의사결정지원을 위한 데이타 웨어하우스 시스템 프로토타입 개발
- 건강진단 시스템을 중심으로 -

Prototype Development of Data Warehouse Systems to Support Decision Making
- focused on a medical examination system -

김 성 언, 이 유 진

대구효성가톨릭대학교 경영학과

요 약

의사결정 지원을 위해 각광받고 있는 데이타 웨어하우스 시스템은 주제 지향적, 통합적, 시계열적, 비휘발적인 데이타 저장공간을 보유하여 사용자가 쉽게 데이타에 접근하여 원하는 분석을 수행할 수 있도록 도와주는 고품질의 정보 제공 시스템이다. 본 논문에서는 구체적인 데이타 웨어하우스 시스템을 소개함에 있어 국내 병원의 건강진단 데이타 웨어하우스 시스템 프로토타입 개발을 시도한다. 이에 데이타 웨어하우스 시스템의 구성에 대해 살펴본 후, 데이타 웨어하우스 시스템 개발 툴인 Cognos사의 PowerPlay를 이용하여 건강진단 데이타 웨어하우스 시스템 개발을 시도하고, 그 구축 방법과 결과를 소개한다.

I. 서 론

데이타베이스 시스템의 출현은 원하는 정보를 쉽게 접근하게 함으로서 정보의 가용성을 증가시키는 효과를 가져왔다. 이 데이타베이스 시스템은 주로 조직을 운용하는 데 필요한 운영(operational) 데이타베이스로 구축되어 조직의 각 부서의 업무를 지원하여 왔다. 그러나 의사결정 지원시스템(decision support system : DSS)이나 중역정보시스템(executive information system : EIS)의 필요성이 증가함에 따라 데이타베이스에도 이러한 시스템을 효율적으로 지원하기 위한 새로운 요구사항이 나타나게 되었다. 즉, DSS나 EIS에서 요구하는 요약 및 분석 작업을 위하여 과거로부터 누적된 이질적이고 방대한 양의 데이타를 통합 관리할 필요성이 증가하게 된 것이다. 기존에는 이러한 요구사항을 운영 데이타베이스에서 직접 연산하여 필요한 정보를 제공해 왔는데 필요한 정보를 추출하기까지에는 엄청난 비용과 시간을 필요로 하거나 경우에 따라서는 불가능하였다.

데이타 웨어하우스는 이러한 운영 데이타베이스가 안고 있는 문제에 대한 해답을 제공한다. 왜냐하면 데이타 웨어하우스는 데이타베이스와는 달리 현재 상태의 데이타 뿐만 아니라 과거의 데이타도 유지하기 때문에 누적된 통합 데이타를 분석하여 필요한 정보를 추출하는 데 효과적이기 때문이다.

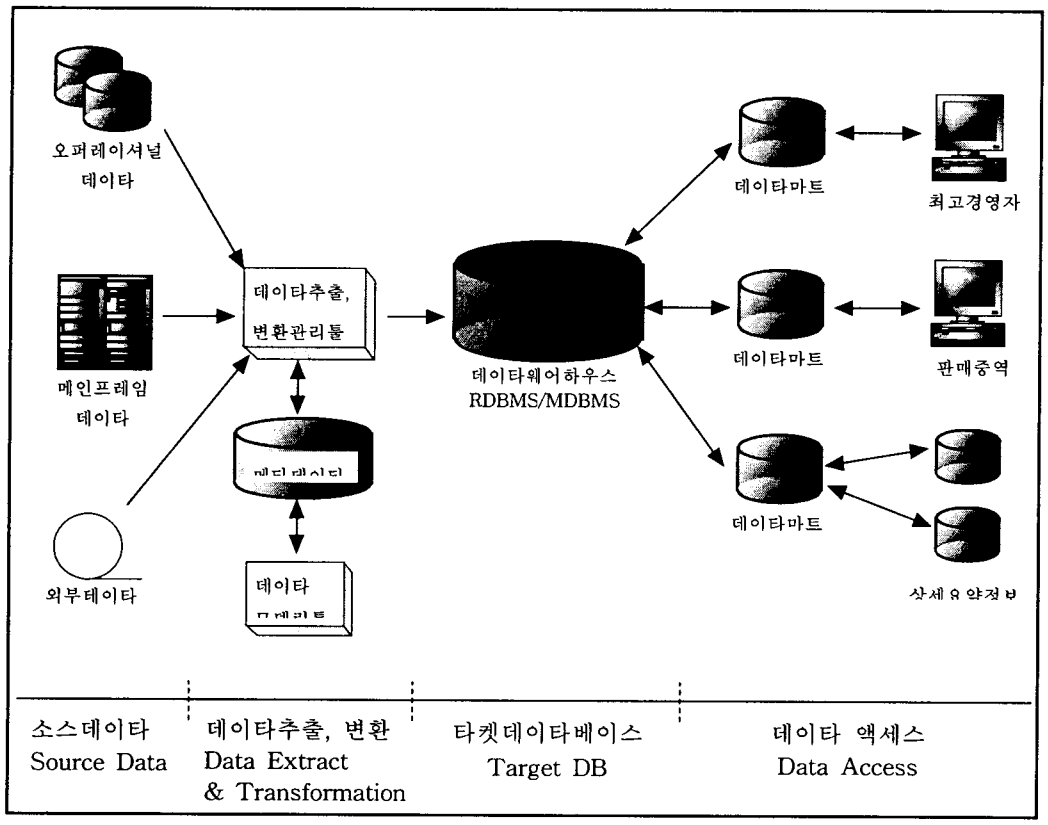
1990년대 초 국내에 데이타 웨어하우스 개념이 소개된 이후 데이타 웨어하우스 시스템 구축은 서비스/유통업계와 금융 및 통신업계 등에서 활발하게 이루어져 왔다. 오늘날 데이타베이스 업체를 선두로 하드웨어 공급업체, 시스템 통합업체, OLAP(On-Line Analytical Processing)툴 공급업체 등 많은 업체들이 데이타 웨어하우스 시스템 구축에 힘을 쏟고 있으며, 일반 기업에서도 이에 대한 관심이 급격히 확산되고 있다. 이와 더불어 데이타 웨어하우스에 관한 연구도 활발해 지고 있으나 이의 구축에 관련된 연구는 미비한 상태로 데이타 웨어하우스의 확산 열기에 보조를 맞추지 못하고 있는 실정이다.

