

Data Access 효율 증대를 위한 Data Hub 에 관한 연구

박은영, 김응모
 성균관대학교 전자전기컴퓨터공학과
 e-mail : searomi@paran.com

A Study on Data Hub for enhancing data access efficiency

Eun-Young Park, Ung-Mo Kim
 Computer Engineering, Sungkyunkwan University

요 약

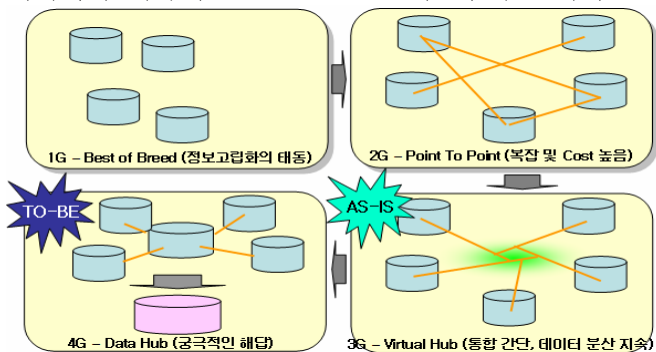
기업의 원활한 활동에 필수적인 요소는 살아 있는 정보의 공급이며, 기업 데이터 베이스는 데이터의 보관/흐름/활용 입장에서의 정보 인프라를 기반으로 하고 있다. 최근 기업 Knowledge Base 구축은 정보의 품질이 좌우하나 기업 정보의 양은 비대해져 데이터의 일관성이 결여되고 DB 의 성능은 점점 저하되고 있다. 본 고에서는 데이터 Hub 의 적용을 통해 얻을 수 있는 DB 성능 향상에 대해 정성적/정량적 관점으로 논의한다.

1. 서론

기업의 원활한 활동에 필수적인 요소는 살아 있는 정보의 공급이며, 기업 데이터 베이스는 데이터의 보관/흐름/활용 입장에서의 정보 인프라를 기반으로 하고 있다. 통합되지 않은 데이터에 의해서 경영 활동 많은 부분에서 비효율적인 대응이 발생하며, 이러한 분리된 데이터는 복합 채널에서의 효율적인 기업 활동을 방해 하게 된다. 기업 Knowledge Base 구축은 정보의 품질이 좌우하나, 일관되지 않은 프로젝트에 의해 기업 정보는 조각조각 관리되고, 그나마 일시적으로 일부의 정보만을 제공 받을 수 밖에 없다.

기업은 전문화된 시스템 기반을 구축해 가야 하며, 그러기 위해서는 통일되고 융화된 질 높은 단일 소스 데이터를 필요로 한다. 기존 업무 프로세스에 혼란을 주지 않고, 이미 구축된 IT 투자에 큰 변동 없이 어떻게 조각난 정보를 통합할 것인가가 당면한 과제이며, 기업 데이터 관리의 희망 사항이다.

데이터의 집중화, 중복 제거 및 질 향상을 완벽하게 통합하여 기업 데이터를 관리하고, 더 중요한 것은 데이터가 지속적으로 모든 소스 시스템과 동기화



(그림 1) 기업 Knowledge Base 구축을 위한 IT 기술 변천사

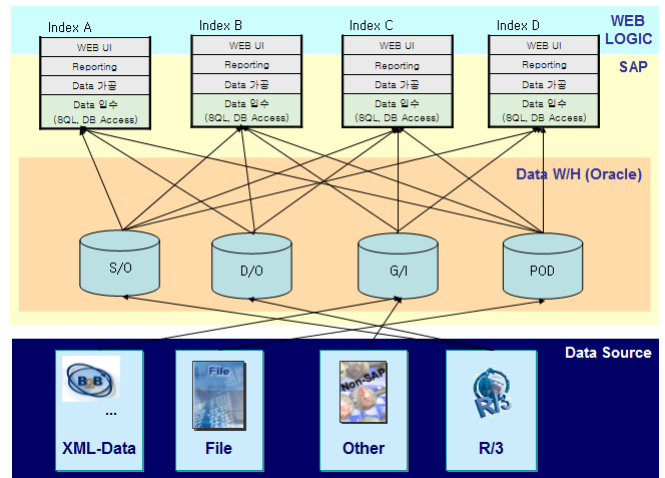
되고, 모든 기간계 시스템 및 분석 시스템과 일치 되어야 한다.

