

Star Schema

Chapter 8. more slow change techniques

컴퓨터과학과

우종선

# Contents

- Time-stamped dimension
  - Point-in-time status of a dimension
  - The time stamped solution
- Type 3 changes
  - Study all facts with old or new dimension values
  - The type 3 solution
- Hybrid technique
  - Conflicting requirements
  - The hybrid response
  - Evaluating and extending the hybrid approach

# Time-stamped dimension

- Type 2 change로는 point-in-time analysis를 수행할 수 없다.
  - 특정 시점에 dimension이 어떻게 보이는지 알 수 없다.
- Time stamped dimension permit 3 forms of point-in-time analysis
  - 쉽게 change history를 알 수 있다.
  - 특정 날짜에 영향을 주는 dimension row를 빠르게 고를 수 있다.
  - 현재 영향을 주는 dimension row를 빠르게 알 수 있다.

# Time-stamped dimension

- Type 2

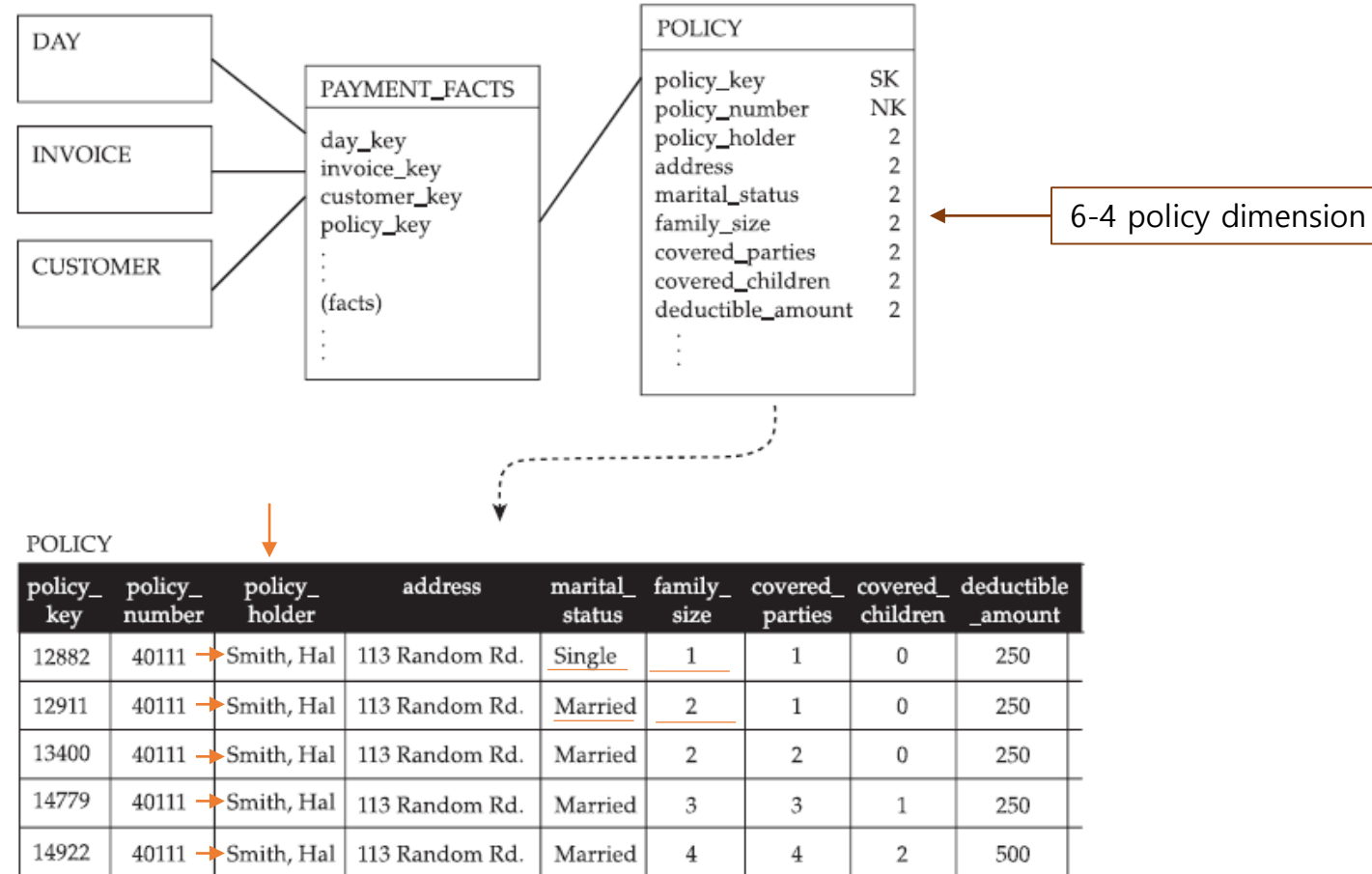
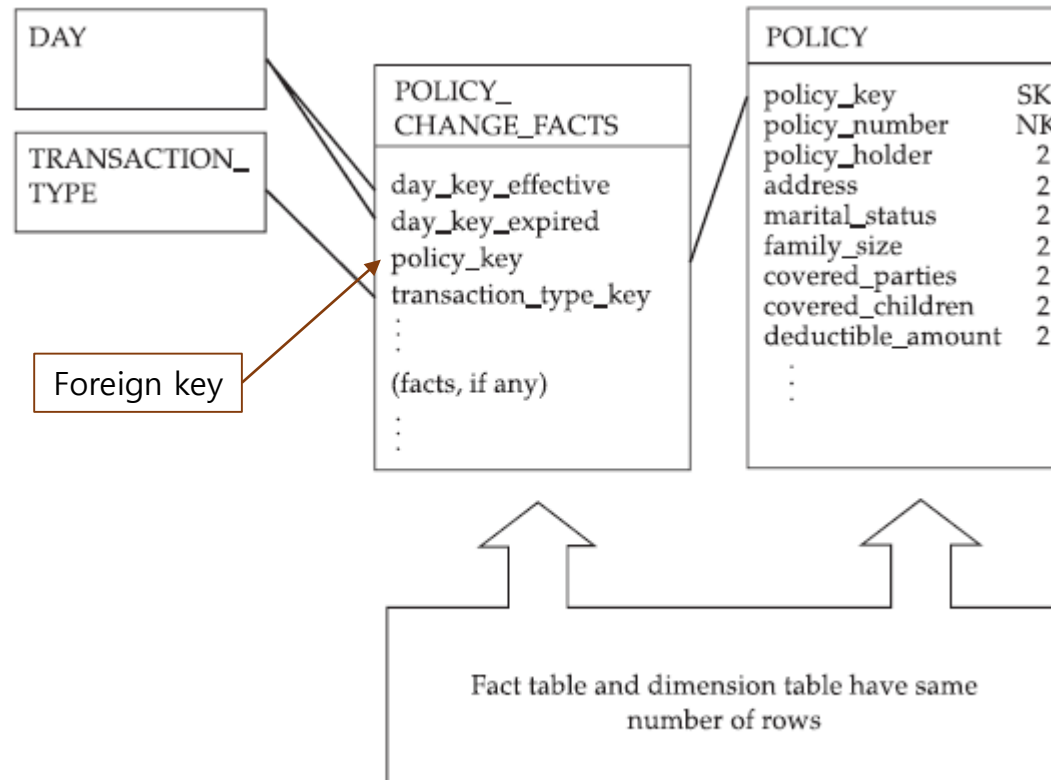


