'98 International Conference Digital Library & Knowledge
September 22, 1998, Seoul Korea

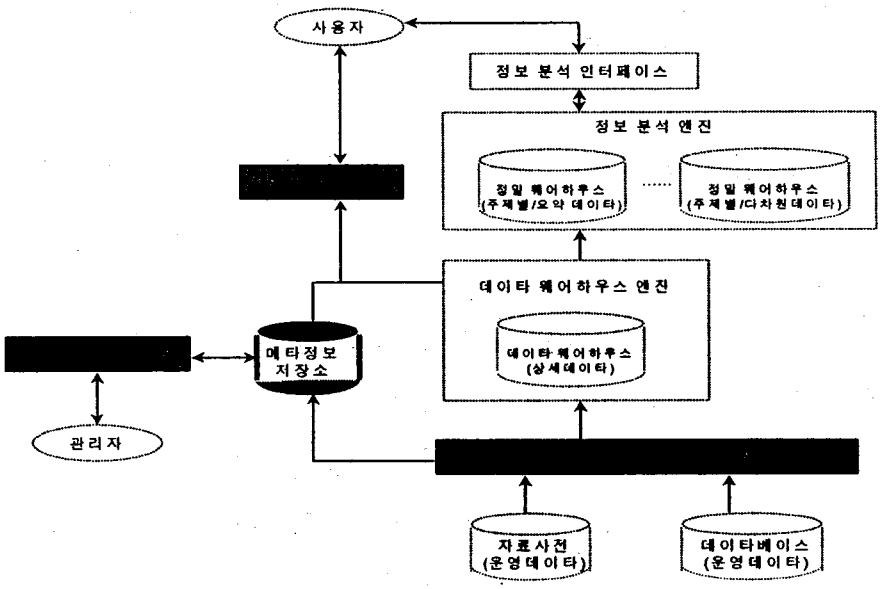
중대형 컴퓨터를 위한 데이터 웨어하우스 구축도구 개발*

이국철 (국민대학교 교수)
이춘열 (국민대학교 교수)
안연식 (국민대학교 대학원)
안효성 (국민대학교 대학원)
공세호 (싸인시스템)
김용희 (아남정보기술)

1. 서론

데이터 웨어하우스는 일상업무 처리를 위한 운영 데이터(Operational Data)의 관리를 목적으로 구현되는 전통적인 데이터베이스 시스템에 대응하는 새로운 개념으로서, 업무분석 및 의사결정을 위한 '정보 데이터(Informational Data)'를 효과적으로 제공하는 것을 목적으로 하고 있다. 최근 가트너그룹의 조사에 의하면, 데이터베이스 보유기업중 대부분이 향후 5년 이내에 데이터 웨어하우스를 구축할 것으로 전망하고 있어 국내외적으로 급격한 수요증가가 예상되고 있다. 데이터 웨어하우스에 대한 수요 증대와 더불어, 국내 연구 기관에서도 독자적인 데이터 웨어하우스 관리 시스템 또는 구현 도구들을 개발하기 시작하고 있으나, 주로 UNIX 시스템을 대상으로 하고 있다. 그러나, 현업에서 사용중인 대부분의 운영 데이터들은 AS/400과 같은 중대형 컴퓨터에 탑재된 것이 다수이며, 이들을 위한 데이터 웨어하우징 도구는 전량 외국업체로부터 수입에 의존하고 있다. 따라서 국산화된 중대형 컴퓨터를 위한 데이터 웨어하우스 구축 도구의 자체 개발은, 외국업체에 대한 의존도를 낮추어 수입대체 효과를 기대할 수 있다. 일반적으로 데이터 웨어하우징 시스템은 [그림 1]과 같이 총 6개의 세부 모듈로 구성되어 있다. 구체적으로, 운용 데이터들을 변환/처리하여 데이터 웨어하우스를 생성하는 「데이터 변환/처리 모듈」, 데이터 웨어하우스를 관리하는 「데이터 웨어하우스 엔진」, 정보 분석을 위한 데이터들을 웨어하우스로부터 추출, 조작관리하는 「정보 분석 엔진」, 사용자들이 분석 및 의사결정을 효과적으로 수행하도록 지원하는「정보 분석 인터페이스」, 데이터 웨어하우스에 저장된 데이터들을 설명하는 메타 데이터 정보 및 이들의 변환/처리에 해당하는 프로세스 정보를 저장 및 관리하는「메타 정보 저장소와 메타정보 관리도구」, 그리고 사용자들이 데이터 웨어하우스 자원을 정보 분석 엔진을 이용하지 않고 직접 검색하는「사용자 검색도구」로 구성되어 있다.

* 본 기술개발은 산업기술연구소의 공업기반기술개발사업으로 개발 중임.



[그림 1] 시스템의 구조도 및 개발범위

2. 국내외 기술개발 동향

가. 제품동향

데이터 웨어하우스 수요가 급증함에 따라, 데이터베이스 시스템 공급업체들은 경쟁적으로 데이터 웨어하우스 구현 도구들을 출시하고 있는 실정이다. 대표적인 국외 관련 우수업체들의 제품을 기능별로 열거하면 [표 1]과 같다.