본 연구에서는 신속하고 정확한 의사결정 지원으로 경영개선에 도움을 주는 데이타 웨어하우스 시스템을 실제로 응용할 수 있도록 프로토타입을 구축함으로써 데이타 웨어하우스 시스템의 구축효과와 효과적인 활용 방안을 제시하고자 한다.

II. 데이타 웨어하우스 시스템 구성

데이타 웨어하우스의 진정한 가치는 데이타를 수집하여 모으는 것보다 정보의 흐름을 관리하는 것이다.[1] 이러한 관점으로 볼 때 데이타 웨어하우스 시스템의 구성도는 업무 환경에 따라 다양하게 구성할 수도 있으나 일반적인 관점에서 <그림 1>과 같이 데이타 웨어하우스의 처리 과정에 따라 소스 데이타, 데이타 추출 및 변환, 타겟 데이타베이스, 데이타 액세스 부분으로 구성되고 있다.

어하우스에 대해서는 일반적인 관계형 데이타베이스나 분석업무에 적합한 다차원 데이타베이스가 사용된다. 데이타 액세스 부분에서는 구축된 데이타 웨어하우스에 데이타 액세스 도구를 사용하여 데이타 마트를 구축하고, 최종 사용자가 데이타 마트에 접근하여 의사결정에 도움을 받도록 한다. 데이타 마트는 전사적인 데이타 웨어하우스의 부분 집합으로서 특정 사용자나 대상에게 가치가 있는 데이타들을 포함하는 소규모의 데이타 웨어하우스이다.



<그림 1> 데이타 웨어하우스 시스템 구성도[2]

소스 데이타 부분은 현재 사용하고 있는 일반화일 시스템, 관계형 데이타베이스, 계층형 데이타베이스 및 외부 데이타를 포함하는 오퍼레이셔널 데이타 환경을 말한다. 데이타 추출 및 변환 부분에서는 소스 데이타에서 타겟 데이타베이스로 데이타를 추출하고, 표준화하여 통합, 요약, 저장하기 위한 특수한 도구를 사용한다. 타겟 데이타베이스 부분에는 소스 데이타를 주제 지향적이며, 통합적으로 집적한 데이타 웨어하우스가 구축되며 이 데이타 웨

III. 데이타 웨어하우스 시스템 프로토타입 개발

III.1 프로젝트 개요

국내 병원의 경영환경이 급속도로 악화되고

1) Hackathorn, R., "Data warehousing energizes your enterprise", Datamation, 1995. 2. pp.38-44

2) 문상룡, "정확한 의사결정의 새주역(데이터 웨어하우징)," 포스 데이터소식, 통권 64호, 1996. 4, pp.38-41

있는 가운데, 경쟁력 강화를 위해 많은 병원들이 데이타 웨어하우스 시스템 구축을 적극 고려하고 있다.

데이타 웨어하우스 시스템은 원본 데이타를 추출하여 주제 중심적이고, 통합적이며 시간성을 가지는 비휘발성 자료의 집합인 데이타 웨어하우스를 통해 다차원 분석을 가능하게 하는 시스템이다. 그러므로 병원의 데이타 웨어하우스 시스템을 구축하면 정확한 경영분석을 통한 각종 비용 발생 통제와 적절한 투자에 의한 경영 상태 개선, 또한 다양한 형태의 임상 분석 작업 등의 효과를 기대할 수 있다.

데이타 웨어하우스 시스템을 병원에 적용할 수 있는 분야는 크게 다음과 같은 두 가지로 나타난다. 첫째, 경영을 위한 분석으로 기존의 O.C.S (Order Communication System : 처방전달시스템)에서 발생된 데이타를 이용하여 이를 다차원으로 분석하여 원가분석 시스템 및 경영분석 시스템에 활용하는 것이고, 둘째, O.C.S나 진단시스템의 데이타 등을 활용하여 다양한 형태의 분석 작업을 하는 것으로 이것은 환자의 증상별 분석 뿐만 아니라 의사, 간호사의 연구자료, 임상분석, 의무기록자료의 통계 등에 이용될 수 있다.

본 논문에서는 병원정보시스템 상 확보할 수 있는 데이타가 한정되었기 때문에 데이타 웨어하우스 시스템의 구축으로 보여 줄 수 있는 분야는 후자에 국한하기로 한다. 이에 건강진단 자료를 이용하여 다차원적인 분석작업을 통해 의사결정 지원을 달성할 수 있는 데이타 웨어하우스 시스템 프로토타입을 개발하기로 한다.

III.2 적용도구

1995년까지만 해도 데이타 웨어하우스 시스템 구축시 Visual Basic, Power-Builder, Developer2000 등 4GL을 많이 사용하였다. 그러나 최근 들어 데이타 웨어하우스 시스템 구축에 대한 관심이 증대되면서 비용과 시간면에서 더욱 효과적이고 좀 더 다양한 분석을 쉽게 할 수 있는 데이타 웨어하우스 구축 도구가 발전하였다. 이에 본 연구는 보다 효율적인 데이타 웨어하우스 시스템 개발을 위하여 Cognos사의 Impromptu와 PowerPlay를 사용하여 데이타 웨어하우스 시스템 구축을 시도하였다.

Impromptu는 SQL Query를 활용하여 데이타베이스에 직접 접근해서 다양한 보고서를 손쉽게 작성할 수 있는 질의 및 보고서 작성 도구이며, PowerPlay는 이차원 구조의 데이타를 다차원의 구조로 재구성하여 검색하고 분석을 용이하게 하는 MOLAP (Multidiemensional On-Line Analytical Processing) 도구로서 다양한 차원 분석과 빠른 응답시간의 제공이 그 장점이다.

III.3 건강진단 데이타 웨어하우스 시스템 개발

(1) 프로젝트의 범위 정의

데이타 웨어하우스 시스템의 개발에 있어 가장 첫 번째 단계는 프로젝트의 범위를 결정하는 것이다. 프로젝트의 범위는 우선 최종 사용자를 확인하고, 요구사항을 결정한 다음 요구사항 정의서의 초안을 작성함으로써 정해진다.

본 연구는 임상자료로서 의미가 있으므로 최종 사용자는 의사와 간호사, 의료 연구원 등으로 본다. 요구사항의 결정은 대개 적절한 최종 사용자 팀을 구성하여 인터뷰나 설문지 형식을 통하여 그들의 의견을 종합하여 결정한다. 그러나 건강진단 시스템은 진단자료를 통한 환자들의 다차원적 증상 분석이라는 동일한 요구사항을 지니기 때문에 이 과정은 생략하였다. 요구사항이 결정되고 나면 요구사항 정의서 즉, 차원 맵을 작성해야 한다. 차원 맵은 <표 1>과 같이 차원과 카테고리로 이루어져 있다.

