

Chapter 7

# Hierarchies and Snowflakes

김주현

---

Lab: 지능정보시스템

E-Mail: rlawngus36@gmail.com

Phone: 010-4798-7525

## Chapter 7

# Hierarchies and Snowflakes

## Contents

---

1. Attribute Hierarchies and Drilling
2. Snowflakes
3. Outriggers

## Chapter 7

# Hierarchies and Snowflakes

## Contents

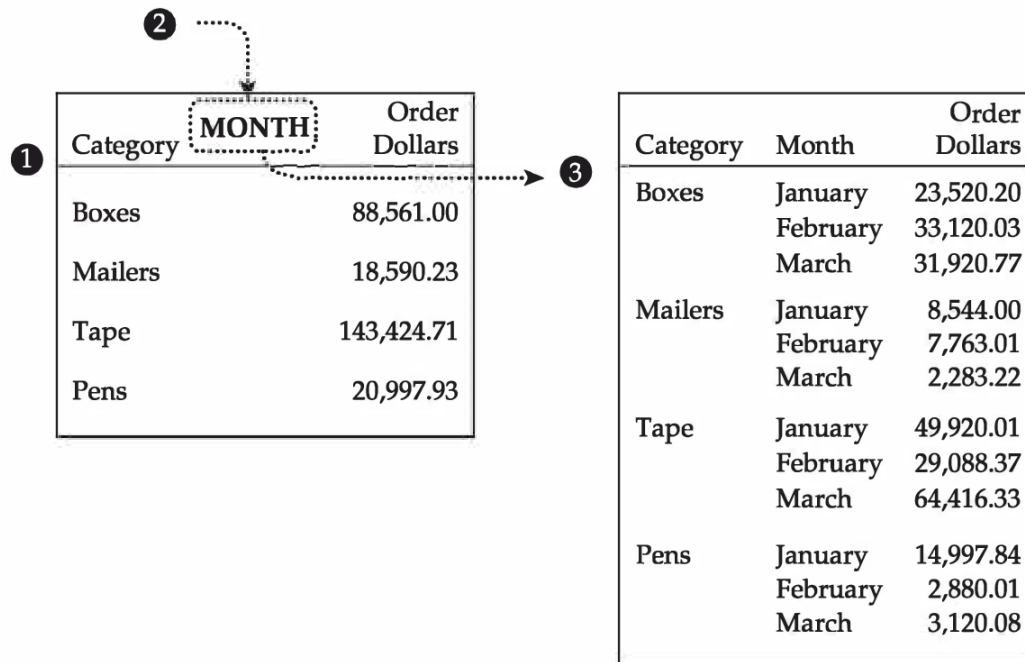
---

1. Attribute Hierarchies and Drilling
2. Snowflakes
3. Outriggers

## [1] The Concept of Drilling

## Drill-Down, Drill-Up, Drill-Thorough

- 드릴 다운(Drill-Down) : 드릴다운은 요약된 형태의 데이터 수준에서 보다 구체적인 내용의 상세데이터로 단계적으로 접근하는 분석기법을 말한다.
- 드릴업(Drill-Up) : 드릴업은 드릴다운과 반대의 방향으로 사용자가 정보를 분석하는 것을 말한다.
- 드릴쓰루(Drill-Through) : 드릴쓰루는 OLAP시스템으로부터 데이터 웨어하우스, 혹은 OLAP시스템에 존재하는 상세데이터에 접근하는 것을 말한다



