

# 데이터 웨어하우스 메타데이터 스키마 구현: H 회사 사례

## Developing A Data Warehouse Metadata Schema: H Company Case

|                       |                  |
|-----------------------|------------------|
| 이 희 석 (Heeseok Lee)   | 한국과학기술원 테크노경영대학원 |
| 조 남 철 (Nam Cheal Cho) | 한솔텔레콤 SI사업팀      |
| 손 명 호 (Myung Ho Sohn) | 한국과학기술원 테크노경영대학원 |
| 김 태 훈 (Taehun Kim)    | 한국과학기술원 테크노경영대학원 |

### 목 차

|                      |                           |
|----------------------|---------------------------|
| I. 서 론               | IV. 데이터 웨어하우스 메타데이터 구축 사례 |
| II. 데이터 웨어하우스와 메타데이터 | V. 결론                     |
| III. 데이터 웨어하우스 메타데이터 |                           |

**Keywords:** Data Warehouse, Metadata, 통신 데이터 웨어하우스 구축 사례

## I. 서 론

1990년대 이후 급속한 경영 환경의 변화와 범 세계적 경쟁의 심화는 기업의 경영 활동에 대한 근본적인 변화를 초래하였다. 더욱 치열해지는 경쟁에서 생존하기 위해서는 경영환경에 대한 정확하고 신속한 분석과 그 결과에 따른 빠른 대응이 필수적이다. 그러나 온라인 거래처리 시스템 (On Line Transaction Processing)으로 대별되는 기존 정보시스템은 분석적인 처리에 부적절하였다. 분석적인 업무를 지원하기 위해서는 방대한 양의 자료를 통합하고 효율적으로 접근 분석할 수 있는 기능이 필요하다.

이와 같이 분석적 업무의 필요성에 의해 등장한

것이 데이터 웨어하우스 (Data Warehouse) 이다. 데이터 웨어하우스는 의사 결정자를 위한 흠이 없고 일관되며 적절한 자료를 제공하기 위하여 등장하였다. 데이터 웨어하우스는 한정된 일부의 사용자를 대상으로 하는 종래의 의사결정 지원시스템 (Decision Support System) 과 달리 일반 사용자에게도 의사결정을 위한 정보를 제공할 수 있다. 의사결정의 권한이 기업의 모든 단계로 확장되는 추세로 인하여 폭 넓은 사용자에게 의사결정을 지원할 수 있는 정보를 제공할 수 있는 데이터 웨어하우스는 최근 몇 년간 여러 산업 분야에 걸쳐 구축되고 있다.

그러나 이러한 장점에도 불구하고 데이터 웨어하우스의 효과적 구축을 위하여는 다음의 문제가 해결

되어야 한다. 첫째로 데이터 웨어하우스는 분석적인 본래의 기능을 수행하기 위해 방대한 양의 과거 이력 정보를 유지하고 관리해야 한다. 둘째로 여러 곳에 분산되어 있는 이질적인 원천 정보를 통합하여야 한다. 셋째로 막대한 개발 및 유지 보수 비용이 소요된다. 마지막으로 적절한 응답 시간을 보장하여야 한다. 즉 기존 시스템에 분산되어 있던 정보를 통합 저장하는 데이터 웨어하우스의 성격상 다양하고 이질적인 다량의 정보가 포함되므로, 그 관리적 측면에서 필요한 복잡하고 많은 작업의 효율적인 수행이 필수적이다. 또한 데이터 웨어하우스에 저장된 정보를 활용하는 사용자들에게 정보의 상세한 내용을 설명해 줄 수 있어야 한다.

위와 같은 기능을 수행하는데 필요한 대부분의 정보는 '데이터에 관한 데이터' (Data About Data) 인 메타데이터 (Metadata) 로 해결이 가능하다. 그러므로 효율적인 메타데이터의 관리가 데이터 웨어하우스의 성패를 좌우한다고 볼 수 있다. 그러나 데이터 웨어하우스 관련 여러 제품의 메타데이터가 서로 상이하고 호환이 되지 않는 문제점이 남아있다.

본 연구는 데이터 웨어하우스의 구축과 운영에 중요한 메타데이터 스키마에 초점을 맞추고 있다. 나아가서, 데이터 웨어하우스 관련 제품간의 상이한 메타데이터 스키마에 관계없이 정보를 최대한 공유할 수 있도록 개념적인 차원에서의 메타데이터 스키마를 제안한다. 또한 설계된 메타데이터 스키마를 바탕으로 메타데이터 데이터베이스를 구축하고 사용자가 필요한 메타데이터를 입력 조회할 수 있는 프로토타입 시스템의 개발을 통하여 제안된 메타데이터 스키마의 실용성을 검증한다. 마지막으로 실제로 데이터 웨어하우스를 구축 운영하고 있는 국내 주요 통신업체의 메타데이터 스키마와 비교 분석하여 개선책을 제시하고자 한다.

## II. 데이터 웨어하우스와 메타데이터

데이터 웨어하우스는 Inmon(1993)에 의해 처음으로

제시된 이래 여러 가지 정의가 발표되었다. 데이터 웨어하우스라는 용어를 처음으로 제시한 Inmon 은 '의사결정을 지원하기 위한 주제 지향적으로 통합된 여러 시간대의 갱신 삭제되지 않는 데이터의 집합'이라 정의하였다. Poe(1996) 는 '의사결정 지원시스템의 기초로 사용되는 읽기 전용의 분석적 데이터베이스'라고 정의하고 있다. 본 연구에서는 '기업의 다양한 자료 원천으로부터 의사결정에 필요한 자료를 추출하여 주제지향적으로 통합한 읽기 전용의 데이터베이스'라고 정의하도록 하겠다.

데이터 웨어하우스의 데이터 구조는 압축 및 요약된 데이터, 1차 요약된 데이터, 최근의 상세 데이터, 과거 상세 데이터와 메타데이터의 다섯 유형으로 구성된다. 또한 데이터 웨어하우스 내의 데이터 구조를 형성하는 프로세스는 크게 다섯 가지 유형 (입력흐름, 출력흐름, 상향흐름, 하향흐름, 메타흐름) 으로 나눌 수 있다 (Hackathorn, 1995).

데이터 웨어하우스에 저장되는 데이터 중 가장 기본적인 데이터는 최근의 상세 데이터이다. 최근의 상세 데이터는 다량의 데이터로 구성되며 높은 접근 효율성을 위하여 디스크에 저장된다. 이와 같이 운영 또는 외부 데이터베이스로부터 데이터 웨어하우스의 최근 상세 데이터로 필요한 데이터를 적재하는 흐름을 입력 흐름이라 한다. 데이터 웨어하우스내의 최근의 상세 데이터는 일정 시점이 지난 후 과거의 상세 데이터로 변화된다. 이와 같이 최근의 상세 데이터가 과거의 상세 데이터로 변화되는 일련의 데이터 흐름을 하향 흐름이라 부른다. 최근의 상세 데이터가 가지는 데이터의 방대함으로부터 사용자의 손쉬운 접근과 상위 수준에서의 정보를 파악할 수 있도록 1차 요약 데이터와 압축, 요약 데이터를 추출한다. 이와 같이 최근의 상세 데이터로부터 요약 데이터를 추출하는 흐름을 상향 흐름이라 한다. 그리고 데이터 흐름을 통하여 데이터 웨어하우스에 저장된 정보를 사용자의 다차원 분석 요청에 따라 필요로 하는 시점에 적절한 정보를 요구되는 형태에 맞게 전달할 수 있는 기능이 필요하다. 이를 출력 흐름이라 한다. 마지막으

로 메타흐름이 있다. 메타흐름은 앞서 설명한 네가지 흐름을 통제하고 관리하는 흐름이다.

Kelly(1997)는 메타데이터를 ‘데이터 웨어하우스의 정보에 대한 원천, 값, 용도, 기능과 데이터 웨어하우스 아키텍처의 다양한 계층에서 정보가 어떻게 변화하고 변경되는지를 묘사하는 데이터’라고 정의하였다.

Devlin(1997)은 ‘정보시스템 환경에 대하여 각 구성요소들간의 관계에서부터 비즈니스와 기술적인 구조에 이르기까지 정의하고 설명하는 데이터’라고 정의하였다. 본 연구에서는 ‘데이터 웨어하우스의 정보 원천에서부터 사용자의 정보 사용에 이르기까지 데이터 웨어하우스 시스템의 전부분에서 필요로 하고 유지 관리되는 데이터’라고 정의하도록 하겠다.

여러 데이터 웨어하우스 관련 업체를 중심으로 메타데이터 표준을 제정하기 위한 노력이 진행되고 있으며 그 일환의 하나가 메타데이터 협회 (Meta Data Coalition) 이다. 메타데이터 협회는 1995년 7월에 설립되었으며 업체에 독립적이고 산업계가 정의하고 유지 관리하는 표준 접근 매카니즘과 표준 어플리케이션 프로그래밍 인터페이스를 수립하는 것을 주요 목표로 하고 있다.

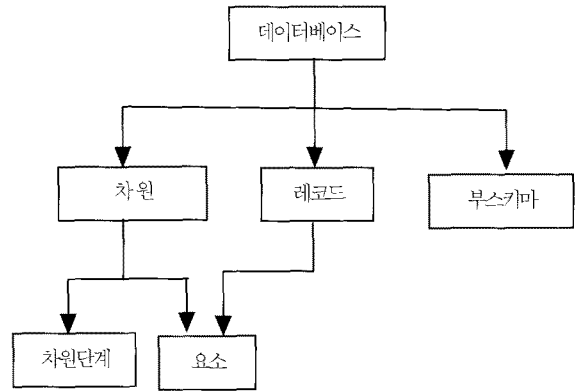
### III. 데이터 웨어하우스 메타데이터

#### 3.1 표준 스키마: 메타그룹 연합 스키마

메타데이터 협회의 표준 스키마인 MDIS (Metadata Interchange Specification) 는 현재 진행중인 표준 스키마이고 업계 표준으로 정립되지는 않았지만 많은 데이터 웨어하우스 관련 업체가 참여하고 있고 가장 널리 알려진 표준이다 (Meta Data Coalition, 1997). 따라서 본 연구에서는 MDIS를 데이터 웨어하우스 메타데이터 스키마의 표준으로 원용하고자 한다.