프로젝트의 범위를 결정하는 것은 시스템 구축에 있어 매우 빠르고 효과적인 수행을 유도한다. 만약 이 단계를 잘 준비하고 계획한다면 시스템 구축 시간을 절감할 수 있고 기대치 않은 결과치들이나 과정 후의 불필요한 조정작업 등을 줄일 수 있게 된다.

(2) 원본 데이타 확보

데이타 웨어하우스 시스템 개발의 두 번째 단계는 원본 데이타를 확보하여 원본 데이타에 필요한 데이타가 있는지, 기존 데이타베이스에 접속이 가능한지 그리고 원본 데이타가 수정되거나 변환될 필요는 없는지를 확인하는 것이다.

본 연구에서 건강진단 데이타 웨어하우스 시스템을 개발하기 위하여 확보한 자료는 A자동차 서비스 직원 420명과 B합섬 죽전공장 직원 137명의 건강진단자료이다.

차원	카테고리 1	카테고리 2	카테고리 3	측정기준
최종진단	진단대분류	진단명		인원수
성별	성별			인원수
나이	연령대	나이		인원수
신장	대분류	신장		인원수
몸무게	대분류	몸무게		인원수
비만도	비만도별			인원수
최고혈압	최고 혈압대별	최고혈압		인원수
최저혈압	최저 혈압대별	최저혈압		인원수
과거병력	대분류	과거병력		인원수
생활습관	대분류	중분류	개선습관명	인원수
외상 및 후유증	외상 및 후유증			인원수
일반상태	일반상태			인원수

<표 1> 건강진단 시스템의 차원 맵

구 분	남	여	합 계
A자동차 서비스	404	16	420
B합섬 죽전 공장	98	39	137
합 계	502	55	557

<표 2> 건강진단 시스템 원본 자료 분류

건강진단 시스템의 원본 데이타는 간단한 환자의 신상과 생활습관 등이 기록되어 있고, 진단결과는 총 7가지의 검사 종목으로 나누어 수치나 유형별로 나타나 있었다. 이 원본 데이타를 이용하여 전 단계에서 설계한 차원 맵에 해당하는 항목들을 추출하였다. 기존 데이타베이스와의 접속은 원본 데이타의 비전산화로 그 가능성이 희박하였다. 원본 데이타의 수정과 변환은 필요에 의해 나이 차원에 국한하여 이루어졌다.

(3) 원본 데이타 변환과 카탈로그 구축

이 단계에서는 원본 데이타를 Impromptu 프로그램이 읽을 수 있도록 DBF 파일로 변환한 다음, 원본 데이타베이스와 연결된 Query 원본 즉, 카탈로그를 생성한다.

원본 데이타 변환과 카탈로그 구축 단계에서 고려할 사항은 다음과 같다. 첫째, 원본 데이타에서 추출한 데이타가 정확한지 확인해야 한다. 올바른 데이타의 사용은 효과적인 카탈로그의 작성을 가능하게 한다. 둘째, 카탈로그는 Impromptu 보고서를 작성하는데 요구되는 모든 데이타 열들을 제공함으로서 최적의 효율성을 가지도록 구성해야 한다. 셋째, 데이타가 제대로 통합되었는지 살펴보아야 한다. 넷째, 관련 측정기준인 인원수가 전 모델에 걸쳐 올바로 대비되는지를 확인하고, 새로운 측정기준이 원본 데이타로부터 필요하지 않는지 고려해야 한다. 이 경우, 본 논문에서는 건강진단시스템 자료 중에서 환자의 신상이나 환경만을 고려하는 데이타 웨어하우스 시스템을 개발하려 하므로 인원수를 제외한 다른 측정기준치는 적용되지 않았다.

(4) Impromptu 보고서 생성과 Query 설계

건강진단 데이타 웨어하우스 시스템 개발의 네 번째 단계는 트랜스포머 원본 데이타를 위한 Impromptu 보고서를 생성하고, 데이타 처리과정 및 배분기능을 최적화 시킬 수 있도록 Query를 설계하는 것이다.

Impromptu 보고서는 크게 2가지 역할을 하는 데, 첫째, 정형화된 보고서의 기능이고, 둘째, PowerPlay에서 인식하는 트랜스포머의 원본 데이타로서의 역할이다. 본 논문에서는 트랜스포머의 원본 데이타로서의 Impromptu 보고서 기능에 그 초점을 둔다.

Impromptu 보고서 생성과정이 끝나면 트랜스포머에서의 원활한 사용을 위한 Query 설계를 한다. 이 과정은 Query 창에서 이루어지는데 Query 창은 데이타 처리 과정이나 배분 기능을 최적화 시킬 수 있도록 데이타, 소트, 그룹, 필터, 액세스 등의 논리적 구조를 설계할 수 있다.

Impromptu 보고서 생성과 Query 설계 단계에서 고려해야 할 사항을 정리하면 다음과 같

다. 첫째, Impromptu 보고서를 작성할 때 원활한 자료추적을 위한 올바른 데이타의 구조를 작성해야 한다. 이것은 올바른 데이타의 여부가 모델의 효율성을 결정함으로 많은 주의를 기울여야 한다. Impromptu 보고서에서 작성된 모든 데이타 항목들이 트랜스포머에 연결되거나 상세사항으로 원활한 Drill Through를 하기 위해서는 올바른 데이타 구조를 작성하는 것이 매우 중요하다. 둘째, 데이타의 통합성을 고려해야 한다. 이 작업은 데이타가 올바른지, 원하는 데이타인지를 재확인하는 과정이다. 셋째, 데이타의 성김성(Sparsity)을 확인해야 한다. 즉, 모델에 영향을 미칠 만큼 하위 자료추적 레벨에 빠진 데이타가 없는지를 확인해야 한다. 넷째, 보고서와 모델이 이상적인 크기로 필터가 되어 있는지를 살펴보아야 한다.