[표 1] 외국 데이터 웨어하우스 제품

제품명	제작사	기능
Data Propagator	미국 IBM	◦ 데이터베이스 자동복제 및 데이터 자동복원 도구
EDA/SQL	미국 Information Builders	◦ 관계형 및 비관계형 원천 데이터 조작 도구
Data Guide	미국 IBM	◦ 메타 데이터 구축 도구
Amis	미국 Haskyns Group	◦ 데이터 변환 및 정제 도구 ◦ OLAP용 도구
MIT	스위스 SAMACS	◦ 다차원 데이터 모델링 도구 ◦ OLAP용 도구
Sales Tracker	미국 Silron Software	◦ 데이터 추출 도구 ◦ OLAP용 도구
ESS Base	미국 Show Case	◦ 데이터 분산/정화 ◦ 메타 데이터 관리 ◦ 원천 데이터 조작 도구 ◦ 다차원 데이터 모델링 도구
LANSA	호주 Aspects	◦ 원천 데이터 조작 도구 ◦ 다차원 데이터 모델링 도구 ◦ 메타 데이터 관리

[표 1]의 제품들의 [그림 1]에 나타난 데이터 웨어하우징 시스템의 구성 모듈별로 재정리하면 [표 2]와 같다.

[표 2] 데이터 웨어하우징 시스템의 구성 모듈별 주요 제품

구성 모듈	대표 제품
데이터 변환/처리 도구	DT propagator, InforRefinder, Platinum, LANSA/AD
데이터 웨어하우스 엔진	DB2, Teradata, Oracle, Informix
정보 분석 엔진	MetaCube Engine, IRI Express
정보 분석 인터페이스	MetaCube Explorer, IRI Explorer, LANSA/Client, Star Tracker
메타정보 저장소 및 관리도구	Data Guide
사용자 검색 도구	Star Tracker, Access, Excel

나. 업체별 동향

외국제품의 현황을 업체별로 분류하면, DBMS 공급업체 와 비DBMS 공급업체, 데이터웨어하우스 엔진 공급업체로 분류 할 수 있다.

먼저, DBMS 공급업체의 제품은 기존의 운영 데이터 관리를 위한 데이터베이스관리시스템 (DBMS : DataBase Management System)을 기반으로 웨어하우스 구현 도구들을 탑재한 형태를 취하고 있다.

- IBM : DB2, Propagator, Data Guide
- ORACLE : ORACLE, IRI Express, IRI Explorer
- Informix : Informix, MetaCube Express, MetaCube Explorer

비DBMS 공급업체들은 DBMS 공급업체들의 데이터베이스관리시스템에 탑재되거나 연계될 수 있는 도구들을 경쟁적으로 개발하고 있다. 이들 제품들은 다양한 시스템들에 탑재될 수 있다는 장점을 지닌다.

- Aspects 사의 LANSA
- SAMACS 사의 MIT
- Show CASE 사의 EssBase

데이터웨어하우스 엔진 공급업체는 기존의 데이터베이스관리시스템 제품들과 독립된 데이터 웨어하우스를 위한 제품이다.

- AT&T 사의 Teradata

국내 기술 및 제품현황은 외국 제품 및 업체들과 비교하여 국내 관련 기술의 현황은 아직 연구개발의 수준이라 할 수 있다. 학교 (서울대학교, 과학기술원 등) 및 국내 연구 기관에서도 데이터웨어하우스 구현을 위한 도구들을 시험적으로 개발하고 있으나, 주로 UNIX 시스템을 대상으로, 개방형 시스템을 위한 제품들이 대부분이다. 국내 업체들은 DBMS 제품 (예: 삼성전자의 CODA)의 경쟁력 취약으로 웨어하우스 구축 도구의 개발에 투자가 거의 없는 실정이다.

3. 본 연구개발의 내용

본 연구는 [그림 1]의 시스템 구조도 상에서 음영으로 나타난 3개의 영역을 개발범위로 하고 있다. 즉

① 사용자 검색 도구 개발

② 데이터 변환/처리 및 웨어하우스 생성도구 개발

③ 메타 정보 모델링 및 관리 도구 개발

이다. 위의 개발내용은 아래와 같은 하드웨어 및 소프트웨어 플랫폼을 기본 개발환경을 전제로 수행되고 있다. 이는 [그림 1]의 데이터 웨어하우스 시스템 구조도 상에서 점선부분으로 나타나 있어서 본 연구의 개발내용에는 포함되지 않는 부분이다.

■ 하드웨어 플랫폼 기본사양

• CPU 개수 (개)	1-4
• 메모리(MB)	512-4096
• 보조기억장치(GB)	500-20,000
• 통신회선(회선)	100-200
• 사용자수(명)	4,000 - 7,000

* 대표적인 기종으로 IBM AS/400 (530)