Figure 8-1 Type 2 changes in a policy dimension

# Time-stamped dimension

- Tracking change history through a fact table



**Figure 8-2** A fact table records the change history of the policy dimension

# Time-stamped dimension

- Time stamped version of policy dimension

```
WHERE 12/31/2006 >= effective_date AND
      12/31/2006 <= expiration_date
```

```
WHERE most_recent_version = "Current"
```

```
WHERE policy_number = 40111
ORDER_BY effective_date
```

Time-stamped version of policy dimension

POLICY									
policy_key									
policy_number									
policy_holder									
marital_status									
family_size									
covered_parties									
spouse_coverage									
covered_children									
deductible_amount									
...									
transaction_type									
effective_date									
expiration_date									
most_recent_version									

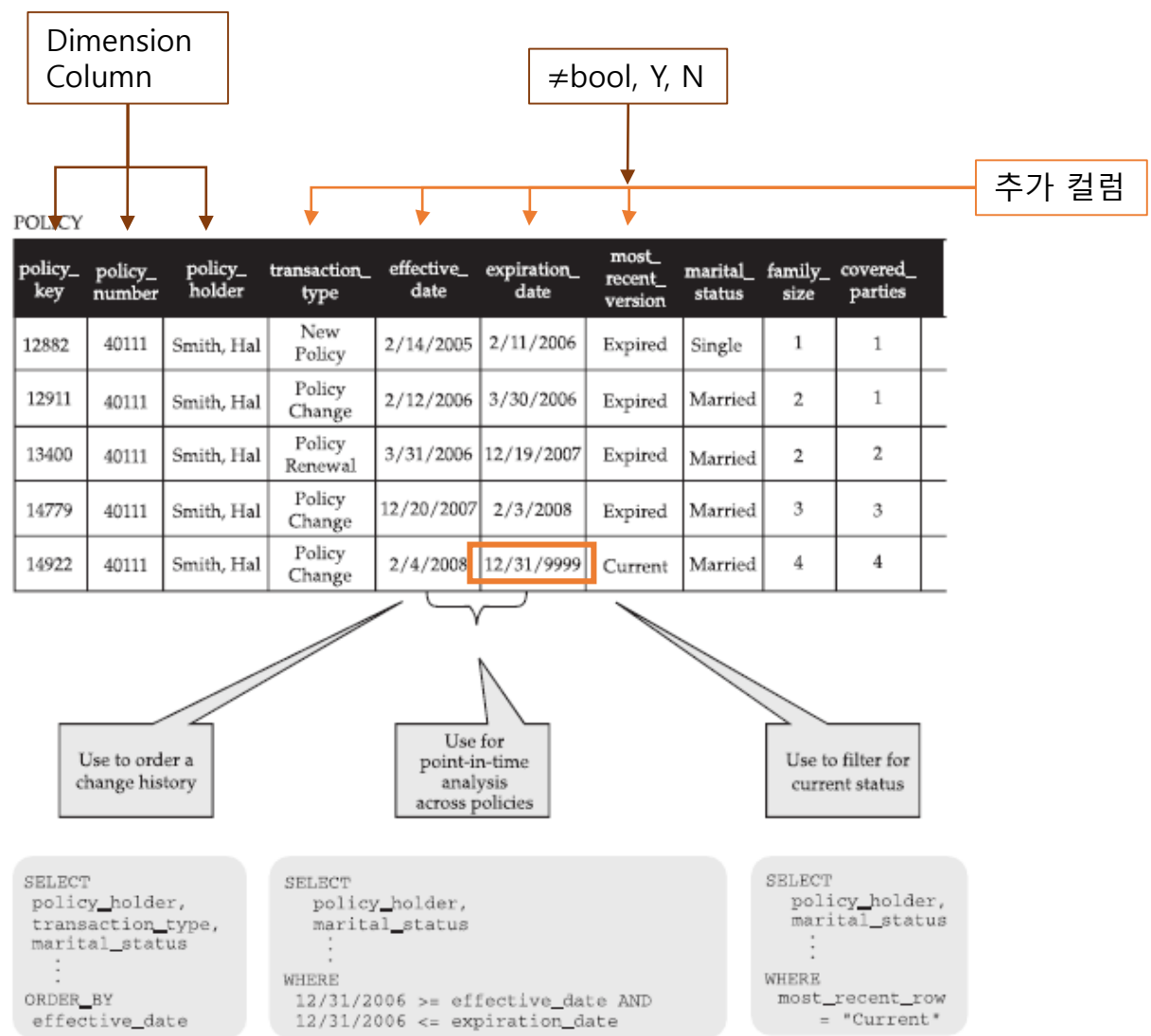


Figure 8-3 A fact table records the change history of the policy dimension

# Time-stamped dimension

- Multiple changes on a single day
  - 시간(24 values), 분(1440 values), 초(8640 values)
  - Expiration = 12:47 , effective = 12:48

```
WHERE
  12/31/2006 >= effective_date AND
  12/31/2006 <= expiration_date AND
  24:00 >= effective_time AND
  24:00 <= expiration_time
```



```
WHERE
  12/31/2006 >= effective_date AND
  12/31/2006 <= expiration_date AND
  last_change_of_day = "Final"
```

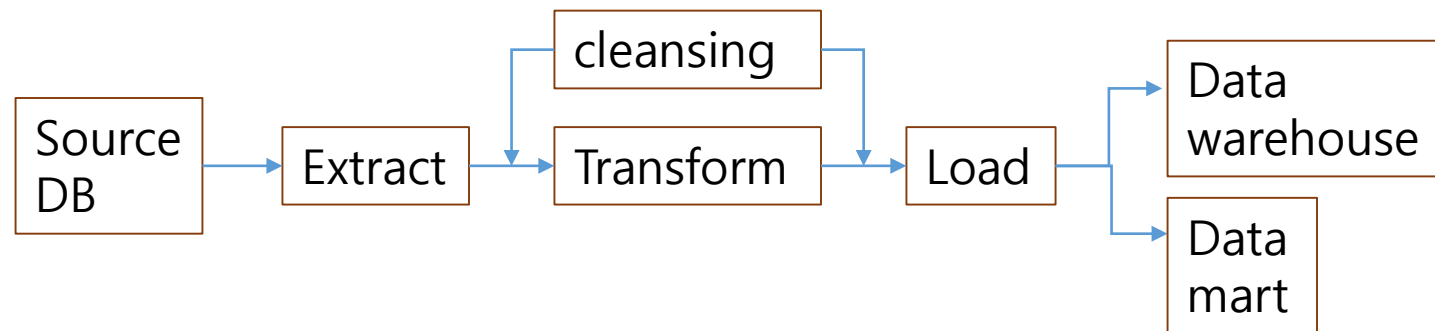
POLICY

policy_key	policy_number	policy_holder	transaction_type	effective_date	expiration_date	most_recent_version	marital_status	family_size	covered_parties
12882	40111	Smith, Hal	New Policy	2/14/2005	2/11/2006	Expired	Single	1	1
12911	40111	Smith, Hal	Policy Change	2/12/2006	3/30/2006	Expired	Married	2	1
13400	40111	Smith, Hal	Policy Renewal	3/31/2006	12/19/2007	Expired	Married	2	2
14779	40111	Smith, Hal	Policy Change	12/20/2007	2/3/2008	Expired	Married	3	3
14922	40111	Smith, Hal	Policy Change	2/4/2008	12/31/9999	Current	Married	4	4

