

9-1. Data Warehouse System

홍형경

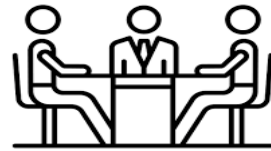
chariehong@gmail.com

2020.01

1. DW (Data Warehouse) 란?

가장 마진이 많이 남는
상품은?

우량 고객은 누구이며
고객들이 가장 많이
구매하는 상품은?



가장 매출이 높은
매장은?

가장 효과적인 판매
채널은?

1. DW (Data Warehouse) 란?

하지만....

- 필요한 데이터를 찾을 수 없어요...
 - 데이터가 여러 곳에 흩어져 있음
- 필요한 데이터를 얻을 수 없어요
 - 전산실에 요청하니 바쁘다고 빨리 안줘요
 - 얻긴 했어도 데이터가 더 필요한데, 또 요청하자니 눈치 보여요
- 데이터를 확보했지만, 사용하기 어려워요
 - 원하는 형태로 만들려면 추가로 엑셀로 작업해야 해요

1. DW (Data Warehouse) 란?

- 데이터 + 창고
- 산재해 있는 데이터를 필요에 맞게 선정해 한 곳에 모아 놓은 것
- 신속한 의사결정을 위해 분석을 쉽게 하도록 데이터 집적
- 운영 데이터베이스(OLTP)에서 선별한 데이터를 가져와 분석에 맞게 테이블 재설계 및 데이터 로딩

1. DW (Data Warehouse) 란?

- OLTP (OnLine Transaction Processing)
 - 데이터 입력, 수정, 삭제 위주
 - 트랜잭션 처리 중심
 - 은행 입출금 시스템, ERP
- OLAP (OnLine Analytic Processing)
 - 데이터 조회 위주 (분석 용도)
 - 특정 주제에 대한 분석과 조회 위주
 - DW (Decision Support System 이라고도 함)
- DW는 OLTP 시스템에서 필요한 데이터를 가져와 통합해 적재

1. DW (Data Warehouse) 란?

한 마디로...

OLTP 시스템은 비즈니스를 **운영**하는 역할

OLAP 시스템은 비즈니스를 **효율화** 하는 역할

2. DW의 간략한 역사

- 1980년대 IBM에서 처음 도입
- 2000년대 초 CRM 열풍으로 DW 시스템 도입 본격화
- 많은 곳에서 CRM 프로젝트 실패
 - 데이터를 우선시 해야 하는데, CRM 시스템 구축을 우선시
 - 시기 상 분석을 위한 충분한 데이터가 쌓이지 않음 (ERP 도입 초반기)
- 이후 ERP 도입이 일반화 되고, 충분한 데이터가 쌓인 후 큰 기업 위주로 DW 시스템 구축
- 최근에는 중.소기업도 DW 시스템 도입 (BI 툴의 활성화)