## [2] The Attribute Hierarchy

### 계층 구조(Hierarchy)

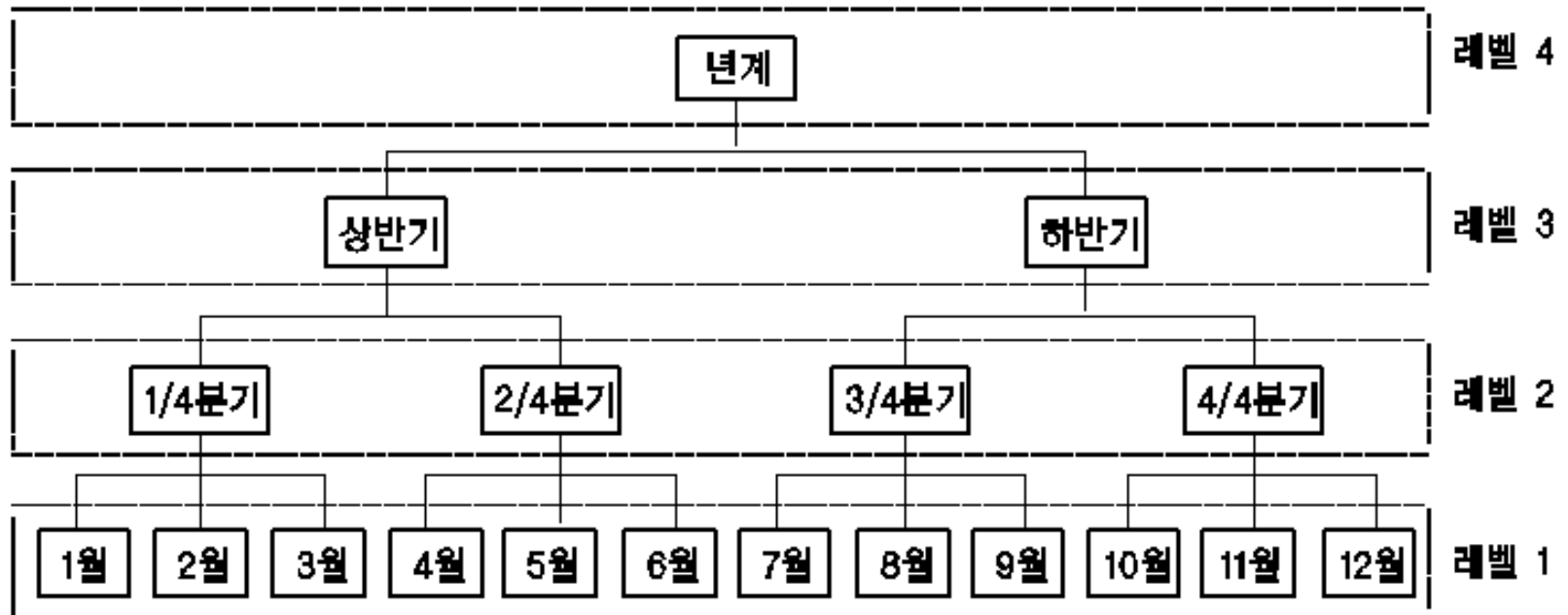
- 레벨들 간에 혹은 항목들 간에 존재 할 수 있다. 항목들간의 가장 기본적인 관계는 Parent-Child 관계이다.
  - Parent : 계층 구조에서 어떤 항목의 바로 상위항목
  - Child : 바로 하위 항목
  - Sibling : 그 항목과 동일한 Parent를 가진 항목
  - Root : Parent가 없는 항목
  - Leaf(Detail) : Child가 없는 항목
  - Ancestor : 계층 구조상의 루트항목 까지 연결되는 가지에서 그 항목의 상위에 나타나는 모든 항목
  - Descendent : Leaf항목까지 연결되는 가지에서 나타나는 모든항목

## [2] The Attribute Hierarchy

계층 구조(Hierarchy)

- 레벨들 간에 혹은 항목들 간에 존재 할 수 있다. 항목들간의 가장 기본적인 관계는 Parent-Child 관계이다.

