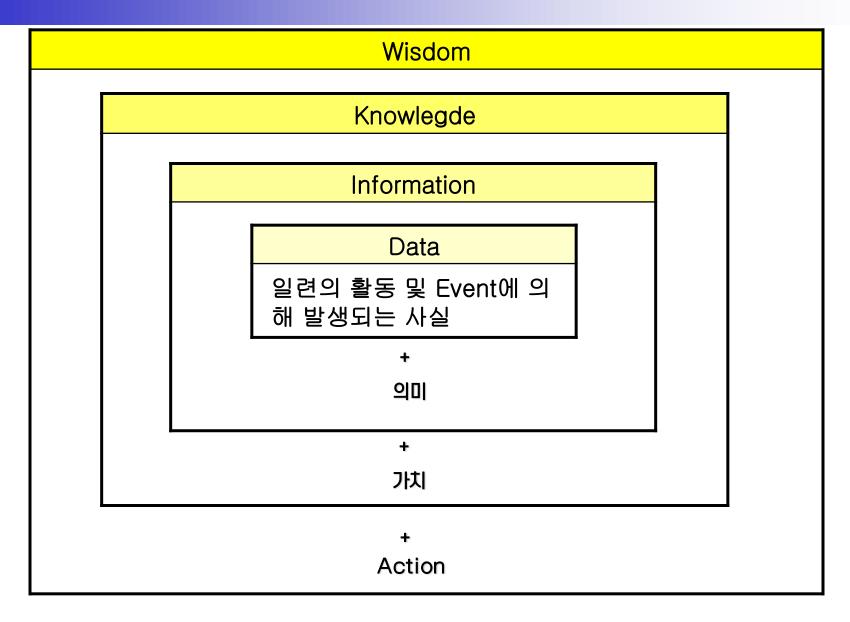
DW / BI 소개

목차

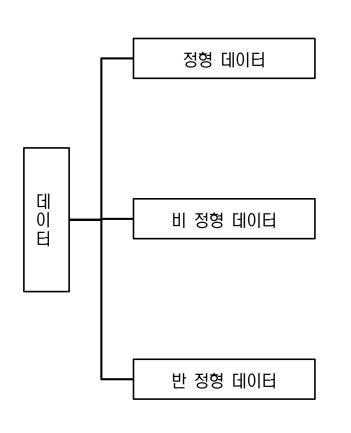
- 1. 개요
- 2. DW
- 3. 적용

1.1 Data, Information, Knowledge and Wisdom

1.개요



데이터는 정형 데이터, 비정형 데이터, 반정형 데이터가 있으며, 기업의 내부 및 외부 데이터로 구분 할 수 있다.



정의 : 데이터의 구조가 알려진 형태의 데이터로

Machine이 처리 가능한 데이터 예 : 고객 Table, 주문 테이블

정의 : 데이터의 일정한 구조가 없는 데이터로

Machine이 이해하지 못하는 데이터

예 : 사진, 문서/ Voice 등…

정의 : 일정한 데이터의 구조는 있느나, 데이터의 해석을

위한 기술적 또는 문맥적인 정보가 없는 데이터

예 : Sensor, Weg Log

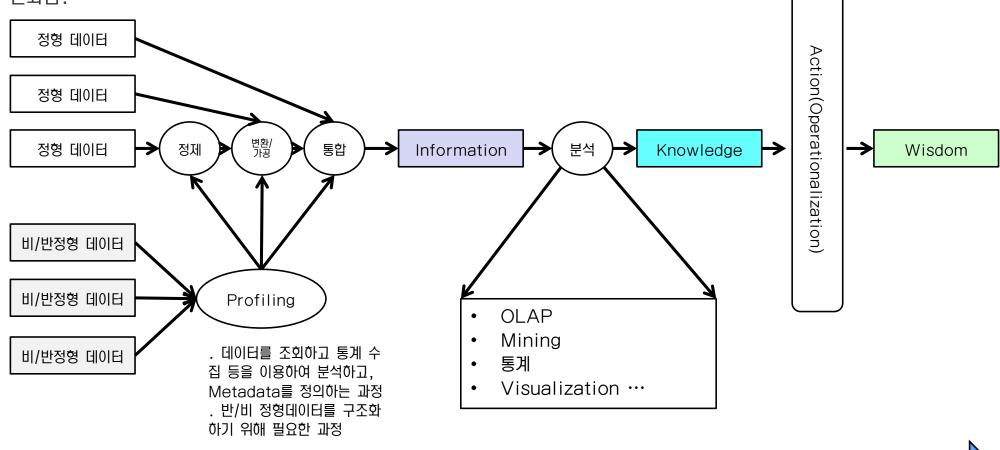
1.1 Data, Information, Knowledge and Wisdom

1.개요

데이터 저장 기술의 발전으로 반/비 정형의 데이터가 수집/관리가 가능해 졌으며, **Big Data로 정의**되어 Biz 개선 및 사업 영역 확대를 위한 분석을 수행함.

Exabytes	지리적 정보 Connected Vehicle	Telematics & Sensor Data 기후 데이터	상품 및 서비스 로그 사물 인터넷	BIG DATA
Petabytes	Web Logs 제안	동적 가격 제안 A/B Testing	WEB	외부의 지리적 정보 업무 데이터의
Terabytes Gigabytes	Offer Details Segmentation 구매/생산 ERP 영업/물류 재무/회계	CRM 고객 접점 Support Contacts	제휴 네트웍 검색 마켓팅 Behavioral Targeting Dynamic Funnels	User Click Stream Speech to Text Social Network 감성
	데이터의 복잡성 및 다양성 증가 데이터 가치의 밀도 감소			

데이터는 가공, 정제 및 통합 과정을 통해 Information으로 Information을 분석함으로써 Knowledge (또는 Insight) 로 가치가 증가하며, 이를 적용 (Operationalization 또는 Action)의 과정을 통하여 Wisdom으로 변화함.

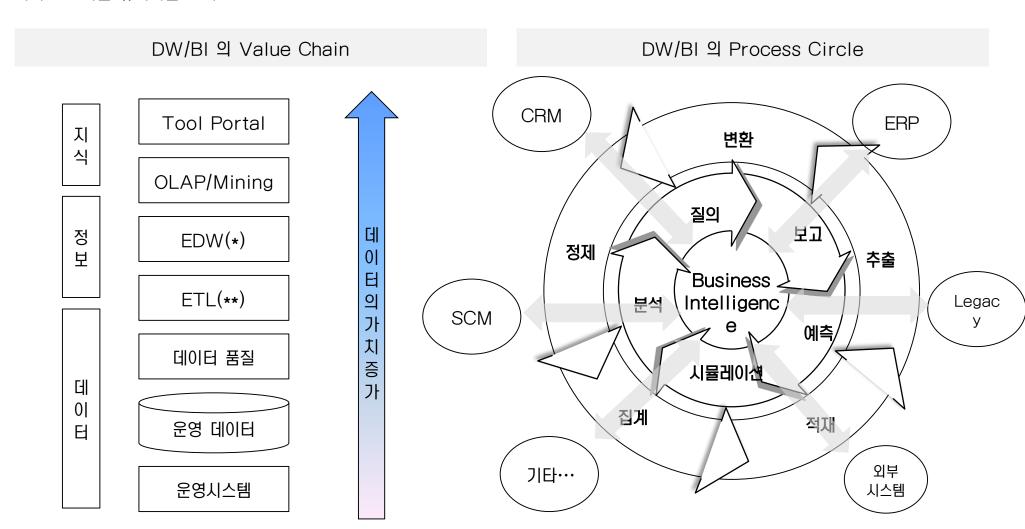


데이터의 가치 증가 (Value Chain)

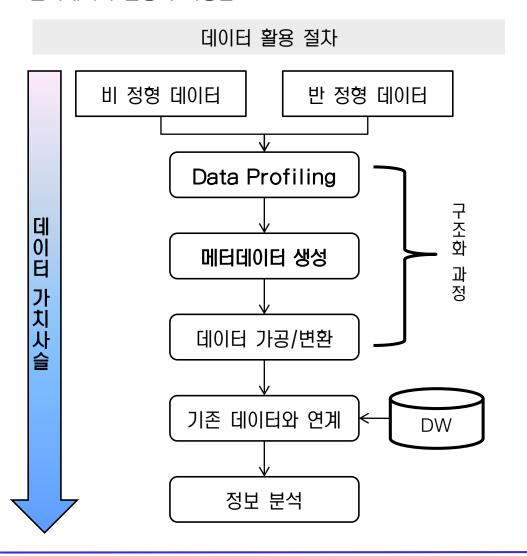
1.1 Data, Information, Knowledge and Wisdom

1.개요

일련의 프로세스를 통하여 서비스 또는 제품의 가치를 향상 시켜 고객의 요구 사항을 충족함으로써, 기업의 가치 또는 이익의 증가를 목적으로 하는 유기적인 조직



비/반 정형 데이터는 데이터의 Profiling 및 분석을 통한 관련 Metadata을 통한 구조화하여 그 의미를 탐핵 후, 분석에서의 활용이 가능함.



철차 개요

Data Profiling

- . 데이터의 활용을 위해 분석하고 데이터에 대한 데이터 (Meta Data)를 수집하는 과정
- . 비/반 정형 데이터의 활용을 위해서는 반드시 필요

메터 데이터 생성

- . Data Profiling에서 수집된 데이터의 데이터 (Metadata)를 기반으로 Meta Data 생성
- . 기존의 생성된 Metadata 들과의 연계

데이터 가공 및 변환

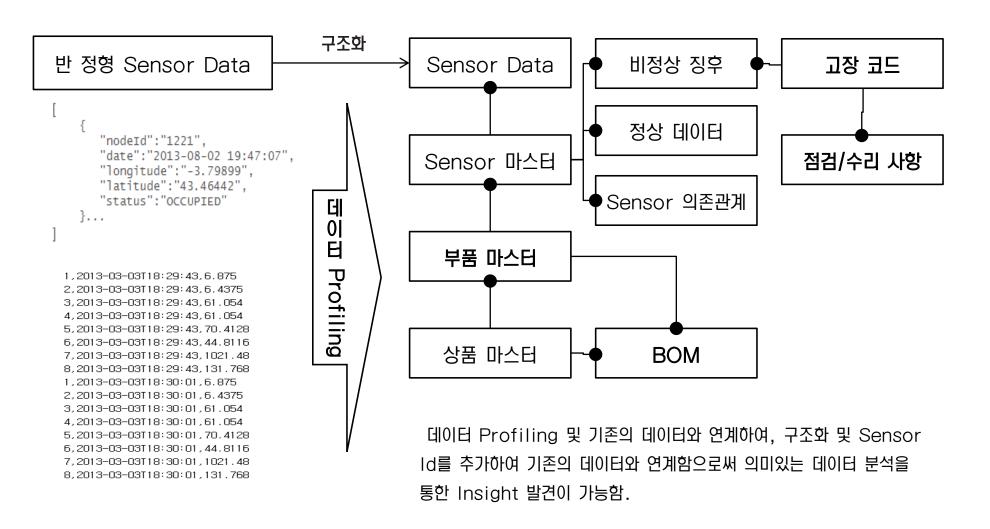
- . 기존의 데이터와 연계 및 정형화를 위한 데이터 가공 및 변 환
- . Data의 Voice To Text 변환, Spliting, Tokenizing, Summarization 등…
- . 새로운 Field 생성