2. DW의 간략한 역사

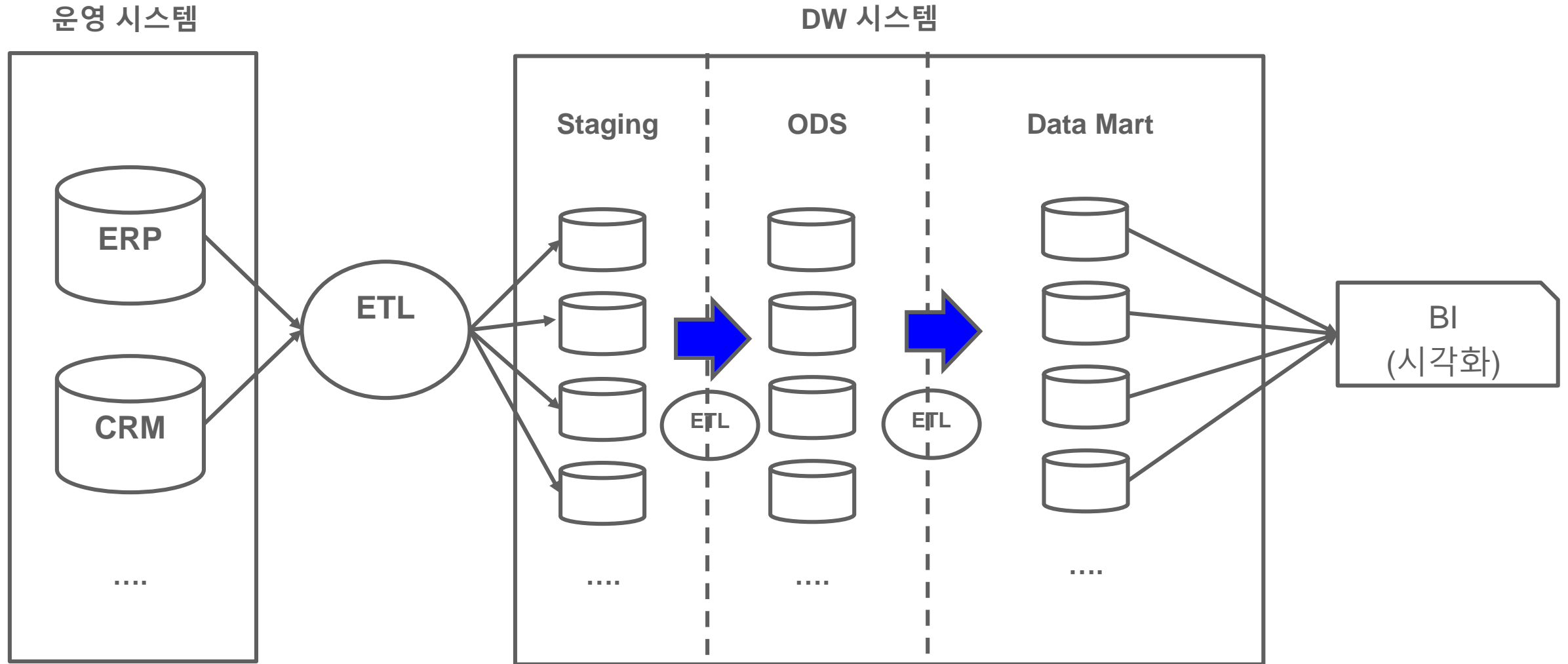
- CRM (Customer Relationship Management) : 고객 관계 관리 시스템
- 기업 입장에서 어떤 고객이 우량 고객인가?
- 파레토(Pareto) 법칙
 - 일명 80대 20 법칙
 - 전체 결과의 80%가 전체 원인의 20%에서 일어난다.
 - 상위 20% 고객이 전체 매출의 80%를 차지한다.
- 충성형 고객을 찾아라
 - 기업에 대한 충성도(loyalty)가 높은 고객
 - 충성형 고객은 매출을 많이 발생시키고, 신규 고객도 끌고 옴(입소문)

2. DW의 간략한 역사

하지만...

- 상위 20% 고객만을 위한 마케팅 집중
- 나머지 하위 고객들을 차별
 - 예, K은행은 50만원 미만 예금주한테 수수료를 물린다고 했음
 - 결국 실행되지 않음
- 롱테일(Long tail) 법칙
 - 상위 20% 제외한 나머지 80%도 엄연한 고객
 - 장기적으로 보면 나머지 80% 고객도 매출 증대에 현격한 기여
- 이후에는 고객 이탈 방지가 CRM의 큰 목표가 됨

3. DW 시스템 아키텍처



3. DW 시스템 아키텍처

(1) 운영 시스템

- ERP, CRM 등 기업 업무 운영을 위한 시스템
- 트랜잭션 중심 (SQL 관점)
 - INSERT, UPDATE, DELETE 중심
 - SELECT는 좁은 범위에서 실행됨
- 매일, 매시간, 매분 신규 데이터 입력
- 데이터 분석 관점 시스템이 아닌 업무 운영 중심 시스템

3. DW 시스템 아키텍처

(2) ETL (Extract Transform Load)