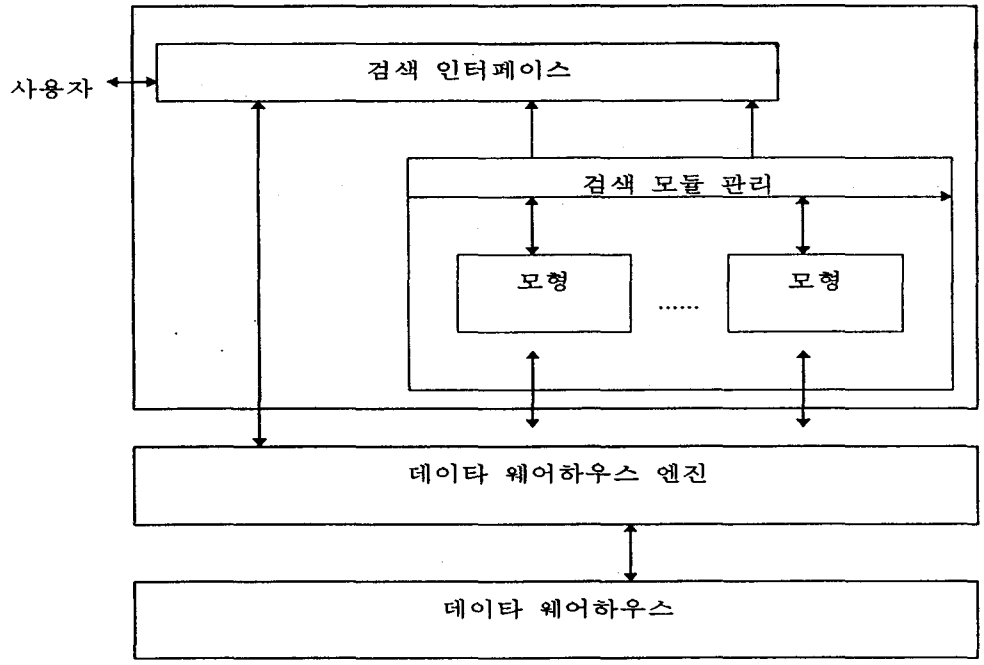
■ 소프트웨어 플랫폼 기본사양

• 데이터 웨어하우스 엔진	• 관계형 DBMS	• DB/2
• 정보분석 엔진	• 데이타베이스 시스템 • 스프레드 쉬트	• Essbase/400

가. 사용자 검색도구

1) 개요

사용자 검색 도구는 데이터 관리 및 조작 기능을 기반으로 하여, 질의 및 보고서 작성 기능 등을 포함한다. 사용자 검색 도구 개발부분은 검색 인터페이스 와 검색 질의 모듈 관리, 검색 질의 기능확장으로 구성된다. 사용자 검색도구의 구조는 [그림 2]과 같다.



[그림 2] 사용자 검색도구의 구조도

2) 상세기능

개발중인 사용자 검색도구 부분은 검색 인터페이스, 검색 질의 모듈관리, 검색질의 확장 기능으로 나누어지며 세부기능은 아래와 같다.

	세부 기능
검색 인터페이스	• 사용이 용이 • 멀티 윈도우에서 질의의 설정과 검색 결과가 동일 화면에 표시 • 질의의 설정과 검색 결과를 시각적으로 추적이 가능
검색 질의 모듈 관리	• 질의 모듈의 특성 정의 및 입력 • 질의 모듈의 버전 관리 • 복합 기준의 설정
검색 질의 기능 확장	(웨어하우스 엔진의 검색 질의 기능 확장) • 질의 설정 메크로 • 가상 필드 • 사전 정의 죠인

3) 개발결과의 예시

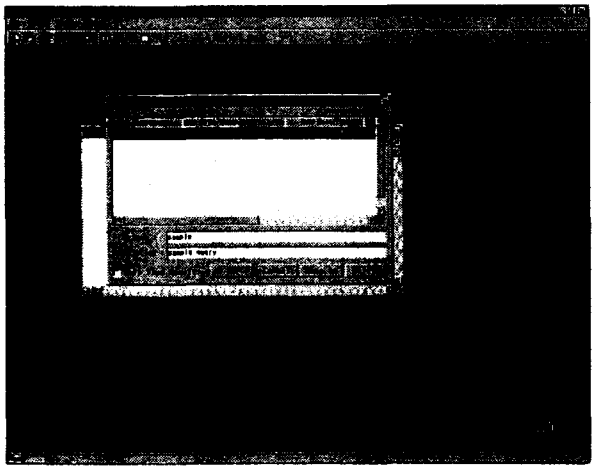
지금까지 개발완료된 사용자 검색도구(Query)의 개발결과를 화면구성을 중심으로 예시하면 아래와 같다.

① [그림 3] : 필요한 검색리스트 선택기능으로 사용자는 사전에 정의된 검색 정의를 선택하여 사용할 수 있고 새로운 정의를 만든다.

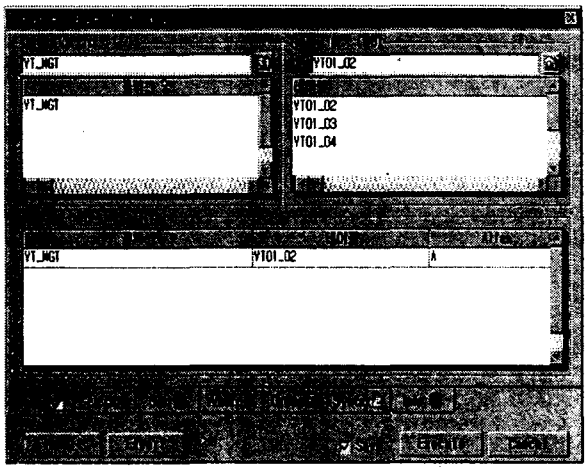
② [그림 4] : 사용자 검색 기본화면이 나타나고 Collection 과 Table중 검색하고자 하는 Collection 과 Table을 선택한다.

③ [그림 5] : 검색하고자하는 항목을 선택하고 선택된 항목에 대하여 새로운 Expression을 설정한다.