(그림 1) 은 기업 Knowledge Base 구축을 위한 IT 기술 변천사를 보여 주고 있는데, 현재 기업들의 IT 기술 수준은 가상 허브를 구축하여 비교적 간단한 통합 기술로 운영되고 있지만 데이터의 분산 문제는 지속 발생되고 있는 수준에 머무르고 있다.

2. 데이터 Hub 도입의 필요성

2.1 現 ERP 시스템 구조 및 문제점

기업 ERP 시스템 중 SAP 환경 내에 구축된 물류 시스템의 구조를 살펴보자.



(그림 2) ERP 시스템 구조 [Logistics]

물류 시스템은 동일 환경의 SAP 시스템, 다른 시스템 등 다양한 데이터 소스로부터 필요한 정보를 인터페이스 받는다. 인터페이스 된 정보는 SAP 데이터 웨

어하우스 내 저장되며 이 데이터를 근간으로 UI 상에 인덱스(또는 지표)를 개발하여 사용자는 기업 활동에 필요한 정보를 얻게 된다. 일반적으로 SAP 환경에서의 Index 는 다음과 같이 4 단계로 개발된다.

데이터 입수 (SQL, DB Access) → 데이터 가공 → 리포팅 → Web UI 연계

현 구조는 동일 데이터 풀로 용도가 다른 인덱스를 개발할 경우, 각 인덱스 개발 때마다 데이터 입수 (SQL, DB Access) 부분 구현이 신규 진행되어야 한다. 또한 각각 인덱스 개발 때마다 데이터 입수/가공 부분이 필요하기 때문에 UI 개발자가 데이터 입수/가공 단계를 개발해야 하는 상황이 종종 발생하며, 반대의 경우도 발생된다.

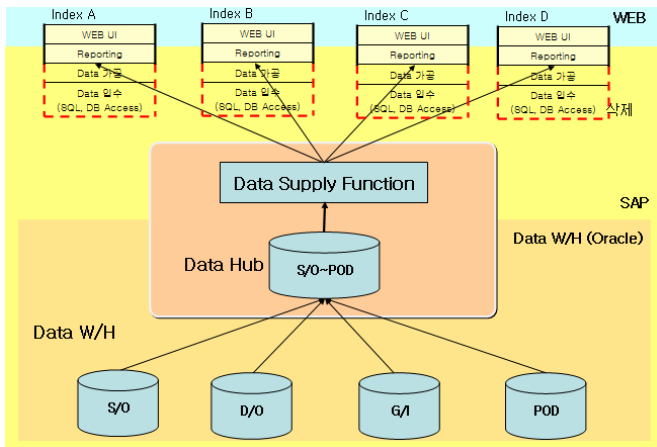
위와 같은 시스템의 구조로 다음과 같은 문제점이 발생되는데, 첫째, 신규 개발되는 인덱스에 따라 DB Access 빈도수가 증가되어 시스템 Traffic 이 높아지며, 둘째, 높아지는 Traffic 에 따라 사용자 조회 효율이 저하된다. 마지막으로, UI 와 DB 간 개발자 전문화 영역에 따른 개발 수준 편차가 발생되어 시스템 개발/운영 생산성 저하를 가져온다.

2.2 데이터 허브 도입 기대 효과

위와 같은 문제점을 해결하기 위해서는 배치 잡 분산, 데이터 입수 부분 개선 등 다양한 방법이 있겠으나, 가장 큰 문제점인 DB Traffic 과 사용자 응답성을 근본적으로 개선하기 위해 데이터 허브 도입을 채택하였다.

물류 데이터 허브는 어플리케이션 층과 데이터 층 사이 데이터 관리 층에 속하며, 물류 데이터 허브는 각 시스템 별 분산된 물류 원천 데이터를 통합 하고, 공급자와 공급 받는 자, 물류비용과 재고 관점으로 정보를 세분화/통합화 후 필요한 물류 정보를 사용처에 공급하는 DB 의 총칭이다.