- 원천(source) 운영 DB(시스템)에서 데이터 추출 → 변경 및 가공 → 대상(target) DW DB에 로딩
- 원천 DB에서 대량의 데이터를 추출(Select)해 DW에 로딩(Insert)
- 수 많은 테이블과 데이터를 추출, 로드 하는데 있어 매우 빠르게 처리
- 원천과 대상 DBMS가 다를 수 있으므로, 여러 DBMS 지원
 - DBMS별 데이터형 자동 변환
 - 대량의 데이터를 빠르게 처리해야 하므로 내부적으로 병렬 처리 기능 탑재

3. DW 시스템 아키텍처

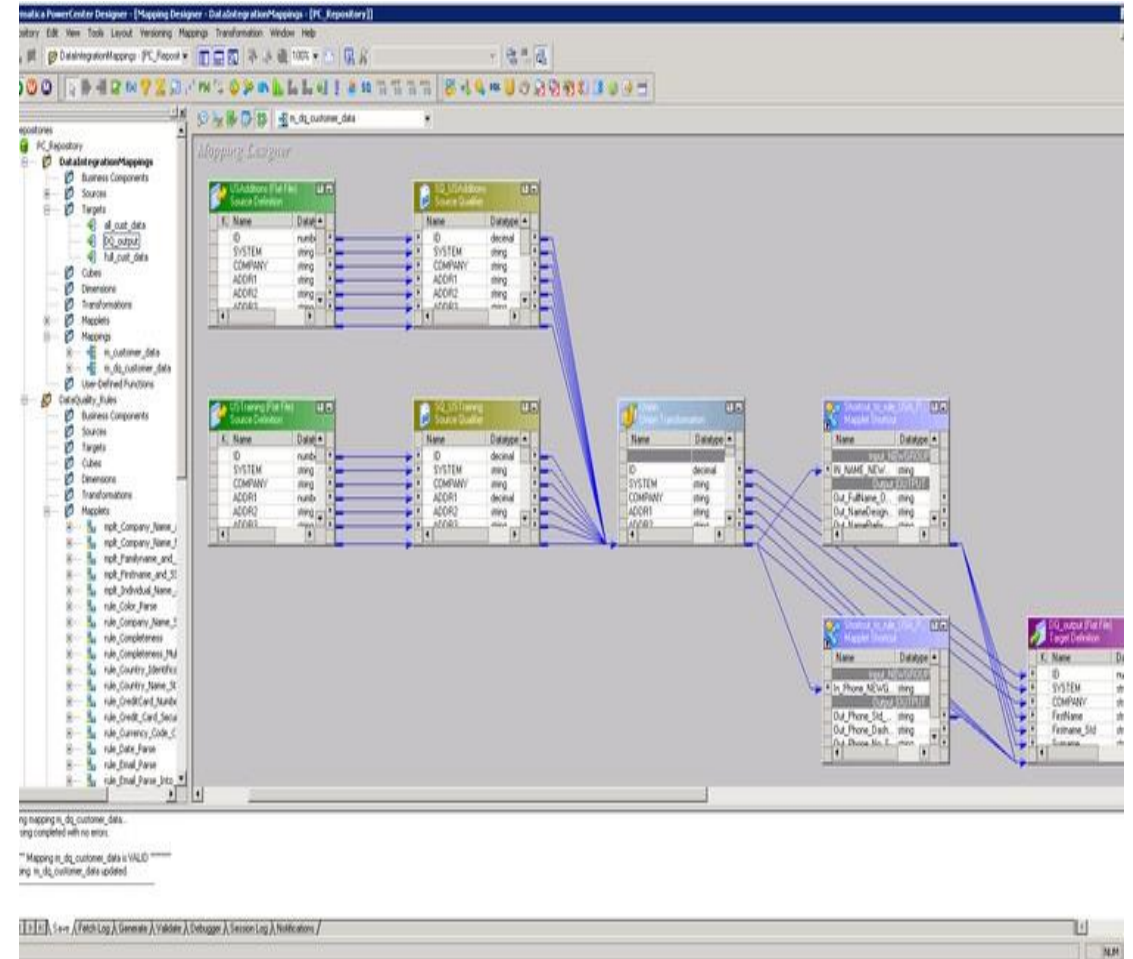
(2) ETL (Extract Transform Load)

