

FIT5195 – Business Intelligence and Data Warehousing

Week 6 – Temporal Data Warehousing

Semester 2, 2022

Developed by:
Dr. Soon Lay-Ki
Soon.LayKi@monash.edu



Learning Objectives

1. To understand the concepts of temporal data warehousing.
2. To be able to implement temporal data warehousing using bridge tables.
3. To understand the different types of Slowly Changing Dimensions (SCD).
4. To be able to implement SCD types using SQL.

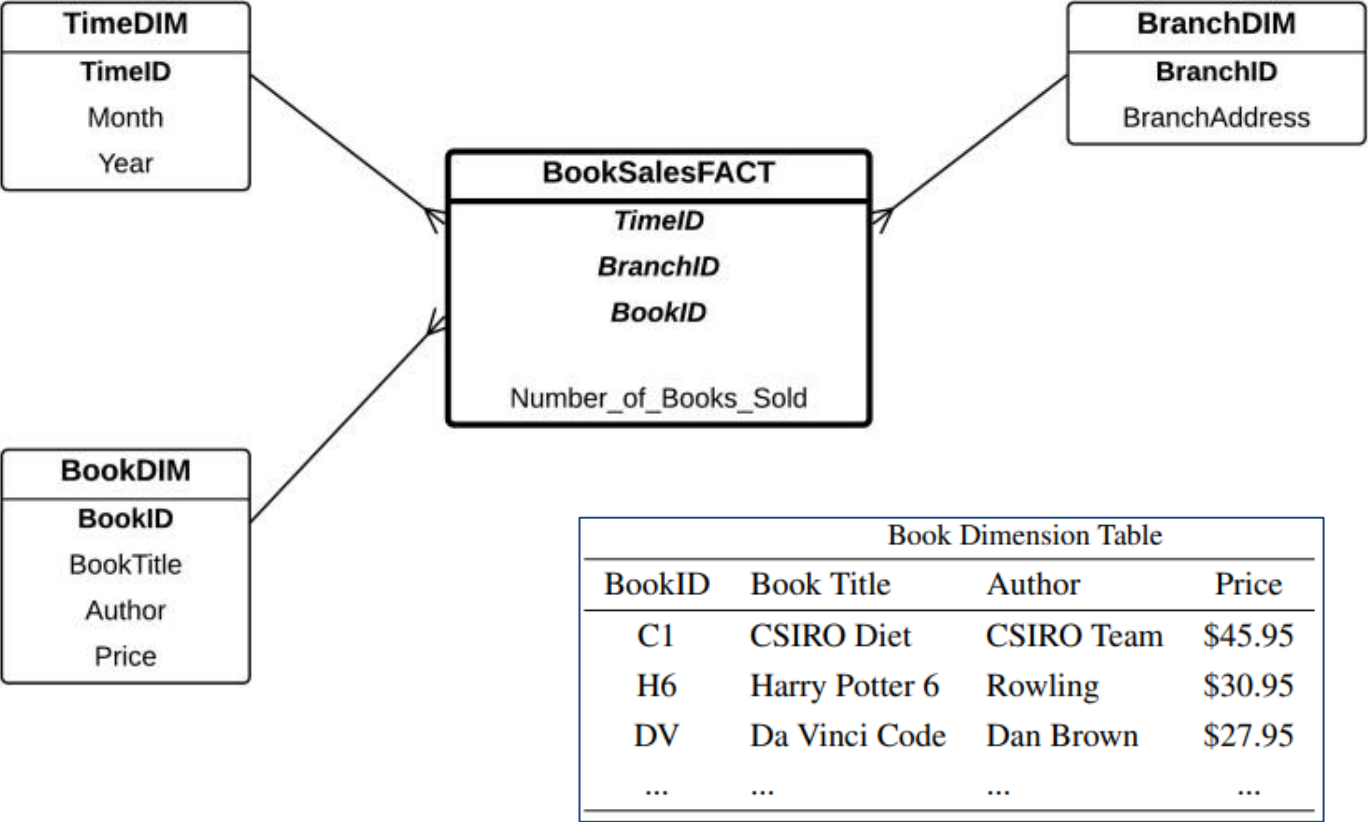
Outline

- Overview of Temporal Data Warehouse - A Bookshop Case Study
- Implementation of Temporal Data Warehousing
- Temporal Attributes
- Temporal Dimensions
- Slowly Changing Dimensions
 - SCD Type 0 and Type 1
 - SCD Type 2
 - SCD Type 3
 - SCD Type 4
 - SCD Type 6
- Implementation of SCD in SQL
- Creating the Fact Tables