기존 데이터와 연계

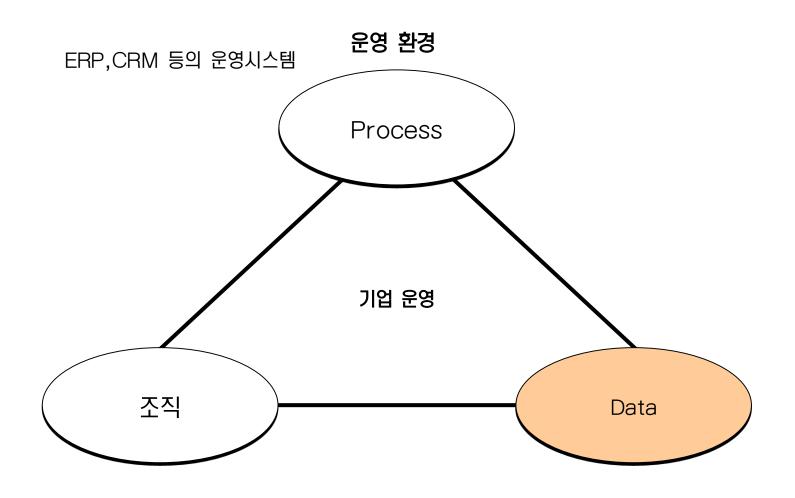
. 기존 DW의 데이터와 연계하여 문맥적 정보 강화

1.1 Data, Information, Knowledge and Wisdom

예시 : 반 정형 데이터인 Sensor Data를 이용한 사전 고장 예측 활용

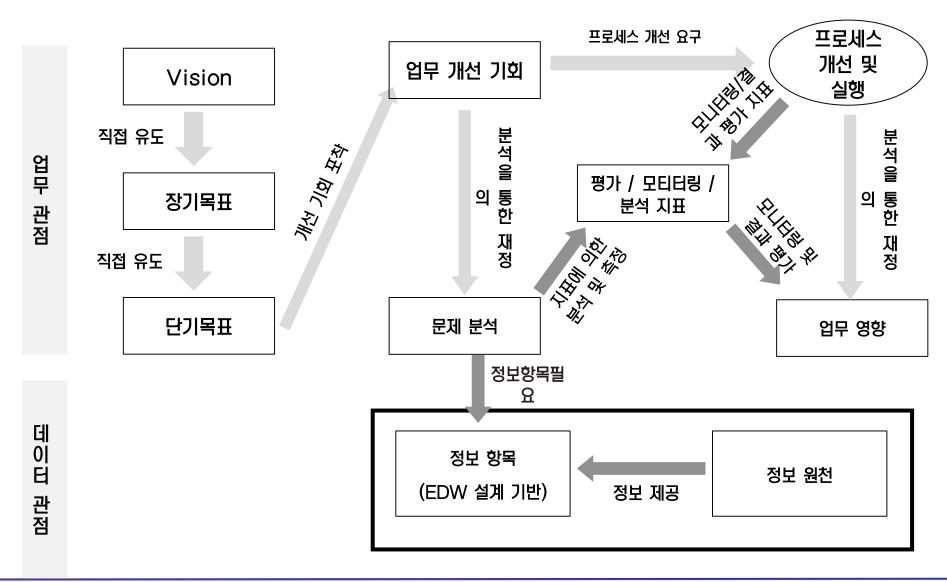


일련의 프로세스를 통하여 서비스 또는 제품의 가치를 향상 시켜 고객의 요구 사항을 충족함으로써, 기업의 가치 또는 이익의 증가를 목적으로 하는 유기적인 조직



1.2 기업과 의사결정 - Vision, Mission, KPI & IT

1.개요



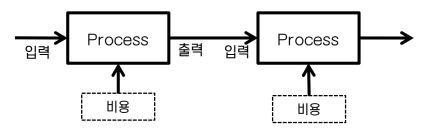
1.2 기업과 의사결정 - Process

1.개요

제품 또는 서비스의 가치를 증가 시키는 일련의 Task 또는 Sub-Process 등을 집합으로 원자재 또는 중간 제품 등의 입력과 처리 및 가치가 증가된 상품 또는 서비스의 출력으로 구성되며, 비용을 동반함

DW/BI 의 Value Chain

제품 또는 상품의 가치를 증가 시키기 위해 일련의 Task로 구성된 과정



- 입력

- . 출력물이 제품인 경우 원자재 또는 중간 자재가 됨
- . 서비스인 경우 인력 또는 재화

- Process

- . 입력물에 대해 가치를 증가시키는 과정
- . 기업 자산이 활용 / 비용이 동반
- . 예 : 제품의 가공 또는 조립 대출 및 회수

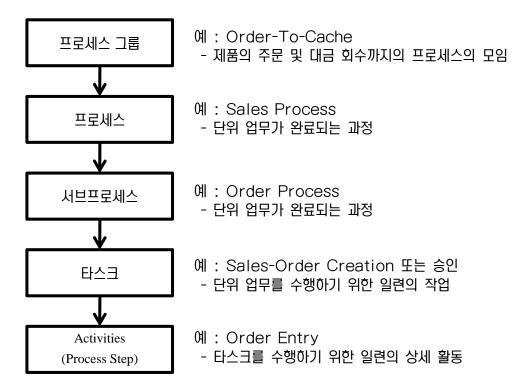
- 출력

- . 가치가 증가된 제품 또는 서비스
- . 예 : 완제품 그의 사표 [[

금융 상품 또는 금융 서비스

DW/BI 의 Process Circle

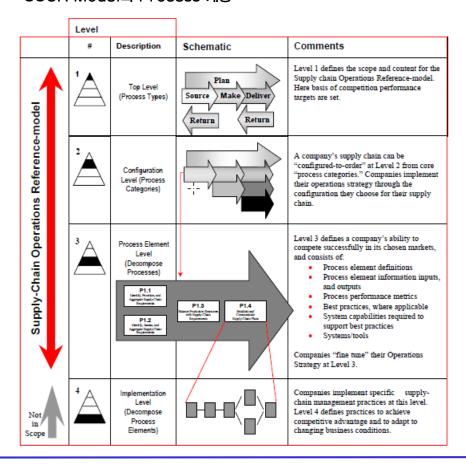
기업이 수행하는 Process의 모임을 Process Group이라고 하며, Process는 Sub-Process로 구성되며 최종으로는 Activities로 이루 어 짐.



원자재 및 재화의 가치를 증가 시키는 프로세스를 Value-Chain이라고도 하며, 이 과정에 동반된 비용을 측정하고, 각 프로세스의 효율성, 적시성 등을 평가하는 것이 Process Performance Management라고 함.

Process 例从

SCOR Model의 Process 계층



Process 별 비용 예시

Land A Marine	Performance Attributes				
Level 1 Metrics	Customer-Facing			Internal-Facing	
	Reliability	Responsiveness	Flexibility	Costs	Assets
Perfect Order Fulfillment	X				
Order Fulfillment Cycle Time		X			
Upside Supply Chain Flexibility		Т	X		
Upside Supply Chain Adaptability			X		
Downside Supply Chain Adaptability			X		
Supply Chain Management Cost				х	
Cost of Goods Sold				x	
Cash-To-Cash Cycle Time					X
Return on Supply Chain Fixed Assets					X
Return on Working Capital					X

주요 평가 관점

- 신뢰성 : 프로세스의 요구 사항을 완벽하게 처리 했는지…

- 책임성 : 프로세스의 요구 조건을 충족 시켜 주었는가

- 유연성 : 비용 절감 및 효율적인 처리를 위한 유성성 평가

- 비용 : 관련된 비용

- 자산 : 사용하는 자산

1.2 기업과 의사결정 - 확장과 복잡성

60년대 말 컴퓨터가 처음 발명된 이후로 Business를 지원하기 위해서 데이터의 저장 및 출력과 주문 입력, 전표 입력 및 입출금 등의 단순 반복적인 업무의 처리를(계정계, 운영계, OLPT) 시작으로 다양한 업무에 적용하였으며, 업무 처리과정에서 누적된 데이터 기반으로 보고서를 작성하거나, 분석하여 의사결정(OLAP, BI, Data Warehouse) 을 하거나 지원하는 영역으로 발전하였음.

구멍가계(마트)

업종: 유통업

주요고객: 동네의 주민

주요프로세스 : 상품 구매, 판매 및 재고 관리와 손익 분석

성공요인 : 저비용 상품 구매, 주민과의 관계, 적절한 재고 관리 및 판

매 상품의 효과적인 선택...



점포가 하나 인 경우

업무 형태

- 주인 and/or 판매 직원이 제품의 구매, 판매 및 재고 관리를 관리 IT 적용
- 판매 제품 종류가 많아지고, 재고가 증가함에 따라 판매 관리를 위한 IT 도입 -> POS 또는 판매 관리 시스템
- 단골 및 고객의 관리는 주인의 기억 속에 있음
- 구매처에 대한 관리 및 평가도 주인의 기억 속에 있음.

지점 개설에 따른 점포 확장

업무 형태

- 주인과 약간 명의 직원이 제품의 구매, 판매 및 재고 관리를 관리 IT 적용
- 구매 단가를 낮추기 위해 통합 구매의 필요성 대두 -> 구매 및 업체 관리 시스템 도입
- 통합 재고 관리의 필요성 대두-> 재고 관리 시스템 도입
 -> 판매 및 구매 관리 시스템과의 연계 필요 -> EAI 도입 또는
 Interface 프로그램 개발 -> 시스템의 복잡성 증가



경쟁이 심화되는 단계

업무 형태

- 주인 and/or 판매 직원이 제품의 구매, 판매 및 재고 관리를 관리 IT 적용
- 판매 제품 종류가 많아지고, 재고가 증가함에 따라 판매 관리를 위한 IT 도입 -> POS 또는 판매 관리 시스템
- 단골 및 고객의 관리는 주인의 기억 속에 있음
- 구매처에 대한 관리 및 평가도 주인의 기억 속에 있음.

1.2 기업과 의사결정 - 운영과 의사결정 차이

운영환경은 활동(또는 업무, 예 : 은행의 입출금, 고객의 주문 처리, 출고 등..)에 필요한 최소한의 정보를 가지고 반복적으로 수행을 지원하는 것으로 활동과 활동 또는 업무와 업무의 연관관계가 적으며, 의사결정 환경은 활동 또는 업무의 중요한 사안(예 : 재고의 효율적관리, 점포의 개설 등)에 대하여 과거의 경험(Where ?)과 외부 정보 등을 이용하여 최적의 선택을 지원하는 것임.

		운영(Transaction) 환경	의사 결정 환경
		• 운영을 위해 매일 매일 발생되는 Biz 활동 • 분석 기반 없는 의사 결정	의사결정을 위한 분석 기반의 정보를 생성하고 활용하는 활동 분석 기반의 의사 결정
		• 구매, 생산 및 판매와 배송	• 자원 계획, 자금 계획, 마케팅 계획
	Business	ල 0 ව	의사결정 지향의 분석
복	BC5000 양사육	• 우유,털과 고기를 위한 양 사육	가축의 확대와 축소 정착과 이동
잡	15C Venice	• 직물과 향료의 구매와 판매	• 가격, 고급업자 및 생산 라인
도	19세기 철도	• 상품과 사람의 이동	• Track의 위치, 노동의 소스
증 가	20세기 은행	• 수신과 여신	• 이자율 • 목표 고객 및 산업 • Portfolio 분석
	20세기의 소매업	소비재 상품의 구매와 판매재고운송	재고의 보유량 결정구매량 결정판매 추이

1.3 의사결정 지원 정보기술(1/2) - IT의 도입과 의사 결정

초기 업무 처리를 위해 도입된 시스템에 데이터가 축척 됨에 따라 이를 이용하여 경영에 필요한 모든 활동(결과 및 원인 분석, 대안 열거 및 의사 결정···)을 IT를 이용하여 처리하고자 하는 시도가 있었으며, IT의 지원 한계에 맞추어 발전하였음.

점

TPS & Data 관리

- Data 관리
- 단순 업무 처리 (Transaction)
- 관리된 데이터를 기반으로 통계 정보 제공

• 의사결정에 필요한 제한적인 정보 제공

MIS

- 경영에 필요한 모든 정보 및 활동 (의사 결정 등)에 IT를 적용 시도
- 인공지능을 이용한 IT기반 의사결정 시도

• 의사결정을 위해 필요한 외부의 상황 경험 및 추론 등의 기능 제공 한계

DSS

- IT에 저장된 데이터를 바탕으로 의사결정을 지원하는 시스템
- 제한된 추론 기능 제공
- 의사결정 대안 제시 시도
- 의사 결정과 운영의 분리 시작
- 추론 기능과 외부 환경을 고려한 대안 제시의 한계

EIS & OLAP

- 경영진에게 의사결정 에 필요한 정보를 제공
- 직관적인 문제점 인지
- 분석을 위한 다양한 정보 및 기능 제공
- 의사 결정과 운영의 분리 가속화
- EUC의 정착
- 결과 중심의 재무 정보 기반
- 전략과 연계된 지표 부족

Performance & Risk
Management

- TPS(Transaction Process System)
- MIS(Management Information System)
- DSS(Decision Support System)
- EiS (Executive Information System)

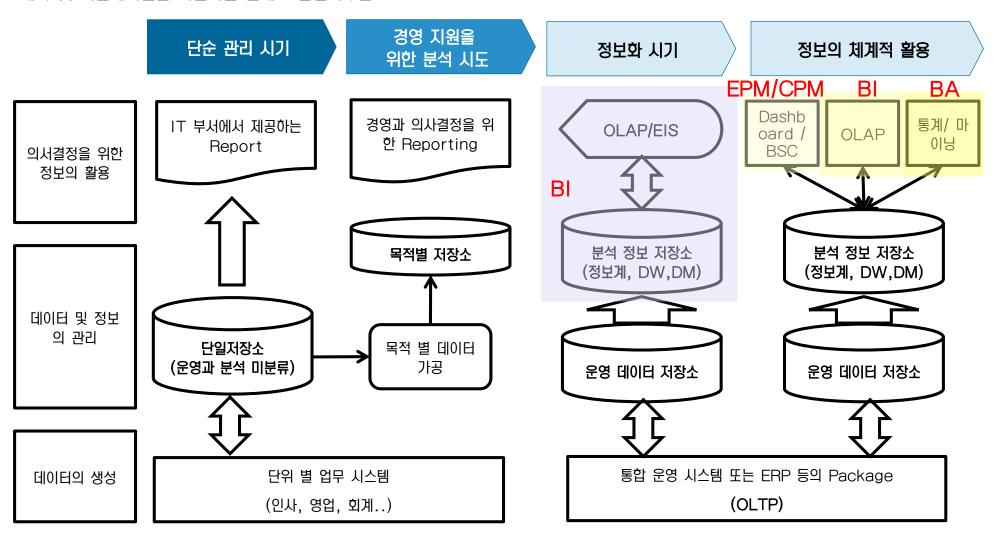
1.3 의사결정 지원 정보기술(2/4) - 시스템 환경 차이

정보 기술(초기 컴퓨터)가 업무에 적용되어 데이터가 축척됨으로써, 이를 기반으로한 최적의 의사결정을 위한 정보의 분석 요구가 증가되었으며, 분석 및 레포팅을 효율적으로 지원하기 위해 어플리케이션 (보고서 관리 시스템, 보고서 추출 프로그램 자동화 시스템)의 도입 및 개발을 하였음. 그러나 어플리케이션으로 는 운영환경과 의사결정을 위한 데이터의 관리 및 처리 패턴이 서로 상이 하여 그 한계가 있었음.