겹치지 않아야 함

# Time-stamped dimension

- Time-stamped dimension with ETL process
- ETL process
  - Extract
    - Source databas로부터 data를 읽음
  - Transform
    - 추출된 data와 최종적으로 원하는 data 구성 및 형태를 연결하는 프로세스
  - Load
    - Target DB에 data를 저장하는 프로세스





# Time-stamped dimension

- Time stamped dimension
  - Ex) family\_size=4, deductible\_amount=500, policy\_number=40111 -> 대체키 14922
  - Dimension table에서 type 2 attribute와 source system을 비교해야 한다.
    - Chapter 17 – look up process

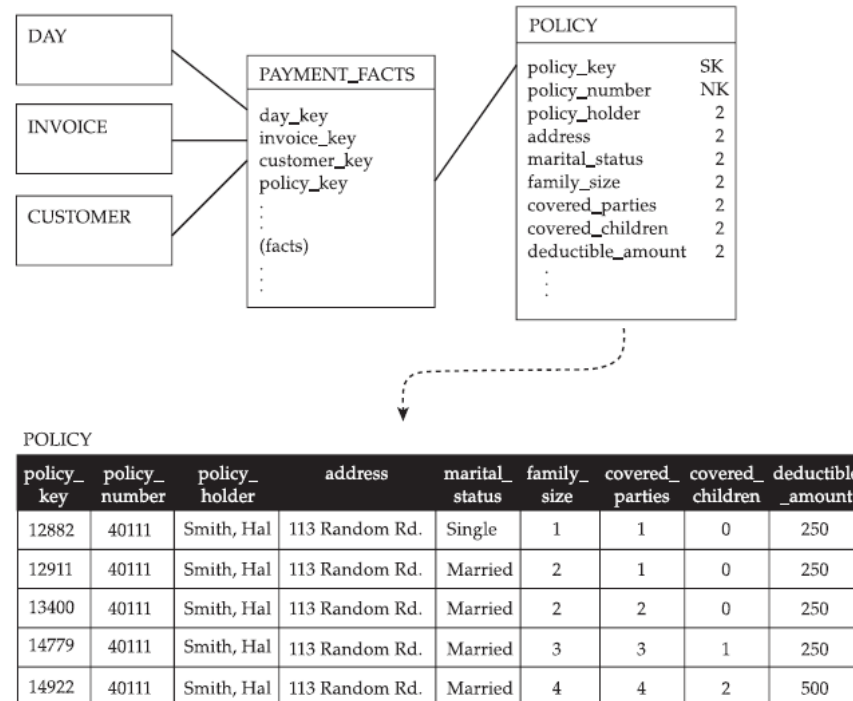
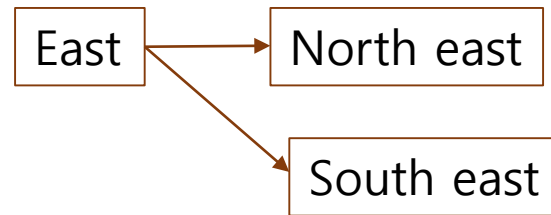


Figure 8-1 Type 2 changes in a policy dimension

# Type 3 change

- 요구사항
  - Change 이후에 기록된 모든 fact를 분석하기 위해 새로운 region을 사용
  - 이전 region을 이용하여 기록된 모든 fact를 분석
- Type 3 change
  - 쿼리에 대해 중복된 결과 방지



Legend:

- SK Surrogate Key
- NK Natural Key
- 1 Type 1
- 2 Type 2
- 3 Type 3 Pair

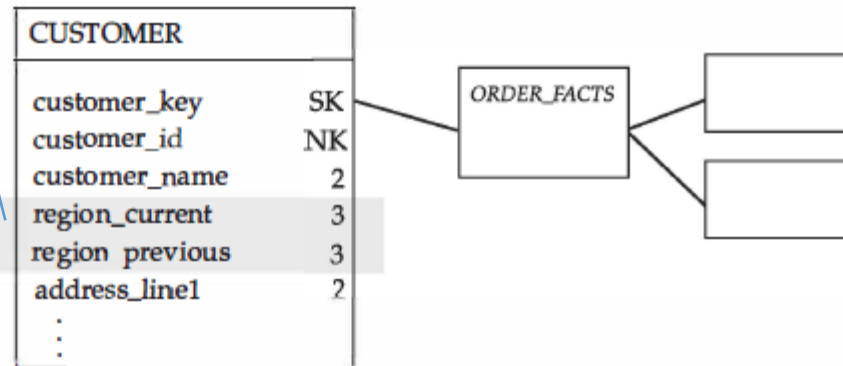


Figure 8-4 A type 3 attribute in the customer dimension

# Type3 changes

- 사용자가 모든 fact를 알고 싶을때 쓰인다.
- Change 이전과 이후의 data를 모두 기록

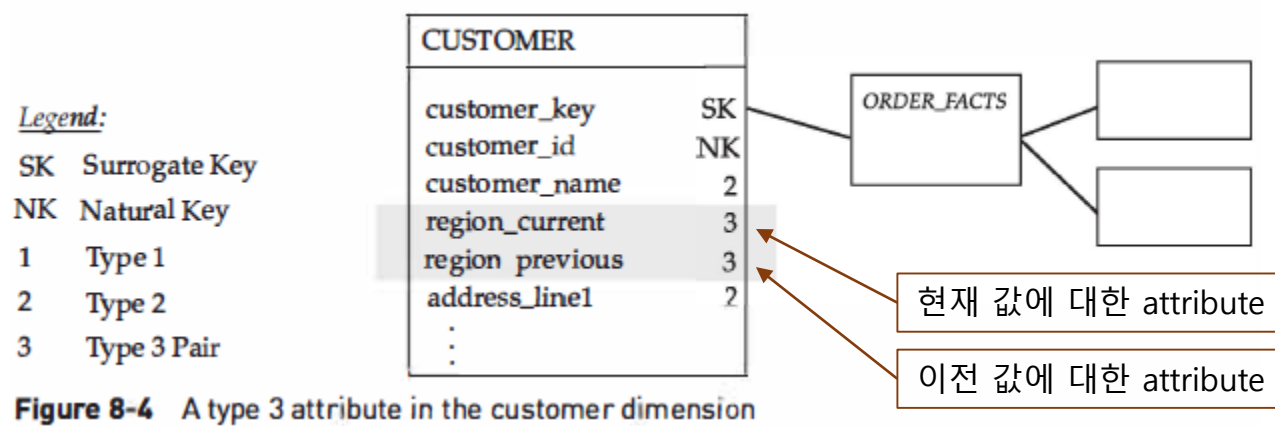


Figure 8-4 A type 3 attribute in the customer dimension

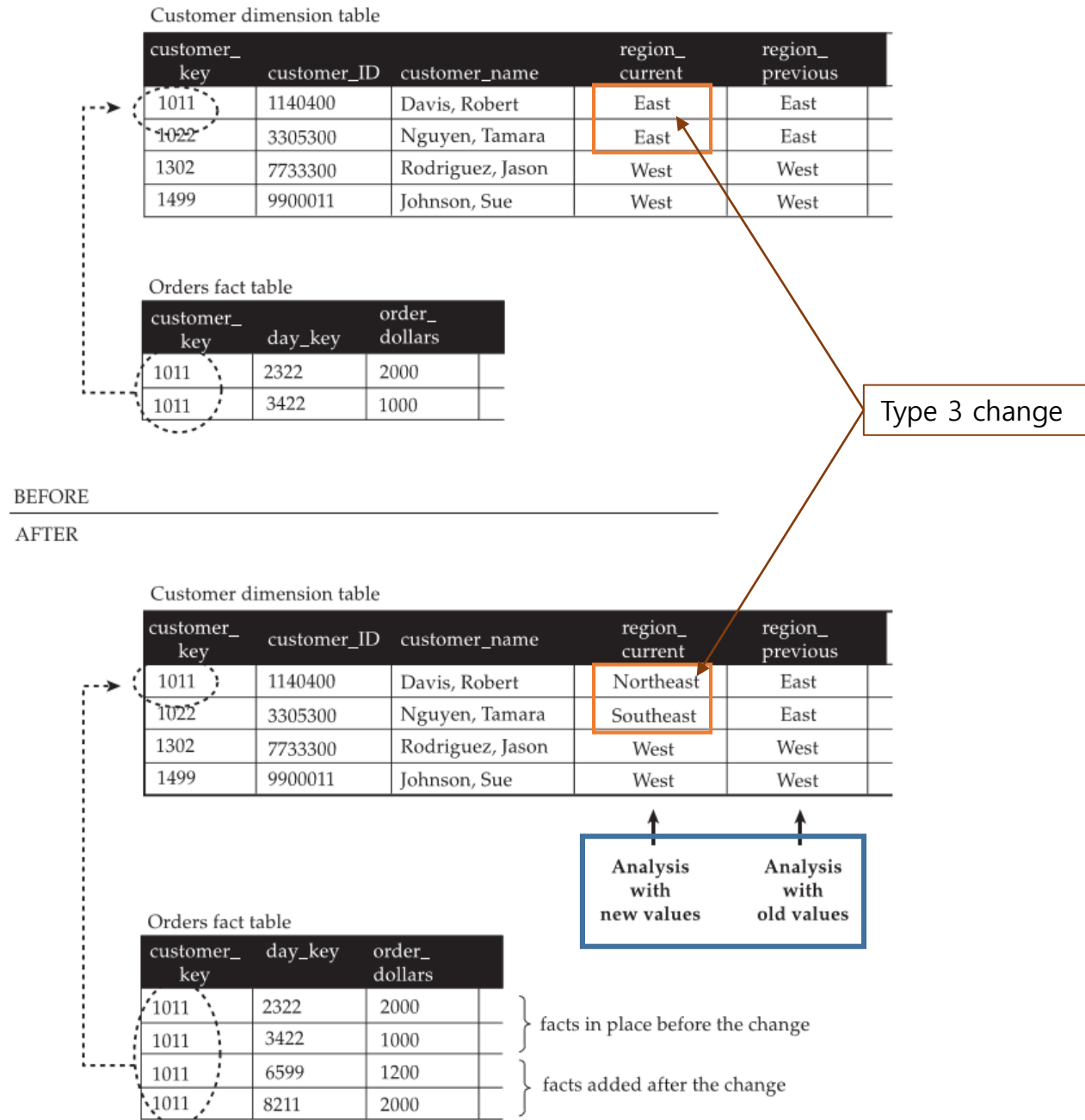


Figure 8-5 A type 3 change in action