Temporal Data Warehouse

- Temporal (or historical) aspect of records is incorporated into the data warehouse
 - E.g. keeping track of book price changes over time
- Also known as **Slowly Changing Dimensions (SCD)**

A Bookshop Case Study



BookSalesFact Table			
TimeID	BranchID	BookID	Number of Books Sold
Mar2008	City	C1	5
Mar2008	City	H6	15
Mar2008	City	DV	23
Mar2008	City
Mar2008	Chadstone	C1	15
Mar2008	Chadstone	H6	3
Mar2008	Chadstone	DV	2
Mar2008	Chadstone
Mar2008	Camberwell	C1	1
Mar2008	Camberwell	H6	1
Mar2008	Camberwell	DV	2
Mar2008	Camberwell
Mar2008
...
...
Dec2007	City	C1	15
Dec2007	City	H6	6
Dec2007	City	DV	6
Dec2007	City
Dec2007	Chadstone	C1	10
Dec2007	Chadstone	H6	8
Dec2007	Chadstone	DV	1
Dec2007	Chadstone
Dec2007	Camberwell	C1	18
Dec2007	Camberwell	H6	3
Dec2007	Camberwell	DV	2
Dec2007	Camberwell
Dec2007
...

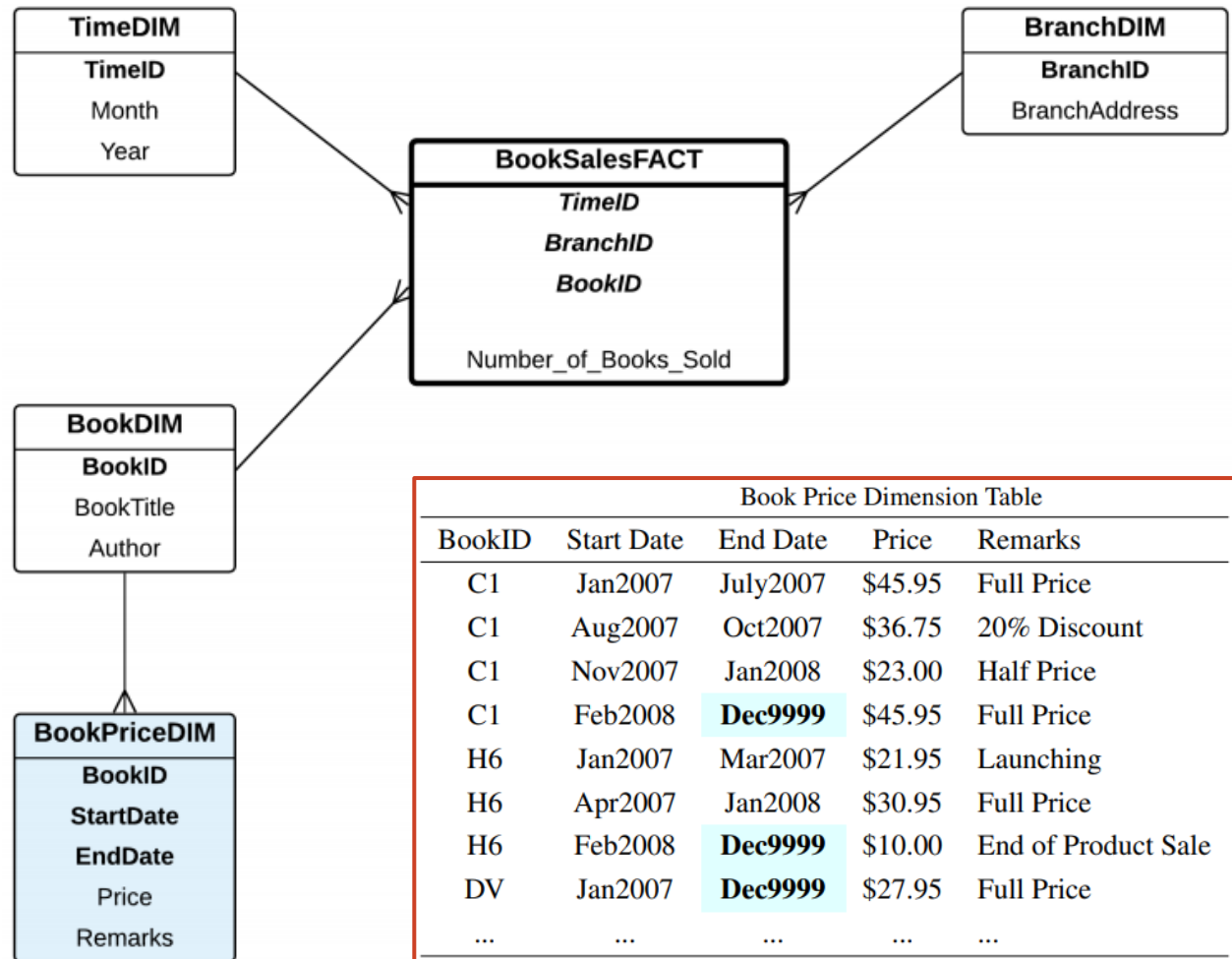
A Bookshop Case Study - Report 1

Report 1 (Book Sales Fact with Book Dimension)						
TimeID	BranchID	BookID	Book Title	Author	Price	Number of Books Sold