- DW 아키텍처 구성에 따라 원천 → DW 뿐만 아니라, DW 내에서도 Staging → ODS → Data Mart 테이블 적재 시에도 사용됨.
- 스케줄링 기능 지원
- 대표적인 ETL 제품
 - Informatica - Powercenter
 - Data Stage
 - SSIS
 - OGG, ODI, Data Sync

3. DW 시스템 아키텍처

(2) ETL (Extract Transform Load)

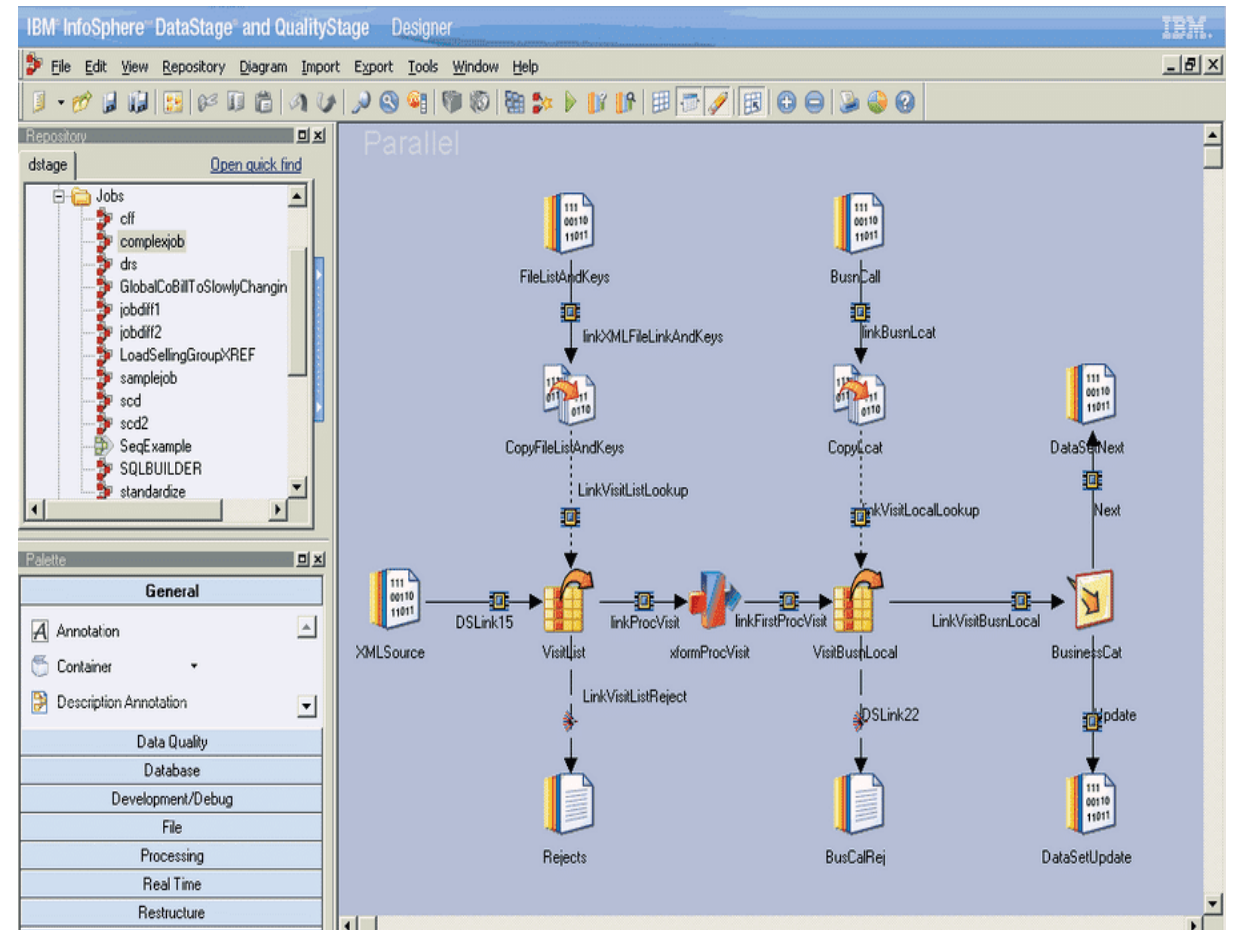
- **Informatica - Powercenter**
 - 가장 많이 사용되는 ETL 제품 중 하나
 - 성능, 사용 편의성 등의 검증을 받아 많이 사용됨
 - 여러 종류의 DBMS 간 변환 기능
 - 병렬 처리 기능



