

실시간 데이터 접근을 위한 ODS 레이어 아키텍처

엄충용*, 정현석**

*숭실대학교 컴퓨터학과

**한국컴퓨터주식회사

e-mail:umcy@selab.ssu.ac.kr, yohaim@hanmail.net

An ODS Layered Architecture for Real-time Data Access

Chung-yong Eom*, Hyun-suk Jung**

*Dept of Computer Science, Soong-sil University

**Korea Computers Inc.

요 약

본 논문에서는 운영 레거시 시스템으로부터 엔드 유저에게까지 실시간으로 데이터를 전달하는 전체적인 프로세스를 정의한 ODS 레이어를 제시한다. ODS의 레이어를 위하여 먼저 데이터 소스가 정의된다. 그 다음 데이터 소스로부터 ODS까지 데이터를 수집하는 방법이 정의된다. 마지막으로 엔드 유저 집단을 위한 ODS로의 데이터 접근을 준비한다. 운영 시스템과 ODS 간의 통합 레벨에 따른 ODS의 타입을 분류하고 특성을 요약한다.

1. 서론

최근 기업에서는 DW와 함께 실시간 데이터 활용을 위해서 ODS를 구축하고 있다. ODS는 DW의 유연성을 확보하고 데이터 정제를 지원하며 실시간 운영 데이터를 획득하고 활용을 가능하게 한다. ODS는 DW의 단점을 보완해 주며 효율적인 데이터 저장소로서의 역할을 한다[1].

이러한 ODS는 현재 DW/DM로 구축되는 의사결정시스템을 위한 초기 데이터저장소로써만 활용되고 있으며 이론적 근간으로 제시될 만한 아키텍처가 존재하지 않는다. 기업의 레거시 운영 시스템으로부터 실시간으로 데이터를 수집하기 위한 체계적이고 계층화된 접근 방법이 필요하다. 또한 운영 데이터의 특성과 성격에 따라 접근을 다르게 할 수 있어야 한다.

본 논문에서는 레거시 시스템으로부터 엔드 유저에게까지 실시간으로 데이터를 전달하는 전체적인 프로세스를 정의한 ODS 레이어 아키텍처를 제시한다. ODS의 레이어를 위하여 먼저 레거시 시스템으로부터 데이터 소스가 정의된다. 그 다음 데이터 소스로부터 ODS까지 데이터를 수집하는 방법이 정의된다.

다. 마지막으로 엔드 유저 집단을 위한 ODS로의 데이터 접근을 준비한다. 제시된 아키텍처를 기반으로 사용자 관점에서 운영 시스템과 ODS 간의 통합 레벨에 따른 타입을 분류하고, 각각의 특성을 요약한다.

제시된 아키텍처를 기반으로 사용 목적에 따라 서로 다른 타입을 가지고 유연하게 ODS를 구성 가능하며, 실시간 의사결정을 위한 각 요구사항 목적에 따라 최적화된 솔루션을 구성하여 활용할 수 있다.

2장에서는 ODS의 특성 및 요구사항에 관하여 설명한다. 3장에서는 레거시 시스템으로부터 ODS까지 데이터를 전달하기 위한 ODS 아키텍처를 각 레이어 별로 제시한다. 4장에서는 제시된 아키텍처를 기반으로 데이터 소스와 ODS 간의 연결 형태에 따라 타입을 분류하고 각각의 특성을 요약한다. 5장에서는 결론 및 향후 연구에 대해 논의한다.

2. ODS의 특성 및 요구사항

2.1 ODS의 역할 및 특성

ODS (Operational Data Store)는 데이터베이스

의 한 유형으로서 데이터 웨어하우스를 위한 일시적인 영역이다[1]. 정적인 데이터를 가지고 있는 데이터 웨어하우스와는 다르게 ODS의 내용은 비즈니스 오퍼레이션의 과정에서 계속적으로 업데이트된다. ODS는 매우 최신의 정보를 짧은 시간 동안 저장하기 때문에, 적은 양의 데이터를 가지고 단순한 쿼리로 빠르게 수행될 수 있도록 설계된다.

[표 1]에서 ODS가 가져야 할 특성을 간단히 요약하였다.

[표 1] ODS 특성

특성	설명
Subject-Oriented	회사의 요구사항뿐만 아니라 특정 어플리케이션의 관점에서 설계
Integrated	전사적 차원에서 운영 데이터의 통합된 이해 제공
Real-time Data Delivery	빠른 업데이트 주기를 통해서 데이터는 항상 up-to-date를 유지
Detailed	해결되어질 비즈니스 문제에 대한 세부적인 데이터를 유지

2.2 실시간 데이터 접근을 위한 ODS의 요구사항

사용자는 ODS를 통하여 다양한 관점(view)에서 운영 데이터를 수집한다. 또한 ODS는 운영 데이터를 계속적이고 실시간으로 접근해야 한다. 이러한 비즈니스 관점과 ODS의 특성을 고려하여 ODS를 설계하기 위해서는 다음과 같이 4가지 범주에서 요구사항을 식별해야 한다[2].