Mar2008	City	C1	CSIRO Diet	CSIRO Team	\$45.95	5
Mar2008	City	H6	Harry Potter 6	Rowling	\$30.95	15
Mar2008	City	DV	Da Vinci Code	Dan Brown	\$27.95	23
Mar2008	City
Mar2008	Chadstone	C1	CSIRO Diet	CSIRO Team	\$45.95	15
Mar2008	Chadstone	H6	Harry Potter 6	Rowling	\$30.95	3
Mar2008	Chadstone	DV	Da Vinci Code	Dan Brown	\$27.95	2
Mar2008	Chadstone
Mar2008	Camberwell	C1	CSIRO Diet	CSIRO Team	\$45.95	1
Mar2008	Camberwell	H6	Harry Potter 6	Rowling	\$30.95	1
Mar2008	Camberwell	DV	Da Vinci Code	Dan Brown	\$27.95	2
Mar2008	Camberwell
Mar2008
...
...
Dec2007	City	C1	CSIRO Diet	CSIRO Team	\$45.95	15
Dec2007	City	H6	Harry Potter 6	Rowling	\$30.95	6
Dec2007	City	DV	Da Vinci Code	Dan Brown	\$27.95	6
Dec2007	City

A Bookshop Case Study - with Bridge Table

- Use a bridge table, BookPriceDIM to store the history of book prices
- Implemented as a **Weak Entity**
 - Composite key: BookID, StartDate, EndDate

Book Dimension Table		
BookID	Book Title	Author
C1	CSIRO Diet	CSIRO Team
H6	Harry Potter 6	Rowling
DV	Da Vinci Code	Dan Brown
...