3. DW 시스템 아키텍처

(2) ETL (Extract Transform Load)

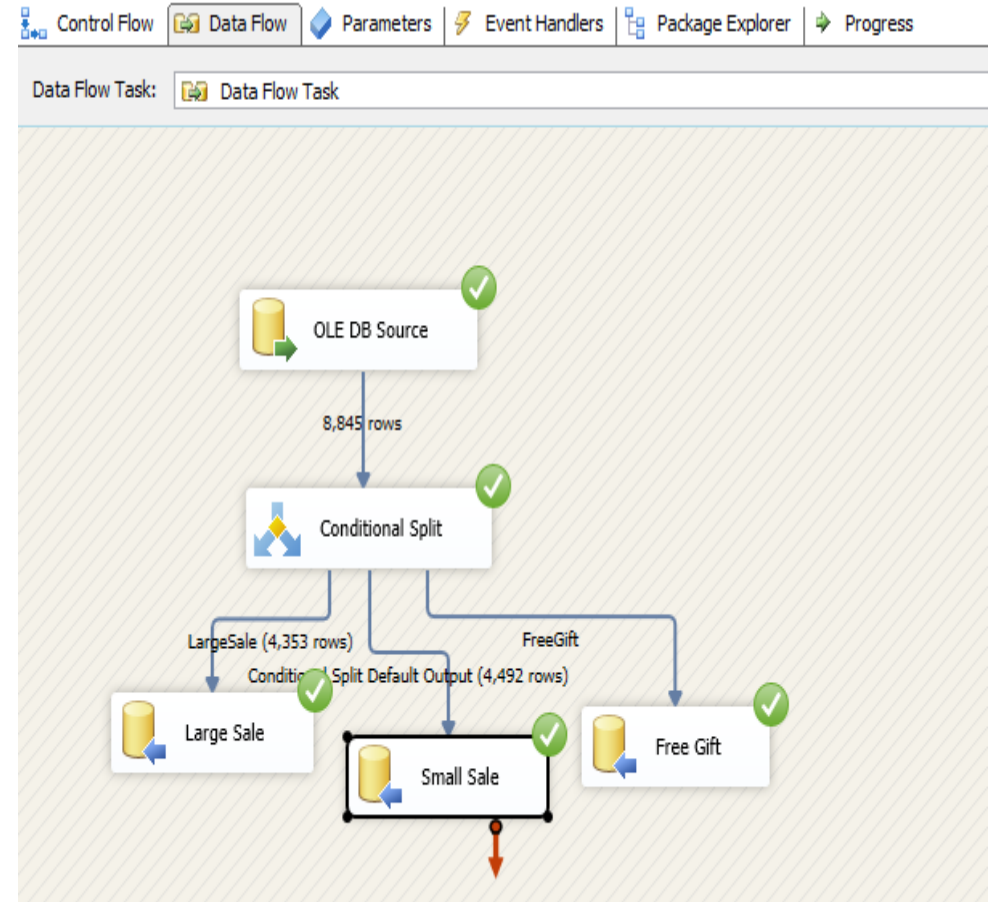
- Data Stage (IBM)
 - InPa와 비교되며 널리 사용되는 제품



3. DW 시스템 아키텍처

(2) ETL (Extract Transform Load)

