

Data Warehouse Crash Course

에샨트 가르그

데이터 엔지니어, 아키텍트, Advisor
eshant.garg@gmail.com



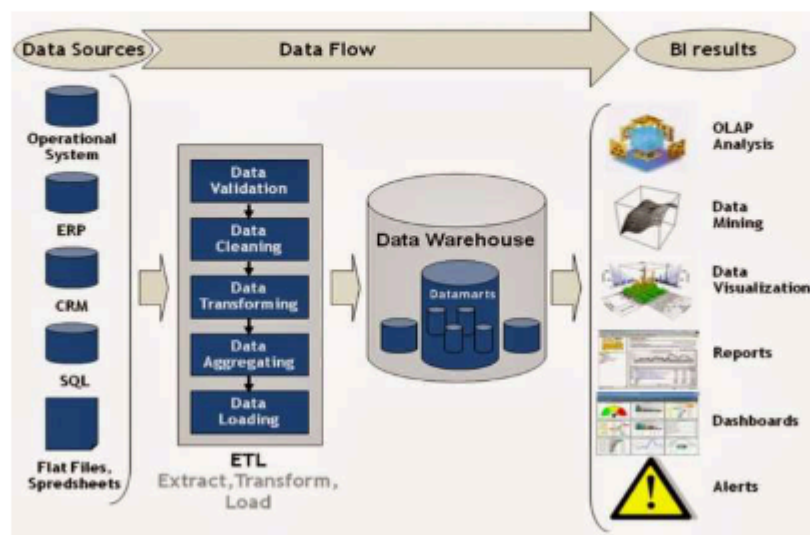
소개



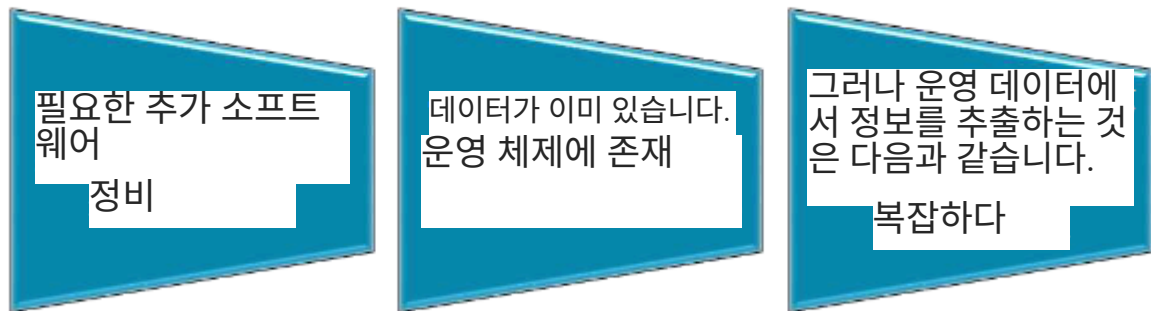
데이터 웨어하우스 기본 사항트랜잭션
데이터베이스에 대한 이점차원 모델링
사실 및 사실 테이블차원 및 차원 Table
Star 대 Snowflake 스키마

Why we need Data warehouse?

데이터 웨어하우스란?



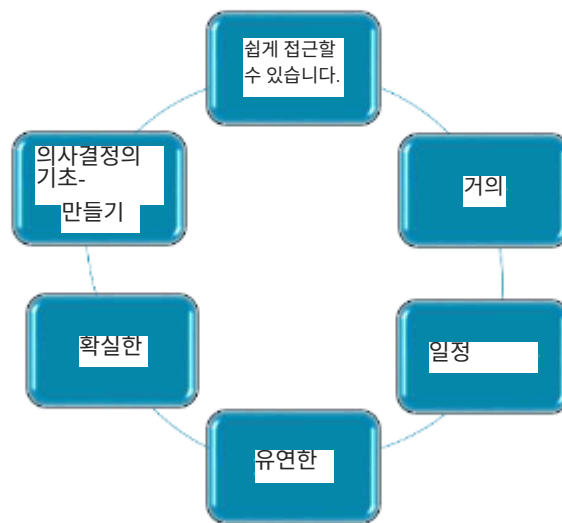
데이터 웨어하우스에 대한 우려 사항



데이터 웨어하우스로 해결할 수 있는 문제

	우리는 너무 많은 데이터를 가지고 있지만 아무것도 만들 수 없습니다
	나는 단지 무엇이 중요한지 알고 싶다
	데이터를 쪼개고 쪼개야 합니다
	비즈니스맨은 데이터에 쉽게 액세스할 수 있어야 합니다.
	부서 간의 숫자는 일치하지 않으며 누가 옳은지 모릅니다
	우리는 사람들이 사실에 근거한 결정을 내리기를 원합니다