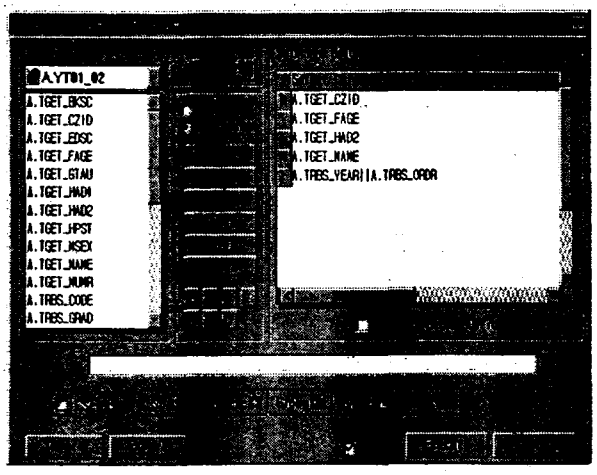
④ [그림 6] : 검색조건을 입력하는 화면으로 검색조건의 Value 항목에 Column과 Function을 지정한다.



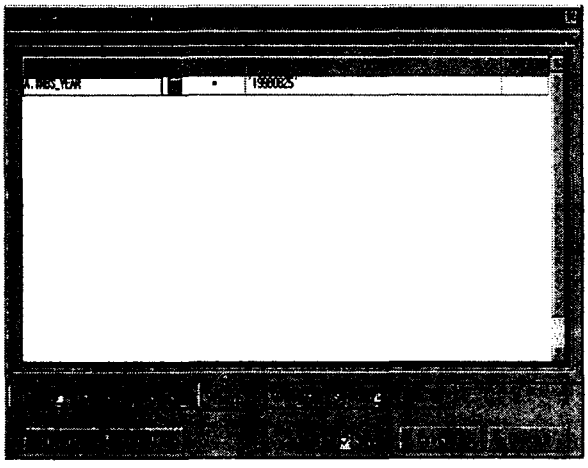
[그림 3]



[그림 4]



[그림 5]

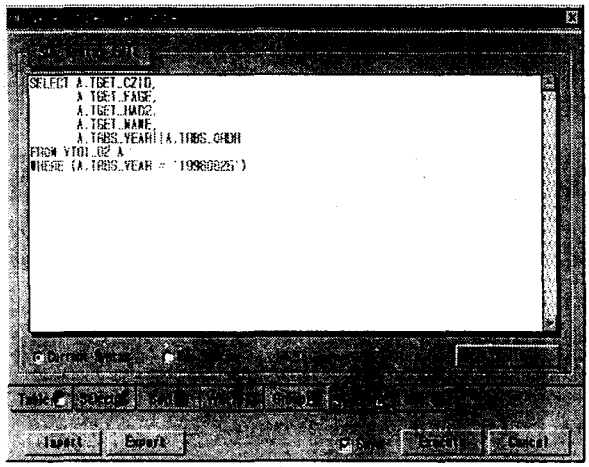


[그림 6]

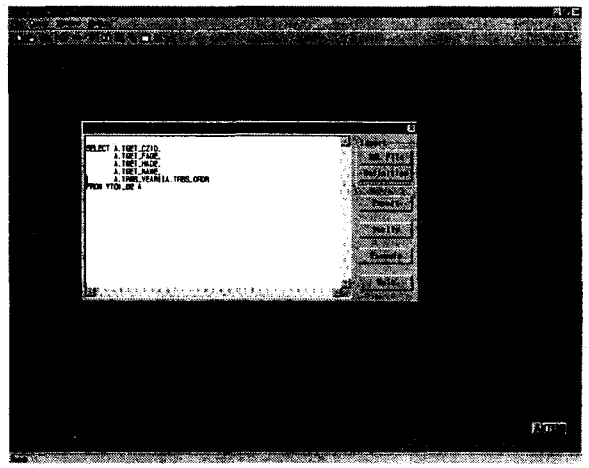
⑤ [그림 7] : 작성된 검색문을 편집할 수 있고 검색문이 유효한지 검증한다.

⑥ [그림 8] : 사용자 검색 기본화면([그림 4] ~ [그림 7]까지)을 사용하지 않고 검색문을 직접 입력하여 실행한다.

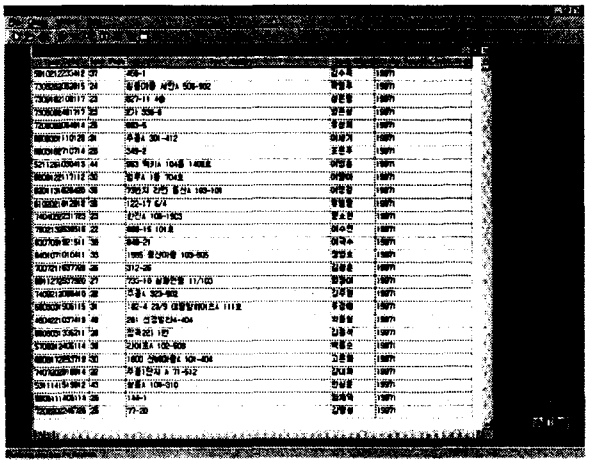
⑦ [그림 9] : 실행된 결과를 화면에 출력하여 보여준다[그림 9].출력된 결과는 Graph Interface [그림 10]를 이용하여 다양한 종류의 분석및 검색보고서를 작성한다. 또한 Query 의 Result set은 MS Excel이나 dBASE3, HTML, Lotus 1-2-3로 변환하여 사용자들이 다양한 보고서를 작성할 수 있다.



