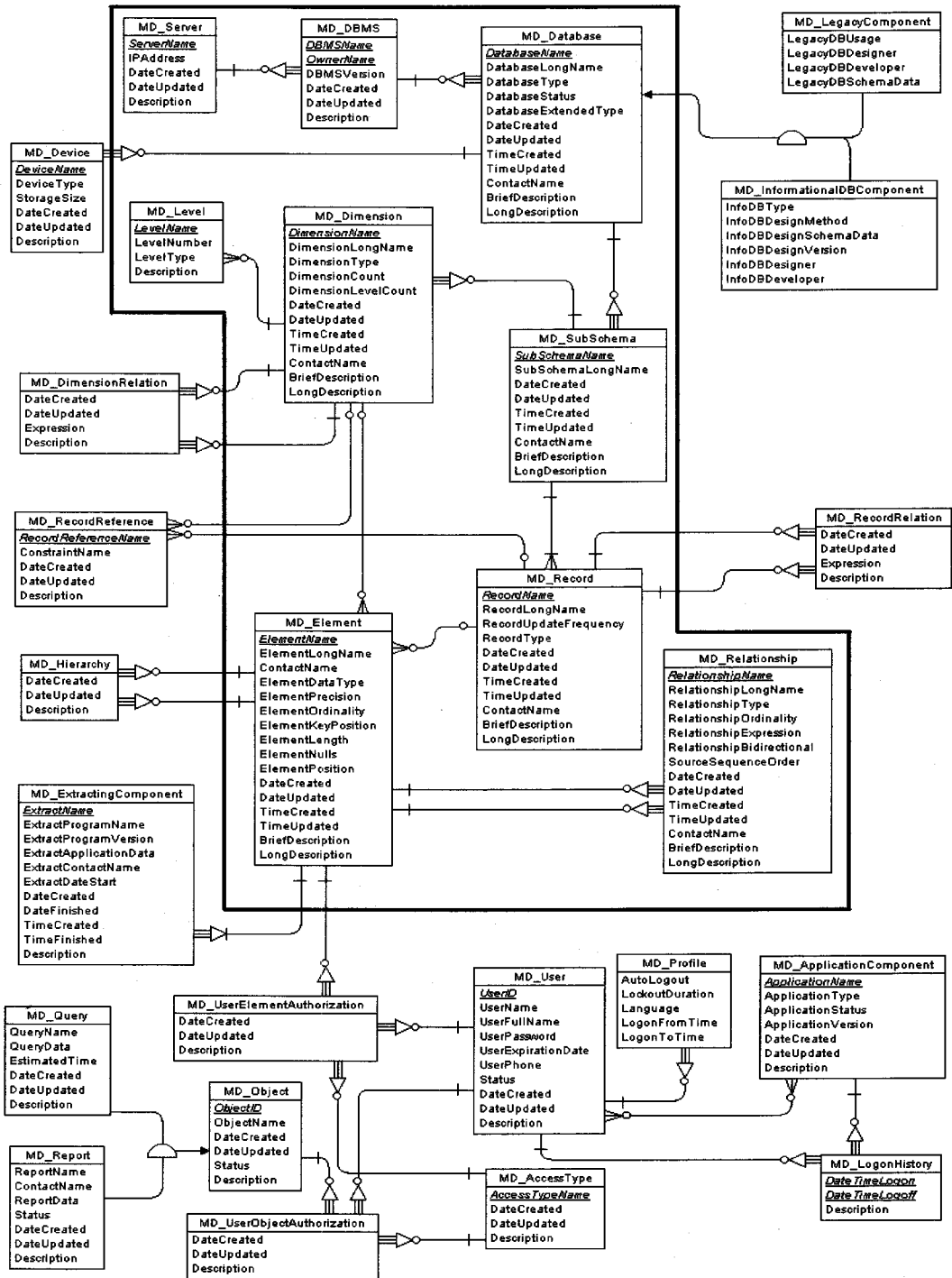
기준 \ 유형	단순접근	OLAP	데이터 마이닝
목적	단순 질의 및 리포팅 데이터 탐색	질의 및 리포팅 데이터 탐색 가설 검증 다차원 시각	패턴 추출 질의 및 리포팅 데이터 탐색 가설 검증 다차원 시각
관점	사용자 관점	사용자 관점	사용자 관점/데이터 관점
사용자 주도	상호작용 (Interactive)	상호작용	상호작용/배치 (Batch)
의사결정 전략	하향식	하향식	하향식/상향식
결과 패턴	정적	정적	정적/동적
분류	Access/Query/Report	ROLAP (Relational OLAP)/ MOLAP (Multidimensional OLAP) /HOLAP (Hybrid OLAP)	데이터 마이닝/정보 에이전트 (Information Agent)
복잡성	저 —————> 고		