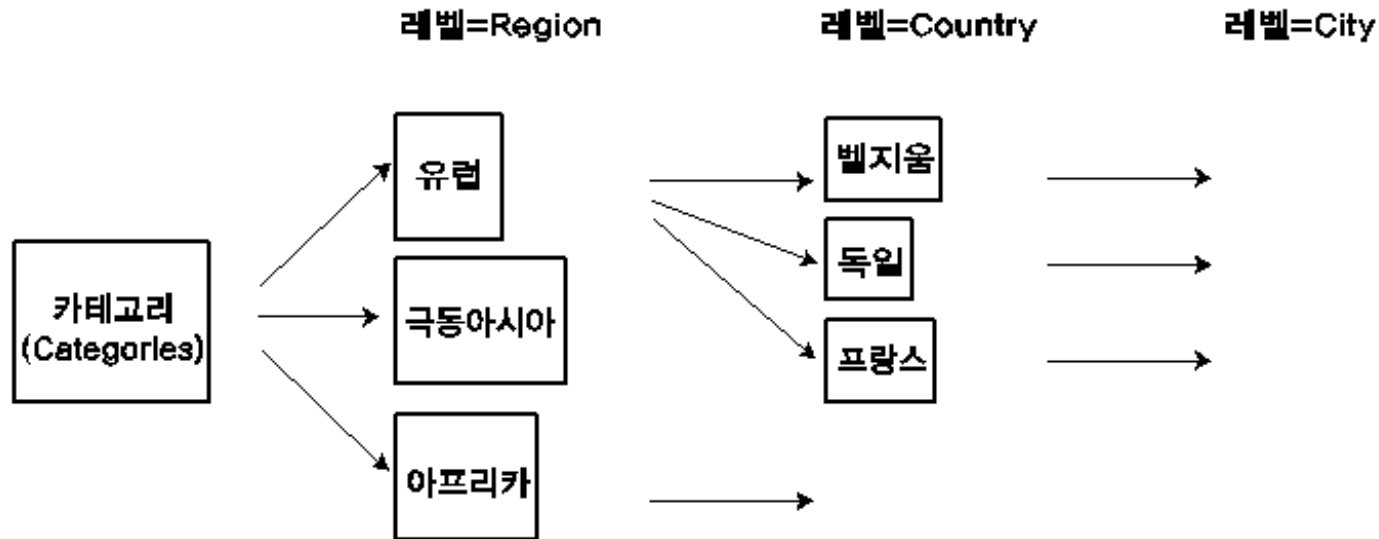
### 계층구조와 레벨



## [2] The Attribute Hierarchy

속성 계층 (Attribute Hierarchy)

- 차원 테이블 내의 정의된 속성들 간에는 일반적으로 잘 정의된 속성 계층(hierarchy)이 존재하는데, 속성계층은 하향식 데이터 구조를 제공한다.



\* 카테고리란 하나의 차원테이블을 구성하고 있는 각각의 데이터 항목을 말한다.

### [3] Multiple Hierarchies in a Dimension

여러 차원의 계층구조(Multiple Hierarchies in a Dimension)

- 하나의 차원에서 여러 속성 계층구조를 갖는 경우가 많다.
- 각 계층 구조는 차원의 모든 특성을 포함하지만 서로 다른 방식을 나타낸다.
- 아래의 7-2의 경우는 브랜드 및 제품의 계층구조로 다양한 제품라인을 관리하기에 유용하다.
- 하지만 제품의 생산을 담당하는 기획자는 아래의 계층구조보단

ALL PRODUCTS(1) -> Business Line(3) -> Manufacturing Locaiton(75) -> Product(8000)  
이런식의 계층구조가 관리하기에 유용하다.