(5) 트랜스포머 모델 및 PowerPlay 화면 작성

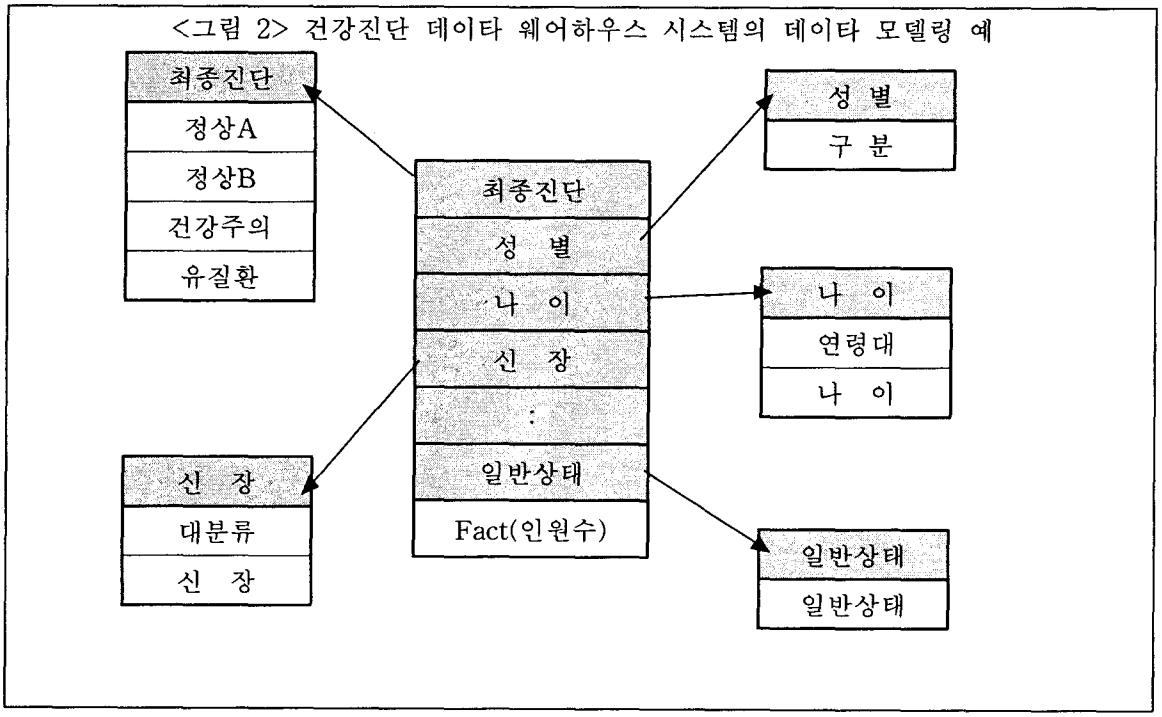
이 단계는 Impromptu 보고서를 토대로 트랜스포머 모델 및 PowerPlay 화면을 작성하는 것으로, 이 단계에서는 최종 사용자들의 데이타 요구사항을 결정하고 최종 사용자의 업무 지표를 만족시킬 수 있는 모델 화면과 측정기준을 작성해야 한다. 이 과정은 첫째, 데이타 모델링, 둘째, 트랜스포머의 차원 작성, 셋째, 데이타 큐브 생성, 넷째, PowerPlay 연결 등의 4가지 단계로 세분할 수 있다. 각 단계별로 구현 방법과 고려 사항을 건강진단 시스템에 비추어 살펴보면 다음과 같다.

① 데이타 모델링

데이타 모델링이란 '데이타구조와 프로세스에 대하여 언어, 그림 숫자, 기타의 매체에 의해 추상화하는 표현'이다[3]. 데이타 모델링은 구조와 프로세스를 잘 이해할 수 있도록 해준다. 또 데이타 설계에서 데이타 구조가 정확하도록 돕는다.

데이타 웨어하우스의 데이타 모델링 기법은 주로 스타 스키마 기법과 스노우플레이크 스키마 기법으로 나눌 수 있다.

<그림 2> 건강진단 데이타 웨어하우스 시스템의 데이타 모델링 예



스타 스키마(Star Schema) 기법은 비정규화의 특수한 형태이다. 스타 스키마는 사실 테이블과 차원 테이블로 구성되는데 사실 테이블이 단일축을 이루고 차원 테이블이 여러 개의 위성처럼 붙게 된다. 사실 테이블은 업무와 관련된 성과 데이타를 포함하며, 단위로써 표현이 가능한 수치적 데이타를 갖는다. 차원 테이블은 제약조건에 해당하는 문자적 값을 포함한다. 스타 스키마의 장점은 사실테이블 내에 있는 데이타에 대한 검색기준으로 차원 테이블의 열을 사용함으로써 응답에 필요한 조인의 횟수를 최소화시키는 것이다. 그러나 이것

3) 이병수 외 2人, "데이타웨어하우스 구축 방법론에 대한 연구", 한국산업정보학회논문지 제4권 제2호, 1999. 6. pp.23-31

은 어플리케이션의 확장성과 유연성을 제한하는 측면도 가지고 있다.

스노우플레이크 스키마(Snowflake Schema)는 스타 스키마의 차원 테이블이 좀더 정규화된 형태이다. 스노우플레이크 스키마는 데이타를 저장하는 데 필요한 저장공간을 최소화하고, 어플리케이션의 유연성을 증가시킨다. 또한 정규화된 작은 차원 테이블들이 조인됨으로써 성능향상을 꾀할 수 있다는 장점을 가지고 있다. 그러나 이러한 장점에도 불구하고 많은 테이블로 인해 그 구조가 복잡하며 사용자 질의에 대한 정확한 결과를 검증하는 데 많은 어려움이 발생한다.

위의 두가지 모델링 설계 기법 중 본 논문의 논리적 데이타 모델링 설계에는 차원 맵의 초안을 기준으로 스타 스키마(Star Schema)기법을 적용하였다.

<그림 2>는 스타 스키마 기법을 이용한 건강진단 데이타 웨어하우스 시스템의 논리적 모델 설계의 예이다. 이 그림에서 보듯이 건강진단시스템에서는 사실테이블이 최종진단, 성별, 나이, 신장 등의 차원과 측정기준인 인원수로 구성된다. 그리고 각각의 차원은 차원 테이블을 형성하고 있다.

② 트랜스포머의 차원 작성

트랜스포머의 차원 작성은 건강진단 시스템에 대한 논리적인 모델링을 완성한 다음 트랜스포머를 실행하여 카테고리, 자료 추적 레벨 및 측정기준 등의 차원을 작성하는 것이다.

이러한 트랜스포머의 차원 작성 시에는 다음과 같은 사항들을 고려해야 한다. 첫째, 측정기준이 기존의 목표와 실적 지표들에 잘 부합되는지를 확인한다. 둘째, 효율적인 자료추적을 위해 원활한 자료추적 경로를 설계해야 한다. 자료추적 경로를 설정할 때는 상위 레벨 하나에 너무 많은 하위 레벨의 카테고리를 설정하지 말고 직관적으로 잘 사용될 수 있는 레벨들을 사용한다. 또한 측정기준 및 지표들을 하위 자료 추적 레벨에 올바르게 배분해야 한다.