# Type 3 changes

- Historic context를 보존하지는 않는다
- Region current 속성은 바뀐 데이터에 대해 type 1을 제공한다.
- Ex) 특정 쿼리에 대해 Robert가 어느 지역인가
  - Type 2 필요 -> hybrid technique

AFTER

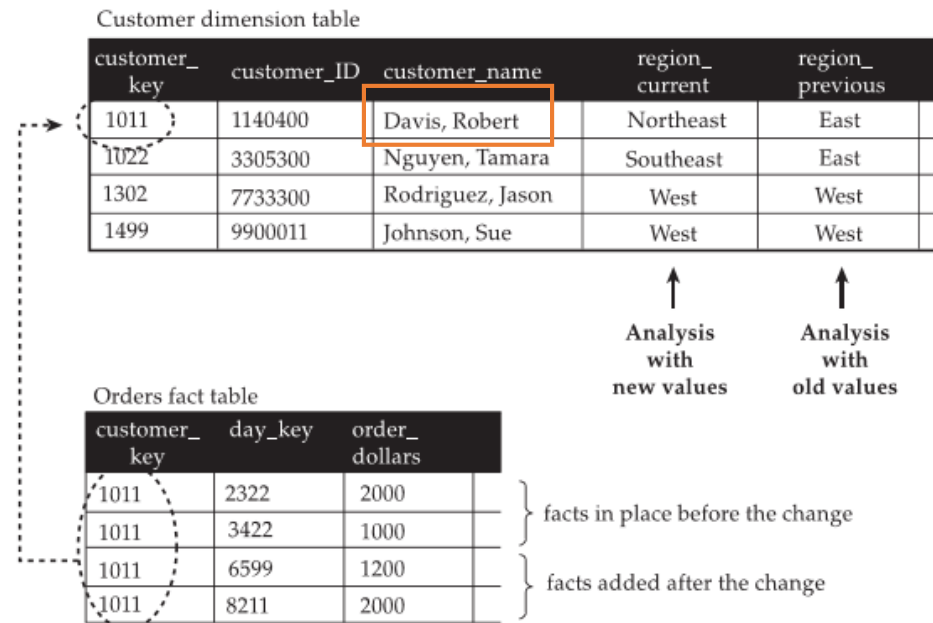


Figure 8-5 A type 3 change in action

# Type 3 changes

- Repeatable process

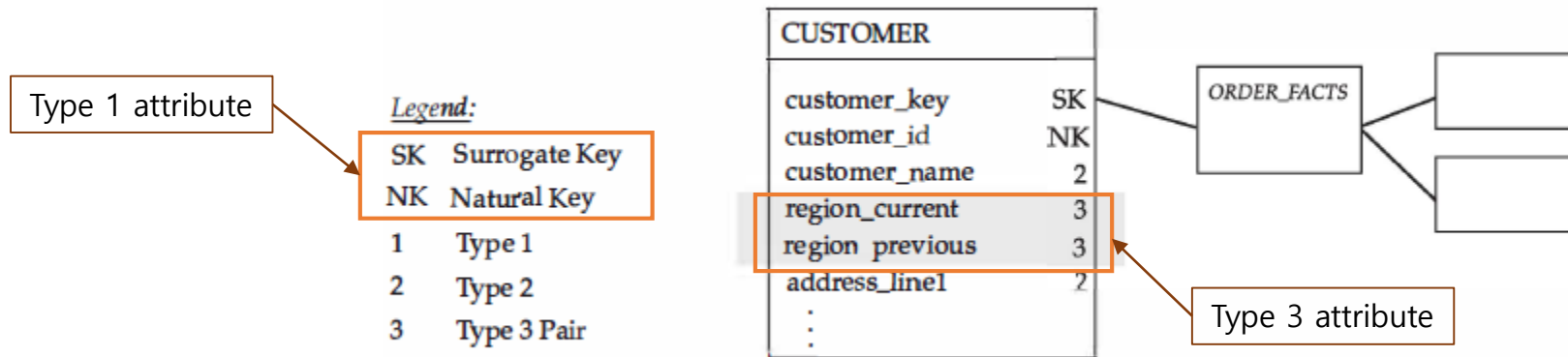
- Region\_current = Northeast -> Region\_previous = Northeast
- Region\_current = Southeast -> Region\_previous = Southeast