이상적인 데이터 웨어하우스 솔루션



Data Warehouse Designer의 책임

비즈니스 사용자 목표 이해
및 목표

정확하고 신뢰할 수 있으며
관련 정보

DWenvironment
지속

OLTP와 OLAP 비교

올티피

- 많은 소액 거래
- 현재 데이터
- 사업 운영에 사용
- 매우 상세합니다.
- 일반적으로 GB 규모
- 처리 성능 한계

올라프

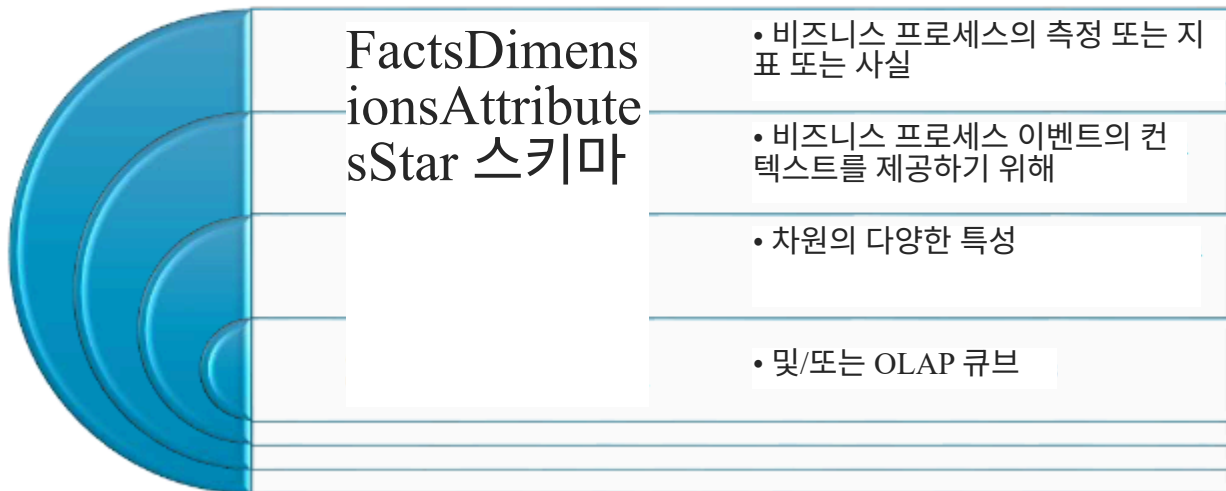
- 볼륨은 적지만 복잡한 쿼리
- 과거, 비휘발성 데이터
- 비즈니스 분석에 사용
- 통합 및 요약
- TB 이상 규모
- 제한 없음, 계산 일시 중지/재개

차원 모델링

데이터 웨어하우스 솔루션에 최적화된 데이터베이스 설계 방법 두 가지 중요한 요구 사항을 해결하기 때문에 널리 사용되는 기술입니다

1. 이해할 수 있는 형식으로 데이터 제공
 2. 빠른 쿼리 성능 제공
- 키워드는 "단순성" DW 환경 유지

차원 모델의 요소



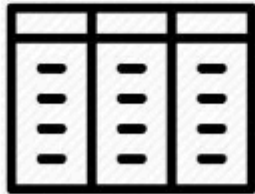
“We sell Cake and other Products in various locations and measure our achievements over time.”