③ 데이타 큐브 생성

데이타 큐브는 다차원 데이타베이스의 다차원 배열에서 시작된 개념으로 여러 가지 다른 그룹 속성을 이용하여 사실 테이블로부터 형성된 다중 집계 뷰를 다루기에 간편한 방법이다.[4] 트랜스포머에서는 데이타 큐브의 한 형태로 PowerCube를 생성한다. PowerCube는 PowerPlay에서 분석을 실행할 수 있도록 압축된 파일의 형태로 저장되어 있는 선택된 측정기준의 모임이다. PowerCube는 다음과 같은 3가지 방법으로 생성이 가능하다.

첫째, 기본 설정으로서 모델이 정의하고 있는 모든 차원과 측정기준들을 포함시켜 작성한다. 둘째, 수정이 가능해서 PowerCube를 차원 일부 및 특정한 측정기준을 제외하고 작성할 수 있다. 셋째, PowerCube 그룹 형태로서 PowerCube를 한 차원 내에서 특정한 레벨로 그룹하여 작성할 수 있다.

건강진단시스템에 대한 데이타 큐브 즉, PowerCube는 이 3가지 방법 가운데 첫 번째 방법인 기본 설정으로 생성하였다. <표 3>은 데이타의 상세수준이 가장 낮은 건강진단 데이타 큐브의 예이다. 이 표에서 알 수 있듯이, 건강진단 큐브는 최종진단, 성별, 나이 등 총 12개의 차원을 가지며 한 개의 Fact 인원수를 가진다. 이 큐브를 통해 분석지원이 가능한 항목은 최종진단결과를 기준으로 한 환자의 생활습관이나 상태이다.

<표 3> 데이타 큐브의 예

CUBE명	주요 분석 보고서	FACT	DIMENSION
건강진단	건강진단자료에 의한 최종진단결과를 기준으로 환자의 생활습관이나 상태를 분석	인원수	최종진단, 성별, 나이, 신장, 체중, 비만도, 최고혈압, 최저혈압, 과거병력, 생활습관, 외상 및 후유증, 일반상태

④ PowerPlay 연결

트랜스포머 모델링이 완성되고 데이타 큐브가 생성되고 나면, PowerPlay로 연결이 가능해진다. PowerPlay 프로그램은 트랜스포머를 원본 데이타로 읽어 들이기 때문에 PowerPlay를 실행한 다음 트랜스포머에서 생성한 데이타 큐브를 찾으면 되므로 연결방법은 비교적 단순하다.

지금까지 건강진단 데이타 웨어하우스 시스템을 5단계로 나누어 그 구축 방법과 고려사항에 대해 살펴 보았다. 이렇듯 데이타 웨어하우스 시스템의 구축이 완성되고 나면 마지막으로 고려해야 할 사항이 있다. 첫째, 사용자

4) V. Hannarayan, A. Rajaraman & J. D. Ullman, "Implementing data cubes efficiently", In Proceedings of ACM SIGMOD 1996 International Conference on Management of Data Engineering, 1989. 6

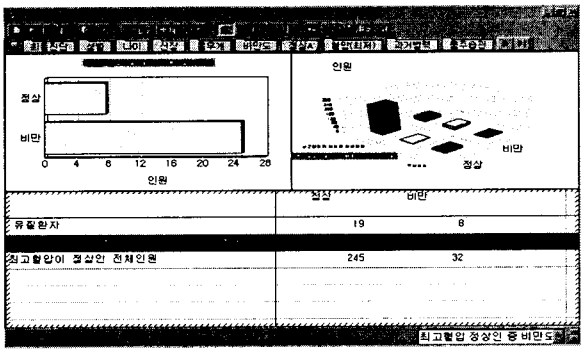
의 요구사항 변경을 항상 확인해야 하고, 둘째, 원하는 데이타의 검색과 분석이 달성될 때까지 수정 및 보완 작업을 계속해야 한다.

V. 건강진단 데이타 웨어하우스 시스템의 기능

건강진단 데이타 웨어하우스 시스템의 기능으로는 추세분석, 상세 수준 분석, 다차원 분석 및 다양한 보고서 제공 등을 들 수 있다. 각각의 기능을 자세하게 살펴보면 다음과 같다.

① 스프래드시트나 그래프를 통한 추세 분석

건강진단 데이타 웨어하우스 시스템은 스프래드시트나 그래프를 통한 추세 분석이 가능하다. 예를 들어, 최고 혈압이 정상인 사람들 중에서 몸무게가 정상인 사람과 비만인 사람이 각각 질환에 걸리는 비율을 분석하고자 한다면, 건강진단 데이타 웨어하우스 시스템에서 <그림 3>과 같은 분석 결과를 신속하게 얻을 수 있다.



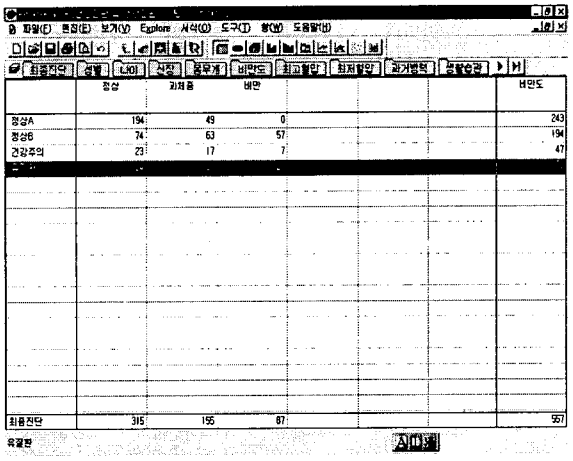
<그림 3> 건강진단 데이타 웨어하우스 시스템의 추세 분석 결과의 예

<그림 3>에서 나타나 듯이 최고 혈압과 몸무게가 정상인 사람 중 유질환자는 8%에 불과하나 최고 혈압이 정상이지만 비만인 사람 중 유질환자는 25%로 나타났다. 또, 사용자의 보다 빠른 이해를 돕기 위하여 상단에 3차원 그래프를 나타내어 추세나 구성비 등을 쉽게 파악할 수 있다.