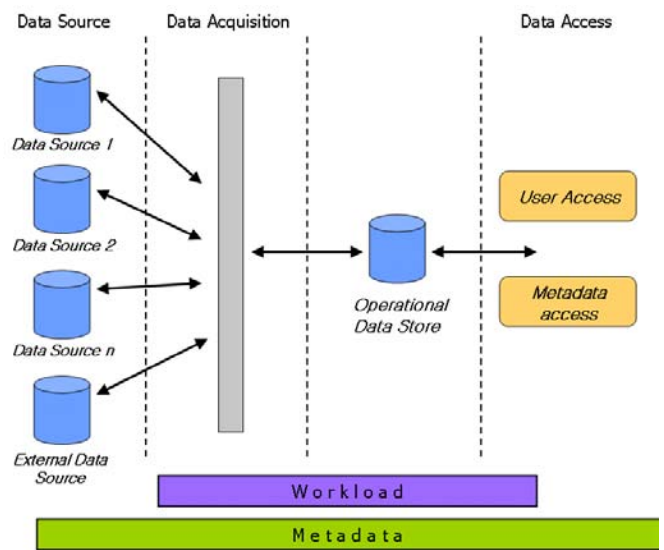
[표 2] ODS 요구사항

범주	요구사항
데이터 전송 (Transferring the data)	다양한 데이터 소스와의 통합 방법 및 데이터 변환 방법
데이터 특성 (Data characteristics)	데이터 업데이트 주기, 소스 동기화, 데이터 세부레벨
ODS 환경 (ODS Environment)	다중 트랜잭션 처리, 유연성, 가용성
ODS 관리 및 유지보수 (ODS administration and maintenance)	관리, 성능, 모니터링, 초기 로드

본 논문에서는 이러한 다양한 요구사항을 ODS의 활용범위에 포함된 레이어화된 아키텍처와 활용 목적에 맞는 ODS타입으로써 상세화하였다.

3. 실시간 데이터 접근을 위한 ODS 레이어 아키텍처

이 장에서는 레거시 시스템으로부터 ODS 엔드 유저 집단에까지 데이터를 전달하는 관점에서 정의된 아키텍처 레이어를 소개한다. 서로 다른 레이어들이 어떻게 연결되어 있는가에 대한 전체적인 개요는 다음 [그림 1]에서 보여 준다.



[그림 1] 전체 ODS 아키텍처 레이어

3.1 데이터 소스 (Data Source) 레이어

이 레이어에서는 비즈니스 목표에 따른 ODS 배치에 필요한 데이터 소스를 정의한다. 기존 레거시 시스템 혹은 애플리케이션에는 많은 양의 복잡하고 분산된 운영 데이터가 존재한다. 특정 비즈니스 목표 관점에서 ODS를 배치하기 위해서는 이러한 데이터들로부터 수집될 소스 데이터가 정의되어야 한다.

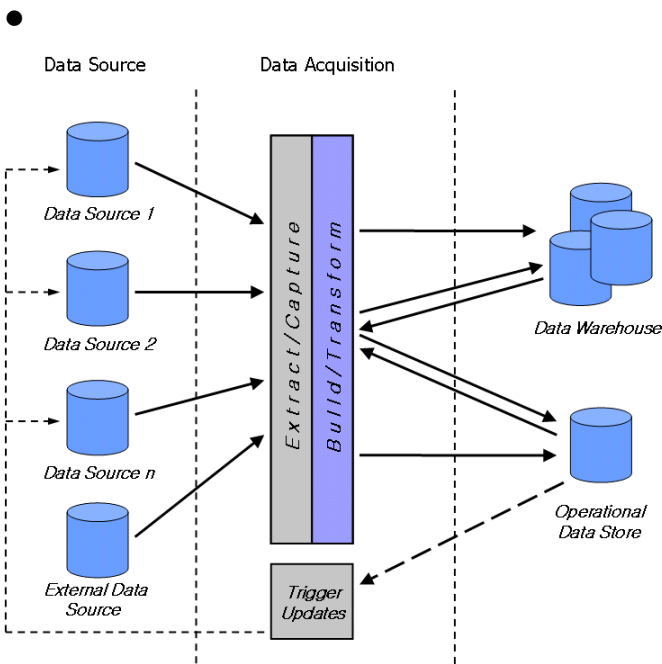
3.2 데이터 수집 (Data Acquisition) 레이어

이 레이어에서는 데이터 소스에서 만들어지는 변경을 캡처하여 요구되는 ODS 데이터를 추출한다. 다음 [그림 2]에서와 같이 데이터 수집 레이어에는 3가지 구성요소들이 포함된다.

