

목 차

1. EDW와 RTE
2. EDW와 EAI의 결합
3. EAI의 특징
4. ETL의 특징
5. 각 솔루션 업체의 EDW 전략
6. EDW의 미래

1. EDW와 RTE

근래의 ETL, 나아가서 EDW의 가장 큰 화두 중의 하나로 Real Time(즉시성)을 얘기할 수 있다. 이전의 Real Time을 이야기 할 때에는 운영계에서 ODS나 DW로의 추출부분의 즉시성을 이야기 했지만 근래에 들어와서는 RTE(Real-Time Enterprise)란 개념과 맞물려 좀더 넓은 의미로서 사용되고 있다.

일반적으로 RTE란 “기업 내/외부를 포함하는 전체적인 관점에서 지속적인 프로세스 개선과 정보를 실시간 전달함으로써 업무 지연을 최소화하고 의사결정 스피드를 높여 경쟁력을 극대화한 기업”이라고 정의되는데, 경영 환경의 변화, 미래의 불확실성, 고객의 요구, 신규 기회를 신속히 파악하고 포착하여 민첩하게 대응을 하는 것을 목표로 한다. 결국 RTE에서 핵심이 되는 부분은 프로세스 개선 및 정보의 실시간 분석 및 전달이라 볼 수 있다. 이러한 관점에서 볼 때 RTE는 구현을 위한 기술 중심의 IT가 아니고 기업의 경영 전략/특성 및 비즈니스 프로세스를 통한 활용의 문제라고 보아

야 할 것이다.

정의에서 보듯이 RTE를 위한 요소 기술들은 업종이나 기업의 특성, 목적에 따라 아주 다양할 수 있다. RTE에서 가장 중요한 IT 관점의 Keyword는 real time과 integration이지만 비즈니스와 시스템, 기술 등의 관점을 고려할 때 RTE를 위해 핵심 요소는 다음의 몇 가지로 볼 수 있다.

- Analytic Intelligence: EDW 및 연관 기술 중심의 활용 분야
- Business Intelligence: CPM, SRM, CRM 등 기업의 BI 솔루션 중심의 활용 분야
- Operational Intelligence: BPM 중심의 활용 분야
- Real Time & Automation & Monitoring: 신속성, 민첩함을 지원하기 위한 기술 요소 (EAI, BPM)
- Integration: EAI, EDW를 통한 데이터, 시스템 통합 기술 요소

여기서 Analytic Intelligence, Business Intelligence, Operational Intelligence는 활용 분야로서의 영역을 나타내며 Real Time & Automation, Integration, Monitoring은 각각을 위

* 상명대학교 컴퓨터소프트웨어공학과 교수
 ** SAS코리아 DW/CRM팀 차장

한 기반 기술 요소로 볼 수 있다.

Analytic Intelligence는 기업의 실시간 정보의 분석 및 공유를 통한 활용으로서 기반기술은 EDW다. A.I.를 위해서는 데이터, 정보의 통합 및 표준화가 필요하며 실시간 정보를 지원하기 위해 EAI와의 연계도 필요하다.

Business Intelligence는 SCM, CPM과 같은 기업의 비즈니스 솔루션을 통한 비즈니스 활용 분야로서 RTE의 주 적용 분야가 될 수 있다. 각 기업은 RTE를 모든 업무 프로세스에 동시에 적용하기 보다는 각 기업의 핵심 적용 분야를 찾아야 한다. 영역에 따라 Customer Intelligence, Suppliers Intelligence, Organizational Intelligence로 구분해 볼 수 있다. 은행의 실시간 위험관리 기능이나 GE의 CPM 중심의 경영 성과 관리(Organizational Intelligence), 델 컴퓨터의 SCM 중심으로 한 주문에 대한 즉각적인 생산을 지원하는 실시간 정보 공유 (Suppliers Intelligence), 유통이나 패션업의 고객의 요구에 대한 실시간 응대 및 요구 사항 반영 (Customer Intelligence)를 예로 들 수 있다.

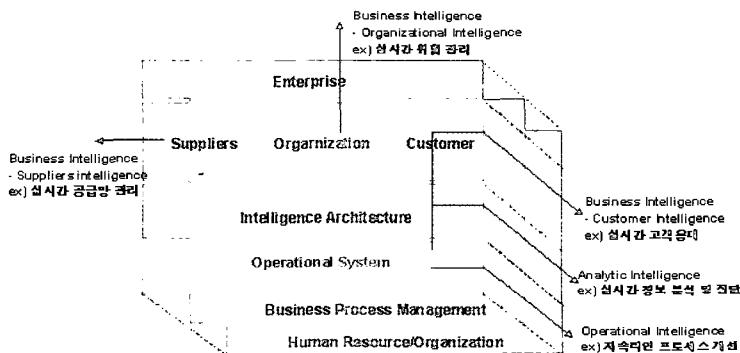
Operational Intelligence는 기업의 운영 업무의 자동화를 통한 기업의 프로세스 개선 및 경영의 신속화를 이루는 것으로 기반이 되는 기술은 BPM (Business Process Management)이다. BPM은 Operational Intelligence뿐만 아니라 Analytic Business Intelligence에서도 핵심 기술이다. 24시간

결산할 수 있는 시스템을 갖춘 시스코 시스템즈를 예로 들 수 있다.