- SSIS (Microsoft)
 - InPa, Data Stage보다 저렴
 - MSSQL 서버 라이선스 있으면 무료 사용 가능(현재는??)
 - Oracle과는 궁합이 잘 안 맞음
 - 배우기 그리 어렵지 않고 시각적 인터페이스 덕분에 사용하기 쉬움



3. DW 시스템 아키텍처

(2) ETL (Extract Transform Load)

- Oracle Golden Gate (Oracle)
 - Oracle to Oracle 의 경우 실시간 동기화 가능
 - 매우 고가
- Oracle Data Integrator (Oracle)
 - OGG보다 저렴, 기능은 떨어짐
 - 다른 툴(Infa, SSIS)보다 배우기 어려움
 - ETL이 아닌 ELT
 - Target DB(DW)의 리소스를 사용해 처리
- Data Sync (Oracle)
 - ODI의 Lite 버전
 - Oracle Cloud에서 ADW, OAC 사용 시 무료로 사용 가능
 - ODI와 내부 동작 방식은 동일

3. DW 시스템 아키텍처

(3) DW 시스템

- Staging DB
 - 원천 DB의 테이블을 그대로 이관
 - 원천 DB와 DW DB간 데이터 불일치 검증 용
 - Staging DB 적재 단계를 생략하기도 함

3. DW 시스템 아키텍처

(3) DW 시스템

- ODS (Operational Data Store)

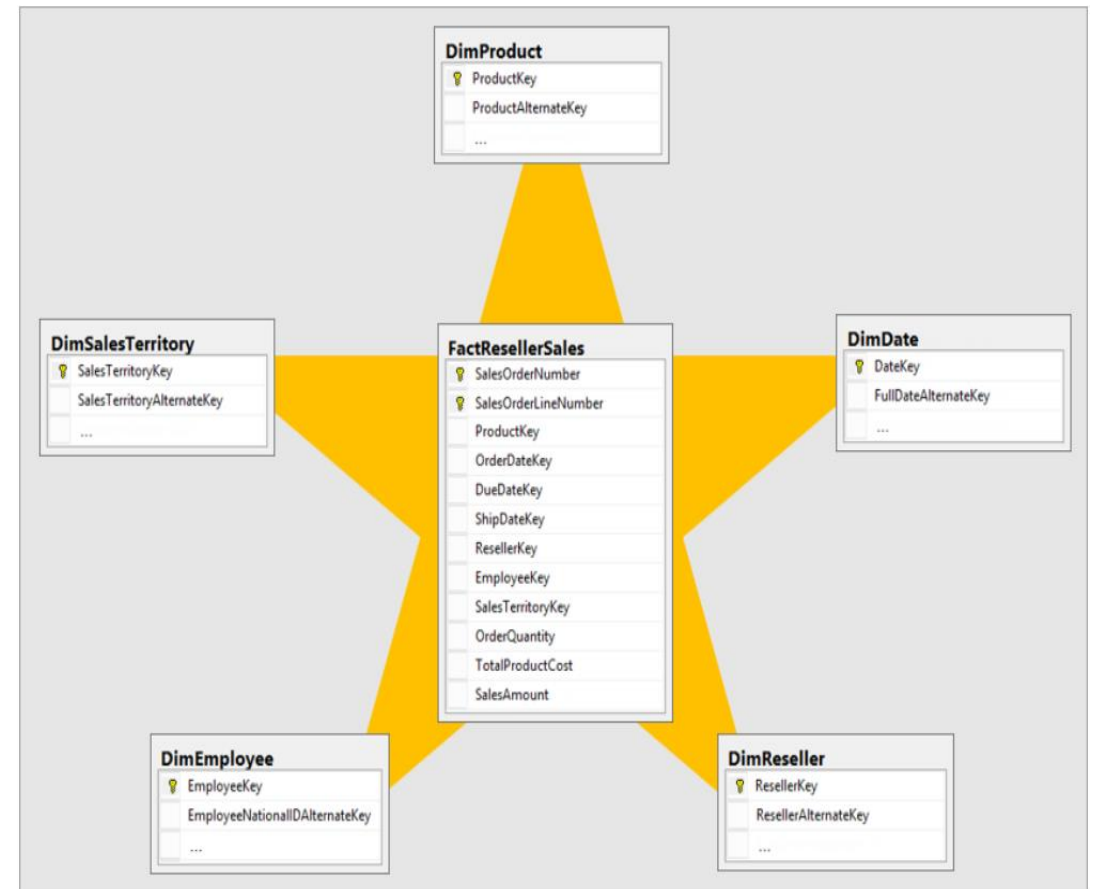
- 원천 DB와 Data Mart 사이의 중단 단계 테이블 집합소
- Staging 테이블을 읽어 데이터 클린징(Cleansing), 데이터 정합성 보장 처리 작업 후 적재
- Staging DB 적재 단계가 없다면 ETL로 추출, 변환, ODS에 로딩