MDIS 메타데이터 스키마는 객체와 관계로 구성되어 있다. 각 객체는 관계의 유형에 따라 여러 가지 관계도를 구성하게 된다. <그림 3-1>은 MDIS 메타데이터 스키마의 객체들간의 포함 관계를 설명한 것

로 객체간의 계층을 잘 설명하여 주고 있다. 이는 사용자가 필요에 따라 선택 또는 사용하게 되는 관계에 따라 메타데이터 스키마가 매우 유연하게 구성될 수 있는 여지를 제공한다.



<그림 3-1> MDIS 메타데이터 스키마의 계층 구조

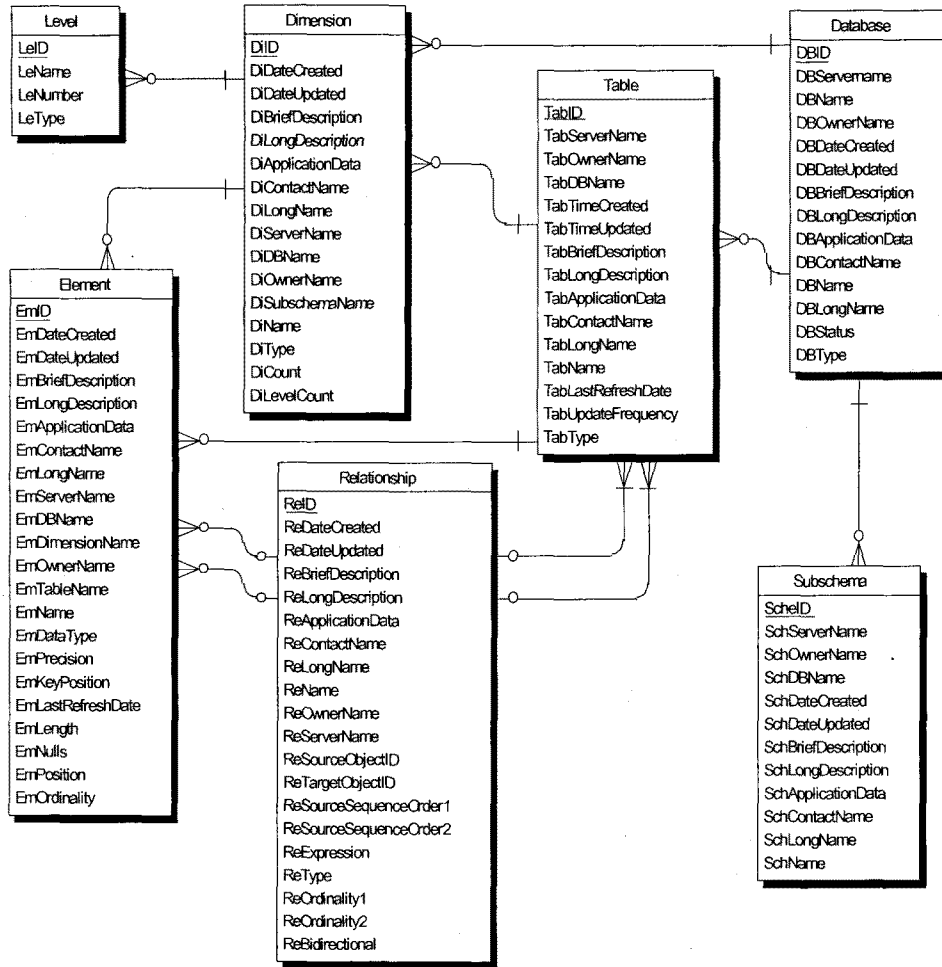
MDIS 메타데이터 스키마에서 사용되는 객체는 데이터베이스 (Database), 부스키마 (Subschema), 차원 (Dimension), 차원 단계 (Level), 레코드 (Record), 요소 (Element) 등 여섯 유형으로 구성되어 있고 각 객체의 연관은 관계 (Relationship) 로 표시된다. <표 3-1>은 각 객체와 관계에 대하여 설명하고 있다.

<표 3-1> MDIS 메타데이터 스키마 목록

| 구분     | 내용                                                                |
|--------|-------------------------------------------------------------------|
| 데이터베이스 | 파일, 관계형 데이터베이스, 망 데이터베이스, 계층 데이터베이스, 다차원 데이터베이스, 객체 데이터베이스 등을 나타냄 |
| 부스키마   | 데이터베이스의 의미있는 부분집합으로 레코드들의 논리적인 묶음                                 |
| 차원     | 데이터의 큐브 (Cube) 에서 고유의 위치를 표현하기 위한 계층을 가지는 요소                      |
| 차원 단계  | 차원을 구성하는 요소들은 하나 또는 다수의 단계 포함                                     |
| 레코드    | 데이터 단위를 설명하는 요소들의 실제적인 묶음                                         |
| 요소     | 설명될 수 있는 최소단위의 데이터에 대한 실제적인 설명                                    |
| 관계     | 개체들간의 관계를 정의하며 MDIS 개체중 가장 유연함                                    |

<그림 3-2>는 위에서 설명한 MDIS 메타데이터 스키마 표준을 개체-관계도 (Entity Relationship Diagram) 를 사용하여 나타낸 것이다. 본 연구에서는 MDIS 메타데이터 스키마를 기본 스키마로 인용하고자 한다.

수반되는 모든 메타데이터를 포함시키는 것은 복잡하다 (Duyn, 1982). 즉 기술 사용자는 메타데이터 내용에 대해서 선택을 해야한다. 모든 메타데이터가 명백하게 나타나 있지 않으며 많은 메타데이터가 암묵의



<그림 3-2> MDIS 표준 기반 메타데이터 스키마

### 3.2 메타데이터 계층 구조

메타데이터 양 (Quantity)과 질 (Quality)을 동시에 보장하는 것은 어려운 작업이다 (Drewry, 1997).

양적인 측면을 보면, 메타데이터의 사이즈와 내용에는 제한이 없으나 메타데이터 생성과 유지보수에는 제한이 따른다 (Sumpter, 1994). 데이터 웨어하우징에

상태로 존재하고 있다 (Spiers, 1997). 이를 위하여 메타데이터 표준이 필요하다.

각 도메인들마다 표준안이 존재하는데, 인터넷 도메인에서는 Dublin-core가 대표적이다. 그런데 이러한 표준안이 그 도메인에서 유일하지는 않다. 왜냐하면 메타데이터 스키마가 유일하게 정의된다면 표준안을 만든 곳이 전세계의 표준을 이끌게 되고 다른 업체들

도 자기 고유의 노하우를 포기하고 표준안을 수용할 수 밖에 없는 것이다.

따라서, 데이터 웨어하우스 도메인에서는 8개 업체들이 모여서 최소한의 메타데이터 교환을 위한 스키마를 만들었으며 스키마는 계속 발전하고 있다. 확정적인 메타스키마는 존재할 수 없으며 계속해서 새로운 요구사항을 받아들여서 발전하게 된다.

현재 MDIS 1.1은 기초적인 정보 데이터베이스의 교환을 위한 구조 메타데이터 (Structural Metadata) 로써, 다른 탐색 등의 비즈니스 메타데이터 또는 어플리케이션 컴포넌트를 위한 메타데이터를 지원하지 않고 있다. 또한 이 표준안은 서서히 발전되고 있으며, 최소한의 교환 메타데이터를 담고 있기 때문에 추가적인 요구사항은 잘 반영이 되지 않고 있다.

즉, 전체 데이터 웨어하우스 환경을 지원하는 통합된 메타데이터 스키마가 구현되지 않고 있다. 그렇지만 구조 메타데이터의 부분은 표준으로 확립되어 있기 때문에 MDIS 표준안을 기반으로 최소한의 메타데이터를 추가하는 것이 필요하다. 본 연구는 Lee and Kim(1998)의 데이터 웨어하우스 아키텍처의 7 컴포넌트에 의한 사용자, 기능 및 위치의 혼합에 의한 분류를 응용하여 데이터 웨어하우스의 프로세스를 4계층으로 구분하여 메타데이터를 분류하였다. 이와 같은 4계층 분류는 Kelly(1996)의 6계층 분류를 4계층으로 통합한 것이다. 즉, 메타데이터를 MDIS 표준의 핵심 웨어하우스 계층 (Core Warehouse Level) 을 포함하여, 적용 계층 (Application Level), 사용자 탐색 계층 (User Navigational Level), 비즈니스 계층 (Business Level) 의 4 계층으로 구분하였다. 이는 <표 2-1>과 같이 요약될 수 있다.

핵심 웨어하우스 계층 (Core Warehouse Level) 의 메타데이터는 데이터 웨어하우스 시스템의 기술적인 구조에 대한 정보이다. 이 핵심웨어하우스 계층은 MDIS 기반의 표준 메타 스키마를 포함한다.

그 목적은 방대한 양의 데이터를 유지 운영하고 사용자에게 적절한 응답 시간을 보장하기 위한 기술적 효율성이다.