- and/or 캡처를 추출 - 데이터 소스로부터 변경을 대규모로 추출하여 캡처함
- and/or 변형을 구축 - 추출된 데이터를 다양한

메소드를 활용해 실시간적으로 변형 및 통합하여 ODS 에 전달함

- 트리거 업데이트 - 데이터 소스에 반영되어야 할 특정 변경을 전달함 (ODS 타입 B만 해당)



[그림 2] 데이터 수집 레이어

3.3 워크로드 (Workload) 레이어

이 레이어에서는 성능상의 병목(bottleneck) 현상을 예방하기 위해 고려해야 할 사항들을 정의하고 워크로드를 관리한다. ODS 워크로드에 영향을 미치는 요소들은 3가지로 분류될 수 있다.

1. 레거시 시스템에서 발생하는 업데이트의 비율 및 로드된 데이터에 적용되어야 하는 변형의 종류
2. 엔드유저로부터 ODS로의 대규모 분석 요청 비율
3. 엔드유저에게 필요한 빠른 응답 시간의 소규모 트랜잭션 혹은 전략(tactical) 쿼리 비율

각 워크로드 요소들에 대한 아키텍처 대응 전략은 다음과 같다.

1. 첫번째 워크로드의 경우 데이터 수집을 스케줄링하여 ODS 업데이트 주기를 감소시킨다.
2. 두번째 워크로드의 경우 대규모 자원을 필요하므로 쿼리를 제한하거나 off-time 시간동안 수행한다.
3. 세번째 워크로드의 경우 트랜잭션의 응답 시간 및 자원을 적절히 고려하여 해당 물리 데이터 모델을 최적화한다.

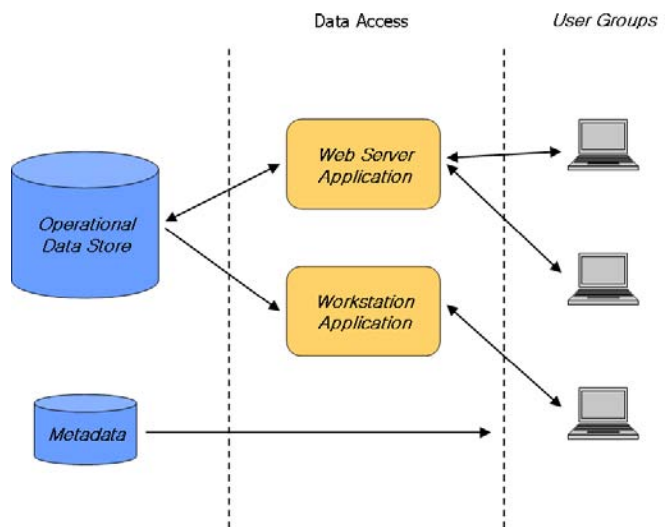
3.4 메타데이터 (Metadata) 레이어

이 레이어에서는 ODS 의 구축, 운영, 비즈니스 목적을 위한 메타데이터를 관리한다. ODS 환경 내에서 존재하는 메타데이터에는 3가지 종류가 있다:

- 빌드타임 메타데이터 - ODS 어플리케이션을 구축하는 동안 생성되고 사용
- 제어 메타데이터 - ODS 환경의 운영을 적극적으로 제어하고 관리하기 위해 사용
- 사용 메타데이터 - 비즈니스 유저에 의해 접근되어져 비즈니스 데이터의 내용, 목적, 근원 등의 정보를 파악하는 데 사용

3.5 데이터 접근 (Data Access) 레이어

이 레이어에서는 비즈니스 사용자가 ODS 에 쿼리를 보내고 데이터를 접근하여 조작할 수 있게 한다. 데이터 접근 레이어의 구성요소는 클라이언트 워크스테이션, 어플리케이션, 커뮤니케이션 표준으로 구성된다. 또한 더 나아가 미들웨어 요구사항 데이터 접근을 위한 하드웨어, 소프트웨어를 포함한다. 이 레이어의 아키텍처는 데이터 접근 요구사항에 매우 의존적이다. 다음 [그림 3] 은 데이터 접근 유형에 따른 데이터 접근 레이어의 구성을 보여 준다.