A Bookshop Case Study

- Report 2

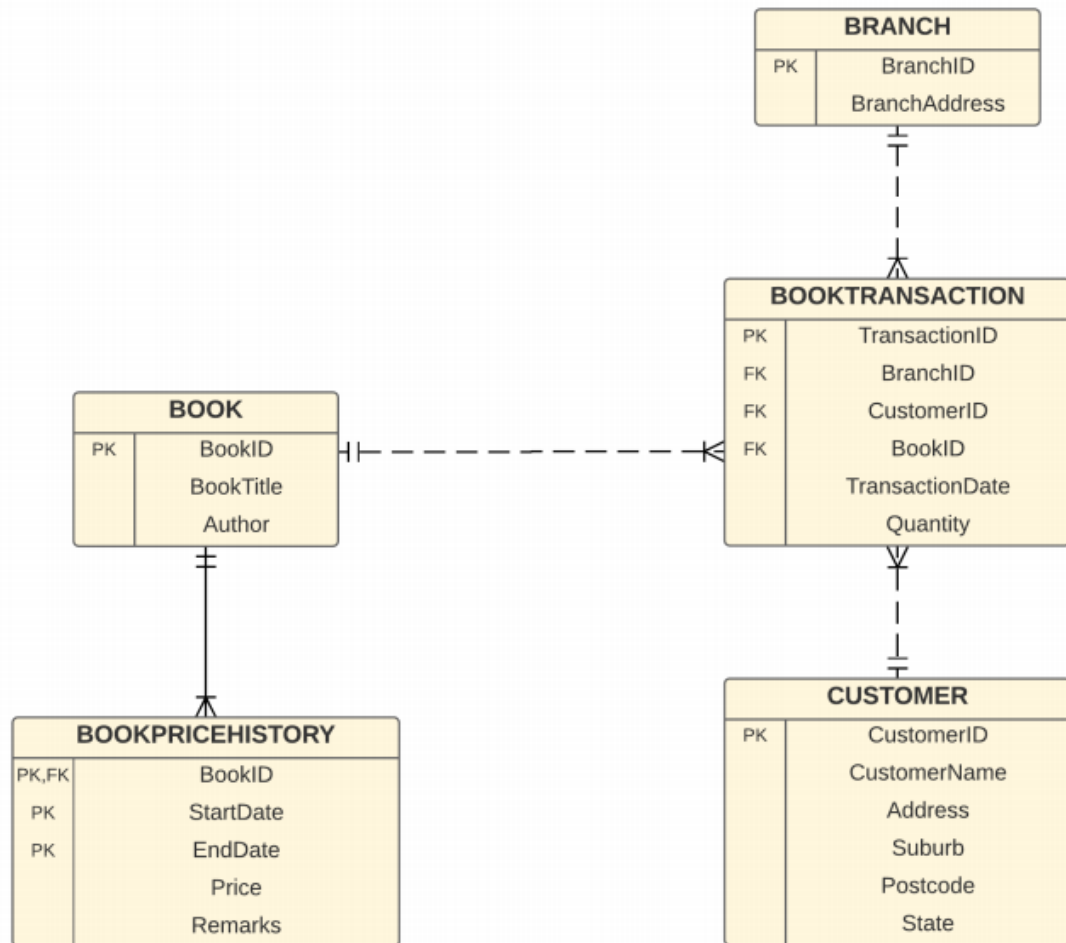
- The correct sales price is indicated according to the TimeID
- Enables more accurate analyses