데이터 허브 도입을 함으로써 달라지는 시스템 구조는 어플리케이션 상에서 각 인덱스 마다 데이터



(그림 3) 데이터 Hub 도입 후 ERP 시스템 구조

웨어하우스에 접속하여 DB Access 를 수행하는 것이 아니라 어플리케이션 층과 데이터 층 사이의 관리 층에 해당하는 데이터 허브 (데이터 허브 코어+ 데이터 Supply Function)을 두는 구조로 변경되며, 각각 인덱스 어플리케이션 개발 시 필요했던 데이터 입수 부분은 데이터 허브 내 데이터 Supply Function 으로 대체 된다.

데이터 허브 도입 시 구조의 변경으로 다음과 같은 효과를 기대할 수 있다.

첫째, 물류 데이터의 중앙화이다. 이는 중앙 집중된 단일 데이터 저장소를 의미하며, 수평/수직 계층 관계 모두에 적용되는 데이터 모델이다.

둘째, 물류 데이터의 품질 향상이다. 데이터 허브를 통해 중복된 데이터 정제 프로세스 제거와 다양한 소스 하드 카피를 감소시킨다. 또한 데이터 정제 라이프 사이클을 위한 단일 통합 데이터 Librarian 을 구축할 수 있으며 중복 제거와 데이터 보완 룰을 위한 단일 집중화된 저장고의 기능을 한다.

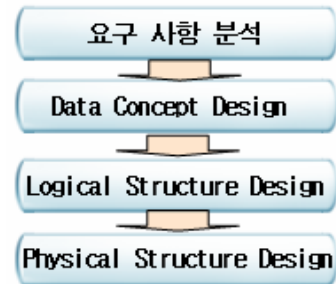
셋째, 소스 데이터와 프로세스 동기화가 가능하다. 데이터 허브에서 물류 데이터 또는 어플리케이션 관리 유연성이 증대되고, 각 인덱스에 최신의 일치된 정보로 물류 정보 관리가 가능하다.

마지막으로 물류 Knowledge Base 활용 극대화이다. 이는 타 시스템에 정제된 데이터를 제공 가능하여 시스템의 신뢰성을 높일 수 있다.

3. 데이터 허브 구현

3.1 데이터 허브 설계

데이터 허브 구현은 물류 시스템 중 제품 물류 리드타임 데이터 산출 인덱스에 적용해 보기로 한다.



(그림 4) 데이터 허브 설계 단계

데이터 허브 구현은 위와 같이 요구사항 분석, 데이터 개념 설계, 논리적 구조 설계, 물리적 구조 설계 4 단계로 구분할 수 있다.

요구 사항 분석은 물류 리드타임 데이터 허브 구현 시 필요한 사용자의 요구 사항을 분석 하는 단계이며, 다음과 같이 기능 중심 4 가지 요구 사항으로 요약하였다.

1. 구간별 리드타임
2. LOPO 별 리드타임
3. Carrier 별 리드타임
4. TPL 별 L/T (TPL : Third Party Logistics)

데이터 개념 설계 단계에서는 위의 요구 사항을 충족시키기 위한 데이터 구상 단계로 각 요구 사항 별 데이터 허브 구성 시 필요한 데이터 범위를 정의하고 항목을 도출해 내는 작업이 필요하다.

데이터 허브를 구성하기 위한 논리적/물리적 구조 설계 단계에 들어간다.

논리적 설계 단계는 데이터 개념 설계에서 정의된 4 가지의 리드타임을 구체화 시키는 작업으로써 상위와 하위의 관계 및 수집해야 할 데이터 테이블을 정의한다. 데이터 허브 코어는 확장성을 위해 상위/하위 2 개 층으로 설계한다. 또한 수집해야 할 데이터 항목들에 대해 L/T 구분자인 Event Code 를 부여한다.