2. EDW와 EAI의 결합

이전 회에서 CDC 방식에 대해 잠시 설명한 바와 같이 현재 출시되어 있거나 알려져 있는 ETL의 방법 가운데 완벽한 Real Time을 구현해 주는 경우는 현재까지는 없다고 보여진다. 물론 Snapshot을 사용하는 방법이나 Application을 수정하는 방법을 사용하고 있는 곳도 있지만 이런 방법들은 DB에 많은 부하를 주거나 전면적인 시스템 재설계를 요구하기 때문에 실효성이 거의 없다고 볼 수 있다.

또한 이전에서 언급한 DB의 로그 캡취 기능을 사용한다고 해도, Near Real Time처리라고 봐야 하고 DB의 고유기능에 의존해야 하기 때문에 DB의 Upgrade시 추출 모듈의 부분적 재설계가 이루어질 수 있다. 이러한 문제점으로 인해 비용의 발생을 감수하고서라도 EAI 솔루션을 도입을 검토하는 업체가 늘고 있으나 ETL 솔루션과 EAI 솔루션의 유사성 때문에 중복투자의 위험성이 제기되는 경우가 발생한다. 또한 이에 맞추어 EAI 업체 측에서도 ETL 툴에 비해 상대적인 약점인 대용량의 데이터 처리와 프로세스 흐름관리 측면 등의 기능을 보완한 제품을 시장에 내놓고 있다. 그러나 현재의 소프트웨어의 기술면에서 볼 때 두 가지의 솔루션은 분명한 차이를 가지고 있으며 이러한 부

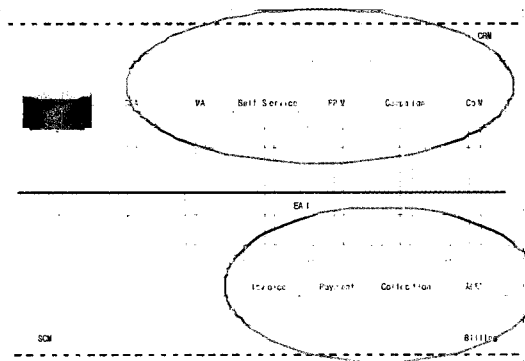


(그림 1) RTE

분들을 극복한 제품이 나오기엔 아직은 시간이 필요해 보인다. 각 솔루션간의 차이를 살펴보면 다음과 같다.

3. EAI의 특징

- Transaction 위주로 데이터를 처리함.
- 일반적으로 CRM, ERP 등 Front-Office단의 Application들의 Data 동기화, 어플리케이션의 통합(EIP)에 사용하고 이들 데이터의 동기화 방식은 Application의 API(Application Programming Interface) 단에서 이루어진다. 즉, DW, ABM, BSC, CRM 등의 근원 데이터를 생성해내는 어플리케이션이 아닌 서로 다른 어플리케이션 간의 결과값의 패싱 등에 사용함.
- Transaction 레벨에서 이루어지는 적은 용량의 Near Real time성 데이터의 처리를 주 기능으로 함.
- 단순한 데이터의 이전 처리에 적합함.



(그림 2) EAI Architecture

4. ETL의 특징

- 대용량의 데이터를 처리하는 병렬처리, Batch Load, Scheduling 기능을 주기능으로 함.
- 서로 다른 소스(2개 이상의 테이블, 서로 다른 이 기종, 다른 Database에 존재하는 테이블들)

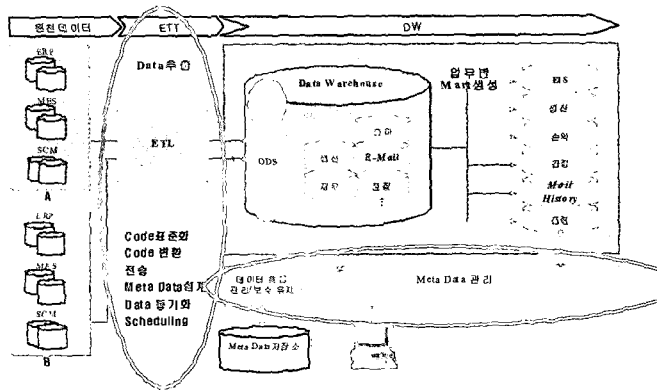
의 데이터를 Merge(통합)해 한 개의 Target 으로 생성해내는 기능을 제공

- 추출에 관련된 일련의 프로세스의 메타 데이터의 관리 기능을 제공하고, CWM(Common Warehouse Metadata)표준을 따르는 도구들의 메타 데이터를 통합 관리할 수 있는 기능을 제공.
- Application의 API를 사용하고 Meta Data를 Handling하여 데이터를 처리하는 기능을 제공 하며, 또한 관계형 데이터 베이스의 로우 레벨 단의 추출을 제공.
- 복잡한 소스와 타겟의 매핑 기술을 제공.