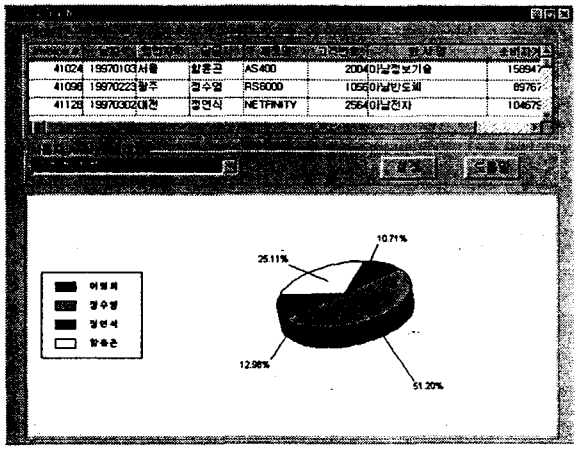
[그림 7]



[그림 8]



[그림 9]



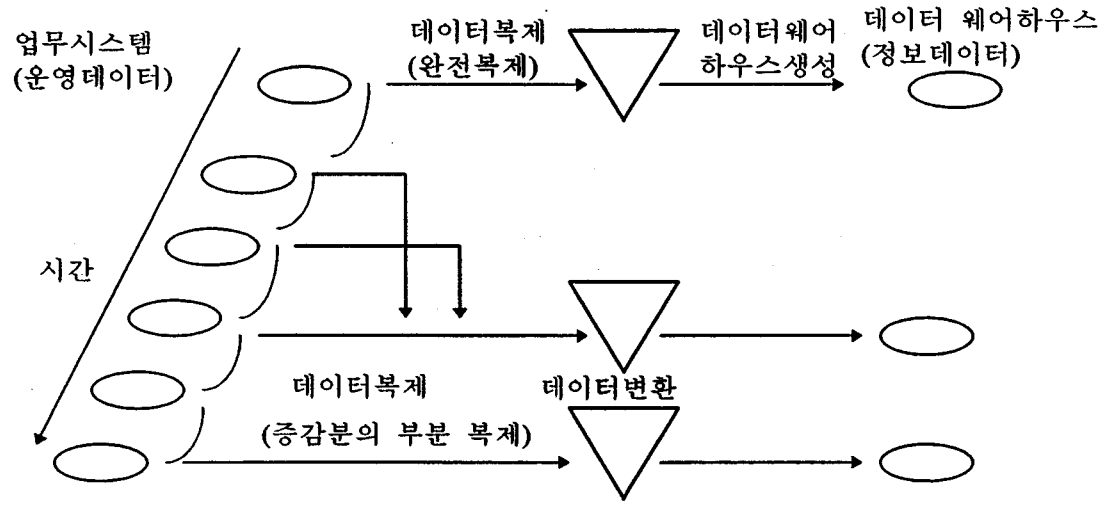
[그림 10]

나. 데이터 변환/처리 및 웨어하우스 생성도구 개발

1) 개요

데이터 변환/처리 및 웨어하우스 생성도구 개발에는 운영 데이터들을 복제하여야 하며, 복제된 자료를 변환하여야 한다. 또한 복제된 데이터들을 정보 분석에 효과적으로 활용하기 위하여서는 변환 작업이 필요하고, 이를 통하여 정보 분석을 위하여 필요한 자료들이 추가로 추출되거나 가공된다. 복제된 자료들이 서로 일치하지 않거나 오류를 포함할 경우에는 이들 자료들을 정제(cleansing)해야 한다.

데이터 변환/처리 및 웨어하우스 생성도구 개발부분은 크게 데이터 복제, 데이터 변환처리 및 복제, 메타정보 갱신으로 구성된다. 데이터 웨어하우스 생성사이클은 [그림 11]과 같다.



[그림 11] 데이터 웨어하우스 생성사이클

2) 상세기능

개발중인 데이터 변환도구 기능에는 크게 데이터 변환기능, 데이터 복제기능, 메타정보 갱신기능, 스케줄링으로 나누어져 있으며 각각의 기능을 설명하면 다음과 같다.

① 데이터 변환 기능

- 변환시스템의 구조는 운영계 시스템 서버에서 웨어하우스 서버로 끌어내는 풀(pull) 방식으로 단일 웨어하우스에 복수의 데이터마트가 있다.
- 데이터를 변환하는 기능은 크게 차원 테이블의 구간을 설정하는 '합계(aggregation)' 기능과 질의 검색을 하는 '선별(filtering)', 테이블 사이의 자료값의 일관성을 비교하고 수정하는 '자료값의 표준화(standardization)', 그리고 자료값의 무결성을 검사하고 오류를 식별하는 오류(cleansing)기능으로 분류된다.

◦ 합계(aggregation)는 차원에 대한 코드를 부여하는 기능과 코드에 따른 합계(질의 처리)를 처리하는 기능이다.

◦ 선별(filtering)은 질의를 검색하여 테이블의 행과 열을 선택하거나 테이블 간의 죠인으로 데이터를 여과하는 기능이다.