	운영 시스템환경	의사결정지원 시스템환경
보관자료	•상세 및 업무 운영에 필요한 자료 •최소한의 이력을 갖는 현재 자료	•집계자료 및 상세 자료 •과거자료
변경주기	•Biz Event 및 업무 처리에 따라 변경	•일정시점에 고정적 •시계열성
자료통합	•업무간의 자료 연결이 최소	•업무간의 자료 연결
데이터모델	•처리 성능 향상을 위한 모델	•조회성능 향상을 위한 모델
자료접근	•수시로 변경/삽입/삭제 발생	•Read Only
보관기간	•60일 ~ 1년	•2 ~ 7년
자료일관성	•보고서 추출 시점에 따라 상이	•보고서 추출 시점과 무관

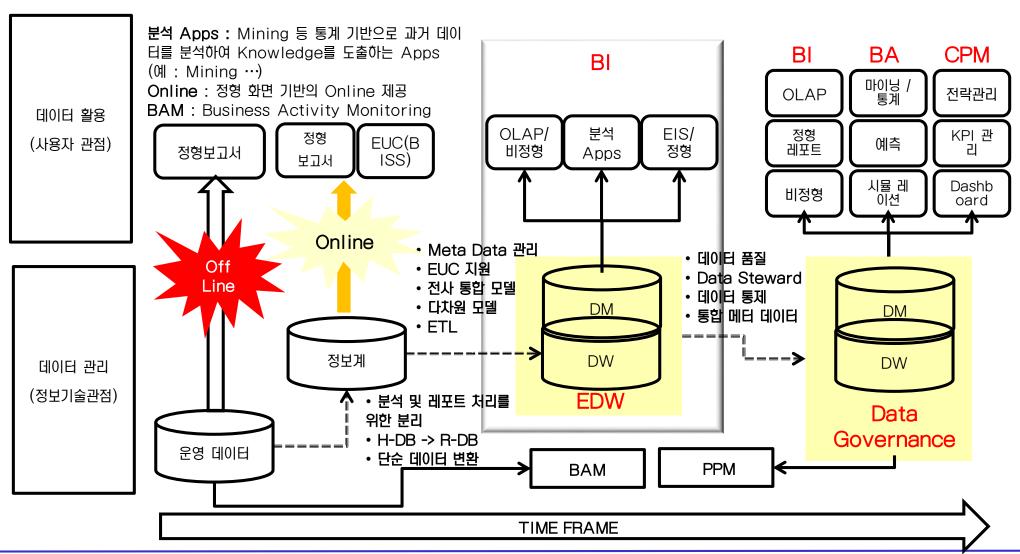
1.3 의사결정 지원 정보기술(3/4)- Architecture의 변화

S/W와 H/W의 발전과 더블어 단순 업무 처리에서 부터 의사 결정을 위한 정보 제공을 거쳐 다양한 통계 기법(Mining 등을)을 이용한 예측 및 시뮬레이션을 지원하는 단계로 발전하였음.



1.3 의사결정 지원 정보기술(4/4)- 주요 용어

초기에 IT 중심의 시스템 전개에 따라 관련 솔루션 및 용어가 정의 되었으며, 정보 사용자 관점의 활용 측면이 강조되어 BI와 BA 솔루션이 분리 발전 되었음.

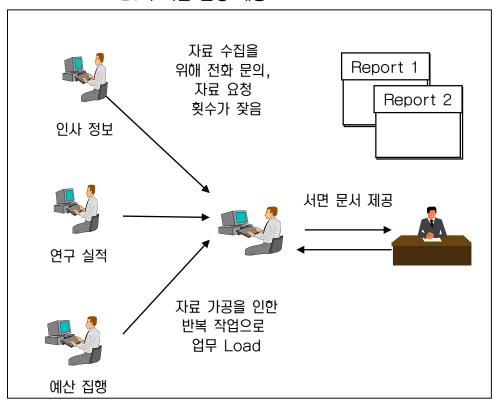


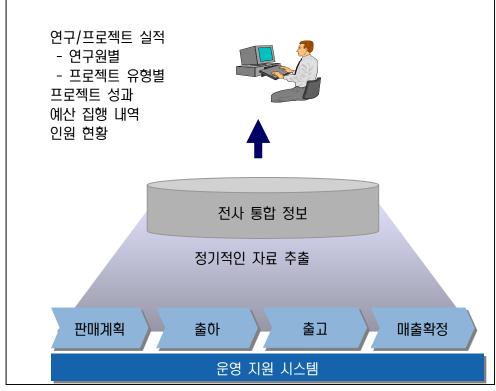
1.4 End User Computing(EUC) 와 DW (1/6)— EUC 정의

보고서의 작성 및 요구가 증가 됨에 따라 보고서의 개발 및 유지 보수가 IT의 큰 비용을 차지하고, 개발에 따른 보고서의 제공 지연을 해소하기 위해 사용자가 직접 보고서를 생성 및 관리하는 시도가 RDBMS의 출현으로 시도되었으나, 이를 지원하는 IT의 제약사항과 개발 방법론의 한계가 있어 초반에 활성화 되지는 안았음.(BISS(*)와 동일)

• 정의

- 정보 사용자가 필요한 정보를 IT 부서의 도움 없이 직접 접근하여 가공 및 분석하는 활동
- BI의 기반 환경 제공





(*) BISS (Business Intelligence Self Service) : EUC와 동일한 개념

1.4 End User Computing(EUC)와 DW (2/6)—DW 이전의 EUC

초기의 EUC 환경은 극히 일부 부서 또는 특정업무를 상대로 어플리케이션의 개발 및 도입 위주로 진행되었으며, RDBMS의 도입으로 범용적인 EUC 도입이 시도 되었으나 기술적 한계가 있어 성공하지는 못했음.

	초기 EUC 환경	
시기	1970년대 ~ 1980년 초반 대	
정보기반기술	• Mainframe 기반 • File/H-DB에 자료 저장 • Program언어 사용	
IT 부서	• 보고서의 작성 및 제공	
사용자 측면	• 필요 정보 IT 부서에 요청 • Sheet형태의 보고서 분석	
• 사용자의 자료 접근이 불가능 • 일부 Vendor에 의존적인 도구현		

정보 Center 시기		
80초반 ~ 80년말		
 Mainframe 가반 PC의 출현 R-DB의 상용화 SQL언어 사용(4GL) 		
보고서 작성 및 제공목적별 데이터 생성 및 제공원시 형태의 정보 접근 도구 지원		
• SQL언어를 사용하여 정보 접근 • Spread sheet 류의 PC S/W사 용		
• PC의 출현 및 사용 • 사용자를 위한 IT 기술 태동		

Client/Server 환경
90년 초반 ~ 90년 중반
 Client/Server 시스템 출현 분산 처리 기술 활용 R-DB의 활용 대중화 MRP 등의 Package 도입 시작
• 복잡한 운영 시스템 관리 • 보고서 작성의 어려움 가중 • 사용자 정보 접근 도구 지원
 다양한 정보 접근 도구 활용 SQL의 기본 지식 불필요 강력한 분석 S/W와 연계(Excel)
 PC의 성능 향상 강력한 분석 도구의 활용 많은 Vendor의 출현 Vendor간의 표준화 시도

1.4 End User Computing(EUC)와 DW (3/6)— DW 출현

최적의 의사결정을 위해서는 대량의 정보 분석, 기업 내부 기억(데이터) 및 외부의 정보가 필요하며, 이는 운영 시스템(OLTP)의 데이터 처리와는 다른 패턴을 가지고 있어, 다른 구조의 데이터 저장 영역이 필요하였음. 이 요구를 충족 시키기 위해 데이터 웨어하우스가 출현하였음.

정보소스로서의 한계

- → 업무처리를 위한 시스템
- → 복잡한 자료 구조
- → 업무간의 정보 연결
- → 보고서 개발이 필요
- → 대용량 자료 저장

기타 시스템과의 통합

→ ERP: 38 %

→ Non ERP: 30 %

→ 기업외부데이터 : 32%

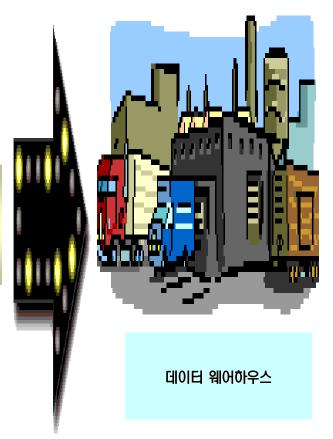
Extended ERP의 대두

- → 분석 어플리케이션을 위한 정보 저 장소가 필요
- → CRM(고객관리)
- → SCM(공급망관리)
- → SEM(전략적기업관리)

***** · · ·



통합된 데이터 저장소 필요



1.4 End User Computing(EUC)와 DW (4/6)-DW의 특징

- ✓ 주제 중심 별 관리
 - 고객, 손익 및 재고 등.
- ✓ 상세 및 요약정보
 - 관리자 및 사용자 요건 중심의 요약
 - 상세 분석 지원
- ✓ 과거이력 및 과거 특정시기의 정보
 - 최소 3년 이상의 과거 데이터
 - 법적 규제 등의 데이터 등…
- ✓ 단위처리 자료량이 많음
 - 대량의 데이터 적재 및 조회
- ✓ 메터 데이터 관리가 필수
 - 사용자, 개발자 및 운영 메터 데이터
- ✓ Single Of Truth
 - 매출액에 대해서는 상세 및 요약 수준에서도 동일한 값 제공

1.4 End User Computing(EUC)와 DW (4/6)-DW의 특징

Metadata는 데이터의 데이터로 DW에서는 필수적으로 관리가 되며, Big Data를 포함한 Logical DW 에는 그 중요성이 더욱 강조됨.

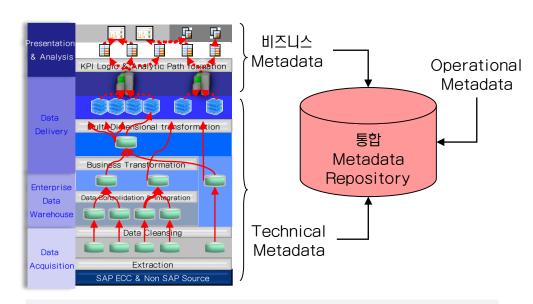
메터 데이터 관리 목적

- 조직 내에 관리하고 있는 데이터 또는 정보에 대한 데이터에 대한 정의 및 관리
- 데이터 항목에 대한 일관된 정의 제공(Data Dictionary)
- 데이터의 생성 및 변경과 관리에 대한 정보 제공
- 변경 이력 및 추적성 제공

메터 데이터 유형

대상	표준화 방안	생성Task
비즈니스 Metadata	 사용자 관점에서의 데이터에 대한 정보 제공 유형: 정의, 출력 방법, Domain Value, 논리적 속성 	데이터 표준화 전사 데이터 Architecture 정의
Technical Metadata	 IT 관점의 데이터에 대한 정보 제공 대상: DB 물리적 속성, 변환 절차, 저장 Table, Code 값 등 	EDW 설계, ETL 설계
Operation Meta Data	 데이터 활용 및 관리에 대한 정보 제공 대상: 데이터 조회 건수, 변경 및 생성 일시, Loading 정보 등 	질의 실행, 데이터 적재 등의 DW 활용

메타데이터 통합 관리



메타데이터 통합 관리 장점

- 데이터 변경 시 쉽게 변경의 영향을 파악 추적할 수 있음
- 통합 메타 데이터에 기반한 시스템 구축으로 시스템과 메타데이터 의 일치성 확보
- 통합 메타데이터를 통하여 전사 사용자가 동일한 데이터에 동일한 접근

1.4 End User Computing(EUC)와 DW (5/6)—DW 발전과정

초기 데이터 웨어하우스는 금융 및 소매업(국내는 은행권이 최초로 구축함) 등의 대량의 운영 업무 처리로 인해 발생된 데이터에 대한 리포트 및 분석 요구를 충족 시키기 위해 **정보계** 의 명명으로 시작되었으며, 1990년 도 중반 데이터 웨어하우스의 정의, 구성 요소 및 방법론 등의 정의되었음.