EAI 도구는 어플리케이션간의 데이터 통합 도구로 90년대 중반에 생겨났으며 Data Warehouse 에서 출발한 ETL 도구와는 출발 배경에 차이를 두고 있다. 이로 인해 ETL 도구는 표준화 규약 - CWM(Common Warehouse Metadata)에 기반 한 통합된 메타 데이터 관리를 제공하는 반면 EAI도구는 Tool에 한정된 메타 데이터의 관리를 제공한다. 물론 각각의 제품의 약점 보완과 포괄하는 업무영역의 확장이라는 개념에서 보면 ETL의 일부 기능을 제공하는 EAI 도구가 있을 수 있고, ETL 도구도 EAI의 일부 기능(Near Real Time Integration)같은 기능들을 제공하려는 추세로 가고 있으나 시장의 인지도나 점유율로 볼 때 아직은 서로의 영역을 대체하는 수준까지는 발전하지 못한 것으로 생각된다. 이러한 위의 여러 이유들로 인해 EDW 내에서 Application단의 데이터 전송과 일부 운영 데이터의 즉시성을 보장하는 최선의 수단으로 EAI가 고려되고 실제로 많은 기업에서 채택, 사용되고 있다.

5. 각 솔루션 업체의 EDW 전략

각 EDW솔루션 업체의 Real Time Processing 에 대한 전략 및 견해는 다음과 같다.



(그림 3) ETL Architecture의 예

5.1 Sybase - Near Real Time

- 데이터 웨어하우스 상의 EAI의 역할을 Application통합 보다는 데이터 전달자의 역할로 봄.
- 비동기 방식의 데이터 전달.
- 실시간 보다는 보다 짧은 주기의 데이터 공급을 목표로 함.
- ETL로의 EAI역할에 관해서는 아직은 ETL을 대체 하기에는 역부족으로 봄.
- EAI는 소량의 데이터를 적기에 DW에 공급함.
- 2PC와 같은 Real Time은 상호 증속되지 않도록 구현하지 않는 것이 바람직함.
- 보다 짧은 주기의 ETL로 해석하는 것이 바람직함.

5.2 Oracle - Oracle Streams(Near Real Time)

- 온라인 로그 파일을 사용하여 변경 분 추출
- DB Trigger를 사용하지 않음.
- 새로운 DB 기능인 오라클 스트림을 사용하여 Redo log파일을 역세스, 변경분을 추출해 냄.
- 일정한 주기에 따라 반복적으로 수행.
- 소스 시스템에 오버헤드를 주지 않음.

5.3 한국 IBM - (Real Time Bi)

- 데이터를 시계열로 구분하여 추출 및 저장 방

식을 달리 가져감.(3년전 데이터-Tape / 3년 전 ~ 전일 ETL Solution/당일 데이터 EAI or EII Solution).

- MQ Series queue를 사용해 지속적인 추출 수행.
- Real Time ETL은 Parallelism 과 Non-stop loading을 수행하여야 함.
- Real Time의 구축은 비싼 비용을 요함.

5.4 한국 NCR - Active Data Warehouse(Near Real Time)

- 즉시적인 Update - real time에 가까움.
- 수초 내에 결과를 얻을 수 있는 짧고 전술적인 쿼리가 많음.
- Event-Driven Activity
- EAI 아키텍처를 사용해 Near Real Time 구현.
- 대량 데이터의 적재는 제3의 ETL Tool 활용.
- Teradata 유틸리티를 활용하여 거의 실시간의 데이터 획득 및 데이터 이행을 위한 패키지 형태의 ADW 어댑터 제공.

(참고 자료 - 2003/9/30 DW의 혁신 - 대용량 및 실시간 EDW 구축방법 세미나)

6. EDW의 미래

기존의 EDW와 비교해 향후 EDW 플랫폼이 가

져야 할 기능 혹은 보완되고 있는 기능들을 보면

- 조회 정보의 실시간 처리.
- 전사 RTE의 정보 제공 기반 제공.
- 기존 SQL에 비해 진보적인 분석 Function 및 언어 제공.
- 플랫폼 내 각 솔루션들의 메타 데이터 통합.
- 시스템 및 보안, 각 솔루션의 통합 관리.