[부록 B] 데이터베이스간의 비교

데이터베이스 컴포넌트	운영 데이터베이스	정보 데이터베이스			메타스토어
	기존 (Legacy) 컴포넌트	ODS 컴포넌트	DW 컴포넌트	DM 컴포넌트	메타데이터 컴포넌트
사용자 수	가장 많음	많음	적음	아주 적음	아주 많음
환경	OLTP	OLTP/OLAP	OLAP	OLAP	정보 탐색
사용자 집단	비 경영층	비경영층/ 경영층	경영층	부서내의 경영층	비경영층/ 경영층
요약 데이터	적음	동적 요약 데이터	저장된 요약 데이터	저장된 요약 데이터	적음
데이터 갱신 빈도	아주 높음	높음	낮은	낮은	아주 높음
데이터 관점	비통합적	통합적	통합적 (과거 추적)	통합적 (과거 추적)	통합적 (과거 추적)
의사결정	매우 단기적/ 기업	단기적/기업	장기적/기업	장기적/부서	부서 혹은 기업내의 사용자
정보제공 관계	ODS/DW/DM에 제공	DW에 제공	DM에 제공	DW에 제공	모든 데이터베이스에서 정보 획득

[부록 C] 핵심 메타데이터 스키마

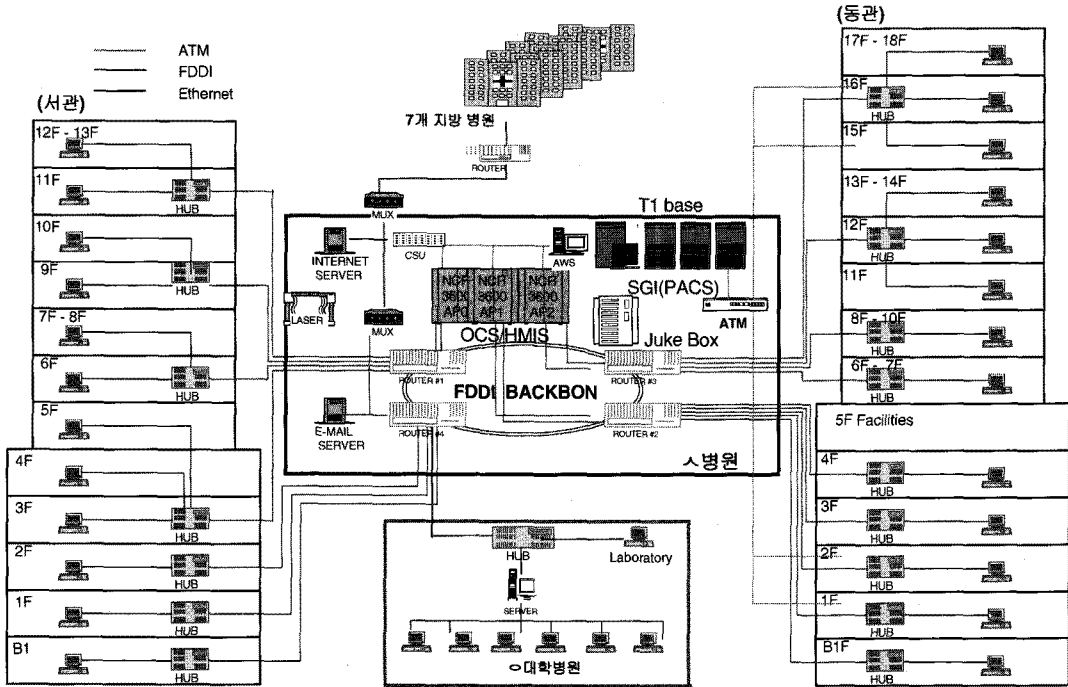
MIS Metamodel



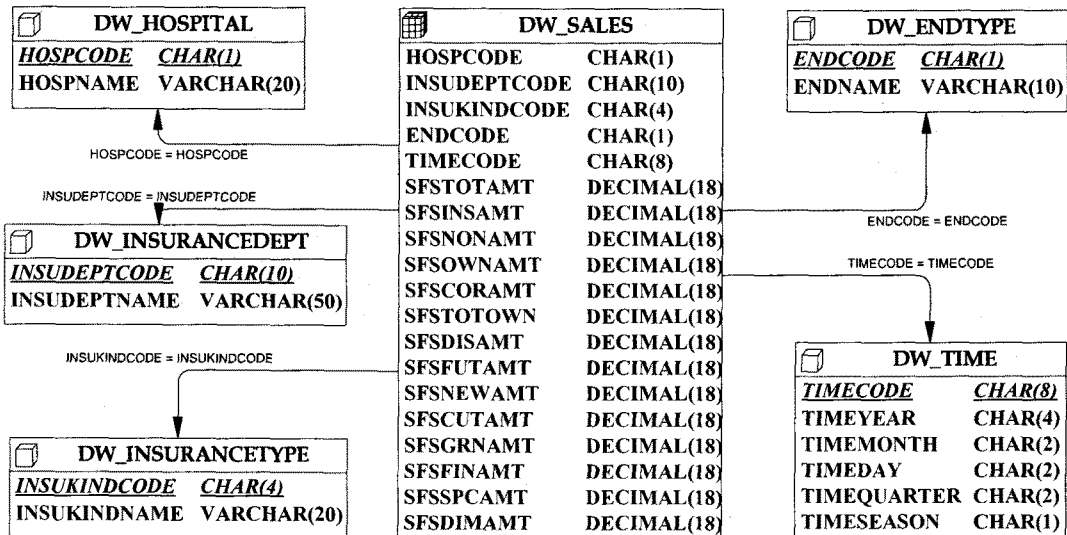
[부록 D] 핵심 메타데이터 스키마 개체 리스트

개체	관련정보
MD_Server	DBMS 엔진이 속하는 서버
MD_DBMS	DBMS 엔진, 버전, 및 관리자
MD_Device	데이터베이스에 대한 장치
MD_Database	파일 그룹, 관계형 데이터베이스, 네트워크 데이터베이스, 계층형 데이터베이스 등
MD_Subschema	관계형 파일의 논리적 그룹과 하나의 데이터베이스의 서브셋을 기술하는 노드의 논리적 그룹 등
MD_Record	관계형 데이터베이스 테이블, 계층형 데이터베이스내의 세그먼트, 네트워크 데이터베이스내의 노드, 다차원 사실 테이블 등