3. DW 시스템 아키텍처

(3) DW 시스템

- Data Mart

- 분석을 위한 최종 테이블
- 한 테이블에 저장된 데이터가 많음
- Fact 테이블, Dimension 테이블로 분리
- Star Schema, Snowflake Schema



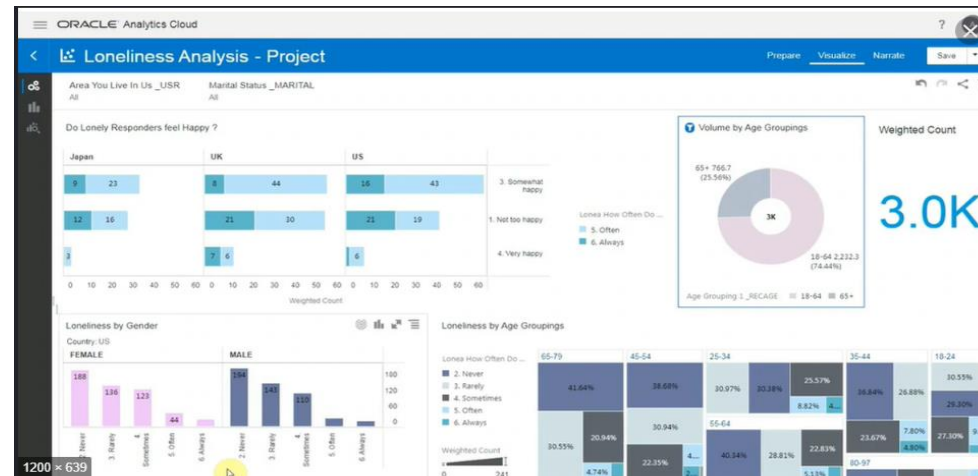
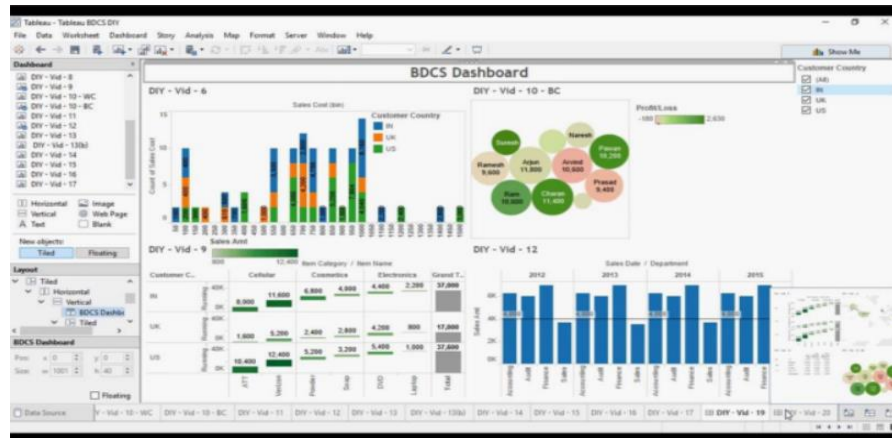
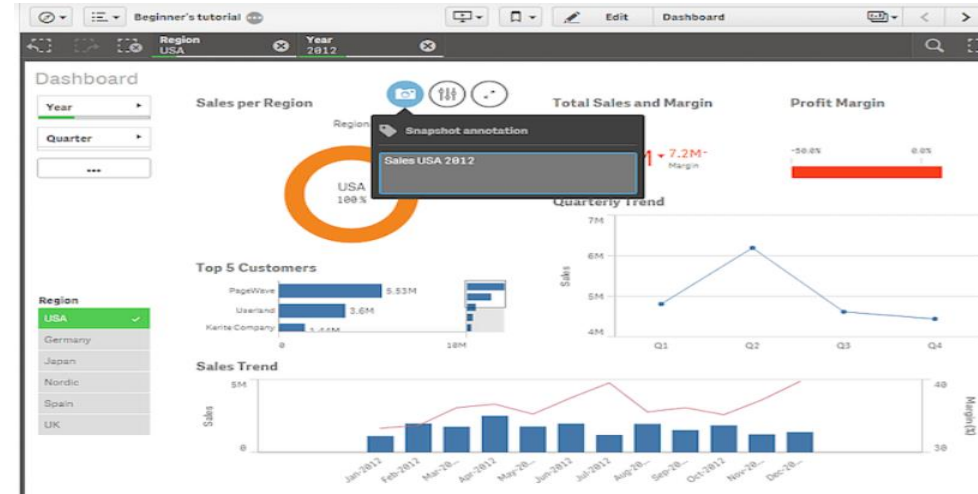
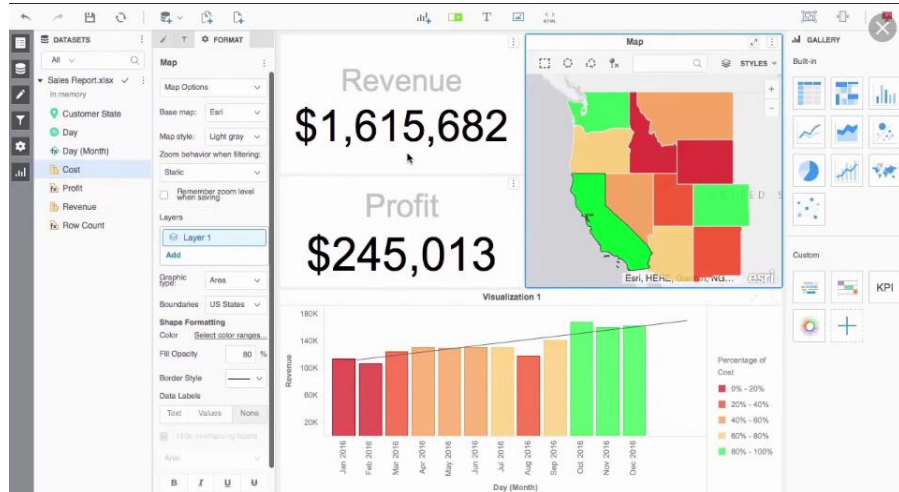
3. DW 시스템 아키텍처

(4) BI (Business Intelligence)

- DW 시스템에 있는 Mart 테이블을 읽어 데이터 분석을 하는 시각화 Tool
- 데이터 분석을 필요로 하는 현업 담당자가 사용
 - SQL 지식 없이도 시각화, 분석이 가능
 - Drag & Drop 으로 손쉽게 사용 가능
- 대표적인(?) BI Tool (상용)
 - MicroStrategy, QlikSense, Tableau, OAC(Oracle Analytics Cloud), ...

3. DW 시스템 아키텍처

(4) BI (Business Intelligence)



4. 데이터 분석의 미래

- DW 시스템은 과거와 현재 데이터를 시각화 해 분석
- DW에서도 Data Mining 기법을 사용해 데이터 분석 (SAS, SPSS)
- 이제는 기존 데이터를 토대로 예측 분석하는 AI가 대세
- 이제는 BI가 아니라 AI다