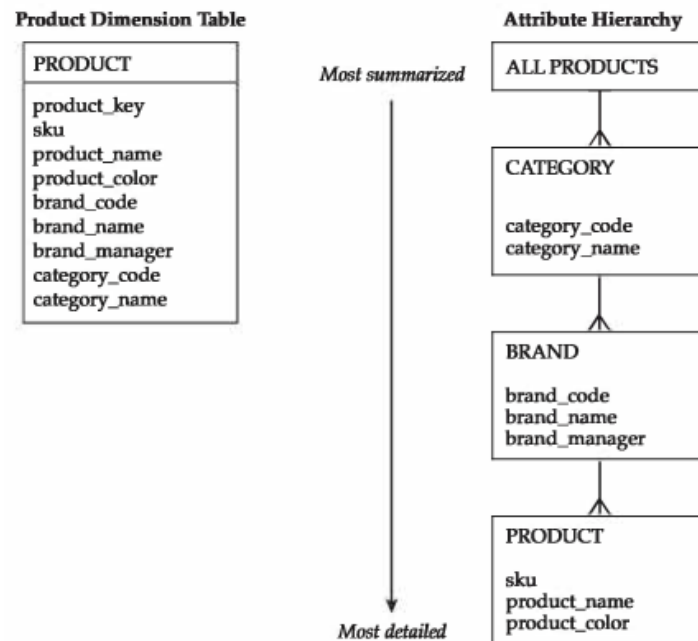


Figure 7-2 An attribute hierarchy in the product table



## [4] Crossing Between Dimensions

차원 간 교차(Crossing Between Dimensions)

- 드릴 경로가 다른 table로 이어질 수 있다.
- 7-1는 Order Dollars를 Category별로 표시한 다음 Month을 추가하여 추가한 세부정보를 작성한 것이다.
- 이 경로는 제품 특성 계층을 무시하고 다른 차원의 세부 정보를 추가한다.
- 5년 동안의 데이터 계층 구조를 다음과 같이 표현할 수 있다.  
All Days(1) -> Years(5) -> Quarters(20) -> Month(60) -> Days(1826)
- 이러한 드릴 경로가 추가되면 모든 날짜에 대한 드릴링, 제품에 대한 드릴링이 가능해진다.

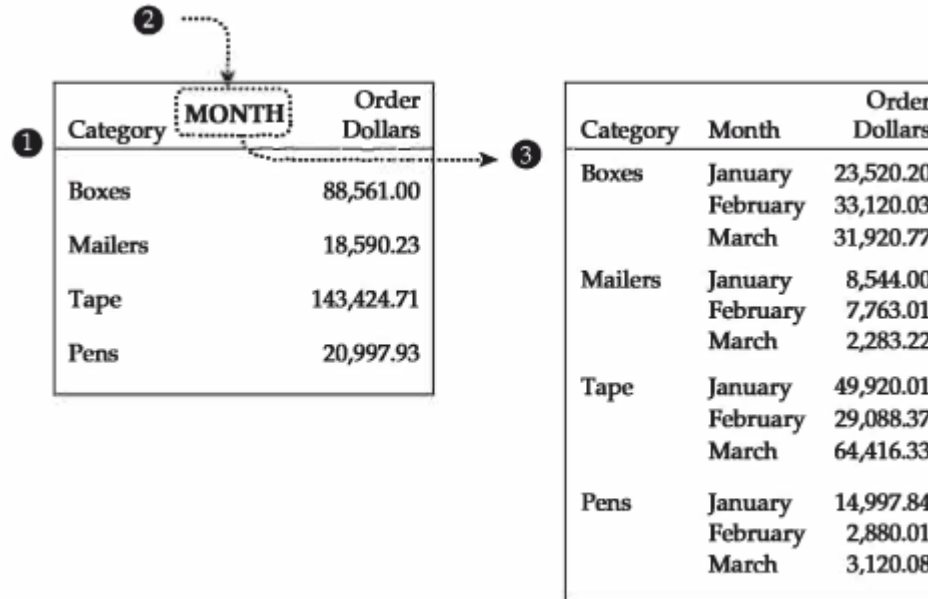


Figure 7-1 Adding dimensional detail

## [5] Instance Hierarchies

### 인스턴스 계층구조(Instance Hierarchies)

- 특성 계층 구조는 차원 특성간의 관계를 설명한다.
- 차원의 인스턴스간에는 다른 계층구조가 있을 수 있다.  
예> 직원은 다른 직원에게 보고를 한다.
- 위의 예에서는 특정 직원을 언급하여야만 표현 가능하다. 또한 관계는 재귀적이다.
- 다른 예로는 부서 내에 속한 부서, 다른 회사를 소유한 회사 및 다른 부서 등이 있다.
- 위의 자세한 내용은 10장 “Recursive Hierarhies and Bridges”에서 설명한다.