등 여러 가지가 있다.

위에서 언급한 내용을 제외하고 이야기를 해 보면

- 기존 SQL에 비해 진보적인 분석 Function 및 언어 제공 : 이전의 RDBMS와 달리 근래의 DB엔진을 보면 기존 관계형 DB에서 제공하지 못했던 MDB의 기능들과, 여러 분석적인 쿼리 및 함수들을 제공하고 있다. 기존의 MDBMS를 사용해야지만 가능했던 신속한 응답이나, 시계열을 핸들링하는 여러 가지 함수들을 폭 넓게 지원하고 있을 뿐만 아니라, 문장 자체 내에서의 새로운 분석적 쿼리문 까지도 지원하고 있다.
- 플랫폼 내에서 각 솔루션들의 메타 데이터 통합 : 일반적으로 국내 기업이 선호하고 있는 Best of Breed 방식(각각의 솔루션중 최고의 제품을 선정해 DW 플랫폼을 구축하는것)의 시스템이 가지고 있는 취약점 중의 하나가 서로의

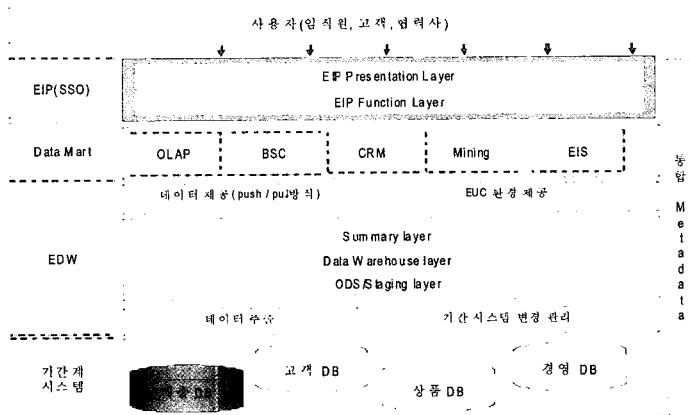
메타 데이터가 통합되지 못해 각자의 정보를 공유할 수 없다는 것이다. 이를 보완하기 위해 한 곳의 토털 솔루션을 사용할 경우 이 또한 한곳의 소프트웨어 제공업체에 대한 의존도가 증가하는 단점이 있다.

이러한 점들을 개선하기 위해 산업에 대한 메타 데이터 표준이 여러 단체에 의해 정의되고 있으며, 점차 이를 따르는 소프트웨어들이 늘고있다.

현재까지는 업체의 이해관계가 상이해 완전히 산업표준에 의해 통합되어 있다고 보기는 힘들지만 가까운 시일에 통합될 수 있으리라 전망된다.

- 시스템 보안 및 각 솔루션의 통합 관리 : 위의 메타 데이터의 통합과 맞물려 사용자 관리, 시스템 성능관리, 비즈니스 메타 관리, 잡 스케줄링 등의 각각의 관리를 담당하는 소프트웨어 들 또한 통합 되어야 하나 이 부분은 기술적인 문제들로 인해 위의 메타 데이터 통합이 완벽히 이루어진 후이나 가능할 것으로 보인다.

위에서 열거한 것들 외에도 각각의 산업별로 보다 구체적이고, 선진 기법을 가미한 BI 솔루션들이 업체별로 나오고 있으며, 기존의 DW 인프라 위에 CRM, ERP 등의 백 오피스, EIP나 캠페인과 같은 프론트 오피스 등이 유기적으로 연결된 플랫폼을 구축하려는 노력이 계속되고 있다.



(그림 4) 연계 구성도

이전에도 그랬으며 앞으로도 EDW는 한때 반짝하고 사라지는 기술이 아닌 새로운 기술의 발전에 따라 모든 BI용 솔루션의 정보 제공자로서 끊임없이

변화되고, 진화되어 기업에서 정보분석의 요구가 있는 한 사라지지 않는 기술로서 자리잡을 것이다.

저자약력



김수홍

1992년 - 현재 상명대학교 컴퓨터정보통신공학부 교수
1998년 - 현재 한국정보처리학회 편집위원
2002년 - 현재 한국정보처리학회 e-Business 연구회
위원장
관심분야 : 정보검색, e-learning, 병렬처리시스템
이 메 일 : soohkim@smu.ac.kr



장윤석

1993년 2월 경희대학교 전산공학과 졸업
1997년~2001년 오라클 코리아 DW Team 컨설턴트
2001년~2003년 삼일 회계법인 CRM 컨설턴트
2003년 - 현재 SAS 코리아 PSD DW/CRM팀 컨설턴트
관심분야 : Business Intelligence / Data Warehouse
모델링 및 구축.
이 메 일 : Yun-Seok.Jang@sas.com)