- 요구사항

- Ex)current version , two previous versions called
- 여러 attribute 이름 지정 Region\_current , Region\_last\_year , Region\_two\_years\_ago

# Type 3 changes

- Documenting the type 3 change



**Figure 8-4** A type 3 attribute in the customer dimension

# Hybrid technique

- Ex) 변경되는 회사명에 대해 어떤 회사명이 가장 매출이 높았는가
- Type 1/2 hybrid
- Type 1
  - Historic company name을 각 주문과 연결 짓지 못한다
- Type 2
  - 모든 기록을 알기 위해 현재 회사명을 사용한다

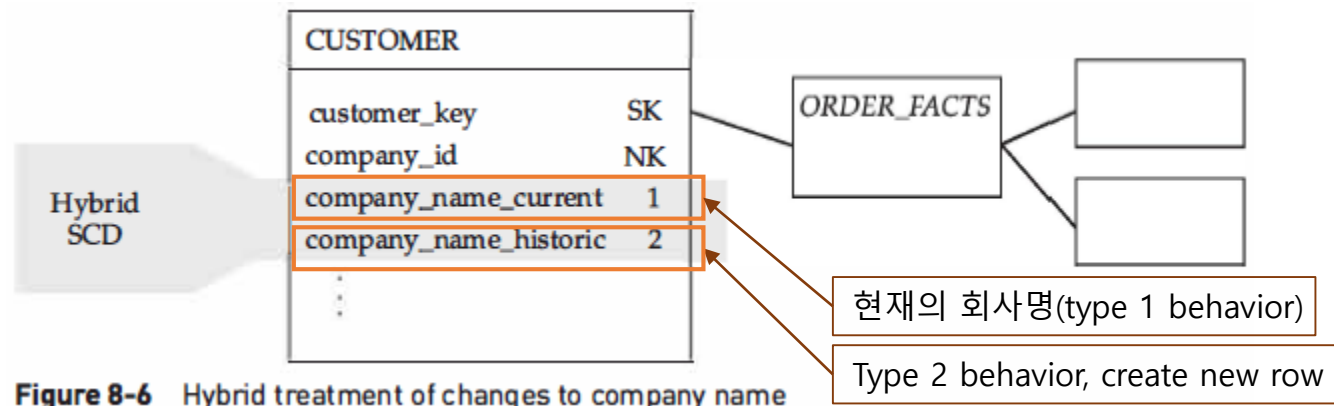
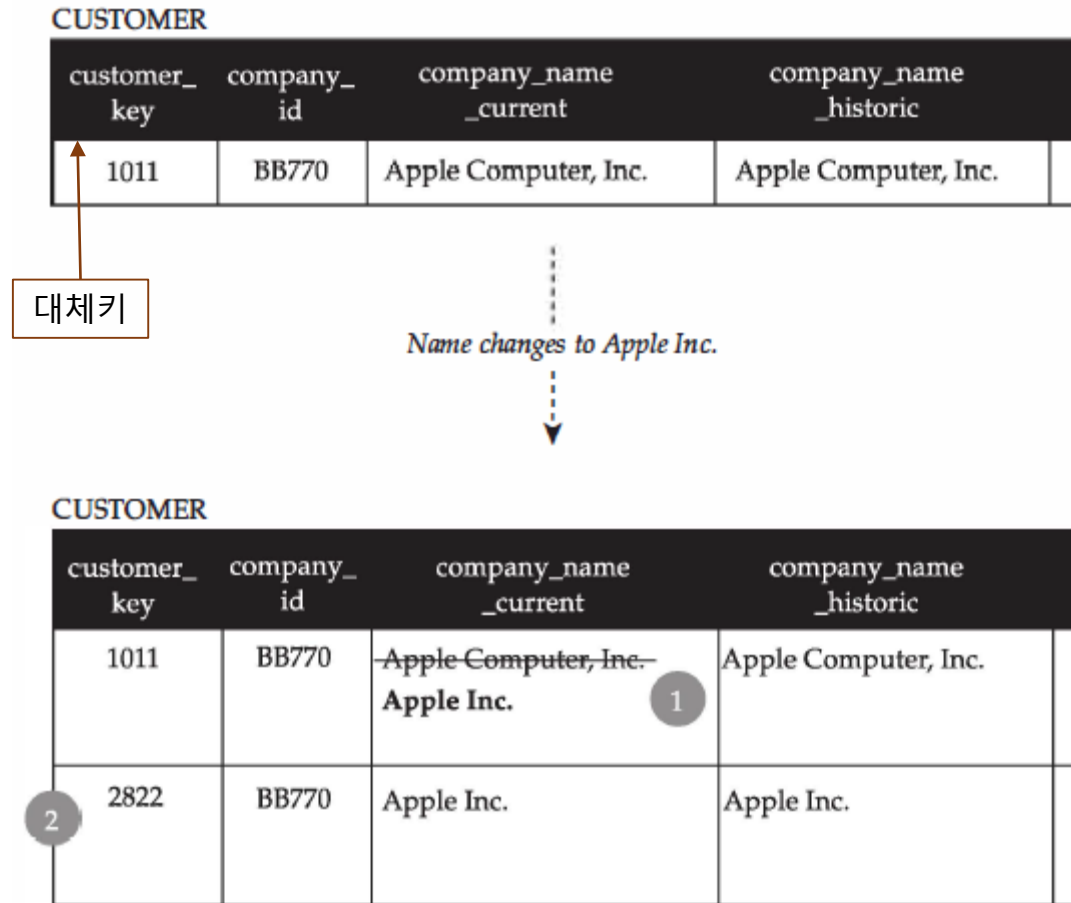