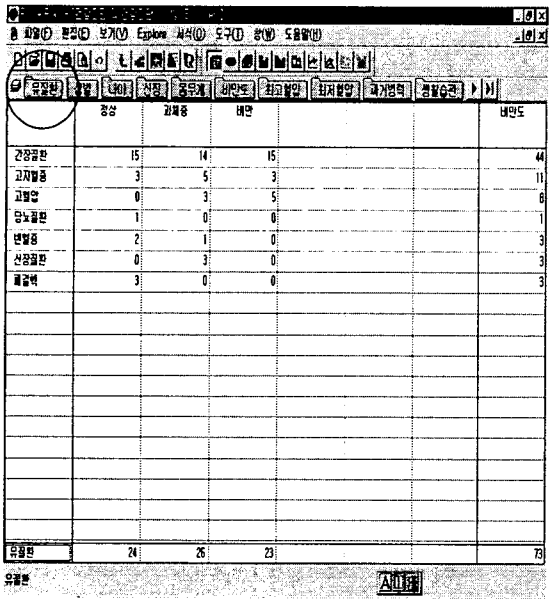
② Drill up/Drill down을 이용한 검색과 분석

사용자는 Drill down을 실행함으로써 집계 레벨로부터 상세 수준으로, Drill up을 실행함으로써 그 반대로 데이타를 자유자재로 검색할 수 있다. 그 예를 들어 보면 <그림 4> <그림 5>와 같다. <그림 4>는 비만도와 최종진단에 대한 분석 결과이다. 즉, 비만도에 따라 몸무게가 정상인 사람, 과체중인 사람, 비만인 사람으로 분류하여 각각의 최종진단 결과를 살펴볼 수 있다. 여기에서 유질환에 대한 더 상세한 정보를 알고 싶으면 drill-down을 원하는 항목인 유질환을 더블 클릭하면 된다

<그림 4> Drill down의 초기 화면 예



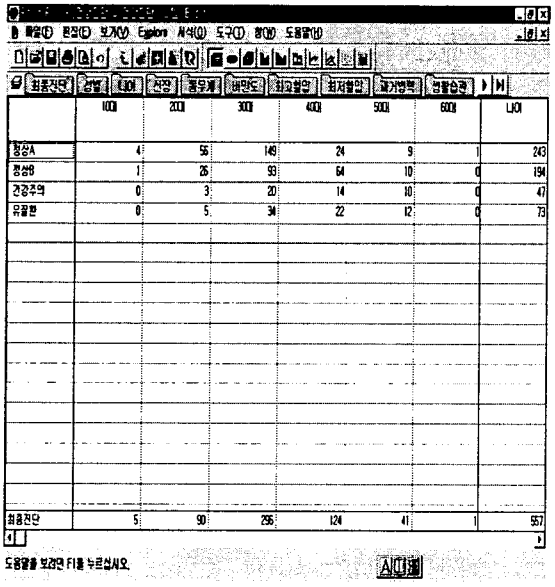
<그림 5>는 그 결과 화면으로서 유질환의 항목별 수치분석이 되어 있다. 이 때, 보고서 상에서 사용자가 무엇을 보고 있는지 알 수 있게 하기 위하여 차원 라인은 사용자가 화면 상에 나타나 있는 데이타의 상세 정도를 나타내주는 지도 역할을 하게 되는 데 최종진단 차원이 열린 폴더 그림의 유질환으로 바뀐 것을 확인할 수 있다. 사용자는 이와 같이 쉽게 drill-up, drill-down을 할 수가 있어 원하는 항목의 분석결과만을 편리하게 볼 수가 있다



<그림 5> Drill down의 예시 화면

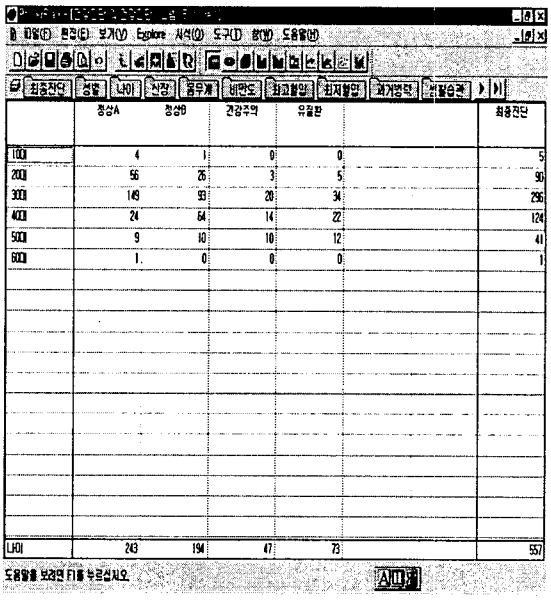
③ 행과 열, 그리고 레이어를 활용한 다차원 분석

행과 열을 자유롭게 바꾸어 분석하는 예가 <그림 6>와 <그림 7>에 나타나 있다.



<그림 6> 행과 열을 바꾸는 분석의 초기화면 예

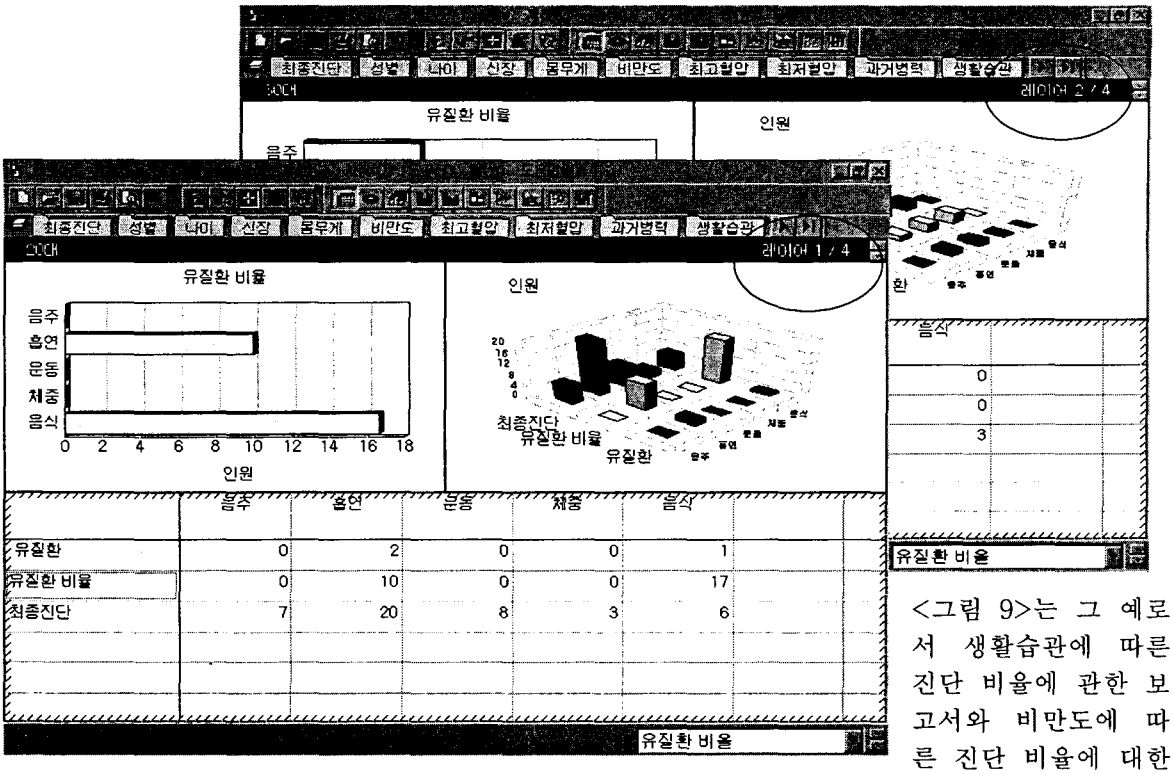
<그림 6>은 행의 항목이 연령대이고 열의 항목이 최종진단이다. 그런데 <그림 7>에서는 최종진단과 연령대가 서로 행과 열이 바뀌어 보여진다. 이러한 기술은 다차원 분석에서의 유연성을 증가시킨다.