MD_Dimension	차원 테이블
MD_Level	각 요소의 애트리뷰트 수준
MD_Element	컬럼, 필드 등
MD_Relationship	데이터 요소간의 관계 (운영 데이터베이스로부터 정보 데이터베이스로의 맵핑)
MD_LegacyComponent	운영 데이터베이스에 대한 정보 및 DW 소스가 되는 외부 데이터베이스
MD_InformationalDBComponent	ODS, DW 및 DM 과 같은 정보 데이터베이스
MD_DimensionRelation	차원간의 관계
MD_RecordReference	차원과 레코드간의 관계
MD_RecordRelation	레코드간의 정보
MD_Hierachy	한 차원내의 데이터 요소의 계층
MD_ExtractingComponent	추출된 요소 및 데이터를 추출하는 소프트웨어의 기술
MD_UserElementAuthorization	특별한 접근 유형을 적용하는 요소로 사용자 접근여부를 구체화
MD_UserObjectAuthorization	특별한 접근 유형을 적용하는 오브젝트의 사용자 접근여부를 구체화
MD_Object	사용자가 접근하는 수동적 요소
MD_Query	정형화된 질의 및 임시적 질의
MD_Report	어플리케이션으로 생성된 리포트 데이터
MD_User	DW, ODS, 및 DM 에 접근을 원하는 사용자
MD_ApplicationComponent	정보 데이터베이스내의 질의 및 리포트 데이터를 위해서 개발되어지거나 구입된 소프트웨어
MD_Profile	자동 로그오프 시간, 자동 로그인 시간, 로그인 시간, 로그오프 시간의 프로필
MD_AccessType	요소 및 오브젝트에 행해질 수 있는 접근 유형
MD_LogonHistory	로그온 시간, 로그오프 시간의 이력