Report 2 with the correct Book Price

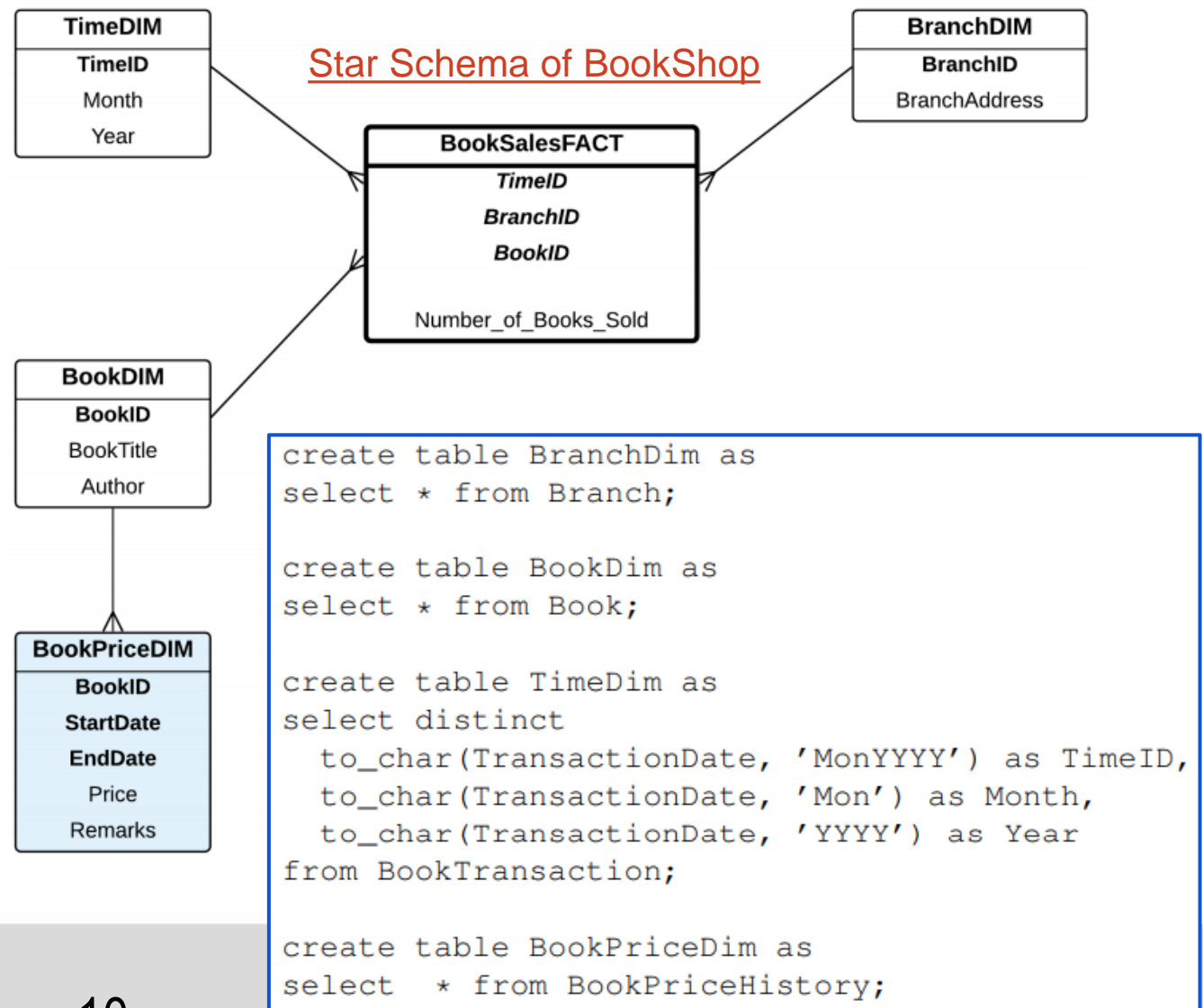
TimeID	BranchID	BookID	Book Title	Author	Price	Number of Books Sold
Mar2008	City	C1	CSIRO Diet	CSIRO Team	\$45.95	5
Mar2008	City	H6	Harry Potter 6	Rowling	\$10.00	15
Mar2008	City	DV	Da Vinci Code	Dan Brown	\$27.95	23
Mar2008	City
Mar2008	Chadstone	C1	CSIRO Diet	CSIRO Team	\$45.95	15
Mar2008	Chadstone	H6	Harry Potter 6	Rowling	\$10.00	3
Mar2008	Chadstone	DV	Da Vinci Code	Dan Brown	\$27.95	2
Mar2008	Chadstone
Mar2008	Camberwell	C1	CSIRO Diet	CSIRO Team	\$45.95	1
Mar2008	Camberwell	H6	Harry Potter 6	Rowling	\$10.00	1
Mar2008	Camberwell	DV	Da Vinci Code	Dan Brown	\$27.95	2
Mar2008	Camberwell
Mar2008
...
...
Dec2007	City	C1	CSIRO Diet	CSIRO Team	\$23.00	15
Dec2007	City	H6	Harry Potter 6	Rowling	\$30.95	6
Dec2007	City	DV	Da Vinci Code	Dan Brown	\$27.95	6
Dec2007	City
Dec2007	Chadstone	C1	CSIRO Diet	CSIRO Team	\$23.00	10
Dec2007	Chadstone	H6	Harry Potter 6	Rowling	\$30.95	8
Dec2007	Chadstone	DV	Da Vinci Code	Dan Brown	\$27.95	1
Dec2007	Chadstone
Dec2007	Camberwell	C1	CSIRO Diet	CSIRO Team	\$23.00	18
Dec2007	Camberwell	H6	Harry Potter 6	Rowling	\$30.95	3
Dec2007	Camberwell	DV	Da Vinci Code	Dan Brown	\$27.95	2