<그림 4-7> 행과 열을 바꾸는 분석의 예시 화면

또, 사용자는 행과 열 또는 그래프 등에 레이어를 추가함으써 여러 기준으로 분석을 실행할 수 있다. 예를 들어, 유질환자를 기준으로 각 연령대 별 생활습관 개선 항목에 대한 분석을 원한다면 <그림 8>과 같이 레이어를 사용하여 분석을 도울 수 있다.

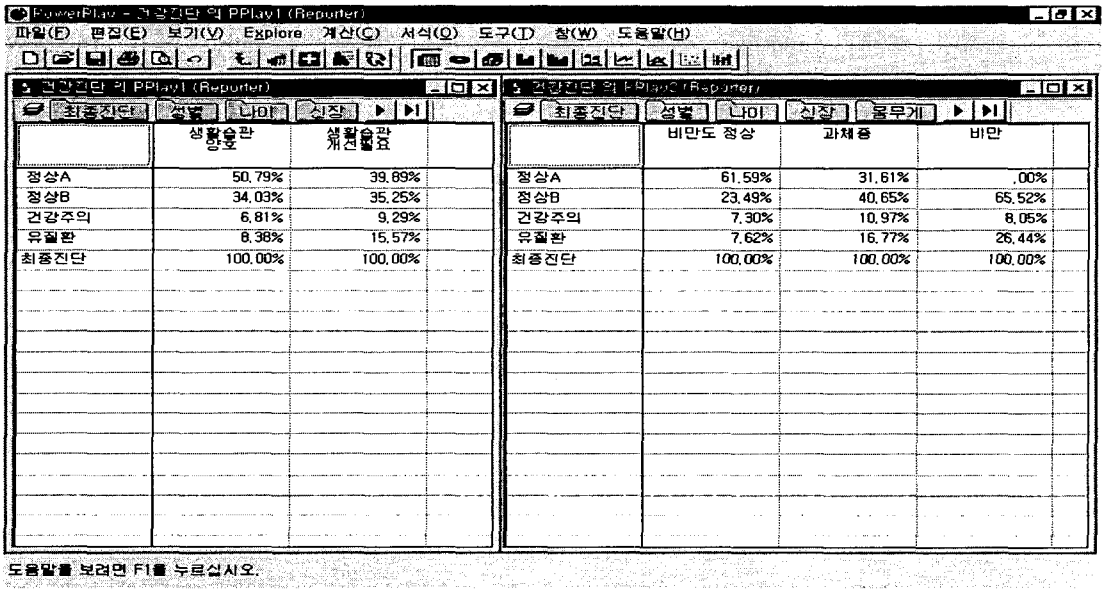
<그림 8>은 각 개선항목에 대한 유질환자 비율이 하단 스프래드시트에 나타나 있고 분석 기준이 되는 연령대가 레이어로 나타나 있다. 이 그림은 총 4개의 레이어 중 첫 번째와 두 번째 레이어로서 20대와 30대 유질환자의 생활 습관 개선항목이 분석되어 있다. 그 분석 상단의 원그래프와 꺾은 선 그래프로 알 수 있듯이 20대 유질환자 중에서는 음식 습관을 개선해야 할 사람들이 가장 많았고 그 다음으로는 흡연자가 많은 것으로 분석되었다. 레이어 분석은 한 번에 한 레이어씩 보여지므로 다음 레이어로 이동을 원하면 화살표를 누르면 된다. 그 결과 <그림 8>에서 보여지듯이 30대 유질환자의 생활 습관 개선항목이 나타난다.



<그림 8> 레이어 예시화면

건강진단 데이타 웨어하우스 시스템은 레이어를 통한 분석을 이용하여 각 연령대 별 유질환자의 생활습관 개선항목을 신속하게 판단할 수 있다.

<그림 9>는 그 예로서 생활습관에 따른 진단 비율에 관한 보고서와 비만도에 따른 진단 비율에 대한 보고서가 동시에 나타난다. 이것은 각기 다른 기준과의 비교 분석을 용이하게 하여 의사결정의 질을 향상시킨다.



<그림 9> 다양한 보고서 형태

④ 신속하고 다양한 보고서 제공

건강진단 데이타 웨어하우스 시스템은 다양한 요구사항에 맞는 보고서를 작성할 수 있으며 한 화면에서 두 가지 형태로 동시에 나타낼 수 있으므로 신속한 비교분석을 지원한다.

지금까지 살펴본 건강진단 데이타 웨어하우스 시스템의 구축 결과를 종합하면 먼저 스프레드시트나 그래프를 통한 추세분석과 Drill up/Drill down 분석이 가능하다. 또 행과 열,

그리고 레이어를 활용한 다차원 분석을 할 수 있고, 신속하고 다양한 보고서가 제공된다. 이 밖에도 차원 필터를 이용한 분석이 가능하고, Impromptu 보고서나 상세한 레벨의 다차원 데이타베이스와 연결되는 Drill through 기능을 통한 데이타 검색을 할 수 있다.