원시 DW

- ▶ Platform과 DBMS 중심
- ▶ 여러 OLTP로부터 관계형 DB로 변환 저장
- ▶ 관계형 DB로부터 보고서 생성
- ▶ 운영환경과 정보지원환경의 분리
- ▶ 예:정보계 시스템

1세대 DW

- ▶ Metadata, ETL 등 DW 정의 확립
- ▶ ETT(추출,변환) 및 OLAP Tools 중심
- ▶ 조회 및 분석 Tools을 이용한 현 황 분석 및 문제점 파악
- ▶ OLAP 중심
 - ROLAP
 - MOLAP
 - DOLAP
 - HOLAP
 - _ ..
- ▶ Data Mart 중심
- ▶ EDW 기반 조성
- ▶ EUC를 위한 기반 마련

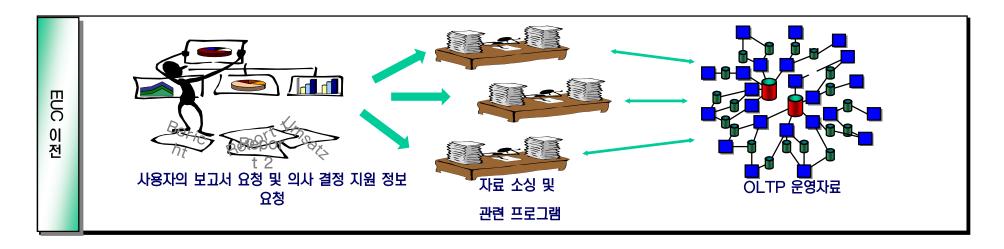
2세대 DW

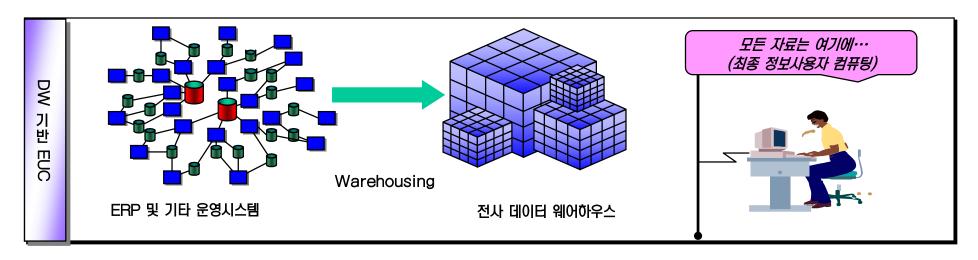
- ▶ 기업내 정보 흐름의 중심
- ▶ 다양한 분석 도구를 이용한 의사 결정 지원 및 OLTP와의 연동
 - 고객관리(CRM)
 - 공급망 관리(SCM)
 - 데이터 마이닝
 - 전략적 기업 관리(SEM)
 - 지식 관리/경영
- ▶ 신기술과의 접목
 - Web 기반 기술
 - e-Commerce
 - WAP•••
- ▶ 정보의 공급망 형성
 - 전사 모델의 체계적 관리



1.4 End User Computing(EUC)와 DW (6/6)

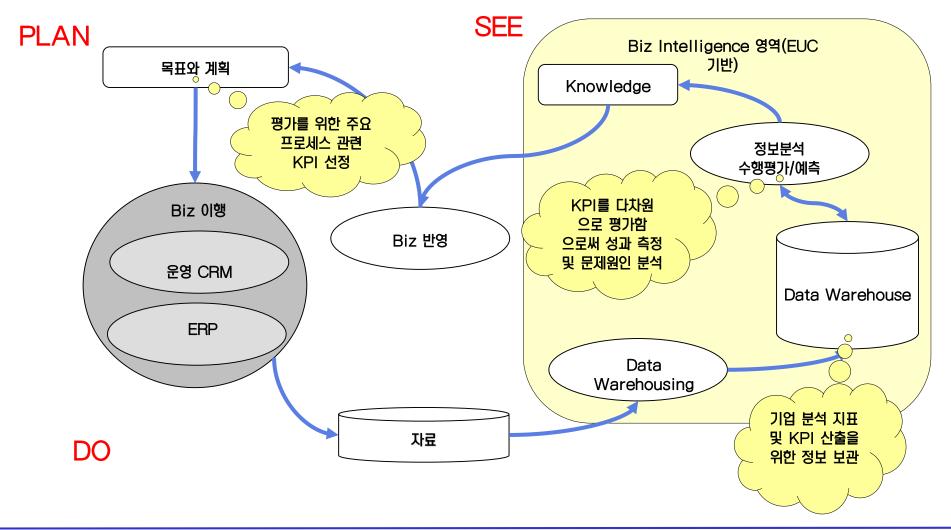
데이터 웨어하우스의 도입과 사용자 도구의 발전으로 사용자가 쉽게 정보에 접근하여, 조회 및 분석 할 수 있는 환경을 마련되어 기업의 EUC 환경 구축의 기반 환경을 제공함.





1.5 Business Intelligence (1/4) - 초기 정의

DW 기반의 EUC 환경은 정형 및 비정형 레포트를 제공하거나 사용자가 필요한 경우 데이터에 접근하여 조회하는 Pull 유형의 정보 활용 패턴이나, 적절한 사람에게 적절한 데이터를 적절한 시점에 적절한 매체를 통하여 제공하는 Push 패턴의 정보 제공과 다양한 분석 App. 의 발전으로 의 BI 환경이 DW 기반에 전개되었음.



1.5 Business Intelligence (2/4) - 협의 및 광의 정의 1.개요

DW는 정보 기술에서 출발하여 사용자에게 최적의 정보 접근 및 환경을 제공하여 주는 솔루션이나, BI는 DW 환경하에 사용자 관점에서 주요 정보와 마이닝 또는 기타 분석 어플리케이션에서 발생된 예외 사항 등을 사용자에게 능동적으로 제공하여 즉시 대응 할 수 있는 환경을 제공하는 솔루션 또는 환경임.

일반적 정의

- 적절한 사람에게 적절한 데이터를 적절한 매체를 통하여 적절한 시간에 제공해 주는 환경

광의의 BI

협의

사용자 관점의 정의

정의

- 기업의 계획의 수립, 예측, 업무의 의사 결정을 위해 적절한 사람에게 필요한 모든 정보를 적시에 적절한 매체로 제공하는 시스템
- 사용자에게 정확하고 문맥적인 정보를 적절한 시간에 제공하여 사실기반의 의사결정을 지원

기능

- Information의 수집
- Information을 Knowledge로 전환
- Knowledge를 Action으로 전환(Wisdom화)
- Action을 Profit 전환

정보 기술 관점의 정의

정의

- 사용자에게 필요한 정보 및 지식을 전달하기 위해 필요한 도구 관련 제품들…

예

- 데이터 웨어하우스
- OLAP 도구
- Dashboard 및 관련 BI 도구
- Data Mining 도구
- 관련 데이터 베이스들…

2000년도 중반부터 시작 되어 기존의 BI로 부터 특징된 질의 도구 중심의 기능 및 솔루션 등을 넘어서 정보 기술의 발전에 기인한 Web 2.0, SOA 및 실시간 데이터 분석 및 실시간 Activities 분석을 통한 예외 및 경고 제공을 업무 환경에 제공해 주는 새로운 BI 환경임

동기

정보 기술의 발전

- Appl. 유연성, 구성 및 민첩성을 제공하는 SOA의 활용의 증가
- XBRL(Extensible Business Report Language) 등 의 표준 기반의 데이터 교환
- Semantic(*) Web 기반 을 통한 외부 데이터의 내부 활용
- SOA 기반의 정보기술 도입 및 전개
 - . Event Driven / Process Intelligence
- Text Mining
- Semantic Search

업무적인 필요성

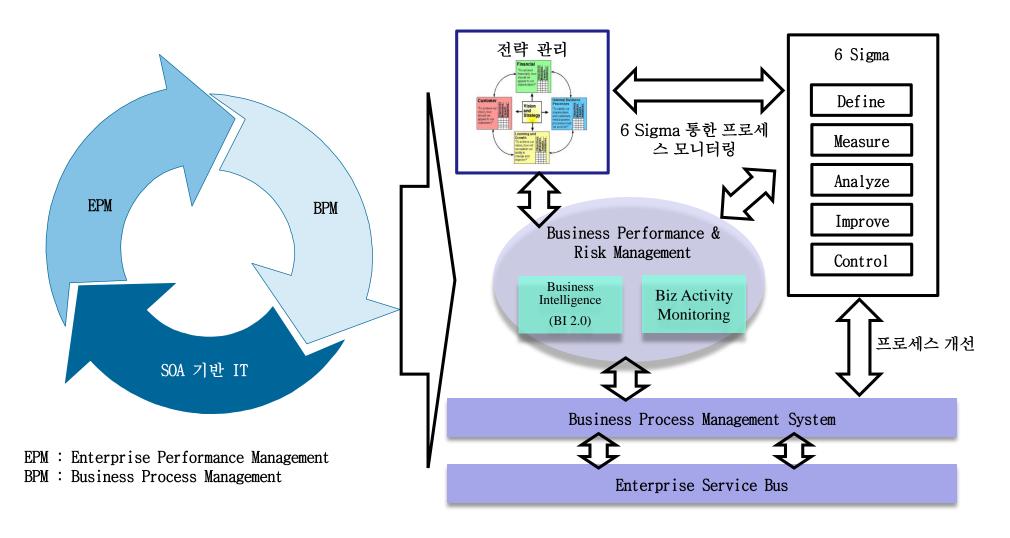
- 결과 중심의 분석 및 대책보다는 사전 예방 중심의 모니터링 지향
- 빠른 의사 결정(일 단위 -> 수분 또는 수시간 단위)
- 정보의 홍수 속에서 검색 시간 단축

특징

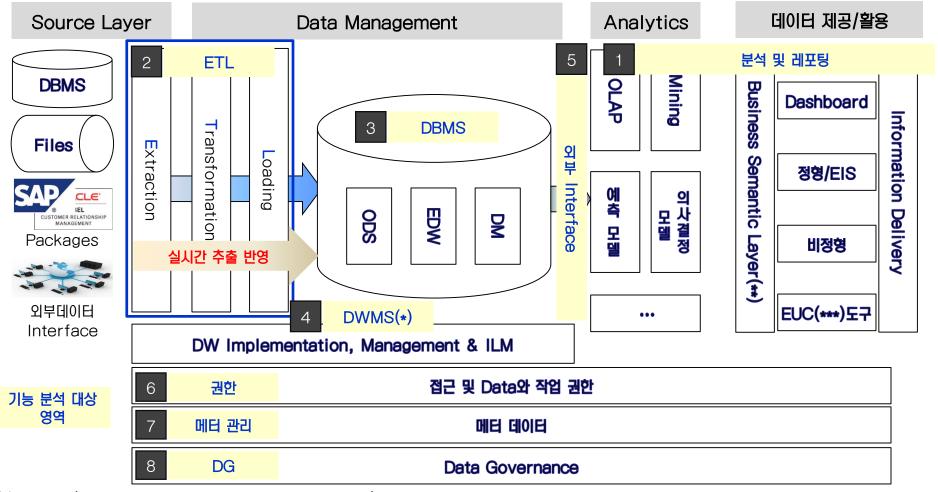
- Event Driven
- 실시간 <-> Right time
- 자동 분석
 - . 실시간에 발생된 데이터를 수집하여 과거의 데이터와 비교 하거나 분석 Appl. 적용함으로써 미래의 성과 및 문제를 예 상
- 예측
 - . Event를 분석함으로써 기대하는 것들에 대한 예상
 - . 예 : 제품이 적시에 도달 할 수 있는가?
- Process 중심
 - . 업무 프로세스의 효율적이고 효과적으로 처리 지원
- 확장성
 - . Event는 예측 불가능으로 확장성이 있어야함.