Implementation of Temporal DW

Implementation of Temporal Data Warehousing



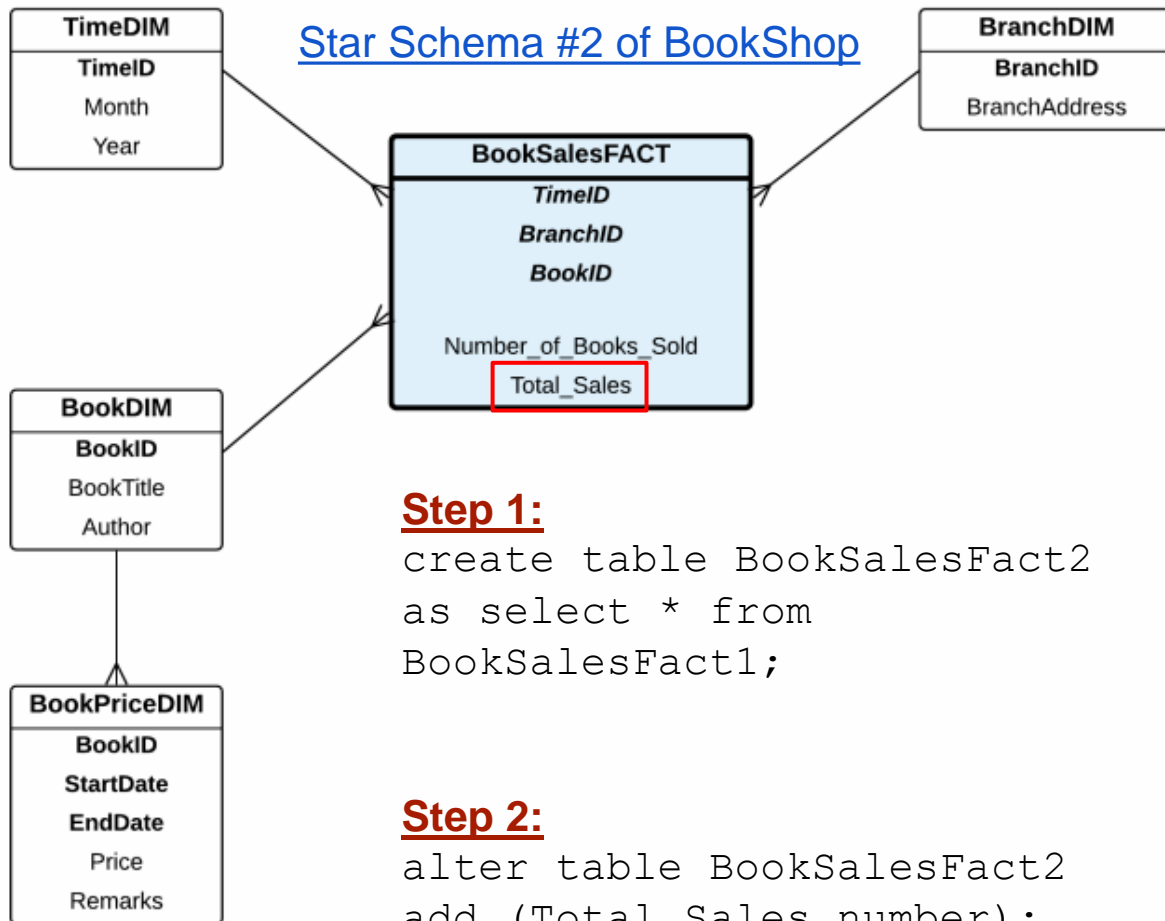
ER Diagram of Bookshop



Implementation of Temporal Data Warehousing - Creating Fact Table

```
create table BookSalesFact1
as
select to_char(T.TransactionDate, 'MonYYYY') as TimeID,
       BK.BookID, BR.BranchID,
       sum(T.Quantity) as Number_of_Books_Sold
from BookTransaction T, Book BK, Branch BR
where T.BranchID = BR.BranchID
       and T.BookID = BK.BookID
group by to_char(T.TransactionDate, 'MonYYYY'), BK.BookID,
BR.BranchID;
```