<표 2-1> 메타데이터 계층

| 구 분         | 내 용                                                        |
|-------------|------------------------------------------------------------|
| 핵심 웨어하우스 계층 | - 데이터베이스 시스템의 데이터 사전<br>- 데이터 웨어하우스 시스템의 기술적인 구조 정보        |
| 적용 계층       | - 기존 운영 시스템에 관한 정보<br>- 데이터의 추출, 세팅, 변환 정보                 |
| 사용자 탐색 계층   | - 질의와 장표 (Report) 작성 도구 지원<br>- 비즈니스 계층과 핵심 웨어하우스 계층의 연결기능 |
| 비즈니스 계층     | - 데이터에 대한 비즈니스 관점<br>- 기업의 전략 및 실행                         |

적용 계층 (Application Level) 의 메타데이터는 기존의 운영 시스템과 데이터 웨어하우스와의 연결 기능을 가진다. 이를 위하여 기존의 운영 시스템에 관한 정보와 기존의 운영 시스템으로부터 데이터 웨어하우스로 데이터가 적재되면서 발생하게 되는 데이터의 추출, 세팅, 변환 정보를 포함한다. 적용 계층의 메타데이터는 데이터 웨어하우스 구축 단계에서부터 매우 중요한 역할을 담당하게 되며 그 특성상 매우 복잡한 구조를 가지고 있으며 정보의 변경에 대한 정확하고 즉각적인 반영이 필수적이다.

사용자 탐색 계층 (User Navigational Level) 의 메타데이터는 일반 사용자가 필요한 정보를 얻기 위하여 사용하는 질의와 장표를 작성하기 위한 도구에 대한 정보이다. 사용자 탐색 계층은 핵심 웨어하우스 계층과 비즈니스 계층을 연결하는 기능을 가진다. 즉 사용자의 비즈니스적인 관점에서 필요로 하는 정보를 해석하고 이에 대응되는 데이터베이스에 저장된 정보를 연결시켜 주는 기능을 포함한다.

마지막으로 비즈니스 계층 (Business Level) 의 메타데이터는 데이터 웨어하우스에 저장된 데이터에 대한 사용자의 비즈니스적인 관점이다. 이 계층의 메타데이터는 특히 의사결정지원 시스템 분석가와 일반 사용자에게 유용하다. 전사적인 데이터 웨어하우스 시스템의 구축을 위해서는 기업의 전략 및 실행 계획과 같은 전사적인 차원에서 공유해야 하는 정보가 포

함되어야 한다.

### 3.3 메타데이터 구현 스키마 비교

메타데이터 스키마의 실제 활용은 학계보다는 업계쪽의 상용 도구 (Tool) 에 적용되고 있다. 메타데이터 스키마는 기업의 자산이며, 이것이 공개되면 기업의 노하우를 잃게 되므로 공개를 하고 있지 않다. 또한 메타데이터 관리 도구 또한 고가의 소프트웨어이기 때문에 평가의 제약이 따르며 스키마는 잘 알려져 있지 않다. 현재 메타데이터 상용도구는 15개 정도의 제품이 있다 (Greenfield, 1999). 연구 논문들 또한 데이터 웨어하우스 메타데이터 분류에 초점을 두고 있으며, 스키마 제시보다는 서술식으로 설명하고 있다.

메타데이터 분류는 데이터 웨어하우스 컴포넌트들의 물리적 장소의 메타데이터인 위치, 기술적 메타데이터와 비즈니스 메타데이터로 분류하는 사용자, 구축시 메타데이터, 제어 메타데이터, 사용 메타데이터로 분류하는 기능으로 나타난다. 위치 (Griffin, 1996; Inmon & Hackathorn, 1994)는 정보 데이터베이스인

운영 데이터 창고, 데이터 웨어하우스, 데이터 마트가 위치해 있는 곳에 따라 메타데이터가 생성되고 관리된다. 사용자 (Hackney, 1997; Marco, 1998; Sachdeva, 1998; Sherman, 1997; Sumpter, 1994; White, 1995)는 개발자와 관리자들을 위한 기술적 메타데이터와 기술적인 능력이 부족한 비즈니스 사용자들이 데이터 웨어하우스를 탐색할 때 이용되는 비즈니스 메타데이터로 구분되어 생성되고 관리된다. 기능 (Devlin, 1997; Hurwitz, 1997; Moriarty & Mandracchia; 1996)은 데이터 웨어하우스를 생성하는 프로세스에서 발생하는 메타데이터를 분류해서 정하고 관리하는 것이다. <표 3-2>는 기존의 연구 결과인 몇 가지 데이터 웨어하우스 구축을 위한 구현 스키마 (Applied Schema) 들과 본 연구에서 제시한 스키마를 비교하고 있다. Lee and Kim (1998)은 메타데이터의 기능을 중심으로 비즈니스 메타데이터와 기술적인 메타데이터로 분류하였고, Sachdeva (1998)은 사용자를 중심으로 일반 사용자용 메타데이터와 데이터 웨어하우스 관리자용 메타데이터, 개발자용 메타데이터로 분류하였다. Burton (1996)은 메타데이터의 분류에 대한 언급이 포함되지

<표 3-2> 메타데이터 구현 스키마 비교

| 구 분             | Lee and Kim [1998]                   | Inmon et. Al [1997]         | Barquin and Edelstein [1997] | Sachdeva [1998]                       | Burton [1996]                     | 본 연구                                           |
|-----------------|--------------------------------------|-----------------------------|------------------------------|---------------------------------------|-----------------------------------|------------------------------------------------|
| 분류              | 사용자, 기능 및 위치의 혼합에 의한 분류              | 사용자, 기능 및 위치의 혼합에 의한 분류     | 기능에 의한분류                     | 비즈니스 메타데이터 (일반사용자) 기술적메타데이터 (관리자,개발자) | 없음                                | 계층별 분류 (핵심웨어하우스 계층, 적용 계층, 사용자 탐험 계층, 비즈니스 계층) |
| 사용자 요구 반영       | 유연함 (대부분 반영)                         | 유연함 (대부분 반영)                | 고정됨                          | 고정됨                                   | 유연함 (전부 반영)                       | 유연함 (대부분 반영)                                   |
| 구성개체            | 데이터 웨어하우스의 (Data Warehousing) 의 7 성분 | Motivation, Entities 등 6 개체 | 질의처리 등 3개체                   | Table, Elements 등 21 개체               | Table, Field, Source, Usage의 4 성분 | Table, Element, Mapping 등 20개체                 |
| MDIS표준          | 전부 포함                                | 일부 포함                       | 거의 없음                        | 일부 포함                                 | 거의 없음                             | 전부 포함                                          |
| 데이터웨어하우스 개발과 관계 | 데이터 웨어하우스 개발과 동시 진행                  | 관련 없음                       | 관련 없음                        | 데이터 웨어하우스 개발과 동시 진행                   | 데이터 웨어하우스 개발 후                    | 데이터 웨어하우스 개발과 동시 진행                            |
| 모델링             | 개체-관계 모델                             | 개체-관계 모델                    | 개체-관계 모델                     | 개체-관계 모델                              | 개체-관계 모델                          | 개체-관계 모델                                       |

않았다. 본 연구에서는 데이터 웨어하우스 시스템의 계층을 핵심 웨어하우스 계층, 적용 계층, 사용자 탐험 계층, 비즈니스 계층 메타데이터로 분류한다.

메타데이터 스키마는 세 가지 구현 스키마가 모두 개체-관계 모델을 사용하였으며 본 연구에서도 관계형 데이터베이스를 효과적으로 표현할 수 있는 개체-관계 모델을 사용하였다.

Lee and Kim (1998)의 스키마는 MDIS 표준 메타데이터 스키마에서 제공하는 여섯 개의 개체와 데이터 웨어하우스 환경의 일곱 가지 성분으로 구성되어 있다. Burton (1996)의 스키마는 Table, Field, Source, Usage 의 네 가지 성분을 가지고 있으며, Sachdeva (1998)의 스키마는 Table, Element 등의 스물 한 개의 개체를 가지고 있다. 각 스키마의 구성 요소는 스키마별로 매우 상이하나 Burton (1996)의 경우 구성요소가 적고 사용자의 요구사항을 유연하게 반영할 수 있는 특성이 있다. Sachdeva (1998)의 경우 구성요소가 많고 자세한 반면 사용자의 요구사항을 반영할 수 있는 스키마의 유연성이 낮은 특성이 있음을 알 수 있다. 즉 스키마의 구성 요소가 갖는 복잡도와 스키마의 유연성사이에는 반비례 관계가 있다고 보여진다.

Inmon et. al (1997)의 스키마는 여섯 가지 성분을 가지고 있으며, Barquin and Edelstein (1997)의 스키마는 질의처리 등의 세가지 개체를 가지고 있다. 그러므로 본 연구에서는 스키마가 갖는 유연성을 최대한 확보하면서도 핵심 메타데이터 스키마를 구성하는 구성요소를 많이 제공할 수 있는 메타데이터 스키마가 이상적이라고 판단된다.

Lee and Kim (1998)은 메타데이터 스키마 표준인 MDIS 표준을 전부 수용하고 있으나, Sachdeva (1998)는 일부만 포함하고 있고 Burton (1996)은 거의 반영하고 있지 않다.

그리고, Inmon et. al (1997)은 일부만 포함하고 있고 Barquin and Edelstein (1997)은 거의 반영하지 않고 있다. 본 연구에서는 MDIS 메타데이터 스키마 표준을 최대한 수용하여 다른 데이터 웨어하우스 시스템간의 메타데이터 호환이 가능한 메타데이터 스키마

를 채택한다. 데이터 웨어하우스 시스템 개발과 메타데이터 스키마를 이용한 메타데이터 관리 시스템의 개발은 Lee and Kim (1998)과 Sachdeva (1998)의 경우, 동시에 진행되도록 설계되어 있으나 Burton(1996)의 경우에는 데이터 웨어하우스 시스템의 개발이 끝난 후에 개발되었다. 메타데이터의 효율적인 관리는 데이터 웨어하우스 시스템의 가장 핵심적인 부분중의 하나이므로 메타데이터 개발은 데이터 웨어하우스 시스템 개발의 중요한 일부분이 되어야 한다. 본 연구에서 제안하고 있는 메타데이터 스키마는 핵심 웨어하우스 계층 메타 데이터에 MDIS 표준의 대부분을 포함하고 있어 다른 데이터 웨어하우스와의 메타데이터의 호환을 가능하게 한다.