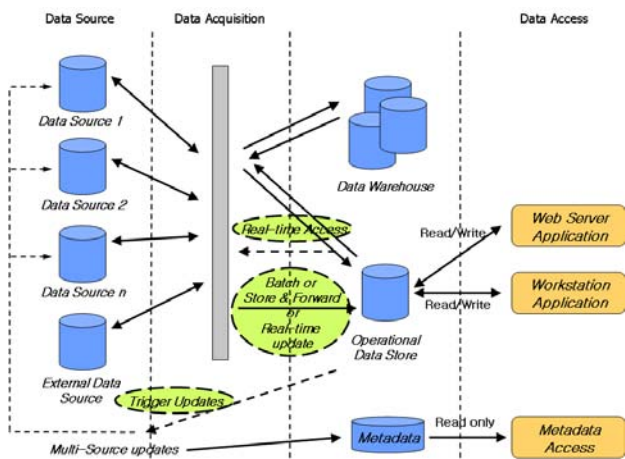


[그림 3] 데이터 접근 레이어

4. 사용 관점에 따른 ODS 아키텍처 타입

이 장에서는 사용 관점에 따른 아키텍처 타입을 설명한다. 일반적으로 ODS 아키텍처는 세가지 타입 (A,B,C) 이 있는데, 그것은 운영 시스템과 ODS 간의 통합 레벨에 따라 구분된다. ODS 사용자는 운영 데이터의 특성과 성격에 따라 접근을 다르게 할 수 있다. 다음 [그림 4]에서는 ODS 아키텍처 타입과

데이터 흐름의 예를 보여 주고 있다.



[그림 4] ODS 아키텍처 타입 A, B, C

ODS 타입 A 는 실시간 레거시 데이터 접근 및 로컬화된 업데이트를 포함한다. 데이터 수정이 레거시 시스템에 피드백되지 않는다. 로컬화된 업데이트는 대개 운영 시스템에서 현재 캡처된 새로운 데이터를 포함하지 않는다.

ODS 타입 B 는 ODS 타입 A 의 특성을 포함하면서 동시에 트리거를 가지고 있다. 트리거는 운영 시스템에 데이터를 피드백하는 메커니즘을 적용한다. 대개 이러한 피드백 요구사항은 데이터 간 충돌을 최소화하는 데 매우 중요하다.

ODS 타입 C 는 레거시 어플리케이션과 완전히 통합되거나 혹은 실시간 업데이트 및 접근을 사용한다. [표 3] 에는 각 ODS 타입의 다양한 특성들을 요약하여 제시하고 있다.

[표 3] ODS 타입 및 특성

특성	ODS type A	ODS type B	ODS type C
데이터 소스	운영 시스템 및 기타	운영 시스템 및 기타	운영 시스템 및 기타
ODS 실시간 업데이트	○	○	○
직접 접근	Read/Write	Read/Write	Read/Write
운영 시스템과 ODS와의 데이터 흐름	단방향	양방향	양방향
운영 시스템의 업데이트 주기	불가능	비동기	실시간

5. 결론

본 논문에서는 기업의 운영 레거시 시스템으로부터 실시간으로 데이터를 접근하기 위한 프로세스를 정의한 ODS 레이어 아키텍처를 제시하고, 각 데이터 흐름과 통합 레벨에 따른 ODS 의 타입 및 특성을 요약하였다. 제시된 아키텍처를 기반으로 ODS 를 구성할 때 다음과 같은 기대효과를 얻을 수 있다.

- 서로 다른 데이터 모델을 가진 레거시 시스템이라도 이에 상관없는 투명한 접근을 가능하게 한다.
- 데이터를 모니터링하기 위해서 DW뿐만 아니라 모든 애플리케이션들이 ODS영역의 데이터를 활용하는데 효과적으로 적용할 수 있다.
- 데이터 아키텍처에 대한 ODS 적용 레이어와 그에 따른 타입을 정의함으로써 활용범위에 대한 정확한 정의 및 구분을 제공한다.

향후 ODS 를 기반으로 한 실시간 데이터 모니터링 도구 구현 및 활용 방안에 대한 연구가 필요하다.

참고문헌

- [1] William H. Inmon, "Building the Operational Data Store", John Wiley & Sons, 1999
- [2] Corinne Baragoin, "Building the Operational Data Store on DB2 UDB Using IBM Data Replication, WebSphere MQ Family, and DB2 Warehouse Manager," "http://www.redbooks.ibm.com/redbooks/pdfs/sg246513.pdf", 2001
- [3] Bill Inmon, "The Operational Data Store", InfoDB, "http://www.evaltech.com/wpapers/ODS2.pdf", 1995
- [4] 이보영, "데이터웨어하우스의 新패러다임이 시작된다", IT Business Journal, 2004
- [5] Ross Altman, "Building an Operational Data Store on Top of Legacy Applications: Sometimes You Do, Sometimes You Don't", Business Integration Journal, "http://www.seebeyond.com/pdf/OperationalDataStore_Altman_BIJ.pdf", 2004