<표 2> 물류 L/T Logical Design

UPPER			LOWER			수집	
Point	Table	Short Text	Table	Short Text	Table	Short Text	
L1	ZTSPV_LT100	구간별 LT	ZTCORE_QT130	LT 별 물품량	ZTEGL_SSP01	PO	
	ZTSPV_LT110	카운터별 LT			ZTEGL_SSP02	Sales order	
	ZTSPV_LT120	TPL 별 LT			ZTEGL_SSP03	SO ITEM	
	ZTSPV_LT130				ZTEGL_SSP04	Schedule line	
					ZTEGL_SSP05	Delivery	
					ZTEGL_SSP06	DO ITEM	
					ZTEGL_SSP07	BILLING	
					ZTEGL_SSP08	Goods issue	
					ZTEGL_SSP09	Goods Receipt	
					ZTEGL_SSP10	BOOKING REQUEST	
					ZTEGL_SSP11	BOOKING CONFIRM.	
					ZTEGL_SEE01	ED	
					ZTEGL_SSP01	SR	
					ZTEGL_SBI01	BILL OF LADING	
					ZTEGL_SAND01	AN	
					ZTEGL_SSD01	ISA	
					ZTEGL_SSP01	STLS GR	
					ZTEGL_SLD01	Outbound DO	
					ZTEGL_SSD01	STLS GR	
					ZTEGL_SSP01	STLS POS	
					ZTEGL_SCM03	CMS	
					ZTEGL_SSD01	HTTS	

<표 3> 물류 L/T 수집 Table 상세화

Business Type	A	C	D	E	F	G	H	I
	구간포기	Event Code	관리일자	약어 (LOPO)	권원 Code	Source	Source	Source
1	수집	PO	1010	PO부서생성일자	L1100	ZTEGLSSP01	ERDAT	
2	수집	PO	1011	PO ETA(의송)일자	L1101.1	ZTEGLSSP01	ZETA	
3	수집	SO	1020	SO부서생성일자	SO	L1020	ZTEGLVBRK	ERDAT
4	수집	SO	1021	SO ETA(의송)일자	SO	L1020.1	ZTEGLVBRP	EDATU
5	수집	BR	1030	BR부서생성일자	BR	L1030	ZTEGLSBR01	ISSUE_DATE
6	수집	BR	1031	BR부서생성일자	BR	L1040	ZTEGLSBR01	ETD
7	수집	BC	1040	BC부서생성일자	BC	L1050	ZTEGLSBC01	ISSUE_DATE
8	수집	BC	1041	BC부서생성일자	BC	L1050.1	ZTEGLSBC01	ETD
9	수집	BC	1042	도착일자(ETA)	BC	L1050.2	ZTEGLSBC01	ETA
10	수집	BC	1043	출발일자(ETD)	BC	L1050.3	ZTEGLSBC01	CY_CLOSE_DATE
11	수집	DO	1050	DO부서생성일자	DO	L1080	ZTEGLVBRP	ERDAT
12	수집	DO	1051	Planned G/I Date	DO	L1080	ZTEGLVBRP	WADAT
13	수집	Invoice	1060	Invoice부서생성일자	IV	L1100		
14	수집	CONTR IN	1061	Container반입일자	CM	L2030.1	ZTEGLSCM01	
15	수집	CONTR IN	1062	Container반입일자	CM	L2030	ZTEGLSCM01	
16	수집	Expert Permit	1070	수출신청/수출일자	ED	L2032	ZTEGLSE001	
17	수집	Expert Permit	1071	수출입일일자	EDA	L2033	ZTEGLSE001	
18	수집	Packing	1080	Container Stuffing일자	PL	L2036	ZTEGLSCM01	
19	수집	Packing	1081	Container Stuffing완료	PL	L2036	ZTEGLSCM01	
20	수집	S/R	1090	S/R부서생성일자	SR	L2070	ZTEGLSSP01	ISSUE_DATE
21	수집	S/R	1091	출발일자(ETD)	SR	L2070.1	ZTEGLSSP01	
22	수집	Factory Out	1100	반출일자	CM	L2060	ZTEGLSCM01	
23	수집	Hub In	1110	지회(CDC) 반입일자	CM	L2075		
24	수집	Hub Out	1111	지회(CDC) 반출일자	CM	L2076		
25	수집	Port CY In	1120	출발일자(항공사)	CM	L2077	ZTEGLSCM01	
26	수집	BL	1130	BL생성/수입일자	BL	L2080	ZTEGLBL01	ISSUE_DATE
27	수집	BL	1131	출발일자(출발일(ATB))	BL	L2080.1	ZTEGLBL01	SHIP_DATE
28	수집	GI	1140	이발일자	GI	L2100	ZTEGLSGI01	ERDAT
29	수집	GI	1141	GI발일자(Posting일자)	GI	L2100	ZTEGLSGI01	ACT_GI_DATE

물리적 설계 단계는 상위 테이블 구성 시 필요한 필드 항목들을 정의하고 각각의 필드항목에 대한 속성을 정의한다.