## IV. 데이터 웨어하우스 메타데이터 구축 사례

### 4.1 프로젝트 개요

ㅇ사는 차세대 첨단 통신 기술의 개발과 관련 산업의 균형적 발전을 꾀하여 국제경쟁력을 높이고 해외시장에도 적극 진출하여 복지 통신 시대를 열어가겠다는 목표를 가지고 1996년에 출범하였다. ㅇ사는 고객에 대한 서비스의 향상과 경쟁력 향상을 위한 정책의 일환으로 1998년초 다음과 같은 목표를 가지고 데이터 웨어하우스를 구축하기로 하였다.

첫째, 사용자에게 직접 데이터를 분석할 수 있는 기능을 제공한다. 둘째, 사용자가 필요로 하는 장표의 처리 시간을 단축한다. 셋째, 프로그램의 개발없이 임의의 정보 처리를 가능하게 한다. 넷째, 분산된 온라인 분석 처리 (On Line Analytical Processing; OLAP)에 대한 통합된 정보를 제공한다. 다섯째, 전사적인 업무를 신속하게 파악한다. 여섯째, 경영 의사 결정, 마케팅 프로그램 개발 및 신상품 개발에 과학적인 지원이 가능하다. 마지막으로, 하나의 정보 원천으로부터 다양한 조회가 가능하도록 한다. 데이터 웨어하우스를 활용할 주요 업무는 <표 4-1>과 같으며 고객 관

련 업무, 통화 관련 업무 및 대리점 관련 업무로 대별된다.

회사의 업무를 대상으로 다음과 같이 세 가지의 데이터 웨어하우스를 이용한 연계 프로그램을 계획하였다. 첫번째 연계 프로그램은 고객 평가 시스템이며, 둘째는 고객 이탈 방지 시스템이고, 셋째는 부정(Fraud) 방지 시스템이다. 부정이란 불법으로 통신 서비스를 사용하는 것으로 이를 방지하기 위해서는 각 가입자의 통화를 상호 비교하는 작업이 필요하다.

회사의 데이터 웨어하우스 시스템은 <그림 4-1>과

같이 구성되어 있다. 운영 데이터 부분을 구성하고 있는 주요 운영 시스템은 CCBS (Customer Care Billing System) 라 불리는 고객 관리 및 요금 관리 시스템과 CCS (Call Center System) 라 불리는 고객지원센터 시스템이다.

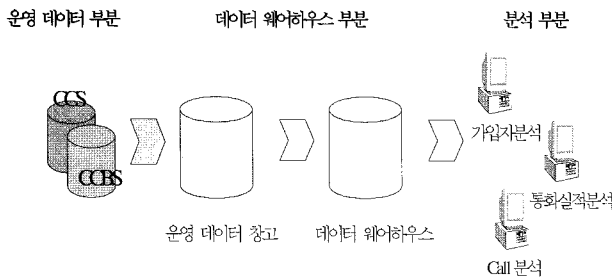
고객 관리 및 요금 관리 시스템은 운영 데이터 부분의 핵심 시스템으로 고객 관리와 요금 관리로 구분된다. 고객 관리는 고객 생애주기 전반에 걸친 정보를 관리하는 시스템이며, 요금 관리는 수납 등의 정보를 관리하는 시스템이다. 고객지원센터 시스템은

<표 4-1> 회사의 데이터 웨어하우스 주요 대상 업무

| 주요업무           | 관리항목        | 내 용                                              |
|----------------|-------------|--------------------------------------------------|
| 고객<br>관련<br>업무 | 가입자 관리      | 영업조직, 지역, 연령, 계약기간, 성별, 직업, 가입유형, 수납 방법 등의 정보 관리 |
|                | 해지자 관리      | 가입자 관리 정보를 포함한 해지 사유, 청구현황 등의 정보 관리              |
|                | 고객 변경 관리    | 가입자 관리 정보를 포함한 변경 사유 등의 정보 관리                    |
|                | 고객 상태 변경 관리 | 가입자 관리 정보를 포함한 상태 변경 사유 등의 정보 관리                 |
|                | 부가 서비스 관리   | 가입자 관리 정보를 포함한 부가서비스, 요금 상품 등의 정보 관리             |
|                | 고객속성별 패턴 관리 | 가입자 관리 정보를 포함한 통화 패턴 등의 정보 관리                    |
| 통화<br>관련<br>업무 | 요금상품통화      | 통화월, 요금 상품, 시간대 등의 정보 관리                         |
|                | 국내 통화 관리    | 통화월, 착신사업자, 시간대 등의 정보 관리                         |
|                | 국제 통화 관리    | 통화월, 국가, 시간대, 국제전화사업자 등의 정보 관리                   |
|                | 부가서비스 통화 관리 | 통화월, 부가서비스, 시간대 등의 정보 관리                         |
|                | 과금 통화 관리    | 과금 통화 항목 등의 정보 관리                                |
|                | 고객 불만 관리 관리 | 가입자 관리 정보를 포함한 불만 유형 등의 정보 관리                    |
|                | 청구/수납/미납 관리 | 가입자 관리 정보를 포함한 청구월, 미납 사유 등의 정보 관리               |
|                | 기지국 트래픽 관리  | 통화월, 기지국, 불완료 요인 등의 정보 관리                        |
| 대리점<br>관련업무    | 유통망 관리      | 영업조직, 대리점 현황 등의 정보 관리                            |



고객지원센터의 업무 전반을 관리하는 시스템이다.



<그림 4-1> 회사의 데이터 웨어하우스 시스템 구성도

데이터 웨어하우스 부분은 운영 데이터를 데이터 웨어하우스로 적재하기 전에 필요한 작업을 수행할 수 있도록 보관하는 운영 데이터 창고 (Operational Data Store) 와 데이터 웨어하우스 자체로 구성되어 있다. 운영 데이터 창고는 다음과 같은 용도로 사용된다. 첫째, 방대한 양의 자료가 보관 운영되고 있는 운영 시스템의 효율이 저하되는 것을 억제하기 위함이다. 둘째, 데이터 웨어하우스로 데이터를 적재하기 위한 작업 중 발생하는 운영 데이터의 변경이 운영 시스템과 독립적이어야 한다. 마지막으로 운영 시스템의 효율을 저해하지 않으면서 운영 시스템의 데이터에 대한 신속한 조회 기능을 제공한다. 분석 부분은 사용자와 데이터를 이용하여 의사결정에 활용할 수 있는 정보를 획득할 수 있도록 지원하는 분석 도구로 구성되어 있다.

회사의 네트워크의 근간은 프레임 릴레이 (Frame Relay) 방식으로 구성되어 있다. 시스템의 주요 부분으로는 데이터 웨어하우스 시스템을 구성하고 있는 서버 (Sequent NUMAQ2000) 와 고객 관리 및 요금 관리 시스템의 서버 (GS8400), 고객지원센터 시스템의 서버 (HP K460), 그리고 일반 사용자용 분석 도구인 Oracle Discoverer 등으로 구성되어 있다.

#### 4.2 데이터 웨어하우스 메타데이터 스키마 현황

회사의 데이터 웨어하우스 시스템의 메타데이터는 시스템 구성과 유사하게 운영 데이터 로딩, 데이터

웨어하우스 적재, 분석관리로 구성되어 있다. 자세한 내용은 <표 4-2>에 요약되어 있다. 운영 데이터 로딩은 운영 데이터를 운영 데이터 창고로 옮기는 작업을 수행하는 부분으로 운영 데이터와 운영 데이터 창고의 필드간 연관 관계와 옮기는 과정에서 수행되는 운영 데이터의 변환 내용이 주된 내용이다. 데이터 웨어하우스 적재는 운영 데이터 창고로부터 데이터 웨어하우스로 옮기는 작업을 수행하는 부분으로 운영 데이터를 데이터 웨어하우스의 용도에 적합한 형태로 변환하는 내용이 중요하며 수행되는 작업의 전 과정에 대한 관리와 모니터링을 자동화하는 것에 주안점을 두고 있다. 분석 관리는 데이터 웨어하우스의 데이터를 사용자의 요구에 적합하도록 제공하기 위하여 필요한 메타데이터를 관리하는 부분이다.

<표 4-2> 회사의 데이터 웨어하우스 메타데이터 현황

| 구분           | 내용                                                                            | 구현 어플리케이션         |
|--------------|-------------------------------------------------------------------------------|-------------------|
| 운영 데이터로딩     | - 운영 데이터에 대한 정보<br>- 운영 데이터 창고에 대한 정보<br>- 로딩 데이터의 매핑 정보                      | Oracle K*Loader   |
| 데이터 웨어하우스 적재 | - 운영 데이터 창고에 대한 정보<br>- 데이터 변환 정보<br>- 적재 작업의 관리 및 모니터링<br>- 데이터 웨어하우스에 대한 정보 | Oracle ETT Module |
| 분석 관리        | - 데이터 웨어하우스에 대한 정보<br>- 사용자 정보<br>- 비즈니스 정보                                   | Oracle Discoverer |

회사의 메타데이터는 주된 작업을 수행하는 어플리케이션들이 서로 독립적으로 수행되어 각각의 메타데이터를 공유하지 못하고 있다. 즉 전사적인 데이터 웨어하우스 메타스키마가 구성되어 있지 않기 때문에 회사의 데이터 웨어하우스에 필수적인 메타데이터가 각 어플리케이션에 분산되어 있어서 관리가 어렵고 데이터 웨어하우스의 성능 저하가 예상된다. 또한 각

부분을 구성하는 메타스키마가 해당 어플리케이션에 내재되어 있는 메타스키마에 종속적이고 메타데이터를 표현할 수 있는 공통의 표준을 가지고 있지 못하다.