## [6] Documenting Attribute Hierarchies

### 특성 계층구조 문서화

- 데이터를 드릴 할 때 특성 계층구조를 사용할 수 있다.
- 특성계층 구조를 사용하여 드릴링을 정의할 때 차원내에서 특성계층 구조를 이해하는것이 매우 중요하다.
- 특성 계층구조를 문서화하면 적합한 차원을 계획하고, 큐브를 설계할 때 이를 참조하여 쉽게 개발자가 이해할 수 있다.
- 속성계층은 그래픽으로 문서화하는 것이 가장 좋다.

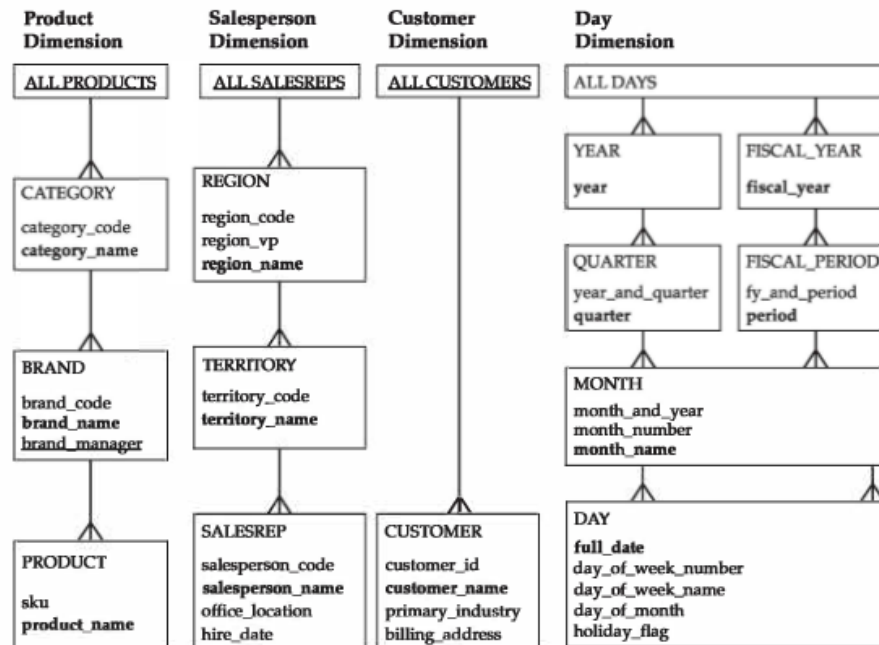


Figure 7-3 Documenting attribute hierarchies

## Chapter 7

# Hierarchies and Snowflakes

## Contents

---

1. Attribute Hierarchies and Drilling
2. Snowflakes
3. Outriggers

## [1] Snowflakes

눈송이 스키마(Snowflake schema)

- 스타 스키마의 사실 테이블 구조와 동일하게 유지하면서 차원 테이블은 정규화된 구조를 형성
- 사실 테이블과 조인되는 차원 테이블이 있으며 이 차원 테이블은 또 다른 테이블의 기본키를 참조하는 외래키를 가진다.
- 스키마 확장에 용이