<표 4> 구간별 L/T(ZTSPV_LT100) Physical Design

Field	Data Element	Data Type	Length	Decimal	Short Text
MANDT	MANDT	CLNT	3	0	Client
ZKEY_SOURCE	ZKEY_TOP	CHAR	38		0 Key Combination, Source Table (Source 데이터 Table 내 Seq 번호)
BIZ_TYPE	ZBIZTYPE	CHAR	3		0 Business Type (ETX,MLD,30)
SALES_TYPE	ZSALESTYPE	CHAR	4		0 Sales Type (Jaga, OEM, 30) (판매 유형 (차기, OEM, 로제트))
ORDER_KEY	ZORDKEY	CHAR	20	0	[Hub] Order Key
LOPO	ZLOPO	CHAR	3		0 Document Type (BR, BC, ED, SR, BL, AN, ID)
CALMONTH	ZCALMONTH	NUMC	6	0	[CAIS] Calendar Year/Month
CALWEEK	ZCALWEEK	NUMC	6	0	Calendar year / week (주회, 주차)
DOC_NO	ZDOC_NO	CHAR	35	0	Document number
LOGI_TYPE	ZLOGITYPE	CHAR	4		0 Logistics Type (물류 Point(내자외물류, 서비스, Normal) 재출발, 반출)
SALES_GUBUN	ZSALES_GUBUN	CHAR	4		0 Sales Gubun (Sales Sub, Direct Sales) (판매 구분(판매입, 반출))
PROD_TYPE	ZPRODTYPE	CHAR	4		0 Product Type (SET, SEMI, HALB) (재출출류 (재출/가제/가제출))
SHIP_TYPE	ZSHIPTYPE	CHAR	10		0 Shipment Type (AIR, VSL, TRUCK, RAIL) (운송수단 (AIR, VSL, TRUCK, RAIL))
TRANS_CODE	ZTRANS_CODE	CHAR	3		0 Transportation Code
MSG_NO	ZMSG_NO	CHAR	35	0	Message No.
REF_DOC_TYPE	ZREFDOCTYPE	CHAR	3		0 Reference Document Type (Reference 문서 Type (BR, BC, SO))
REF_DOC_KEY	ZREFDOC	CHAR	20	0	Reference Document Number (Reference 문서 번호)
REF_DOC_ITEM	ZREFDOCITEM	CHAR	6	0	Reference Document Item (Reference 문서 item)
REF_MSG_NO	ZREFMSGNO	CHAR	35	0	Reference Message Number (참고 메시지 번호)
TPL_CODE	ZTPL_CODE	CHAR	10	0	3PL Code (STLS Code)
KUNNR	ZKUNNR	CHAR	10	0	Ship-to party
ZDIND	ZDIND	CHAR	1	0	리드타임 Indicator
ROUTE	ZROUTE	CHAR	6	0	Route
LZONE	ZLZONE	CHAR	10	0	Transportation zone to or from which the goods are delivered
BUKRS_P	ZBUKRS_P	CHAR	4	0	Plant Level Conversion Code (Company)
BUKRS_S	ZBUKRS_S	CHAR	4	0	Company Code (Sales) (판매법인 Company)
FROM_SITE	ZFSITE	CHAR	20	0	From Site
BUKRS_F	ZBUKRS_F	CHAR	4	0	From Company Code (출발 법인)
FROM_SITE_TYPE	ZFSITETYPE	CHAR	10	0	From Site Type
TO_SITE	ZTSITE	CHAR	10	0	To Site
BUKRS_T	ZBUKRS_T	CHAR	4	0	To Company Code (도착 법인)
TO_SITE_TYPE	ZTSITETYPE	CHAR	10	0	To Site Type
ZFCTRY	ZFCTRY	CHAR	2	0	From Country
ZFPOR	ZFPOR	CHAR	10	0	From Port
ZTCTRY	ZTCTRY	CHAR	2	0	To Country
ZTPOR	ZTPOR	CHAR	10	0	To Port