1.5 Business Intelligence (4/4) – BI 2.0 Use Cases

SOA 기반의 BI 2.0은 실시간 지향적인 사전 예방 중심의 관리를 지원하며, 수행한 프로세스의 데이터를 분석하여 프로세스 성과 관리 등에 수행 하며, 이를 통하여 기업의 전략 관리 및 전략 수행의 모니링을 할 수 있는 환경을 제공함.



1.6 DW와 BI Architecture(1/2) - 개요

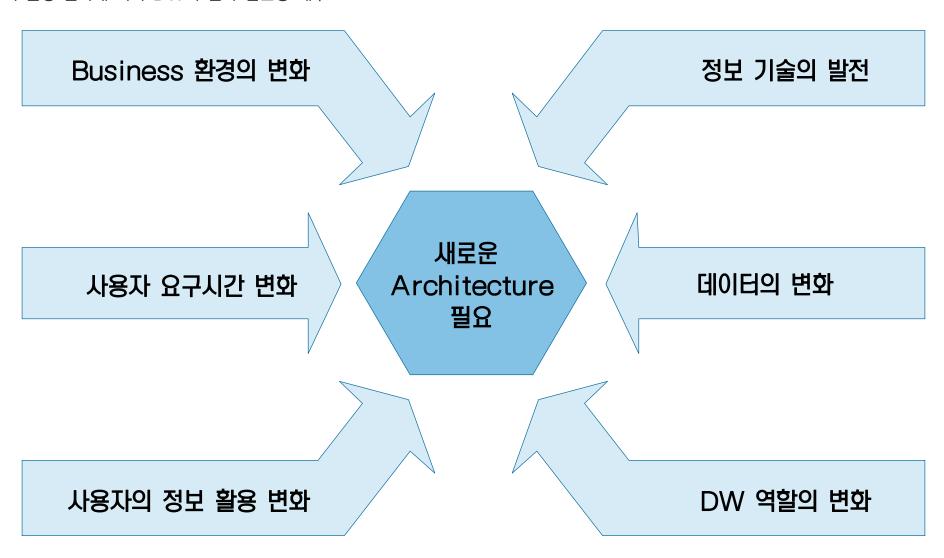


- (*) DWMS(Data Warehouse Management System) : 설계된 DW의 구현, 데이터 관리 및 실행 관리 및 모니터링을 제공하는 시스템 (**) Business Semantic Layer : 사용자를 위한 정보 구조를 제공하는 Layer로 물리적인 구조와의 독립성과 유지 보수의 용이성 제공
- (***) EUC(End User Computing): Business Semantic Layer 기반에 사용자가 직접 데이터를 조회하고 분석할 수 있도록 하는 도구

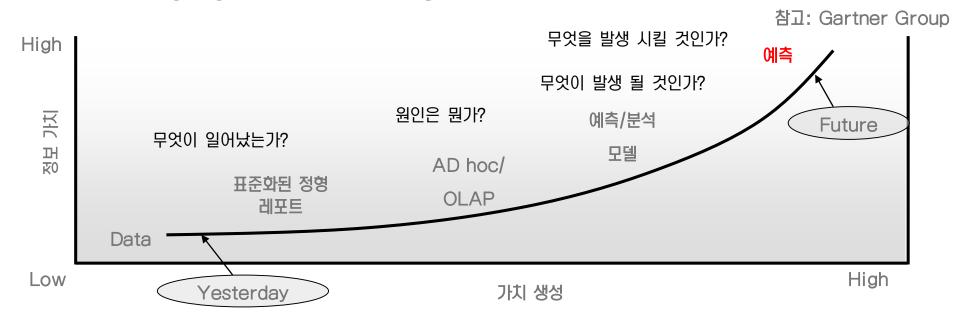
1.6 DW와 BI Architecture(2/2)-영역별 상세

영역	내용	주 필요 기능
1) 레포팅 및 분석	- 정보 전달 기능 - Visualization - 정보를 분석하는 Application 및 OLAP - 사용자 관점의 구조 제공	- Mobile 및 Web 등을 통한 정보 전달 기능 - 정형 및 비정형 작성 및 제공 - 문서 작성 S/W와의 연계 - 사용자 View 작성 및 관리 기능
2) ETL	- 데이터를 추출 - 추출한 데이터의 변환 - 변환된 데이터 적재	- 변경 데이터 추출 기능 - 변환 구현 방식 - 적재 방식
3) DBMS	- DW의 특성에 맞는 데이터 관리	- 대용량 데이터 관리 - 최적의 적재 / 조회 성능
4) DWMS	- EDW 및 DW 구현 - 데이터 관리 - 실행 관리 및 모니터링	- DW 모델 구현 방식(메터 연계, 수작업…) - ILM(Information Lifecycle Management) - 구조 변경 관리 - 실행 스케줄링 기능
5) 외부 Interface	- OLAP 도구 및 Mining Appl.와의 연계 - 외부 마트 및 분석 Appl. 로의 데이터 제공	- 표준 Interface 제공 - 외부 데이터 제공 기능(Push/Pull)
6) 권한	- 시스템 접근 권한 및 SSO - 작업 실행/데이터 접근 권한	- 외부 인증과의 연계 - 사용자별 실행 및 데이터 접근 권한 제어
7) 메터 관리	- 전 영역에서의 메터 데이터 통합 관리	- DW 전영역에 자동 연계된 메터 데이터 관리 - 업무, 기술 및 운영 메터 데이터 관리 - 영향도 분석
8) Data Governance	- Data Governance에 따른 관리를 지원하기 위한 기능 등	- Data Stewaredship 관리 - 품질 및 Profiling 관리

외부의 환경 변화에 따라 DW의 변화 필요성 대두



데이터 기반의 분석 환경이 성숙됨에 따라 BI의 분석 방식이 변화 하였음.



과거의 BI

- 주로 DW(Data Infra)를 구축하는 IT적인 측면 강조
- ETL, DW, OLAP,
 Reporting, Data mining
- 필요한 BI Application은 자체 적으로 BI라 하지 않고 구축

현재의 BI

- 구축된 Data Infra를 활용하는
 Bl application 강조
- 분석CRM, ABC, BSC, CPM, 각 부문 별 application
- 표준화된 tool, solution

미래의 BI

- 모든 BI application 내에 Forecasting, planning, simulation 등 기능 탑재
- Big Data를 포함한 전사적인 통합 BI 강조
- Operational system에
 Auto-feed back

Big Data 기술 기반과 DW의 차이점

	빅데이터 기술	DW
데이터 유형	- 정형, 비정형(SNS, 사진, 동영상, 위치정보, 지리정보, 기타)	- 정형 데이터
데이터 특징	- 데이터의 통합 및 연계 수준 낮음 - 데이터의 품질 수준 다양 - 데이터의 정제 및 Profiling 필요	- 데이터의 통합 및 연계 수준 높음 - 데이터 품질 수준 높음
정보 활용	- 통계 및 시각화를 통한 예측 분석	- Dashboard, 정형 보고서 등. - 성과/OLAP 분석 및 모니터링
정보 접근	- NoSQL 및 MapReduce등의 - 제한적인 SQL	- SQL 기반
기본 기술	- Hadoop, NoSQL, 기타	RDB에 근간
서버 구조	- MPP 기반(Scale-Out 확장)	MPP 또는 SMP 기반 (SMP의 경우 성능 개선의 한계 있음)
서버	- 저비용 의 x86 machine 기반	고비용의 UNIX 선호 (Appliance는 x86 기반임)
소프트웨어	- 공개 또는 공개 기반의 상용 S/W	대부분 모두 고가 상용 소프트웨어
문제점	- 공개 S/W로 인한 지원 - 데이터의 안정성	가격이 비싸고, 데이터처리 양이 늘어나면 한계에 이른다

Big Data 시대 및 외부 환경의 변화에 따라 **DW의 역할이 변화**됨.

Real Time Enterprise(RTE)

- 사전 방지 및 대응을 위한 실시간 분석 및 모니터링 (Operational BI)
- 이를 통한 다양한 비즈니스 대응력 향상

- 데이터량의 증가, 다양한 분석
 작업, 시스템간의 인터페이스
 등으로 시스템 성능 이슈
- 데이터 증가에 따른 확장성
- 정보환경의 가용성
- TCO, IT 자원 통합





- 모바일 환경의 대두로 인한 소설 데이터 생성
- 비정형 데이터 분석을 통한 다양한 분석 정보 확보 필요
- 사물인터넷의 데이터 처리

DW Appliance



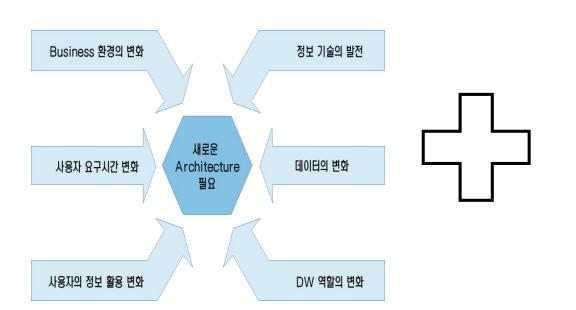
Big Data

1.7 Logical Data Warehouse - 정의

Big Data 시대 및 외부 환경의 변화에 따라 **DW의 역할이 변화**됨.

Logical Data Warehouse

(DV + 데이터 관리/활용 체계 + 분석 Platform)



- ´DW/BI의 체계적 관리 및 지원
- . Data Governance / 품질
- . MDM
- . 활용 통계 및 제어
- . Semantic 제어
- 분석 Platform
- Big Data 기술과의 상호 보완

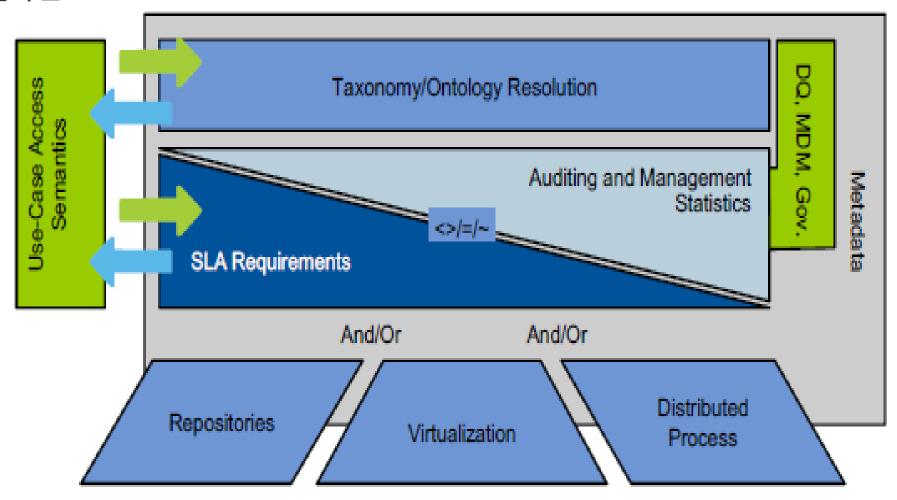
1.7 Logical Data Warehouse - 정의

2011년 Gartner가 발표

- 기존의 DW 구조의 진화
- 정보 관리 메터 데이터의 관리 및 활용 강화
- 3가지 정보 관리 방식
 - Data Virtualization
 - Distribution Process 과 동등 Partner로써의 Repository와의 Join
- 정보 관리 Platform
 - 데이터 품질
 - 마스터 데이터 관리
 - 데이터 Governance
 - 운영 및 감시를 위한 통계 기능
- 분석 BI, CPM 등의 모든 Use Case를 지원

1.7 Logical Data Warehouse - 정의

기본 구조



DQ = data quality; Gov. = governance; MDM = master data management

데이터 가상화 (Data Virtualization)

What Is Data Virtualization?

Wikipedia:

"Data virtualization is [..] an application to retrieve and manipulate data without requiring technical details about the data, such as how it is formatted or where it is physically located."