사실 및 사실 표



사실 = 비즈니스 측정값

- 영업
- 이익
- 음량
- 거래 횟수



팩트 테이블 = 조직의 비즈니스 프로세스 이벤트에서 발생한 성능 측정값을 저장하는 테이블

사실 및 사실 표

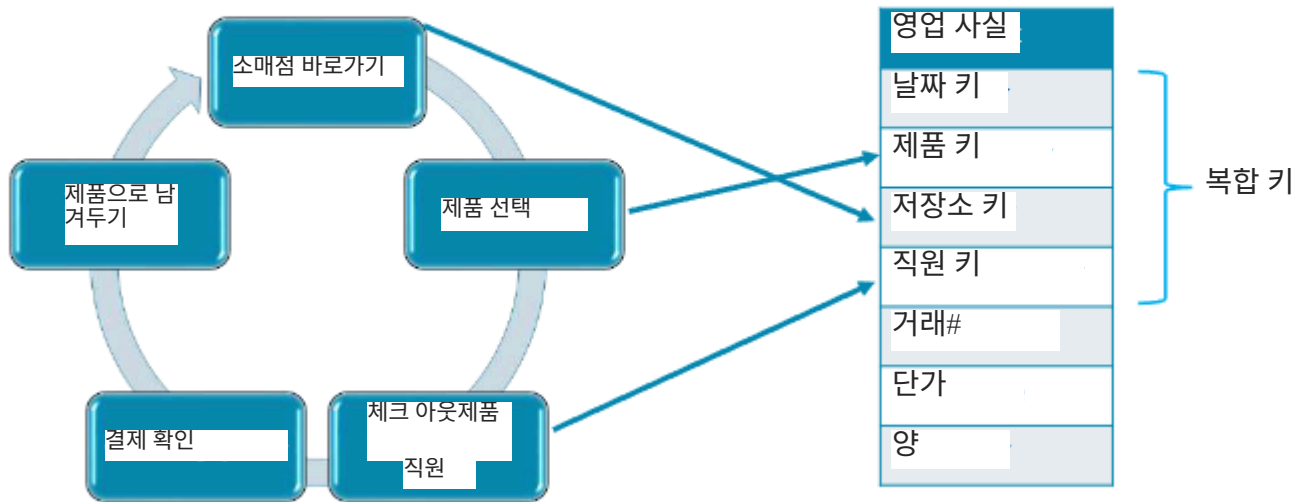


사실은 다음과 같은 질문에 답합니다.

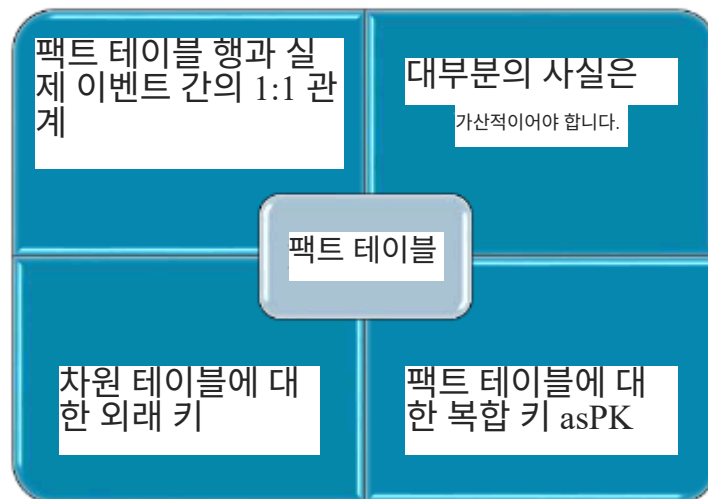
우리는 무엇을 하고 있는 걸까요? (매도, 매수, 개수)우리는 무엇을 성취하고 있습니까? (더 많은 판매량, 더 큰 이익)팩트 테이블의 1개 행은 실생활에서 1개의 측정값입니다.팩트 테이블의 팩트 열은 가산적이어야 합니다.팩트는 차원과 함께 사용할 때 의미가 있습니다.

외래 키와 연결Dat/Time 차원은 대부분의 데이터 웨어하우스에 있습니다.

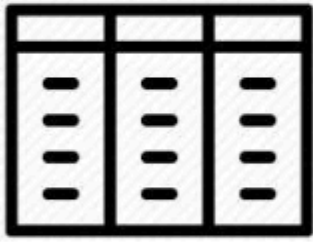
팩트 테이블의 예



팩트 테이블의 특성



차원이란 무엇입니까?