◦ 자료값의 표준화(standardization)는 예를 들어 코드, 문자열, 소문자, 대문자, 숫자, 날짜, 계량단위 등에 대한 자료값의 일관성을 비교하거나 수정하는 기능이다.

◦ 오류(cleansing)는 자료값의 무결성을 검사하고 오류를 식별하는 기능이다.

② 데이터 복제기능

◦ 데이터 복제의 기능은 데이터 복사 간격 지정, 변동 사항 복사, 데이터 갱신, 변동 기록 , 복제 작업 정의 인터페이스기능이다.

③ 메타 정보 갱신기능

◦ 메타 정보 갱신기능에는 메타 정보 모델링에서 정의된 자료 추출과 메타 정보의 입력 및 갱신기능이다.

④ 변환 스케쥴링

◦ 변환주기로는 실시간 변환과 주기적(배치) 변환이 있으며, 갱신방법으로는 증감 갱신(net change basis)와 완전복제가 있다. 그리고 증감 갱신의 방법으로는 로그 데이터를 활용하는 방법과 트리거(trigger)를 활용하는 두가지 방법이 있다.

3) 개발결과의 예시

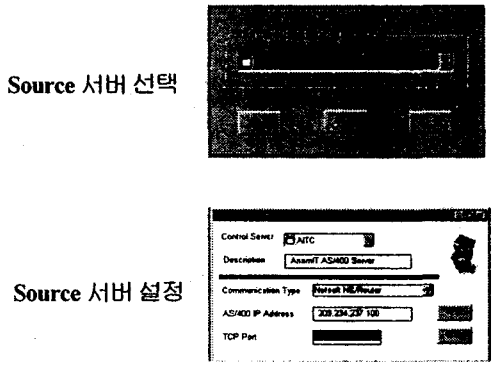
현재 데이터 변환/처리 및 웨어하우스 생성도구(Replica)의 개발결과를 화면구성을 중심으로 예시하면 아래와 같다.

① [그림 12] : 원시 운영계 서버를 선택하는 화면이다.

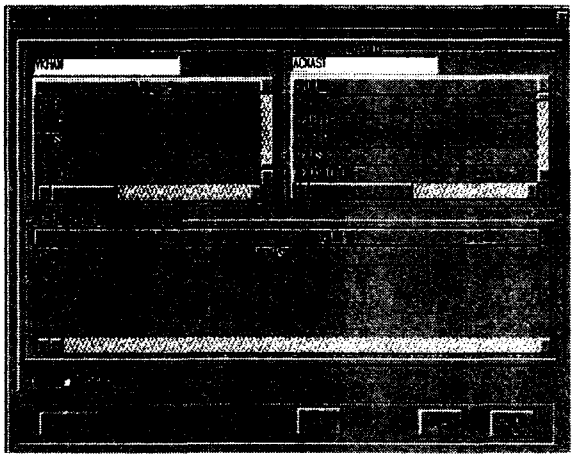
② [그림 13] : 복제하는 원시 운영계 서버의 Collection 과 Table을 선택한다.

③ [그림 14] : 복제하는 원시 운영계 서버의 Collection 과 Table에서의 복제할 항목을 선택한다.

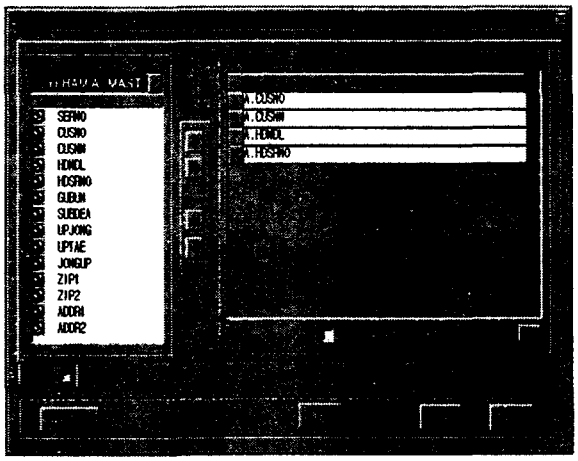
④ [그림 15] : 데이터 웨어하우스 서버를 선택하는 화면이다.



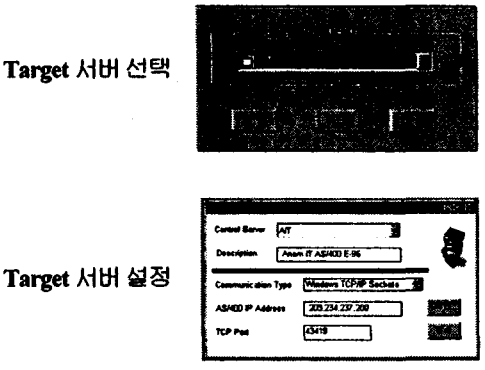
[그림 12]



[그림 13]



[그림 14]



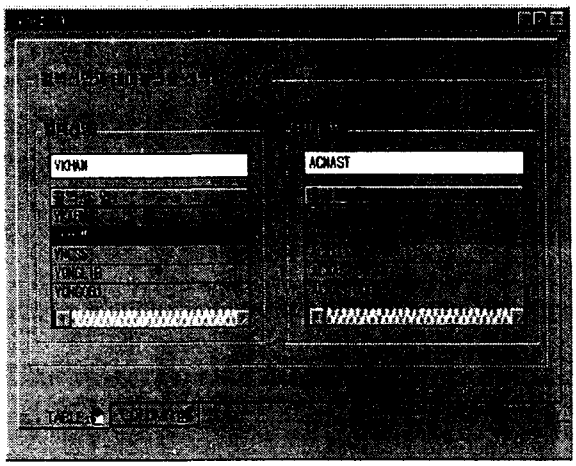
[그림 15]

⑤ [그림 16] : 데이터 웨어하우스 서버의 Collection 과 Table을 선택한다.