따라서 데이터 웨어하우스의 운영을 통하여 발생하게 되는 회사의 고유 업무를 위한 메타데이터의 추가 필요성과 같이 지속적으로 반영되어야 하는 사용자의 요구 사항에 대하여 각각의 어플리케이션이 조화롭게 대처하여 전체 데이터 웨어하우스 시스템에 대한 일관성을 유지하여야 한다. 또한, 위에서 대두된 문제점은 향후 데이터 웨어하우스 시스템의 확장이나

특정 기능을 수행하는 어플리케이션의 변경 등에 제한을 초래할 수 있다.

### 4.3 메타데이터 스키마 개선

본 절에서는 MDIS 메타데이터 스키마 표준과 비교를 통하여 얻은 효율적인 데이터 웨어하우스 메타데이터 스키마의 특성을 반영하여 회사의 데이터 웨어하우스 메타스키마에 대한 개선 방안을 제시해 보도록 하겠다.

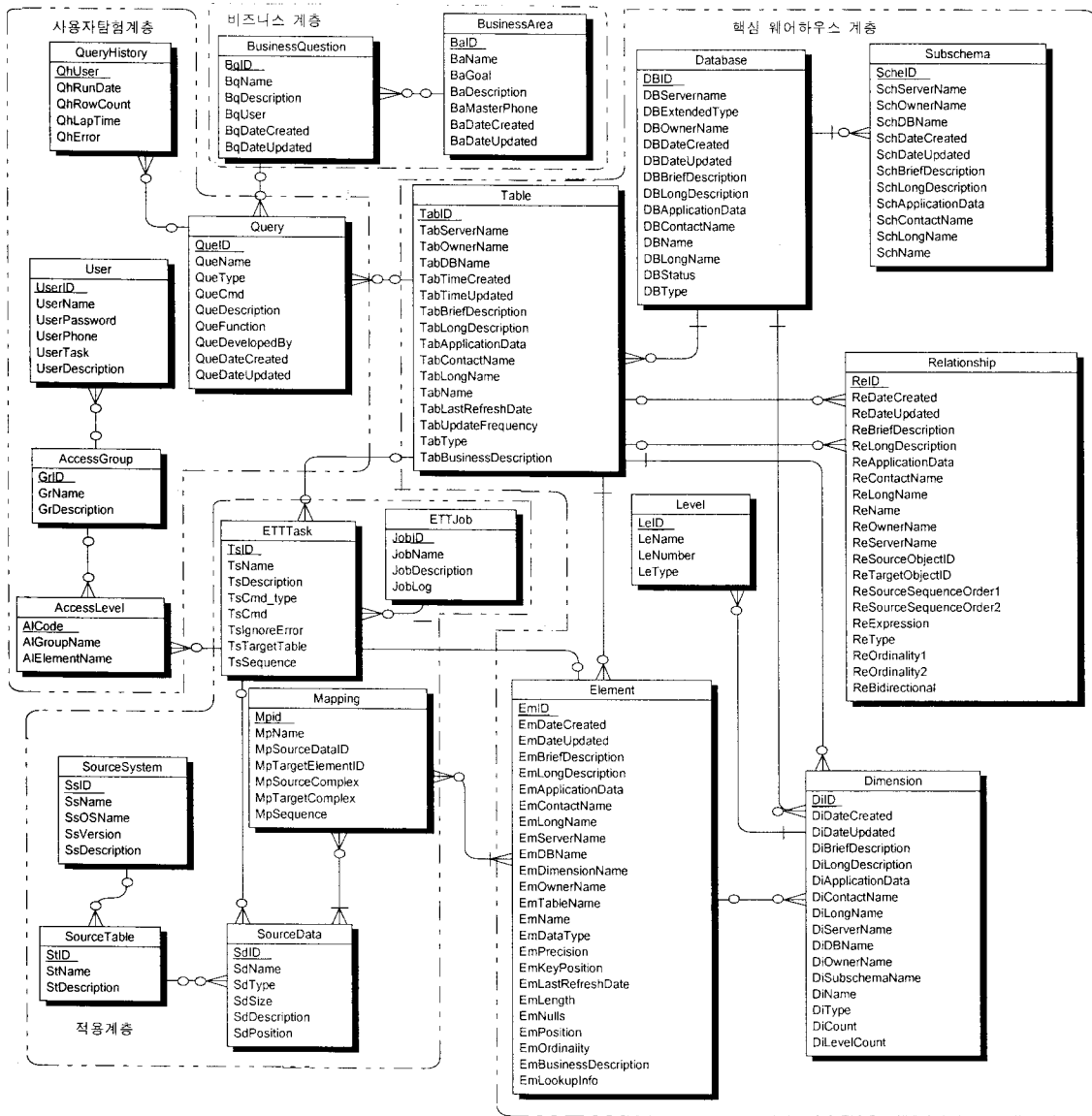
회사의 데이터 웨어하우스 메타스키마에 대한 개선을 위하여 몇 가지 목표를 채택하였다.

첫째로, 데이터 웨어하우스를 위한 전사적 메타스키마이다. 일부의 조직이나 담당 업무만을 위한 것이 아니라 전체적인 시각에서의 메타스키마를 작성해야 한다. 둘째로, 도구에 독립적인 메타스키마이다. 서로 다른 도구간에도 메타데이터가 공유될 수 있어야 한다. 셋째로, 메타데이터의 관리가 용이해야 한다. 넷째로, 데이터 웨어하우스 시스템의 요소들에 대한 확장에 대처할 수 있어야 한다. 다섯째로, 메타데이터를 필요로 하는 사용자의 접근이 쉽고 개방적이어야 한다. 여섯째로 현재 구축되어 있는 메타스키마를 최대한 포함하여야 한다. 일곱째로, MDIS 메타데이터 스키마 표준을 수용하여야 한다. 마지막으로, 2장에서 제시된 메타데이터의 분류에 따른 각 계층의 메타데이터를 모두 저장할 수 있어야 한다.

회사의 데이터 웨어하우스를 위한 개선 후의 메타스키마가 <그림 4-2>이다. 개선 후의 메타스키마는 20개의 개체 (Entity) 를 가지고 있다. 각 개체에 대한 설명은 부록에 상세하게 설명되어 있다.

핵심 웨어하우스 계층은 MDIS 표준을 포함한다. 핵심 웨어하우스 계층 메타데이터는 데이터 웨어하우스를 구성하는 관계형 데이터베이스에 관한 정보를 가지고 있다. 이는 실제의 데이터베이스를 구성하는데 필요한 기술적인 정보들로서 관계형 데이터베이스의 데이터 사전에 해당한다. 제품 특유의 상세하고 다양한 정보 중 데이터 웨어하우스 사용자에게 필요하다고 판단되는 정보를 선택한다. 핵심 웨어하우스 계층의 테이블(Table) 개체는 데이터 웨어하우스를 구성하고 있는 관계형 데이터베이스의 테이블에 대한 정보를 저장하며 MDIS 표준의 레코드 개체에 해당한다. 테이블 개체에 대한 데이터베이스 차원에서의 정보 뿐만 아니라 사용자 관점에서 본 테이블에 대한 비즈니스 차원에서의 정보도 함께 저장한다. 요소 개체는 테이블 개체를 구성하는 열 (Column) 들에 대한 정보를 저장하며 MDIS 표준의 요소 개체에 해당한다. 관계 개체는 테이블 개체간의 관계에 대한 정보를 저장하며 MDIS 표준의 관계 개체에 해당한다. 주로 저장되어지는 정보는 사실 테이블과 색인 테이블과의 관계, 테이블과 테이블 간의 기수 관계 등이다. 차원 개체는 사실 테이블에 대하여 사용자의 필요에 따른 차원에 대한 정보를 저장하며 MDIS 표준의 차원 개체에 해당한다.

적용 계층 메타데이터는 기존의 운영 시스템에 대한 정보와 데이터의 추출, 여과, 변환에 대한 정보가 있다. 기존의 운영 시스템에 대한 정보는 운영 시스템 자체에 대한 정보와 각 운영 시스템의 테이블/파일의 정보, 그리고 각 테이블/파일의 컬럼/필드에 대한 정보가 해당된다. 또한 기존 운영 시스템의 컬럼/필드와 데이터 웨어하우스를 구성하는 관계형 데이터베이스의 필드들과의 사상에 관한 정보가 포함된다. 이는 현재의 데이터베이스 시스템의 운영 데이터 로딩 메타데이터에 해당한다. 데이터의 추출, 여과, 변



<그림 4-2> 개선 후 데이터 웨어하우스 메타데이터 구현 스키마

환에 대한 정보는 데이터의 적재 작업 정보와 각각의 작업을 구성하고 있는 과제들에 대한 정보로 구성되어 있다. 이는 현재의 데이터베이스 시스템의 적재 메타데이터에 해당한다.

적용 계층의 추출 변환 작업(ETT Job) 개체는 운영 시스템으로부터 데이터 웨어하우스로 필요한 데이터를 추출, 여과, 변환, 적재하는 작업에 대한 정보를 저장한다. 작업은 서로 독립적으로 처리될 수 있는

최소 단위로 구성되어 있으며 각 작업의 수행 결과도 함께 저장된다. 추출 변환 과제 (ETT Task) 개체는 추출 변환 작업 개체를 구성하는 작업들을 수행의 순서와 효율 등을 고려하여 분할한 각각의 과제들에 대한 정보에 해당한다. 원천 시스템 (Source System) 개체, 원천 테이블 (Source Table) 개체, 그리고 원천 요소 (Source Data) 개체는 운영 시스템에 저장되어 있는 데이터에 대한 정보를 관리한다. 사상 (Mapping)

개체는 운영 시스템에 저장되어 있는 데이터와 데이터 웨어하우스에 저장되어 있는 데이터간의 관계에 대한 정보를 저장한다. 사용자 개체, 접근 그룹 개체, 접근 단계 개체는 사용자 개인에 대한 정보와 사용자 집단에 대한 정보를 저장한다. 각각의 사용자 집단에 대한 데이터를 사용할 수 있는 권한 정보도 함께 관리한다.