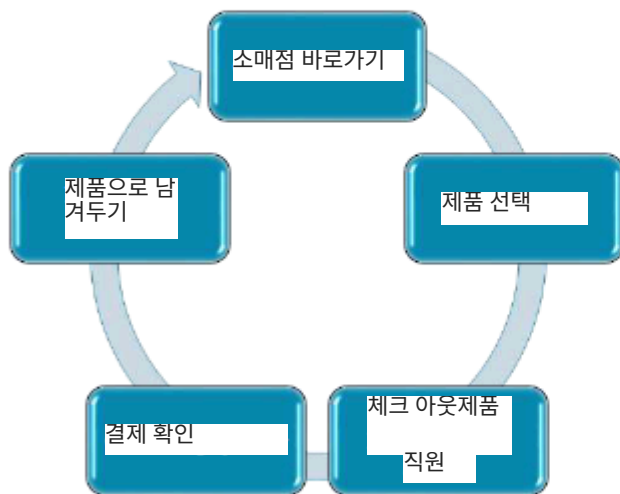


Companions to a fact tableBusinessprocess 측
정 이벤트와 관련된 텍스트 컨텍스트

차원 테이블에서 답변하는 질문



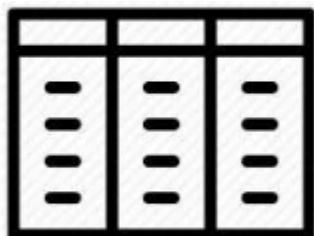
차원 테이블의 예



영업 사실
날짜 키
제품 키
저장소 키
직원 키
거래#
단가
양

제품 차원
제품 키
제품 이름
브랜드 이름
범주 이름
하위 범주 이름
패키지 유형
패키지 크기
무게
중량 측정 단위

차원 테이블의 특징



차원 테이블의 속성 수에 대한 제한 없음

50-100개의 속성이 있는 테이블이 있는 것이 일반적입니다. 일부 차원 테이블에는 소수의 속성만 있습니다. 팩트 테이블보다 행 수가 적습니다.

그러나 훨씬 더 넓을 수 있습니다. 단일 기본 키로 정의됩니다.