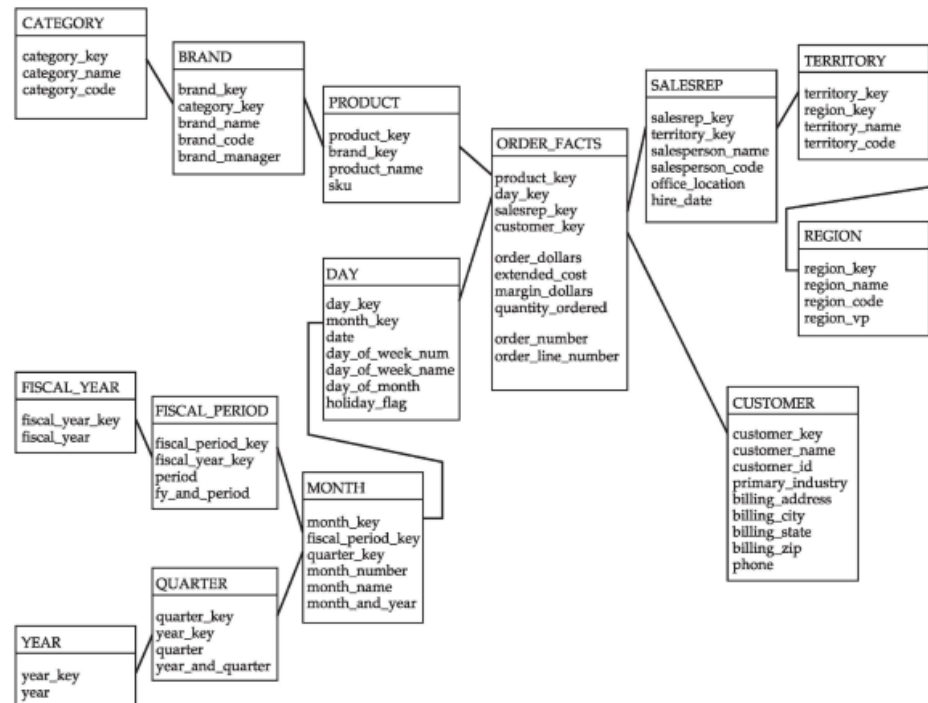


Figure 7-5 A snowflake schema

## [2] Avoiding Snowflakes

눈송이 스키마(Snowflake schema)

- 차원 특성간의 관계를 모델링 하는 것은 유용성을 떨어뜨리고, ETL프로세스를 복잡하게 만들고 성능을 저하시킬 수 도있다.
- 하지만 눈송이 스키마를 사용하는데 좋은 이유가있다.