Or more simply:

A solution that sits in front of multiple data sources and allows them to be treated as a single SQL database

가상화 특징

Encapsulation (Information Hiding)

- 내부 구현의 상세한 감추고 외부에 필요한 속성을 분리하여 표현 예 : Window Objects

추상화(Abstraction)

- 특정한 현상의 상세를 감추고 필요한 항목만 제공하는 방식 예 : 배송에 필요한 고객정보 요구 시 고객 주소와 전화 번호만 제공

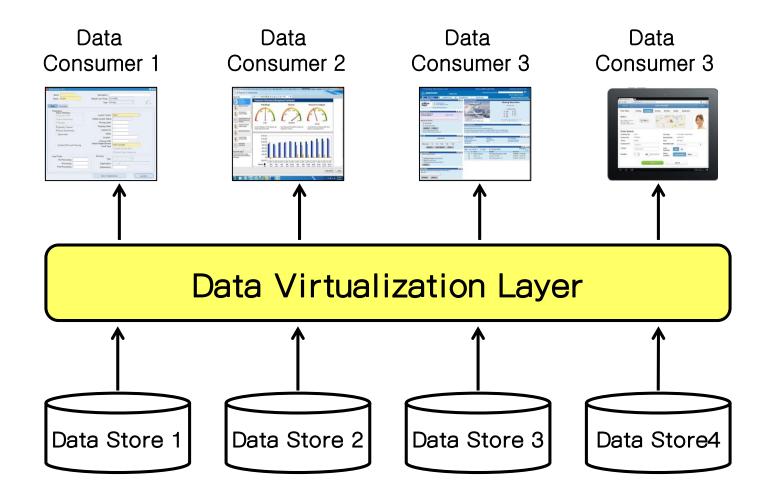
Data Integration

- 이종의 데이터 구조를 통합하여 단일 뷰로 제공하는 과정
- 많은 경우 DI는 DV 기반 하에 구성
- DV는 꼭 DI를 포함하지 않음

Data Federation

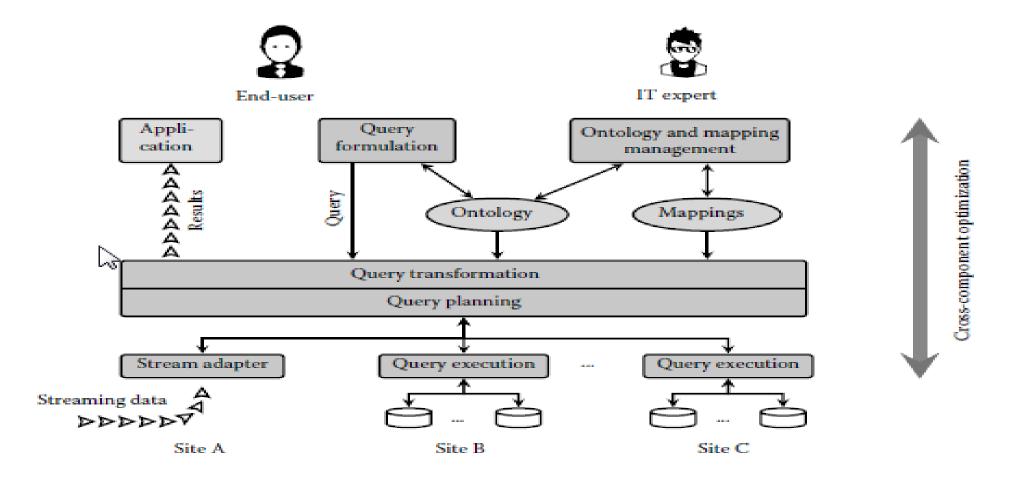
- 둘 이상의 데이터 소스를 통합하여 하나의 단일 뷰를 제공하는 방식
- Data Integration이 반드시 포함됨
- 모든 DF는 DV이나 모든 DV는 DF가 아님.

가상화 기본 구조



Taxonomy / Ontology

- 최종 사용자와 IT와 협의하여 문제 영역의 단어 구성 및 정의, 정의된 단어와의 Query 문장 Mapping을 통하여 사용자가 손쉽게 데이터를 접근 할 수 있는 환경을 제공



Repository Management

- Data mart 및 ODS를 위해 LDW가 실제 데이터를 저장하는 저장소를 관리
- 데이터의 변경 정도, 사용자의 SLA 및 사용 빈도 등을 고려하여, 물리적 관리 여부 결정

Data Virtualization

- Agile 개발과 유연성 등을 위해 반드시 제공
- 사용 통계 제공 필요

• 분산 처리

- 둘 이상의 데이터 저장소 또는 Cluster에 대한 복합 질의에 대한 분산 처리 예 : Hadoop MapReduce 포함
- 최적의 Path 정의 기능 필수

• Auditing 통계와 성능 평가 서비스

- 성능과 Audit과 관계된 통계 저장 및 관리
- 데이터 Profile 및 Repository 화 여부 결정 등에 사용

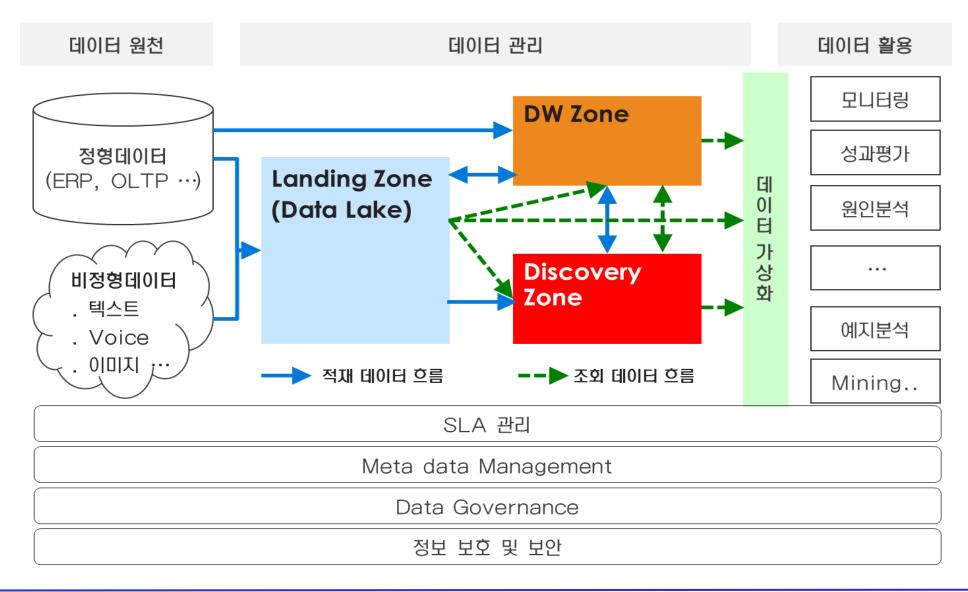
- SLA (Service Level Agreement) 관리
 - 연결된 모든 시스템, 사용자 및 Application 의 기대 수준에 대한 양적/질적 메터 관리
 - 운영/감시 통계를 위해 누적된 데이터와 기대 수준과의 비교하여 시스템의 구성을 수정 보완 . 예 : DV로 구성된 통합 구조가 사용자의 기대에 못 미치는 경우 Repository로 저장

Metadata Management

- LDW의 주요 구성 요소
- 모든 서비스에 의해 공유 및 재사용이 필수

예: MDM의 품질 정보와 DQ의 품질 Rule 정보 공유 DV의 변환 Logic을 Repository화 할 때 재사용

1.7 Logical Data Warehouse - Data Architecture



전통 분석 기법

통계 분석

- 공통 통계 함수 및 계산식을 이용하여 고성능 처리 및 분석
- Sample Data 기반

마켓팅 분석

- 마켓팅 결과를 최적화 하기 위해 고객과의 접촉 이력 분석

Segmentation

- Data Point의 자연적 Groping 발견

Data Transformation

- 향상된 분석을 위한 데이터 변환

새로운 분석 기법

Path 분석

- 사용자가 정해진 결과에 도달하는 과정의 Pattern 분석
- Chunk Path Pattern

Text 분석

- Text에서 Pattern을 분석하여 문제 및 기회 포착에 활용
- VoC를 분석하여 고객의 주요 불만 요인 분석 및 사전 대처

Graph 분석

- 고객, 조직 및 프로세스 간의 관계와 구조를 분석하고 이해
- 고객간의 자금 흐름 / 특정 제품와 Clustering

SQL M/R 시각화

- 분석의 결과를 시각화하여 일반 사용자가 이해하기 쉽게 제공

전통적 통계 기반의 분석 데이터의 Pattern 검색에는 그 한계가 있으며, 시각화 분석을 이용하여 상호 보완해야함.

통계 기반 분석

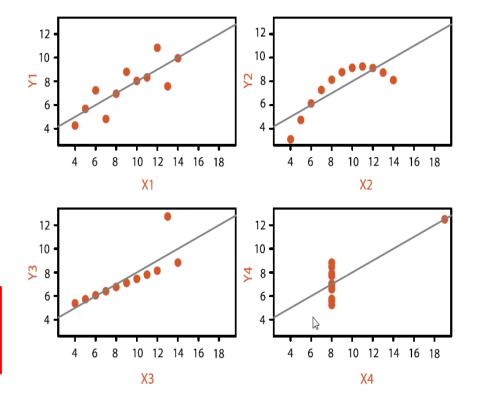
- 서로 다른 데이터 집합이나 모든 데이터 집합이 평균, 변차 및 표준 변차가 동일여 관계를 쉽게 찾을 수 없음.

	1		2		3		4	
	X	Υ	X	Υ	X	Y	X	Υ
	10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
	8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
	13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
	9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
	11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
	14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
	6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
	4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
	12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
	7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
	5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89
Mean	9.0	7.5	9.0	7.5	9.0	7.5	9.0	7.5
Variance	10.0	3.75	10.0	3.75	10.0	3.75	10.0	3.75
Correlation	0.816		0.816		0.816		0.816	

발췌: Anscombe's Quartet

시각화 기반 분석

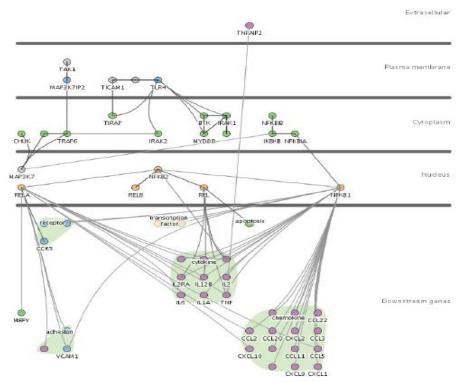
- Graph를 이용한 시각화 시 서로 다른 데이터 패턴이 보임



Graph 등을 이용한 수기 기반의 시각화는 표현 방식, 공간 및 시간에 그 한계가 있어 많이 사용되지는 않았으나, IT Display 기술의 발전으로 활용가치가 점점 높아짐.

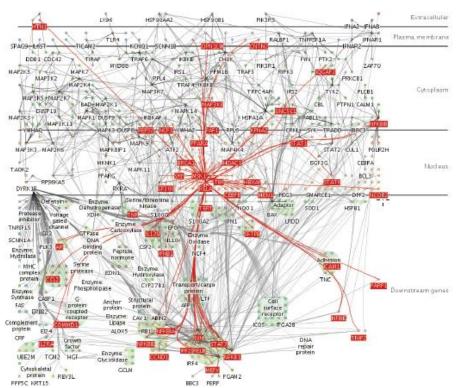
수기 기반 시각화 분석

- 수기 Diagram은 표현 가능한 데이터의 공각적인 한계가 있어 분석이 제한적임.



Computer 기반 시각화 분석

- 대량의 데이터를 **화면에 적절한 Graph로 출력하여 손쉬운** 데이터 탐색 및 Insight 발견을 지원함.



통계와 시각과 분석은 상호 보완 관계가 있음.