사용자 탐험 계층 메타데이터는 크게 두 가지로 구분되는 정보를 포함한다. 첫째는 사용자의 질의와 장표를 작성하기 위한 정보이고, 둘째는 사용자에 대한 정보이다. 질의와 장표는 데이터 웨어하우스의 개발 단계에서부터 사전에 작성 저장되어 있는 것과 필요에 따라 사용자에게 의해 즉시 작성 실행되는 것이 있다. 여기에서 주로 관리되는 정보는 사전에 작성 저장되는 질의와 장표로 데이터 웨어하우스의 성능에 매우 중요한 요소이다. 사용자 정보에는 해당 사용자가 데이터 웨어하우스를 사용할 수 있는 적법한 사용자 인지를 확인하고 해당 사용자가 데이터 웨어하우스의 어떤 데이터에 접근할 수 있는 권한을 부여 받았는지 등에 대한 정보가 포함된다. 사용자 탐험 계층의 질의 개체와 질의 이력 (Query History) 개체는 데이터에 대하여 사전에 설정되어 있는 질의에 대한 정보와 해당 질의의 실행에 대한 이력 정보가 저장된다. 질의의 내용이나 결과, 소요 시간 등은 데이터 웨어하우스의 성능 관리에 매우 중요한 정보이다.

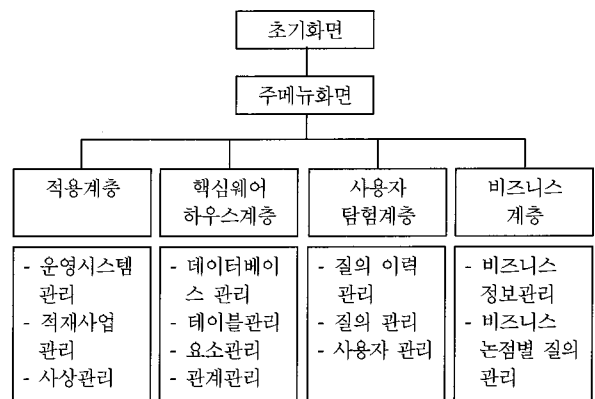
비즈니스 계층 메타데이터는 데이터 웨어하우스에 저장되어 있는 데이터에 대한 사용자의 비즈니스 관점에 대한 정보가 포함된다. 또한 데이터 웨어하우스를 중심으로 연관되는 기업의 전략과 실행 목표와 같은 정보도 포함된다. 이 계층의 메타데이터는 상대적으로 사용자의 요구에 따라 변경되거나 추가 확장될 여지가 매우 높은 부분이다. 비즈니스 계층의 비즈니스 영역 (Business Area) 개체는 데이터 웨어하우스를 이용하는 주요 조직과 해당 조직의 이용 목적 등에 대한 정보를 저장한다. 데이터 웨어하우스가 기업의 전략적인 목표와 유기적으로 결합되기 위한 중요한 정보를 제공한다. 비즈니스 논점 (Business Question)

개체는 비즈니스 영역에서 데이터 웨어하우스를 사용하여 구체적으로 획득하고자 하는 보다 상세한 업무적인 목표들에 대한 정보를 저장한다. 일반 사용자들이 가지고 있는 비즈니스 차원의 시각이 가장 많이 활용되는 부분이고 질의 개체와의 유기적인 결합이 데이터 웨어하우스 사용자들의 효율 향상에 필수적이다.

#### 4.4 메타데이터 관리 시스템 프로토타입

회사의 메타데이터 관리 시스템의 프로토타입을 개발하기 위한 시스템을 살펴보면, 서버는 펜티엄급 퍼스컴을 사용하였고 서버에 설치된 운영시스템은 윈도우 95이다. 데이터베이스는 퍼스널 컴퓨터용 Oracle DBMS (Database Management System) 가 이용되었다. 클라이언트에 설치된 어플리케이션 개발은 Powerbuilder 를 사용하였다.

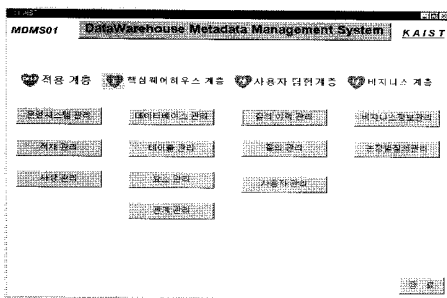
프로토타입 시스템은 사용자를 확인하기 위한 초기 화면과 주메뉴 화면을 비롯하여 메타데이터의 계층별로 구성된 4 개의 모듈로 구성되어 있으며 각각의 모듈에서 제공하는 기능을 기준으로 세부 모듈로 상세화 하였다. 프로토타입의 전체 시스템 구조도는 <그림 4-3>과 같다.



<그림 4-3> 프로토타입 시스템 구조도

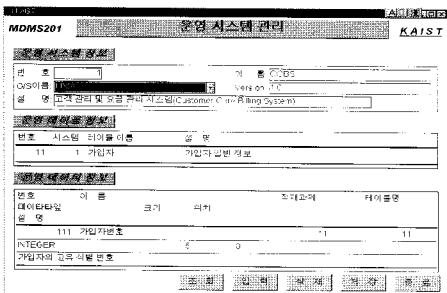
초기 화면은 해당 사용자가 데이터 웨어하우스의 합법적인 사용자인지 여부를 확인하고 해당 사용자의

데이터에 대한 접근 권한을 사용자 정보로부터 읽어들이는 화면이다. 사용자가 자신의 사원번호를 입력하면 해당 사원번호에 해당하는 사용자의 이름을 화면에 표시해준다. 사용자의 비밀번호는 숫자와 문자가 모두 허용되며 보안을 위하여 화면에는 표시하지 않는다. 주 메뉴화면은 프로토타입 시스템에서 제공되는 모든 서비스를 접근할 수 있게 해준다. 각 모듈의 상세 모듈을 모듈 계층에 따라 동일 화면에 구성하여 사용자로 하여금 필요한 모듈에 해당하는 버튼을 클릭하여 빠르고 편리하게 실행시킬 수 있게 하며 <그림 4-4>와 같다.



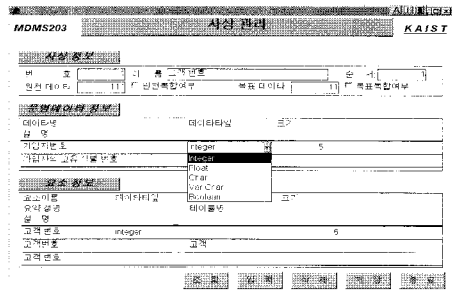
<그림 4-4> 프로토타입 주메뉴 화면

적용 계층의 메타데이터와 관련된 모듈은 운영 시스템 관리, 적재 작업 관리, 사상 관리 등 세 개의 세부 모듈로 구성되어 있다. 운영 시스템 관리는 데이터 웨어하우스에 저장되는 데이터의 원천이 되는 기존의 운영 시스템에 대한 정보를 제공해주는 화면으로 각각의 운영 시스템에 대하여 연관된 테이블/파일 정보와 컬럼/필드 정보까지 하나의 화면으로 파악할 수 있으며 <그림 4-5>와 같다.



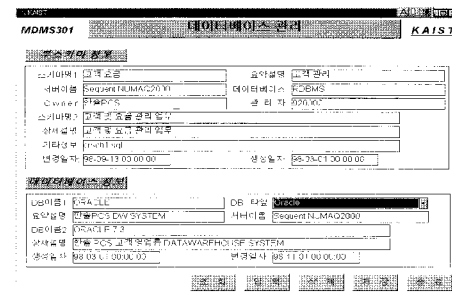
<그림 4-5> 운영 시스템 관리 화면 (적용 계층)

적재 작업 관리는 운영 시스템으로부터 데이터 웨어하우스로의 데이터 적재 작업과 하부 과제들에 대한 정보를 제공해주는 화면으로 적재 작업과 연관된 하부 과제들의 정보를 알 수 있다. 사상 관리는 운영 시스템의 컬럼/필드와 데이터 웨어하우스의 필드의 사상에 대한 정보를 제공한다. 사상 관리는 해당 사상과 연결되어 있는 운영 시스템의 컬럼/필드와 데이터 웨어하우스의 필드가 단일 필드인지 복합 필드인지가 중요한 정보가 되며 <그림 4-6>과 같다.



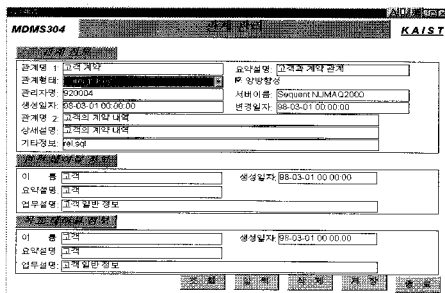
<그림 4-6> 사상 관리 화면 (적용 계층)

핵심 웨어하우스 계층의 메타데이터와 관련된 모듈은 데이터베이스 관리, 테이블 관리, 요소 관리, 관계 관리 등 네 개의 세부 모듈로 구성되어 있다. 데이터베이스 관리는 부 스키마와 데이터베이스에 대한 정보를 제공하는 화면으로 동일한 데이터베이스로 구성되어 있고 <그림 4-7>과 같다. 고객 영업용 데이터베이스만이 구축되어 있는 현재의 데이터 웨어하우스 시스템에서는 필요성이 적다.