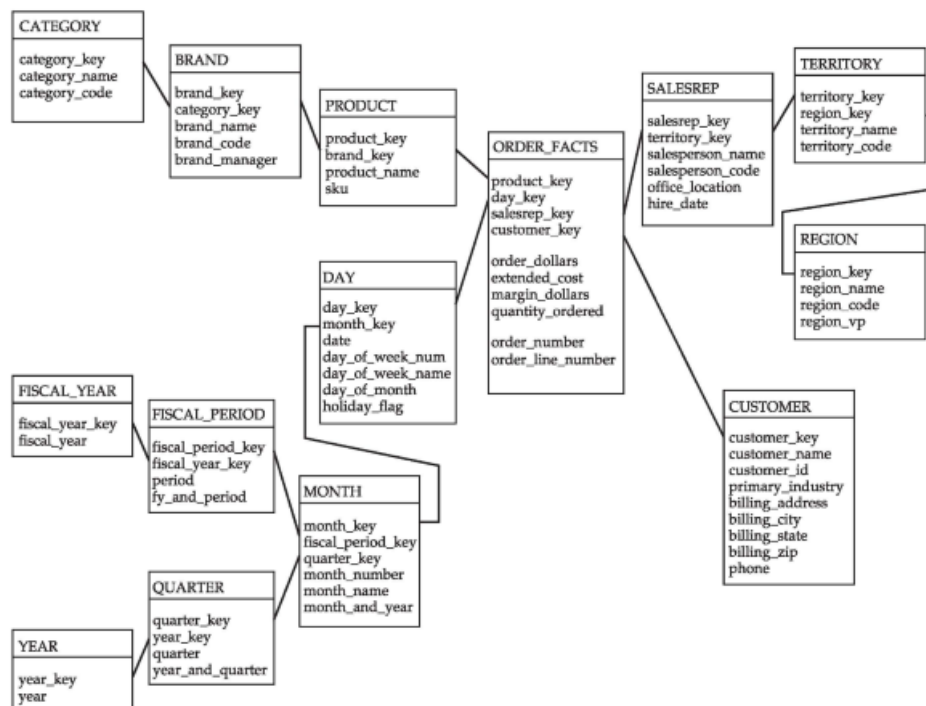


Figure 7-5 A snowflake schema

### [3] Normalization Is UseFul in an Operational Setting, Is Not Useful in an Analytic Setting

Normalization is useful in an operational setting

- ER 모델링은 OS또는 OLTP 시스템을 지원하는 데이터베이스를 디자인할 때 종종 사용됨.
- 이 모델링 양식은 눈송이와 같은 방식으로 속성간의 관계를 중점으로 둔다.
- 정규화는 데이터를 수정할 수 있는 다양한 트랜잭션을 지원하는 운영시스템에 적합하다
- 정규화 프로세스를 통해 중복 모델에서 제거한다.
- 반복되는 그룹은 자체 테이블로 이동하고 기본 키 / 외래 키 관계 또는 조인으로 표시 한다.

Normalization is not useful in an analytic setting

- 분석 데이터베이스에서는 정규화 수준을 요구하지 않는다.
- 데이터의 무결성을 보장 할 필요가 없다.
- 유용성을 떨어 뜨리고 ETL을 복잡하게 만들고 성능에 부정적인 영향을 줄 수 있다.

## Chapter 7

# Hierarchies and Snowflakes

## Contents

---

1. Attribute Hierarchies and Drilling
2. Snowflakes
3. Outriggers



## [1] Outrigger

### Outrigger(아웃트리거)

- 반복되는 속성 집합은 일관성이 없다.
- 반복 속성은 기본 키/오래 키 관계를 통해 차원에 링크 된 새 테이블에 배치된다.
- 아웃 트리거라고 불리는 테이블은 단일 ETL 프로세스와 일관성을 보장하지만 유용성과 성능에 영향을 미친다.
- 아웃트리거가 있으면 스타 스키마를 쿼리하는데 사용할 수 있는 조인 수가 늘어난다.  
이는 복잡성이 증가하고 잠재적으로 성능이 저하 될 수 있다.