3.2 구현

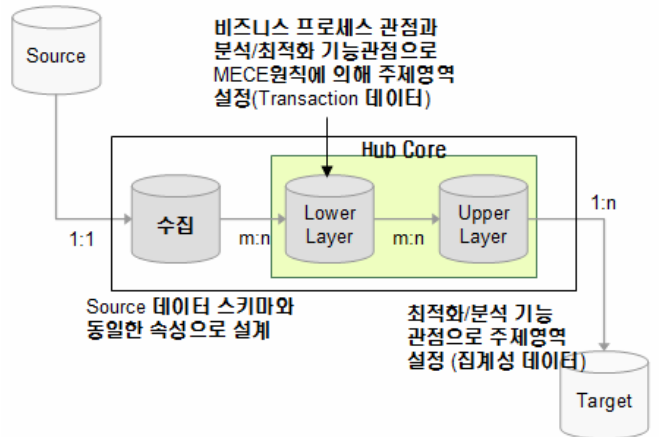
데이터 허브 구현은 데이터 수집, 데이터 허브 구현, 리포팅, UI 개발 4 단계로 구성된다.

우선 물류 리드타임의 DB 설계 후 미리 정의해 놓은 항목대로 데이터를 수집한다. 대부분 기존 확보된 데이터이나 신규 정의 된 데이터의 경우에는 소스(국내/해외 지사 ERP 데이터 및 업체 데이터) 로 부터 데이터를 인터페이스 해야 한다.

데이터 수집이 완료된 후 데이터 허브 코어는 데 디자인 시 정의했던 바와 같이 2 개 층으로 구현한다.

- 하위 층은 비즈니스 프로세스 관점으로 분류되어 MECE 하게 설계하여 다양한 물류 업무에서 요구하는 데이터를 적재할 수 있도록 함

- 상위 층은 시스템 내 분석/최적화 기능 관점으로 분류되어 다양한 솔루션에서 요구하는 데이터를 효율적으로 공급할 수 있도록 하는 목적을 두고 있다.



*MECE (Mutually Exclusive and Collectively Exhaustive) :: 중복되지 않고 각각의 합이 전체를 포함할 수 있는 요소의 집합 (그림 6) 데이터 허브 구조

데이터 허브 코어 개발 후 데이터 허브의 Supply Function 을 개발한다. Supply Function 은 어플리케이션 에서 허브 데이터의 Access 를 용이하도록 코어의 데이터를 가공하는 기능을 한다. 구간별 리드타임 예를 들어 설명하면 기 정의해 놓은 Event Code 를 기본으로 리드타임 산출을 요하는 구간의 리드타임 을 아래 프로그램과 같이 구성하여 놓는다.

```

Program 시용서

● 저리 상세 요건(로직)
Lead Time 저리 로직
공용 : ZTCORE_QT100 에서 CALWEEK , BUKRS_P 를 조건으로 데이터를 읽어서,
ZTCORE_CM100 에서 ZTCORE_QT130 의 ZKEY_SOURCE 를 읽는다.

반출 ~ 선적
1. ZTCORE_QT130-EVENT_CODE = '1100' ( 반출일자 )
2. ZTCORE_QT130-EVENT_CODE = '1131' ( 출발항상지출발일 ) 을 읽어와서

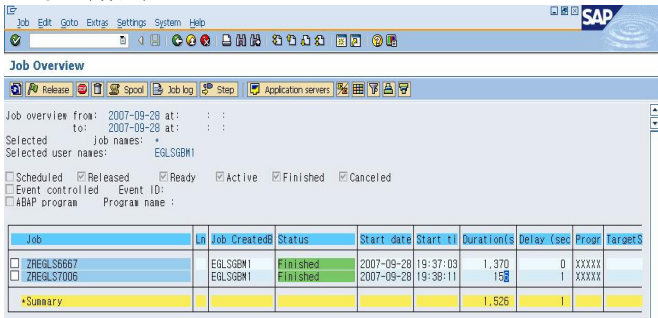
CALL FUNCTION 'Z_F_EGL_S9102'
EXPORTING
P_BUKRS = 회사코드
P_DATELS = 1. 날짜
P_DATELE = 2. 날짜
P_WERKS = Plant 코드
PLOPT = ''
IMPORTING
P_LLT = 날짜 Lead Time 이 계산 되어서 나온다.
P_LHD = 0. 날짜
P_WLT = 0. 날짜
    
```