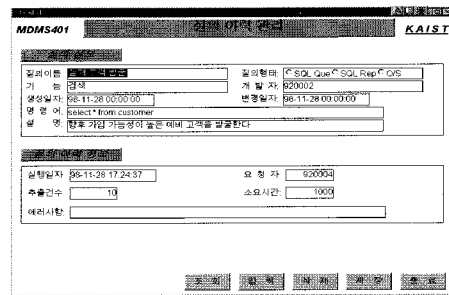
<그림 4-7> 데이터베이스 관리 화면 (핵심 웨어하우스 계층)

테이블 관리는 테이블에 관한 상세 정보와 테이블에 연관된 차원, 차원 단계 등에 대한 정보를 제공하는 화면이다. 요소 관리는 각 테이블의 요소들에 대한 정보를 제공한다. 요소 정보에는 요소에 해당하는 데이터의 값과 관련된 타입, 크기, 위치 등과 같은 정보가 포함되며 이해를 돕기 위해 해당 요소를 포함하고 있는 테이블의 정보도 함께 표시한다. 관계 관리는 테이블간의 관계에 관한 정보를 제공하며 <그림 4-8>과 같다. 데이터 웨어하우스의 메타데이터 관리 시스템의 기능 향상을 위해서는 관계 관리 기능의 향상이 필수적이다. 관계 관리는 관계의 성격을 알려주는 설명과 함께 해당 관계를 구성하는 원천 테이블 정보와 목표 테이블 정보를 함께 표시해준다.



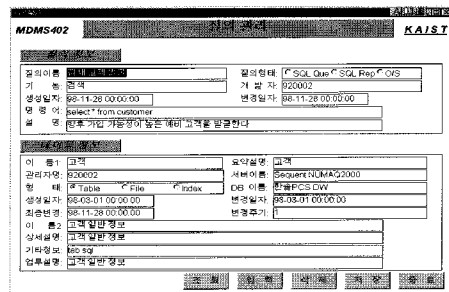
<그림 4-8> 관계 관리 화면 (핵심 웨어하우스 계층)

사용자 탐험 계층의 메타데이터와 관련된 모듈은 질의 이력 관리, 질의 관리 및 사용자 관리의 세 개 세부 모듈로 구성되어 있다. 일반 사용자들이 가장 빈번하게 활용하게 될 부분으로 향후 사용자들의 요구가 증가함에 따라 많은 향상이 필요한 모듈이다. 질의 이력 관리는 데이터 웨어하우스에 저장되어 있는 질의들의 상세한 실행 이력을 제공하는 화면으로 데이터 웨어하우스의 성능 확인과 불필요하거나 잘못 작성된 질의를 색출하기 위한 정보를 제공할 수 있으며 <그림 4-9>와 같다.



<그림 4-9> 질의 이력 관리 화면 (사용자 탐험 계층)

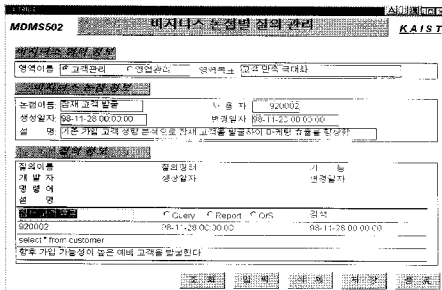
질의 관리는 데이터 웨어하우스에 저장된 질의에 대한 상세한 정보를 제공할 수 있는 화면으로 질의내용 뿐만 아니라 관련된 테이블, 차원, 요소들에 대한 정보도 함께 제공하며 그림 <4-10>과 같다. 사용자 관리는 데이터 웨어하우스를 사용할 수 있는 사용자에 대한 정보를 제공하는 화면으로 사용자 암호, 사용자 그룹 정보, 그룹별 데이터 접근 권한 정보 등을 알 수 있다.



<그림 4-10> 질의 관리 화면 (사용자 탐험 계층)

비즈니스 계층의 메타데이터와 관련된 모듈은 비즈니스 정보 관리와 비즈니스 논점별 질의 관리로 구성되어 있다. 비즈니스 정보 관리는 비즈니스 영역 정보와 각 영역별로 중요시되는 비즈니스 논점들에 대한 정보를 제공한다. 기업 전략, 영업 목표 등과 같은 회사의 경영 환경 변화에 밀접하게 민감하게 반응하고 얼마나 잘 표현할 수 있는지가 매우 중요하다.

회사가 데이터 웨어하우스를 이용하여 얻고자 하는 목표가 잘 드러나있다. 비즈니스 논점별 질의 관리리는 사용자가 데이터 웨어하우스를 이용하여 얻고자 하는 여러 가지 비즈니스 논점들에 대하여 데이터 웨어하우스가 어떠한 질의를 통하여 필요한 데이터를 제공하는지를 알 수 있는 화면으로 <그림 4-11>과 같다.



(그림 4-11) 비즈니스 논점별 질의 관리 화면 (비즈니스 계층)

이상과 같이 회사의 데이터 웨어하우스 메타데이터 관리 시스템의 프로토타입을 통하여 앞에서 제시된 목표에 부합하는 결과를 기대하게 되었다.

첫번째 기대효과는 현재 사용중인 데이터 웨어하우스 시스템이 기업의 경영에서 차지하는 역할과의 미를 잘 이해하게 될 수 있다는 점이다. 예를 들면, <그림4-11>의 비즈니스 논점별 질의 관리와 비즈니스 정보관리를 통하여 비즈니스 정보에 대한 총괄적인 윤곽을 파악하게 됨으로 사용중인 데이터 웨어하우스 시스템의 역할을 효과적으로 이해하는 것이 가능하다. 현재 운영중인 데이터 웨어하우스가 일부 부서 또는 일부의 업무를 위한 목적으로 개발 운영되고 있지만, 비즈니스 계층의 메타데이터를 통해 비즈니스 정보를 용이하게 파악할 수 있다.

둘째는 현재의 데이터 웨어하우스 시스템을 구성하고 있는 어플리케이션에 독립적으로 운영될 수 있다는 점이다. 현재, 회사의 데이터 웨어하우스 시스템을 구성하고 있는 어플리케이션의 메타데이터는 어플리케이션에 종속적이고 공유가 불가능하였다.

따라서, 프로토타입을 구축하기 위하여 메타데이터를 수작업으로 입력하였다. 이 과정에서 시간이 많이 소요되었고, 각 어플리케이션의 메타데이터를 해석 변환하여야 되는 문제점이 발생되었다. 그러나, 도구나 어플리케이션에 독립적인 메타스키마를 구축을 통하여 어플리케이션에 독립적인 시스템을 구현할 수 있었다.

세째는 메타데이터의 관리가 용이해진다는 점이다. 간단하면서도 효과적으로 메타데이터를 조회하고 수정하는 것이 가능하게 되었다. 예를 들면, <그림 4-5>의 운영시스템 관리를 통하여 각각의 운영시스템에 연관된 테이블/파일 정보와 컬럼/필드 정보를 하나의 화면으로 파악하게 해줌으로 효과적인 관리를 기대하게 되었다.

네째는 사용자의 접근에 대한 통제가 용이하고 관리 가능하게 된다는 점이다. 예를 들면, <그림4-3>의 초기화면을 통하여 해당 사용자가 합법적인 사용자인지를 확인하고 데이터에 대한 접근 권한을 부여하게 된다. 데이터 웨어하우스의 메타데이터 전체에 대하여 사용자를 몇 개의 집단으로 분류하여 각 집단별로 접근 권한을 부여하였다. 이를 통하여 사용자가 필요로 하는 영역의 메타데이터를 쉽게 제공 받을 수 있게 되었다.

다섯째는 MDIS 표준기반 메타데이터 스키마를 모두 반영하여 표준에 부합한다는 점이다. <그림 3-2>에서 제시된 MDIS 표준 기반 메타데이터 스키마는 <그림 4-2>의 개선후 메타데이터 스키마에 모두 포함되고 있다.

마지막으로 현재 데이터 웨어하우스 시스템의 메타데이터를 모두 포함하고 있다는 점이다. 현재의 메타데이터를 프로토타입으로 이전하여 시험하는 과정에서 <그림4-3>의 각 영역이 현재의 데이터 웨어하우스 시스템의 메타데이터를 모두 포함하고 있는 것을 확인할 수 있었다. 이를 통해 종전 데이터와의 연속성이 보장되는 것을 기대할 수 있다.

## V. 결 론

본 연구는 데이터 웨어하우스의 구축과 운영에 중요한 메타데이터 스키마에 대한 연구이다. 메타데이터 스키마는 데이터 웨어하우스의 구축 초기 단계에서부터 구축 후 운영의 단계에까지 전 과정을 통하여 사용되고 중요하게 관리되어야 하는 데이터 웨어하우스의 메타데이터를 관리할 수 있는 체계를 제공한다. 따라서 본 연구는 성공적인 데이터 웨어하우스 메타데이터를 포함하기 위해 메타데이터를 네 개의 계층으로 분류하였고, 이를 바탕으로 전사적인 데이터 웨어하우스를 위한 메타데이터 스키마를 개발하였다.

제안된 메타데이터 스키마는 MDIS 표준을 포함하므로 호환성을 보장하는 스키마이다. 본 스키마는 국내 구축중인 실제 통신업체 데이터 웨어하우스 메타데이터에 성공적으로 접목 중이다. 또한, 메타데이터 관리를 위한 프로토타입 시스템 개발을 통하여 사용자 유용성을 제고하였다.

향후에는 본 연구의 스키마를 실제 프로젝트에 활용하여, 문제점을 보완하는 연구가 진행될 예정이다. 또한 이렇게 보완된 스키마를 다른 유사 프로젝트에도 적용해보는 연구도 필요할 것이다.