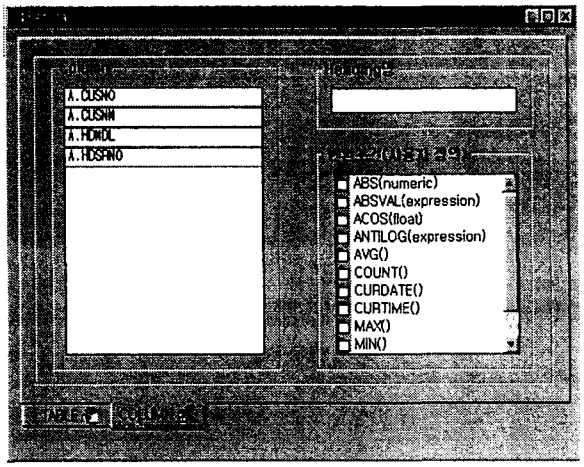
⑥ [그림 17] : Column의 명과 변환조건 및 Column의 설명을 입력할 수 있다.

⑦ [그림 18] : Dimension을 선택하고 실행할 수 있다.

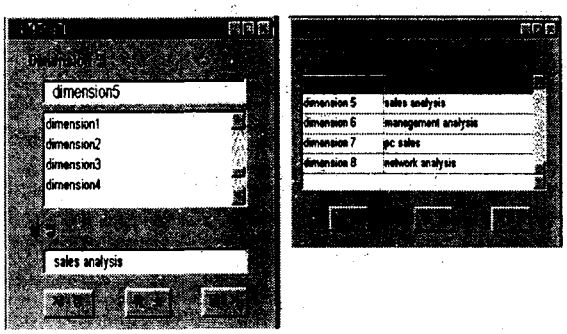
⑧ [그림 19] : 선택된 Dimension을 샐행한 후의 결과이다.



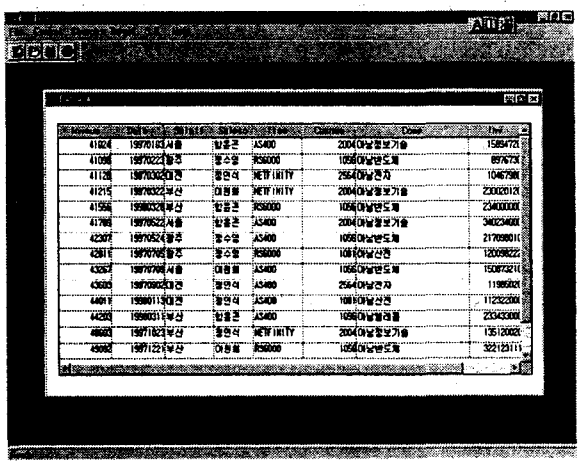
[그림 16]



[그림 17]



[그림 18]

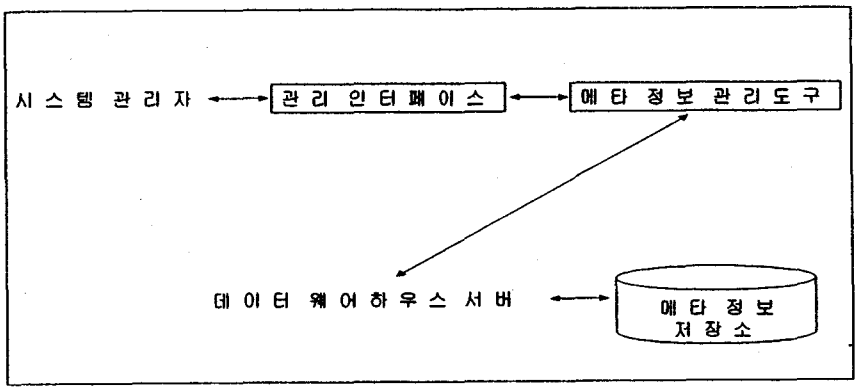


[그림 19]

다. 메타 정보 모델링 및 관리 도구 개발

1) 개요

메타 정보 저장소는 데이터 웨어하우스에 저장된 데이터에 대한 메타 정보를 포함한다. 메타 정보로는 데이터의 명칭, 길이, 의미, 포함된 파일이나 테이블 등 데이터를 이해하기 위하여 필요한 모든 항목들인 메타 데이터 정보와 데이터가 변환되고 처리된 과정을 설명하는 메타 프로세스 정보를 포함한다. 메타 정보 모델링 및 관리 도구 개발부분은 메타 정보 저장소 구축, 메타 정보 관리 모듈, 관리 인터페이스로 구성된다. 메타 정보 저장소 및 관리모듈의 구조도는 [그림 20]과 같다.



[그림 20] 메타 정보 저장소 및 관리 모듈의 구조도

2) 상세기능

개발중인 메타 정보 모델링 및 관리 도구의 기능에는 메타 정보 저장소 구축, 메타정보 관리 모듈, 관리 인터페이스 기능으로 나누어지며 세부기능은 아래와 같다. 현재 각 상세 기능들이 개발중인 관계로 인하여, 결과물 예시는 포함되어 있지 않다.