# [1] Outrigger

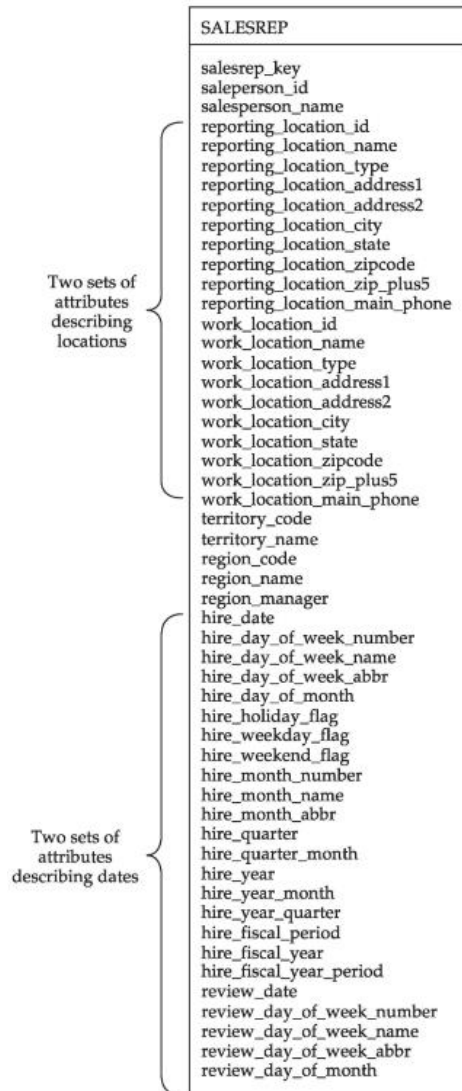


Figure 7-6 Repeating location and date attributes in a dimension table

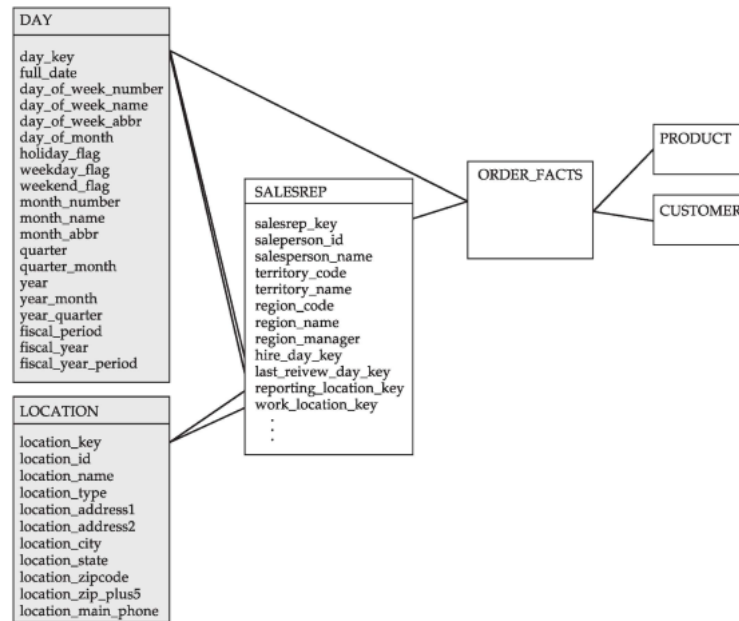
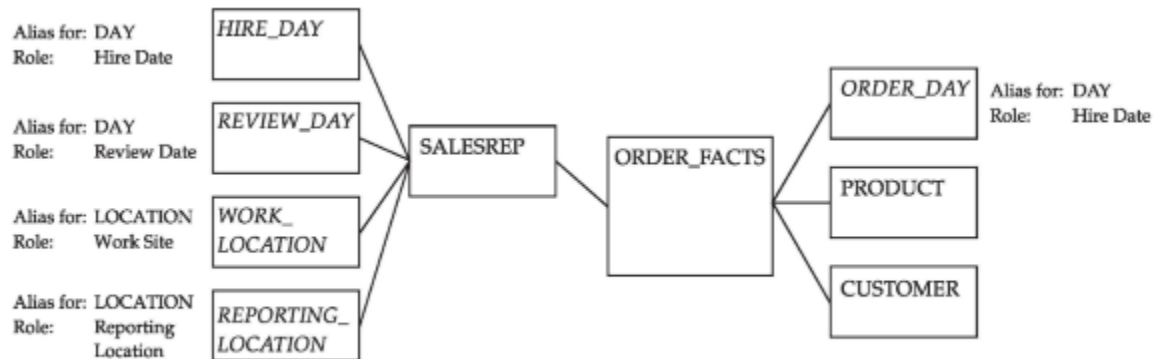


Figure 7-7 Two outriggers for the salesrep dimension

## [1] Outrigger

## Outrigger(아웃트리거)

- 아웃트리거가 있으면 스타 스키마를 쿼리하는데 사용할 수 있는 조인 수가 늘어난다.  
이는 복잡성이 증가하고 잠재적으로 성능이 저하 될 수 있다.
- 아래 예제에서 아웃트리거는 ORDER FACTS를 쿼리할 때 처리할 조인 수를 두배로 늘린다.  
아웃 트리거가 없으면 스타스키마는 최대 네 번의 조인이 필요하다.  
아웃 트리거와 함께 8개의 조인이 필요해진다.



**Figure 7-8** Querying a star with outriggers