개요

- . 통계 기법을 이용한 분석 모델 생성
- . 샘플링, Training set, test set 등의 데이터 적용 필요

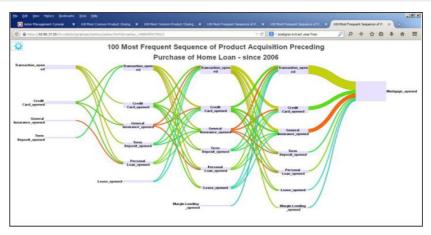
장점

. 수치화된 결과로 적용이 용이

단점

- . 통계적 지식 요구
- . 모델 변경 시 많은 시간 소요

시각화 기반 분석



개요

- . 지리적 정보를 이용한 분석
- . Graph/Event Path 등을 이용한 시각화 분석
- . Data의 Pattern을 탐색

장점

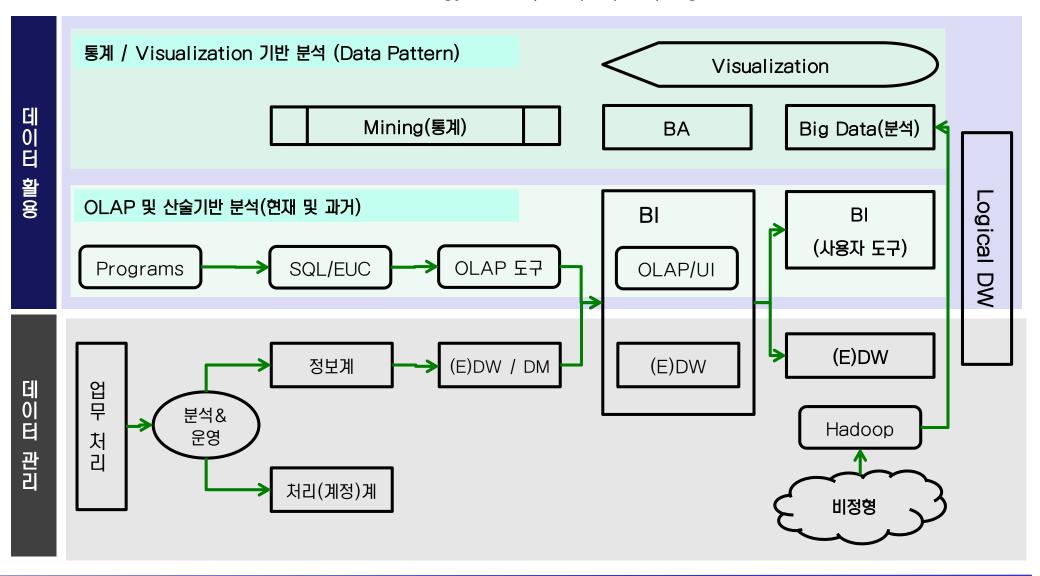
- . 통계적 지식이 없어도 Insight 도출 가능
- . 상호 대화식의 분석 가능 / 변경이 용이

단점

적용을 위해서는 통계 등등이 필요

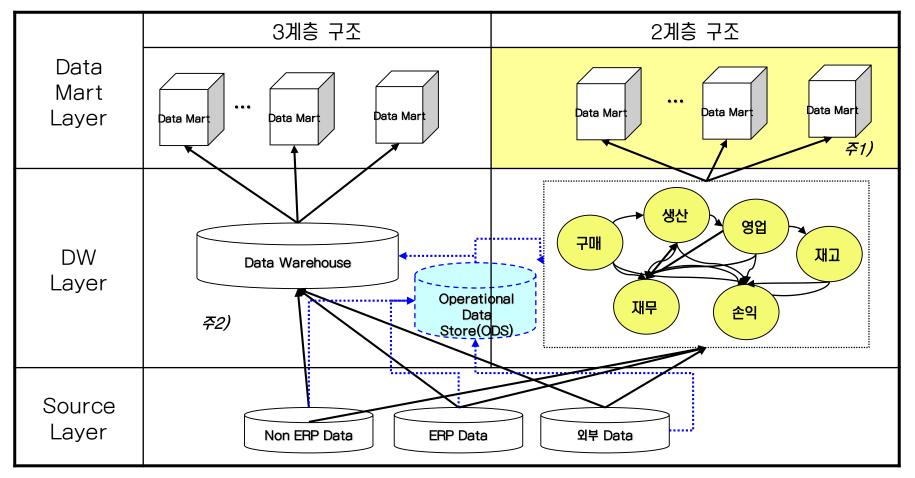
1.8 용어 및 개념의 변화 요약

의사 결정을 지원하는 Architecture의 Terminology는 정보계, DW, BI, BA, Big Data 및 Data Science로 변화하였음.



2.1 Data Architecture

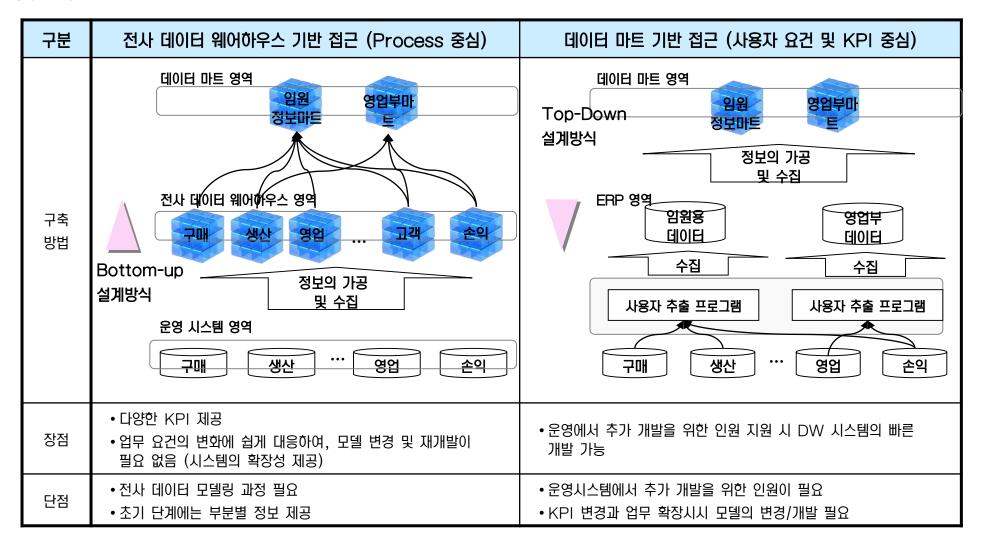
DW Architecture는 Bill Inmon이 주창하는 3계층과 Ralph Kimball이 주장하는 2계층 Architecture가 있으며, 대용량 및 Tool set을 이용한 In-house 개발에서는 3계층을 중소 용량 및 DW Package (SAP BI) 도입 시는 2계층을 추로 채택함.



- 주1) 2계층에서 DM는 선택적 영역이며 주로 분석 및 보고서 구현은 DW영역에서 일어남.
- 주2) 3계층 구조에서 DW 영역은 ER (Entity Relation) 모델 기법을 활용하며, 사용자의 직접 정보 접근은 허용하지 않는다. DM 영역은 MD (Multi Dimensional) 모델 기법을 활용한다.

2.2 설계 유형

EDW와 DM는 정보를 설계하는 방식의 차이이며 개발 기간 및 시스템의 유연성에 영향을 주며, 아래의 그림은 2계층 구조를 선택 하는 경우의 구조임..



2.3 개발 방법론(1/2)

EDW와 DM는 정보를 설계하는 방식의 차이이며 개발 기간 및 시스템의 유연성에 영향을 주며, 아래의 그림은 2계층 구조를 선택 하는 경우의 구조임..

Bottom-up 방법론

- D/W 구축 후 각 각의 Mart를 구축하는 방법론
- _ 장점
 - 전체 DW에 대한 데이터의 흐름을 일관되게 유지 할 수 있다.
- 단점
 - 개발 기간이 길다(2-5년)

Top Down 방법론

- 별도의 Mart를 구축 후 이를 중심으로 DW구성
- 장점
 - 개발 기간이 짧아(3-10개월) 사용자의 요구에 빠른 대응이 가능하다.
- 단점
 - 전체 DW를 위한 데이터의 흐름이 복잡해 진다.

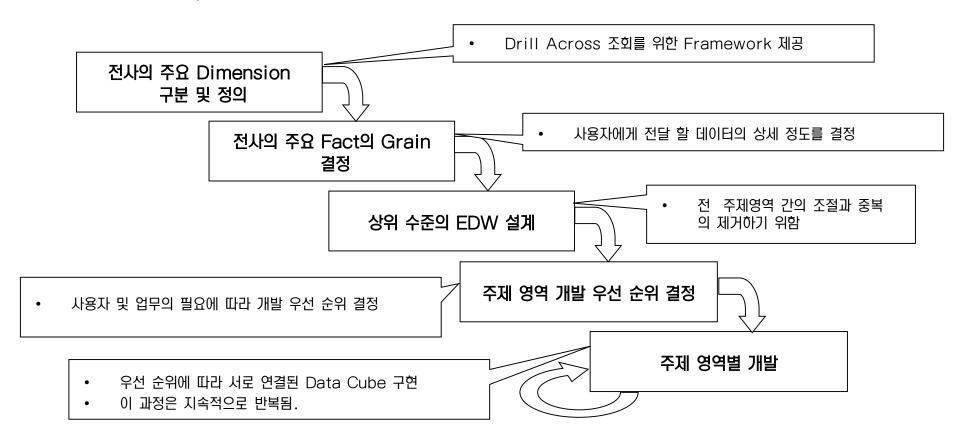
Mixed 방법론

- DW의 전사 모델은 기업 EDW를 위한 데이터 모델 구성 후 Data Mart별로 개발
 - 사용자 주요 요구 사항 및 주제 영역별 개발
- 장점
 - 데이터의 모델 및 자료의 흐름을 일관성을 유지 할 수 있다.
 - 사용자의 요구에 빠르게 대응 할 수 있다.

2.3 개발 방법론(2/2) - 설계 절차

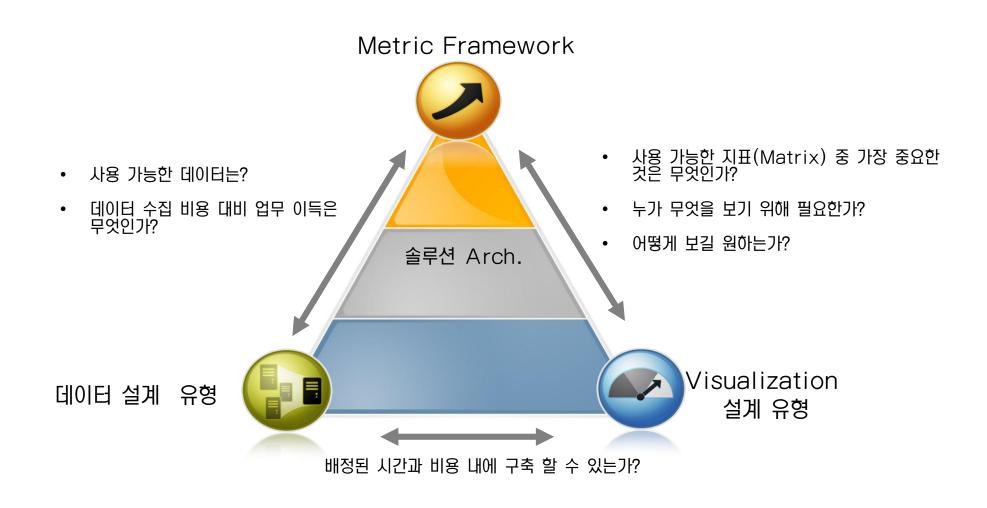
2계층 구조에서는 3계층 구조의 개발 방법론 중 Mixed 방법론을 활용한다.

Think Big, But Build Small



3.1 분석 Framework(2/4) - 주요 Components

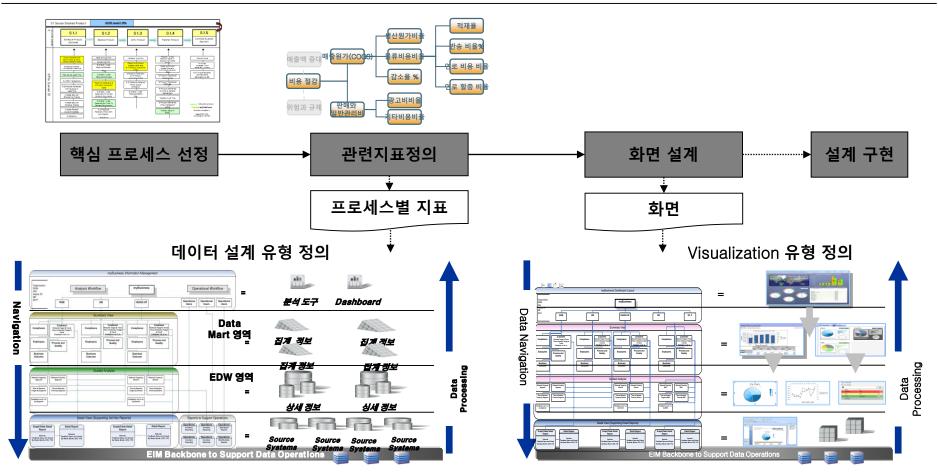
성과 관리의 체계적인 관리를 위해 체계적인 Data Warehouse의 설계 및 구축과 모니터링 및 분석 등의 지표를 제공하는 Metrix Framework과 직관적인 Visualization기반의 UI를 제공 해야함



3.1 분석 Framework(3/4) - 지표 기반 설계

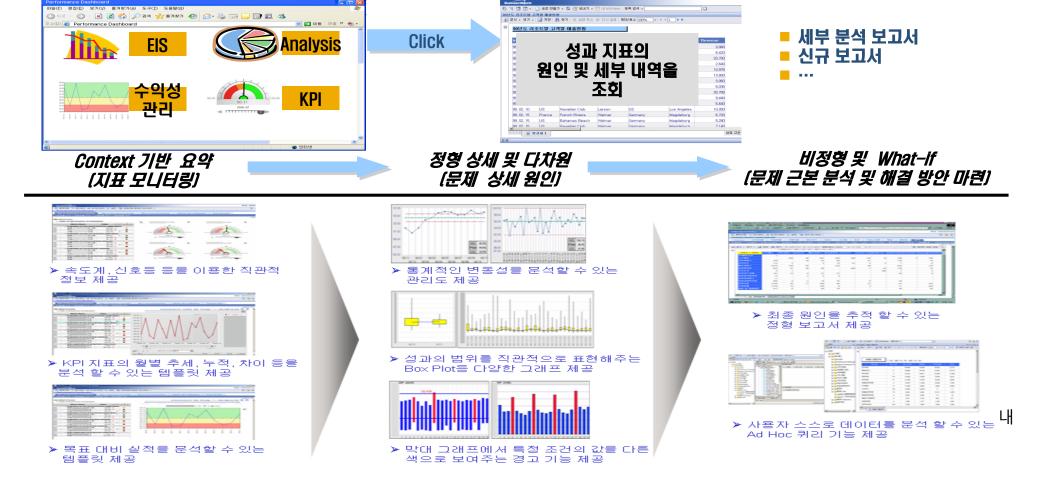
기업의 프로세스 별로 안정성,신속성,정확성,신뢰성 등을 평가하는 성과 지표(Performance Metrics)를 정의하고 이러한 Framework에 기반하여 지표간의 연관 관계를 정의하고 영향을 주는 하위 지표를 식별함

핵심프로세스 평가지표 도출 방법



3.1 분석 Framework(4/4) - 분석 절차 예시

성과 분석 출고에 수고에 기반을 둔 화면 설계에 따라 직관적인 생선으로 기계 제공과 상세 및 근본 원인 분석 화면 또는 도구와의 용이한 연계를 제공



3.2 지표와 데이터 모델(1/2) - Layer 별 설계

KPI의 연관 및 계층 관계 및 지표 별 Analysis Path에 따라 지표와 특성을 Navigation 할 수 있는 정보모델을 각 Data 저 Architecture의 Layer의 목적과 특성에 맞추어 설계.