메타 정보 저장소 구축	- 정보 카다로그 생성. - 사용자 샘플정보 셋업. - 관리자와 사용자의 업데이트 권한 지정. - 정보 자원의 그룹화 - 사용자 정의 명칭으로의 변환
메타 정보 관리 모듈	- 정보 자원 탐색 - 데이터의 의미 - 데이터의 생성 및 전달 경로 - 정보 자원 현황 관리 - 데이터 생성의 효율성 - 데이터 품질 - 데이터 영향 분석
관리 인터페이스	- 사용의 용이성 - 추적 가능성

라. 개발사양 및 진행결과 요약

1) 개발사양

위의 3가지 개발내용에 대한 구체적인 개발사양을 개발언어, 개발도구 및 적용알고리즘으로 요약해보면 아래와 같다.

개발대상 모듈	주요기능	개발언어	개발도구 (자동화 도구)	적용 알고리즘 /개발모형
사용자 검색도구	데이터 검색기능	PowerBuilder	System Architect	• 질의 처리 (최적화) 알고리즘
데이터 변환/처리 및 웨어하우스 생성도구	데이터 변환 및 복제기능	C++, JAVA, PowerBuilder, C, RPG	System Architect	• 무결성 검증 알고리즘 • 변환 알고리즘
메타 정보 관리도구	정보자원 현황 관리기능	PowerBuilder	ER-Win, S-Designer	• 데이터 제조공정 모형

2) 진행결과 요약

또한 각 모듈별로 현재까지 개발된 내용과 개발중인 내용을 요약해보면 아래와 과 같다.

개발대상 모듈	개발완료	개발중
사용자 검색도구	검색 인터페이스 검색 질의 모듈 관리 검색 질의 기능 확장	질의 분석기능
데이터 변환/처리 및 웨어하우스 생성도구	데이터 복제기능	데이터 변환 기능 메타 정보 갱신기능 스케쥴링
메타 정보 관리도구	미 완료	메타 정보 저장소 구축 메타 정보 관리 모듈 관리 인터페이스

4. 결론

본 연구개발은 중대형 컴퓨터를 위한 데이터 웨어하우스 구축도구개발을 목적으로 사용자 검색도구, 데이터 변환/처리 및 웨어하우스 생성도구, 메타 정보 관리도구부분으로 나누어 개발이 진행중이다.

현재 우선적으로 사용자 검색도구의 개발을 선행하여 추진한 결과 거의 완료단계에 있다. 특히 사용자 검색도구는 중대형 컴퓨터를 위한 데이터 웨어하우스 구축도구와는 상관없이 활용이 가능하도록 설계되어 독립적인 프로그램으로도 사용이 가능하게 구현 되어있다.

한편 하드웨어, 소프트웨어 플랫폼등 여러 개발환경 정보에 대한 인프라 구축 후 메타 데이터 모델링과 데이터 변환/처리를 위한 연구, 적용 및 운용을 위한 프로그램 시험중에 있으며 1998년 12월 말까지 기술개발과제를 완료할 예정이다.

본 연구결과인 중대형 컴퓨터를 위한 데이터 웨어하우스 구축도구가 개발될 경우, 데이터 웨어하우스 구현 도구에 대한 국내 기술력의 확보 및 기술 자립도를 달성할 수 있게 된다. 또한 국내 사용환경에 맞는 데이터 웨어하우스 구현 도구의 개발을 통해, 장기적으로 한국형 데이터 웨어하우징 시스템 구축이 가능해지고, 경쟁력을 갖춘 국산 데이터 웨어하우스 구현 도구를 개발함으로써, 외국 제품에 대한 의존성 탈피 및 외화 절감효과를 유도하고, 연관 응용시스템 구현 도구개발을 촉진할 수 있다고 판단된다..

참 고 문 헌

IBM, AS/400 DB2/400 SQL Programing, version 3, 1994,9

IBM, Data Warehousing Concepts for AS/400, 1995

Inmon, W, H., Building the Data Warehouse. Second Edition, John Wiley & Sons, Inc., 1996

Inmon, W, H., R. D. Hackathorn , Using the Data Warehouse, John Wiley & Sons, Inc., 1996

Sheryl Larsen, 웨어하우스 정밀분석, DB2, 1996

데이터 웨어하우스 구축, 월간 랜 타임즈, 1996,6

데이터 웨어하우징을 위한 메타데이터의 구조, 데이터베이스월드, 1998,7

메타데이터 관리, 데이터베이스월드, 1998,6

엔터프라이즈 데이터 관리 방법과 솔루션, 월간 랜 타임즈, 1998,7

이상구, Data warehouse 개요, 한국정보처리전문가협회 전산관리자 세미나

이영재, 조용찬, 은행 정보계시스템의 데이터웨어하우스 구현 방향, KIMS 97춘계 학술대회

이희석,김태훈, 데이터 웨어하우스와 의사결정시스템, 컴퓨터월드, 1997.6

장길동, 시퀀트 데이터 웨어하우스, ORACLE MAGAZINE, 1997

장동인, 실무자를 위한 데이터 웨어하우스, 한국오라클

조재희, 박성진, 데이터 웨어하우징과 OLAP, 대청 컴퓨터월드, 1996