팩트 테이블을 사용한 참조 무결성의 기초 Denormalized

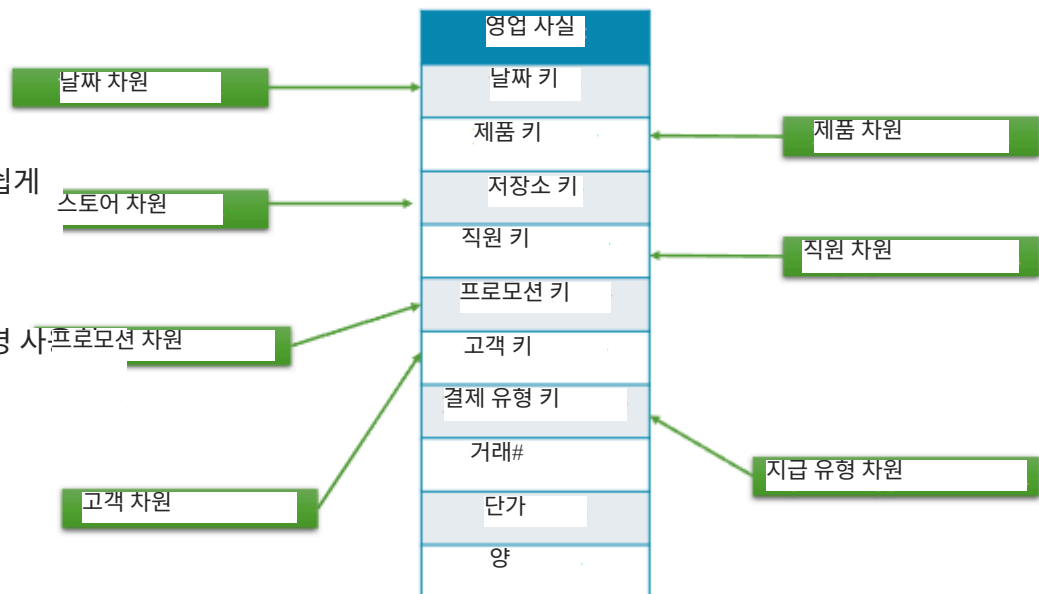
single dimension 테이블 내에서 평면화된 다대일 관계

차원 테이블의 예

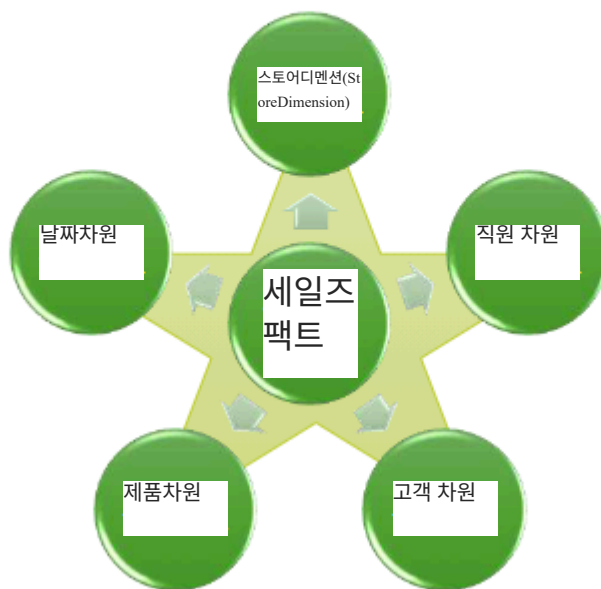
특성

SimpleSymmetric비즈니스 사용자가 쉽게
알아볼 수 있습니다.

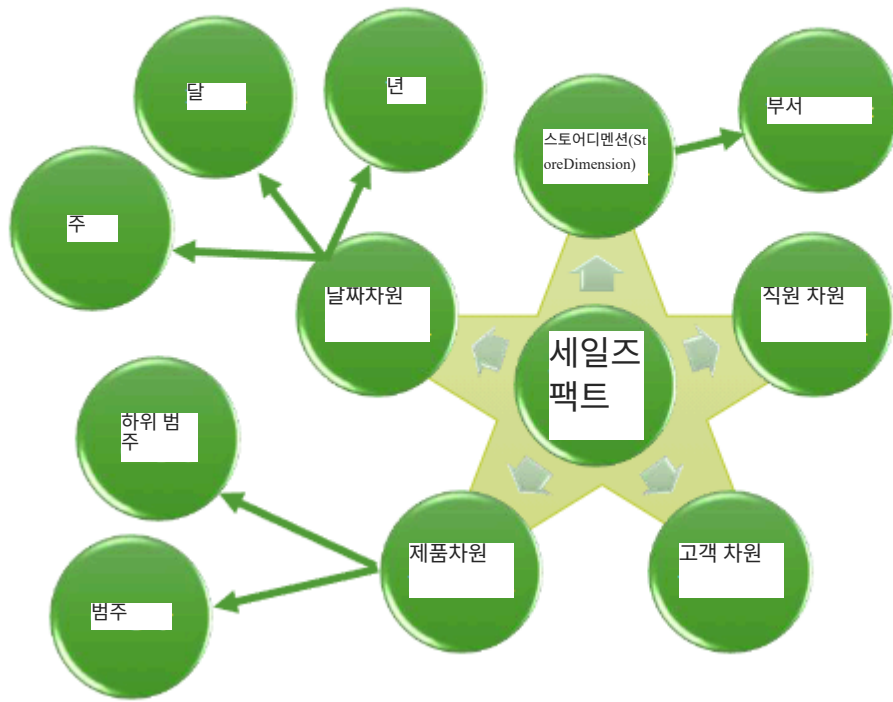
테이블 수 감소의미 있는 비즈니스 설명 사
장 가능



Star 스키마로서의 차원 모델



눈 조각으로서의 차원 모델



요약

데이터 웨어하우스란? 데이터 웨어하우스가 필요한 이유
데이터 웨어하우스(OLAP) 대 트랜잭션 데이터베이스(OLTP)
차원 모델링
사실 및 팩트 테이블
차원 및 차원 테이블
스타 대 눈송이 스키마