Data Mart 영역 집계 정보 집계 정보 EDW **영역**

EIM Backbone to Support Data Operations

Source

Systems

Source

Source

지표와 모델 관계

계출별 모델 특징

데이터 마트 영역

- 사용자 요구 사항 중심의 데이터 구성
- 분석의 용도에 맞게 목적성 다차원 구조
- 성능요건이 반영된 분리된 데이터 구조

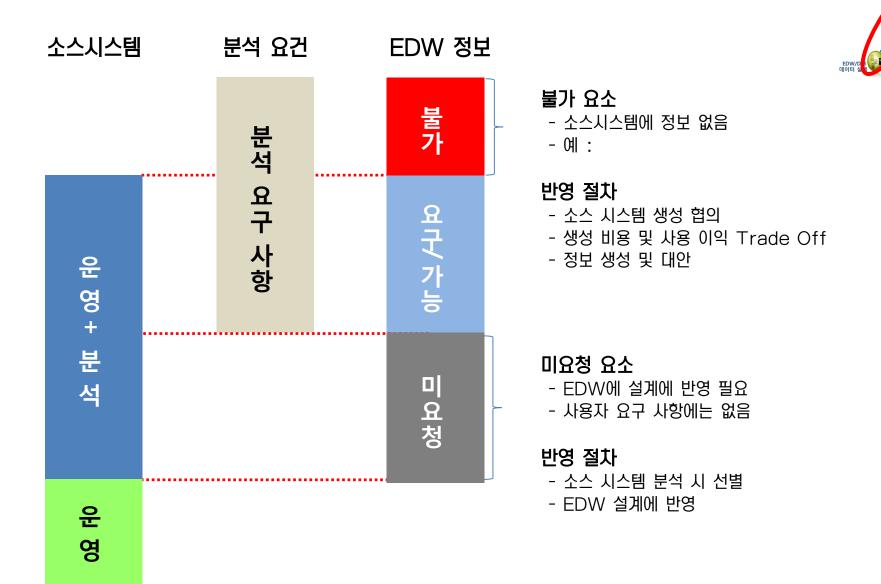
EDW 통합 영역

- 주제중심의 데이터 구조 기간시스템의 변화에 영향을 받지 않도록 제품,수주,설계 등과 같이 주제 중심적으로 데이터를 구성.
- 정규화 구조 하나의 정보는 한곳에 위치하게 하는 원칙하에 정규화하여 중복을 제거하고, 데이터의 정합성을 유지
- 표준화 및 정제된 데이터 전사적인 차원의 통합된 데이터 공간으로 특정 요건에 치 우치지 않은 표준화된 데이터를 제공.

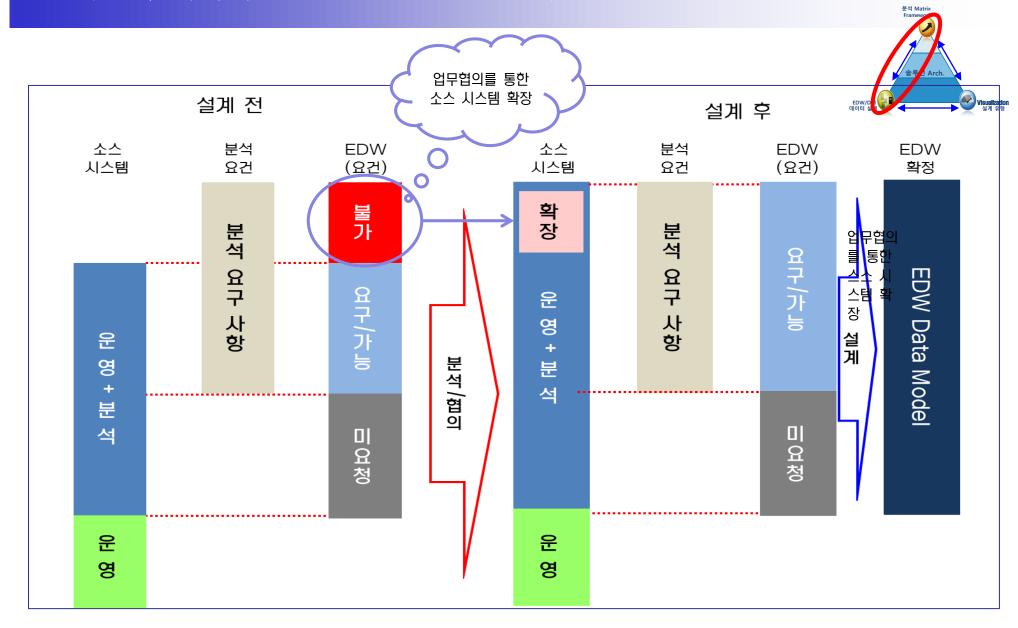
데이터 수집 영역

- 데이터 연계 버퍼 역할
- 데이터 history 기능 및 Cleansing 기능 제공
- 운영계와 유사한 데이터 구성

3.2 지표와 데이터 모델(1/2) - EDW 설계

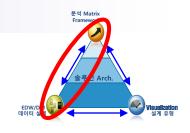


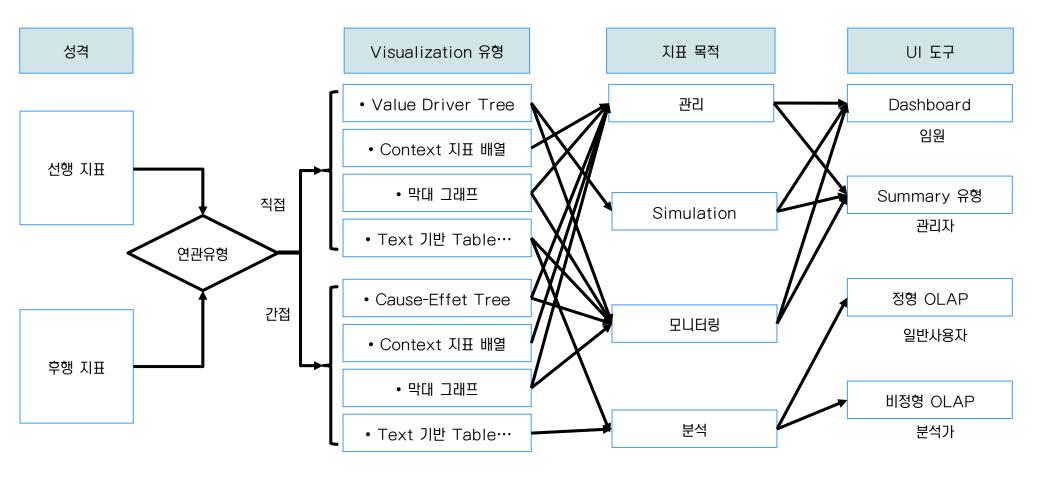
3.2 지표와 데이터 모델(1/2) - EDW 설계



3.3 지표와 Visualization(1/3) -지표유형과 UI

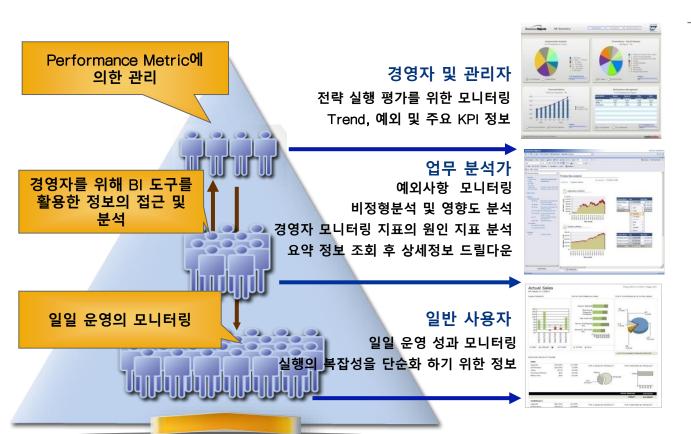
지표에 따라 연관 관계(선행 및 후행 지표)와 연관 유형(직접과 간접) 정의하고, 지표의 목적(관리, 모니터링, 분석 및 Simulation)등에 따라 사용자 계층 별 지표와 관련 Visualization 및 분석을 위한 UI 도구 활용





3.3 지표와 Visualization(2/3) - 사용자 별 유형

업무의 모니터링, 분석 및 성과 관리를 위해 다양한 사용자 계층에게 필요한 지표를 적절한 매체를 통해 제공함



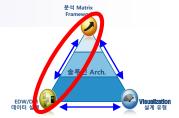
Enterprise Data Warehouse

적용 Use Case

- 1. 경영/관리자를 위한 대쉬보드 시스템
- 핵심성과 및 모니터링 지표 위주의 Dashboard 제공
- 단순 지표 Display가 아닌 Context 기 바
- 모니터나 LCD TV로 제공되며 별도의 클릭이 필요 없는 시스템(No Scroll No Click)
- 원인분석 정보 제공을 UI 도구 제공
- 2. 업무 분석 및 일반사용자를 위한 분석 및 레포팅 포털 구축
- 현재 정형화된 리포트 및 다차원 분석 가능 리포트(Performance Matrix 기 반)
- 비정형 분석 지원
- What-If Analysis

3.3 지표와 Visualization(3/4) - Dashboards

사용자 인터페이스를 통하여 제공함으로써 전략 운영화기업의 경영 상태 및 발생된 문제점을 즉시 인지하여 즉각 조치 환경 제공 출위 상태 및 발생된 문제점을 즉시



정의

역할 중심적으로 기업의 외부 및 내부의 정보를 하나의 Page에 출력하여 대처 가능한 정보를 제공함으로써 제시된 일련의 단계를 통하여 빠른 이해 및 의사결정을 가능하게 하는 화면.

목적

- 경영진을 위한 맞춤 정보 제공
 - 전략의 관리와 실행을 모니터링
 - 기업의 핵심 역량 실행의 모니터링
 - LOB(Line Of Business)와의 Communication
- 문맥적인 핵심 정보 제공
 - 성과 분석 Framework 기반의 KPI
 - 선행과 후행 KPI와의 연관 제공
 - One Screen, One Mean.
- 직관적인 사용자 Interface
 - 문맥적인 화면 배열
 - 시각적인 예외 사항 인지
 - No Click, No Scroll

설계 고려 사항

의사소통 과 명확한 정보 전달

• 정보는 의도한 대상을 위한 주요 의사결정에 중 점을 두어야 함.

항상 상호 대화적

• 지표의 Framework에 기본을 두어 상호 연관을 발견 할 수 있는 분석 경로를 활용하여 일관되고 상세한 정보를 제공해야 함.

지능형 화면

• 정보는 논리적인 방식으로 구성되어 쉽게 이해할 수 있고 바로 응대할 수 있어야 함.

협업

• 정보조회 중의 의문사항에 대해 관련 담당자와의 질의 및 답변기능 구현

개인화

• 필요에 따라 사용자가 자신의 화면과 내용을 구성할 수 있어야 함..