## 참 고 문 헌

- 박 석, 김희원, 김말희, "데이터 웨어하우스의 메타데이터," *정보과학회지* 제15권 제 5호, 1997. 5.
- 서우종, 이희석, "Metadata and Meta-Information System for Hypermedia Documents," *경영과학회지*, pp.89-92, 1998.
- 안연식, 이춘열, 이국철, "데이터 웨어하우스 구축을 위한 메타프로세스의 표현 방법," *경영과학회지*, 1998.
- Atre, A. and Storer, P., "Data distribution and warehousing," *DBMS*, pp. 54-62, October 1995.
- APT Data Group, "Briefing paper: what is metadata," *Working Paper*, 1996.
- Barquin, R.C. and Edelstein, H.A., *Planning and Designing the Data Warehouse*, Prentice Hall, 1997.
- Brown, S., "Preparing data for the data warehouse," *DCI's Data Warehouse World Conf.*, pp. D1-D36, 1997.
- Burton, P. and Green, S., "Meta data: the key to data warehouse design," *Working Paper*, 1996.
- Corey, M.J. and Abbey, M., *Oracle Data Warehousing*, McGraw Hill, 1997.
- Demarest, M., "Building the data mart," *DBMS*, vol.7, no. 8, pp.44-47, 1994.
- Dempsey, L. and Heery, R., "A review of metadata: a survey of current resource description formats," working paper, 1997.
- Devlin, B., "Data Warehouse from architecture to Implementation," Addison & Wesley, 1997.
- Drewry, M., H. Conover, S. McCoy and S.J. Graves, "Metadata: quality vs. quantity," *2<sup>nd</sup> IEEE Metadata Conf.*, 1997.
- Duyn, J.V., *Developing a Data Dictionary System*, Prentice Hall, 1982.
- Gray, P. and Watson, H.J., *Decision Support in the Data Warehouse*, Prentice Hall, 1998.
- Greenfield, L., The Data Warehousing Information Center, (<http://pwp.starnetinc.com/larryg/catalog.html>), 1999.
- Griffin, J., "Metadata: capturing the heart of the data warehouse," *Application Development Trends*, 1996.
- Hackathorn, R., "Data warehousing energizes your enterprise," *Datamation*, pp.38-42, February 1995.
- Hackney, D., *Understanding and Implementing successful Data Marts*, Addison Wesley, 1997.
- Hackney, D., "Metadata in 850 (or so) words," *Data Warehouse Forums*, 1997.
- Hurwitz, J., "The evolution of metadata," *DBMS*,



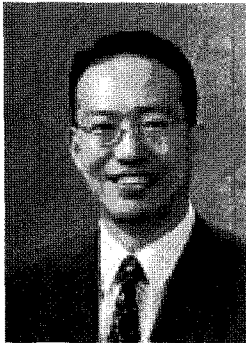
- pp.12-14, July 1997.
- Inmon, W.H., *Building the Data Warehouse*, John Wiley & Sons, 1993.
- Inmon, W.H. and R.D.Hackathorn, *Using the Data Warehouse*, John Wiley & Sons, 1994.
- Inmon, W.H., Welch, J.D., Katherine and L. Glassey, *Managing the Data Warehouse*, John Wiley & Sons, 1997.
- Inmon, W.H., Zachman J. A., and Geiger, J. G, 1997, *Data Stores, Data Warehousing, and the Zachman Framework*, McGraw-Hill.
- Kelly, S., *Data Warehousing in Action*, John Wiley & Sons, 1997.
- Kelly, S., *Data Warehousing: the Route to Mass Customization*, John Wiley, 1996.
- Kimball, R., "Meta meta data data," *DBMS*, March 1998.
- Kimball, R., *The Data Warehouse Toolkit*, John Wiley & Sons, 1996.
- Lee, H. and Kim, T., "A Metastore-based data warehouse development methodology," *International Conf. Digital Library & Knowledge*, Sep 1998.
- Marco, D., "Managing meta data," *Data Management Review*, April 1998.
- Metadata Coalition, "Metadata Interchange Specification (MDIS) Version 1.1," working paper, 1997.
- Meyer, D. and Cannon, C., *Building a better Data Warehouse*, Prentice Hall, 1998.
- Moriarty, T. and Mandracchia, C., "Heart of the warehouse," *Database Programming and Design*, pp.70-74, December 1996.
- Orfali, O., Harkey, D., and Edwards, J., *The Essential Client/Server Survival Guide*, John Wiley & Sons, 1996.
- Ozsu, M. T. and Valduriez, P., *Principles of Distributed Database Systems*, Prentice Hall, 1991.
- Poe, V., *Building a Data Warehouse for Decision Support*, Prentice Hall, 1996.
- Poe, V., "Data warehouse: architecture is not infrastructure," *Database Programming and Design*, July 1995.
- Sachdeva, S., "Meta data architecture for data warehousing," *Data Management Review*, April 1998.
- Sherman, R.P., "Metadata: the missing link," *DBMS*, August 1997.
- Spiers, J., "Metadata: maximizing the value of information," *Working Paper*, 1997.
- Subramanian, A., Smith, L.D. and Nelson A.C., "Strategic planning for data warehousing in the public sector," *Proc. of the 29th Annual HICSS*, pp.54-61, 1996.
- Sumpter, R. M., "Whitepaper on data management," *Working Paper*, 1994.
- Taylor, D. A., *Business Engineering with Object Technology*, John Wiley & Sons, 1995.
- White, C., "The key to a data warehouse," *Database Programming and Design*, pp. 23-25, February 1995.

## 부 록

<회사의 개선후 메타스키마 개체 상세 설명>

| 계층구분            | 개 체 명            | 내 용                                                         |
|-----------------|------------------|-------------------------------------------------------------|
| 핵심<br>웨어<br>하우스 | Database         | 파일, 관계형 데이터베이스, 망 데이터베이스, 계층 데이터베이스, 다중차원 데이터베이스, 객체 데이터베이스 |
|                 | Subschema        | 데이터베이스의 의미있는 부분집합으로 레코드들의 논리적인 묶음                           |
|                 | Level            | 데이터의 다중입방체에서 고유의 위치를 표현하기 위한 계층을 가지는 요소                     |
|                 | Table            | 데이터 웨어하우스를 구성하고 있는 관계형 데이터베이스의 테이블                          |
|                 | Element          | 테이블을 구성하고 있는 컬럼                                             |
|                 | Relationship     | 테이블간의 관계에 대한 정보, 사실 테이블과 색인 테이블과의 관계, 테이블과 테이블 간의 기수 관계     |
|                 | Dimension        | 사실 테이블에 대하여 사용자가 지정한 차원                                     |
| 적용              | ETTJob           | 운영시스템으로부터 데이터 웨어하우스로 필요한 데이터를 추출, 여과, 변환, 적재하는 작업           |
|                 | ETTTask          | 추출 변환 작업을 구성하는 작업                                           |
|                 | SourceSystem     | 운영 시스템                                                      |
|                 | SourceTable      | 운영 시스템의 테이블/파일                                              |
|                 | SourceData       | 운영 시스템의 컬럼/필드                                               |
|                 | Mapping          | 운영 시스템의 데이터와 데이터 웨어하우스의 데이터간의 관계                            |
| 사용자<br>탐험       | User             | 데이터 웨어하우스 사용자                                               |
|                 | AccessGroup      | 데이터 웨어하우스 사용자 집단                                            |
|                 | AccessLevel      | 데이터 웨어하우스 사용자 그룹의 접근 권한                                     |
|                 | Query            | 데이터 웨어하우스에 저장된 사용자 질의                                       |
|                 | QueryHistory     | 사용자 질의의 수행 이력                                               |
| 비지<br>니스        | BusinessArea     | 데이터 웨어하우스를 이용하는 조직                                          |
|                 | BusinessQuestion | 데이터 웨어하우스를 이용하여 얻고자 하는 비즈니스적인 질의                            |

◎ 저 자 소 개 ◎



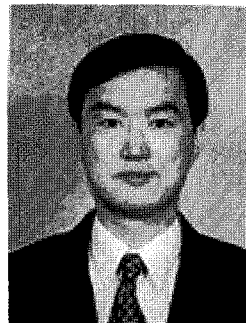
**이 희 석 (dbdc@unitel.co.kr)**

공동저자 이희석은 서울대 산업공학학사, KAIST 산업공학 석사, 그리고 University of Arizona에서 경영정보시스템을 전공하여 경영학박사학위를 취득하였으며 University of Nebraska at Omaha에서 MIS 강의를 하였다. 현재는 KAIST 테크노경영대학원에서 경영 정보분야를 연구하고 있다. 주요 관심분야는 지식관리, 데이터베이스, 기업정보시스템, 정보전략, ERP 등이다.



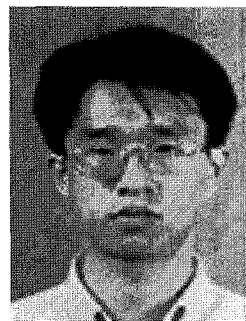
**조 남 철 (chonc@hansol.co.kr)**

공동저자 조남철은 서울대 수학과를 졸업하고 KAIST 테크노경영대학원에서 경영학석사 학위를 취득하였으며 현재 한솔텔레콤 SI팀에서 근무하고 있다. 주요 관심분야는 데이터 웨어하우스 구축, 분산객체 기술, 데이터베이스 설계 등이다.



**손 명 호 (totalsol@chollian.net)**

공동저자 손명호는 서울대 산업공학과를 졸업하고 KAIST 테크노경영대학원에서 경영학 석사 학위를 취득하였으며 현재 KAIST 테크노경영대학원 박사과정을 이수중이다. 주요 관심분야는 지식경영, 데이터 웨어하우스 구축, 분산객체 기술 등이다.



**김 태 훈 (kdbdc@chollian.net)**

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