이와 같은 구축 결과를 통해 얻어낸 건강진단 데이타 웨어하우스 시스템의 효과를 정리하면 크게 4가지로 나누어 진다. 첫째, 최종 사용자의 정확하고 신속한 의사결정 지원이 가능하다. 둘째, 환자의 증상분석이나 임상자료 또는 연구자료로서의 올바른 정보 제공이 실현된다. 셋째, 데이타를 효율적으로 관리 할 수 있다. 마지막으로 넷째, 환자들의 신뢰감을 통한 경영의 안정화 등이다.

5. 결 론

본 논문은 사용자들이 의사결정을 하기 위해 필요로 하는 정보를 한곳에 저장해 두고 이 정보를 필요로 하는 사람에게 적시에 적절한 형태로 제공하는 데이타 웨어하우스 시스템의 프로토타입 개발을 목적으로 한다.

데이타 웨어하우스는 의사결정을 지원하기 위한 주제 지향적, 통합적, 시변적, 비휘발성 데이타의 집합이라 할 수 있다. 정보화, 지식화 시대인 현재를 살아가고 있는 기업들이 경쟁에서 우위를 점하고 보다 빠르고 정확한 의사결정을 지원 받기 위해서는 데이타 웨어하우스 시스템의 구축은 필수라고 할 수 있다.

본 논문에서는 경영 안정화와 경영환경 개선에 힘을 쏟고 있는 병원을 대상으로 윈도우 98의 운영체제하에 Cognos사의 Impromptu와 PowerPlay를 이용하여 건강진단 데이타 웨어하우스 시스템 프로토타입을 개발하였다.

구축된 건강진단 데이타 웨어하우스 시스템은 다차원 분석이 가능하고, 원하는 행과 열을 바꾸어 분석할 수 있다. 또 이 시스템은 수치분석과 추세분석 등을 위한 스프래드시트와 다양한 그래프를 제공할 수 있으며 원하는 여러 형태의 보고서도 신속하게 제공할 수 있다.

건강진단 데이타 웨어하우스 시스템의 효과로는 첫째 최종 사용자의 정확하고 신속한 의사결정 지원, 둘째 환자의 증상분석이나 임상자료 또는 연구자료로서의 올바른 정보 제공, 셋째 건강진단 데이타의 효율적 관리, 마지막으로 환자들의 신뢰감을 통한 경영의 안정화 등이다.

본 논문은 건강진단 데이타 웨어하우스 시스템을 개발하고 그 효과를 살펴 봄으로써 데이타 웨어하우스 개발자와 연구자에게 조금이나마 기여를 하였다고 본다. 우선, 데이타 웨어하우스 개발자에게는 실제 데이타 웨어하우스 시스템의 개발 단계를 자세히 소개함으로써 구현 시 장·단점과 문제점들을 파악할 수 있게 하였고, 데이타 웨어하우스에 관심이 있는 연구자들에게는 건강진단 데이타 웨어하우스 시스템을 소개함으로써 앞으로의 연구 방향에 대한 하나의 이슈를 제공하였다.

본 연구는 이러한 결과와 효과를 얻으면서도 몇 가지 한계점이 지적되는데 우선 건강진단 데이타 웨어하우스 시스템은 원본 자료가 건강진단자료로서 동일 인물에 대한 축적된 자료를 구할 수 없었기 때문에 데이타 웨어하우스의 큰 특징 중의 하나인 시간 차원에 대한 분석을 할 수 없었다. 또한 시스템 사용 결과에 대한 통계 분석이 제공된다면 데이타 웨어하우스 시스템 효과의 신뢰성을 더욱 뒷받침하리라 기대된다.

현재 데이타 웨어하우스에 대한 연구는 데이타 웨어하우스와 OLAP에 인트라넷까지 결합하는 연구와 다차원 분석과 OLAP에서 나아가 데이타 마이닝과 Business Modeling, Forecasting의 연구가 진행되고 있다. 그러므로 앞으로 이 분야에 대한 깊이 있는 연구가 필요하다고 볼 수 있다.

참 고 문 헌

강동진, "병원의 경쟁우위전략과 정보시스템 구축환경의 관련성 연구," 대구효성가톨릭대학교 대학원 박사학위논문, 1997.

박미선 외 2人, "보험산업의 데이타웨어하우스 구축방안에 관한 연구", 한국경영정보학회, '99 춘계 학술대회 논문집, 1999. 6. pp.397-411

신현신, "국내은행의 의사결정지원을 위한 데이타 웨어하우스 구축에 관한 연구, 성균관대학교 대학원 석사학위논문, 1996. 11.

이병수 외 2人, "데이타웨어 하우스 구축 방법론에 대한 연구", 한국산업정보학회논문지 제4권 제2호, 1999. 6. pp.23-31

이성수, 이영재, "전자제조업에서 데이타 마트 구현에 관한 사례연구", 한국경영정보학회, '99 춘계 학술대회 논문집, 1999. 6. pp.387-396

이재식 & 전용준, "데이타 웨어하우스 설계를 위한 개념적 모델링 접근 방법", 한국데이타베이스학회, '96 데이타베이스 심포지

움 및 학술대회, 1996. 11. pp.217-228
정 철, "의사결정 지원을 위한 데이타 웨어하우스의 구축", 한국외국어대학교 대학원 석사학위논문, 1997. 8.
조용찬, "은행정보계 시스템의 데이타 웨어하우스 구현 방향에 관한 연구", 동국대학교 대학원 석사학위논문, 1997.
조재희 & 박성진, "데이타 웨어하우징과 OLAP", 대청정보시스템, 1996.
Cognos, Introduction to Impromptu & PowerPlay, 1998.
Devlin, B., "From Data Model to Warehouse in Bite-Sized Chunks!", InfoDB, Vol 9, No 5, 1995. 10. pp.1-10
Hackathorn, R., "Data warehousing energizes your enterprise", Datamation, 1995. pp.38-44
Hammergren, T., "Data Warehousing : Building the Corporate knowledgebase," International Thomson Computer press, 1996
Inmon, W. H., "Building the Data Warehouse(2nd Ed.)," John Wiley & Sons, 1996.
Kimball, R. & Inmon W. H., "The Data Warehouse Toolkit," John Wiley & Sons, Inc., 1996.
McElreath, J., " An Architectural Perspective of Data Warehouse," Information Strategy : The Executive's Journal, Vol 12, No 4, 1996. 7. pp.30-41
Mundy, J., "From the Ground Up : Building a Data Warehouse", Sybase Server Magazine, 1995.