MD_UserMD_ ApplicationComponent	사용자 어플리케이션 (엔터티화된 관계 (Entitized Relationship))

[부록 E] 사병원 네트워크 구성도

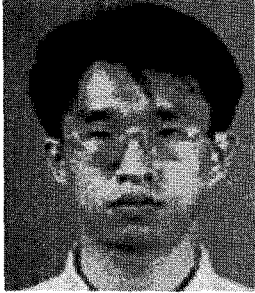


[부록 F] 비임상 데이터 웨어하우스를 위한 스타 스키마 일부



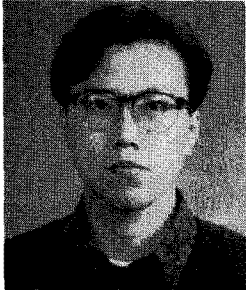
◆ 이 논문은 2000년 3월 25일 접수하여 1차 수정을 거쳐 2000년 7월 24일 게재확정 되었습니다.

◆ 저자소개 ◆



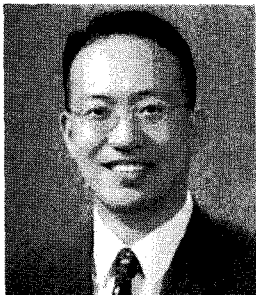
김태훈 (Kim, Taehoon)

공동저자 김태훈은 고려대 통계학학사 및 동 대학원 석사를 취득하고 현재는 KAIST 테크노경영대학원 경영공학 박사과정을 이수중이다. 주요 관심 분야는 데이터 웨어하우스 구축, 분산 데이터베이스 설계, 의료정보시스템 아키텍처, 데이터베이스 설계 등이다.



김종호 (Kim, Jongho)

공동저자 김종호는 KAIST 경영정책학사 및 동 대학원 MIS 석사를 취득하고 현재는 KAIST 테크노경영대학원 경영공학 박사과정을 이수중이다. 주요 관심 분야는 하이퍼미디어 설계, 의료정보시스템 아키텍처, 데이터 웨어하우스 구축 등이다.



이희석 (Lee, Heeseok)

공동저자 이희석은 서울대 산업공학학사, KAIST 산업공학 석사, 그리고 University of Arizona at Tucson 에서 경영정보시스템을 전공하여 경영학박사 학위를 취득하였으며, 1992년에서 1994년까지 University of Nebraska at Omaha 에서 MIS 강의를 하였다. 현재는 KAIST 테크노경영대학원 부교수로 재직하고 있다. 주요 관심분야는 인터넷 비즈니스, 지식경영, 정보전략, ERP, 데이터베이스이다.