Figure 8-6 Hybrid treatment of changes to company name





① Old row(s) updated with new company\_name\_current

② Row is added with new name in both positions

Figure 8-7 A type 1/2 hybrid change in action

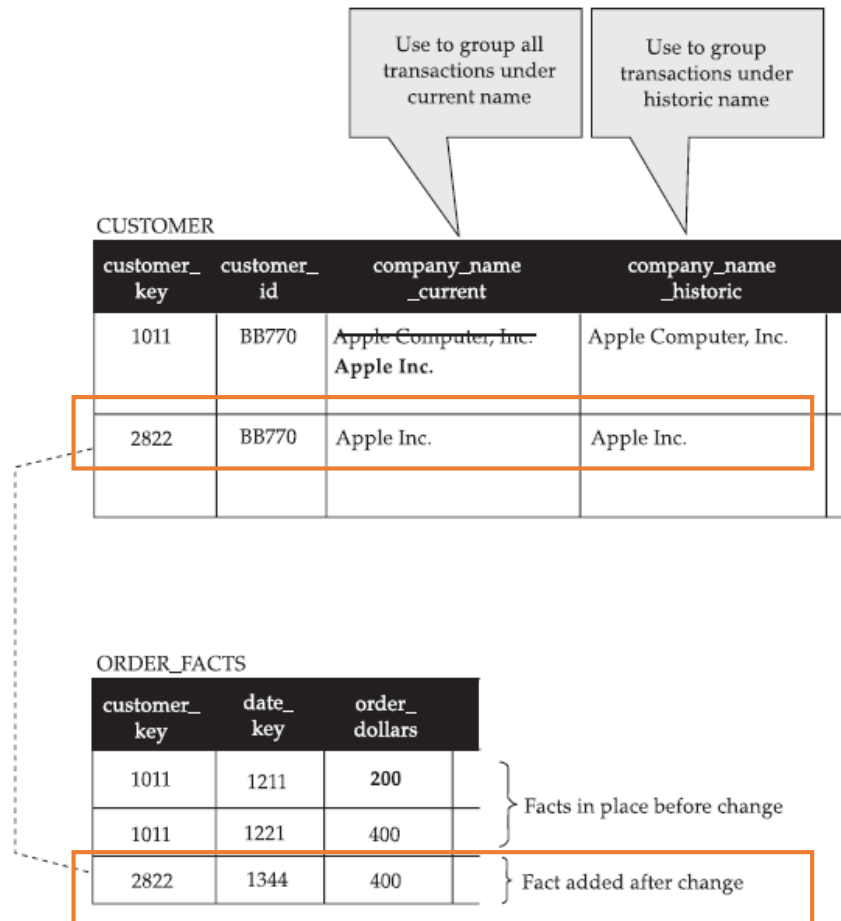


Figure 8-8 A hybrid response offers two analytic possibilities

CUSTOMER

customer_key	company_id	company_name_current	company_name_historic
1011	BB770	<del>Apple Computer, Inc.</del> <del>Apple Inc.</del> iApple Inc.	Apple Computer, Inc.
2822	BB770	<del>Apple Inc.</del> iApple Inc.	Apple Inc.
3100	BB770	iApple Inc.	iApple Inc.

ORDER\_FACTS

customer_key	date_key	order_dollars
1011	1211	200
1011	1221	400
2822	1344	400
3100	1411	1000

Figure 8-9 The company's name changes again

# Hybrid techniques

- Type 1/2/3 hybrid
  - 아주 드물게 필요
  - 3 columns
    - Current\_value
    - Previous\_value
    - Historic\_value
  - Complexity가 증가하므로 요구 사항에 맞게 사용한다
  - Type 1/2
- Type 1/3
  - Type 3 includes type 1 behavior column

# Summary

- Type 1, Type 2를 기반으로 한 type 3, time-stamped dimension, hybrid change에 대해서 다루었다.
- Time-stamped dimension
  - Point-in-time analysis
- Type 3 change
  - Study facts , change before, after
- Hybrid change
  - Different requirements
  - Type 1/2 hybrid