(그림 7) 데이터 Supply Function 로직

데이터 Supply Function 까지의 개발이 완료되면 리드타임 산출 시 필요한 기본 구성이 완료되었으며, 이후에는 기존 SAP 어플리케이션 개발 순서와 동일하게 리포팅 (ODS Loading, Query, Template), UI 로딩의 순서로 구현한다.

4. 성능 평가

데이터 허브의 성능 평가는 SAP 어플리케이션에서 지원하는 배치 잡 수행 시간 측정 기능을 사용하여 진행 하였다.



(그림 8) 성능 측정 - Batch Job 수행 시간

성능 평가는 물류 구간별 리드타임 중 고객 배송 지시~ 고객 배송 완료 리드타임 산출을 데이터 허브를 사용하지 않은 경우와 데이터 허브를 사용한 경우의 두 가지 배치 잡 수행 시간을 측정하는 것으로 진행하였다.

실험은 특정 주차(1 주차) 데이터로 진행되었으며 실험 결과는 다음과 같다.

(단위 : 초)

횟수	기존 방법 (데이터 허브 未사용)	데이터 허브 사용 시
1	1,505	175
2	1,440	148
3	1,370	156
4	1,485	165
5	1,398	152
평균	1,440	159

<표 5> Batch Job 수행 시간 실험 결과

최종 실험 결과는 기존 방법으로 물류 리드타임 산출 시 약 1,440 초 수준이나, 데이터 허브 사용 시 약 1/9 수준인 159 초로 나왔다.

5. 결론

데이터 허브는 최근과 같이 기업 데이터가 점차 방대해져 관리의 중요성이 커지고 또한 기업 의사 결정에 신속하게 정확한 데이터를 제공해야만 하는 상황에서는 날로 중요성이 커지고 있는 상황이다.

데이터 허브의 도입은 기업 데이터의 중앙화, 품질 향상과 같은 정성적 효과 외에도 데이터 산출 성능 향상과 정량적인 효과를 기대할 수 있다.

데이터 허브는 지속적인 데이터의 추가로 허브의 규모가 비대해질 경우 데이터 발취 시 시간이 오래

소모될 수 있으며, 또한 데이터 허브 운영 중 변경 (추가/삭제) 시 유연성이 떨어지므로 초기 설계 시 향후 확장성을 고려하여 신중히 설계해야 하는 단점도 가지고 있다.

그러나 데이터 허브는 위와 같은 단점에도 불구하고 현재 기업 정보 시스템 구축에 있어서 많은 장점과 가능성을 제시하고 있으므로 많은 기업 정보 시스템에 활용될 것으로 보인다.

참고문헌

- [1] SDS ERP Consulting, "SAP BW Overview", 삼성 SDS, 2005
- [2] SDS ERP Consulting, "Warehouse Management", 삼성 SDS, 2005
- [3] 이진권 "정보 인프라의 근간, 고객 데이터 허브", 한국 오라클, 2005
- [4] Ramez Elmasri, Shamkant B. Navathe, "Data Base System 4th edition", Addison Wesley, 2003
- [5] Tan Steinbach, Kumar, "Introduction to Data Mining", Addison Wesley, 2006
- [6] Han Kamber, "Data Mining Concept and Techniques 2nd Edition", Morgan Kaufmann, 2006
- [7] 김성수, "물류 정보 시스템", 대영사, 2001
- [8] 김정환, "물류 시스템 설계론", 문영각, 1999
- [9] SKDG, "Application Development ABAP Workbench", 한국 SAP 개발자 동호회, 2002