Implementation of Temporal Data Warehousing - Star Schema #2



Step 3:

```
declare
cursor PriceCursor is
    select * from BookPriceDim;
begin
    for Item in PriceCursor loop
        -- update value for Total_Sales in
        -- BookSalesFact2
        update BookSalesFact2
        set Total_Sales =
            Number_Of_Books_Sold * Item.Price
        where BookID = Item.BookID
        and to_date(TimeID, 'MonYYYY') >=
            to_date(Item.StartDate,
                'MonYYYY')
        and to_date(TimeID, 'MonYYYY') <=
            to_date(Item.EndDate, 'MonYYYY');
    end loop;
end;
```

Implementation of Temporal Data Warehousing - Star Schema #2 (cont.)

```
-- if BookSalesFact1 has not been created
create table BookSalesFact2
as
select to_char(T.TransactionDate, 'MonYYYY') as TimeID,
       BK.BookID, BR.BranchID,
       sum(T.Quantity) as Number_Of_Books_Sold,
       sum(T.Quantity * BP.Price) as Total_Sales
from BookTransaction T, Book BK, Branch BR, BookPriceHistory BP
where T.BranchID = BR.BranchID
      and T.BookID = BK.BookID
      and BK.BookID = BP.BookID
      and T.TransactionDate >= to_date(BP.StartDate, 'MonYYYY')
      and T.TransactionDate <= to_date(BP.EndDate, 'MonYYYY')
group by to_char(T.TransactionDate, 'MonYYYY'), BK.BookID, BR.BranchID;
```

Temporal Attributes

Temporal Attributes

- Temporal attribute
 - An attribute in which the value of that attribute has a life-span
 - In this example, each *book price* has a life-span, and it is determined by the *StartDate* and *EndDate* attributes in the *BookPriceDIM* table
 - *BookPriceDIM* table is a RDBMS implementation of a temporal data warehousing

Book Dimension Table		
BookID	Book Title	Author
C1	CSIRO Diet	CSIRO Team
H6	Harry Potter 6	Rowling
DV	Da Vinci Code	Dan Brown
...

1:m

Book Price Dimension Table				
BookID	Start Date	End Date	Price	Remarks
C1	Jan2007	July2007	\$45.95	Full Price
C1	Aug2007	Oct2007	\$36.75	20% Discount
C1	Nov2007	Jan2008	\$23.00	Half Price
C1	Feb2008	Dec9999	\$45.95	Full Price
H6	Jan2007	Mar2007	\$21.95	Launching
H6	Apr2007	Jan2008	\$30.95	Full Price
H6	Feb2008	Dec9999	\$10.00	End of Product Sale
DV	Jan2007	Dec9999	\$27.95	Full Price
...

Generating Report 2

Report 2 with the correct Book Price

TimeID	BranchID	BookID	Book Title	Author	Price	Number of Books Sold
Mar2008	City	C1	CSIRO Diet	CSIRO Team	\$45.95	5
Mar2008	City	H6	Harry Potter 6	Rowling	\$10.00	15
Mar2008	City	DV	Da Vinci Code	Dan Brown	\$27.95	23
Mar2008	City
Mar2008	Chadstone	C1	CSIRO Diet	CSIRO Team	\$45.95	15
Mar2008	Chadstone	H6	Harry Potter 6	Rowling	\$10.00	3
Mar2008	Chadstone	DV	Da Vinci Code	Dan Brown	\$27.95	2
Mar2008	Chadstone
Mar2008	Camberwell	C1	CSIRO Diet	CSIRO Team	\$45.95	1
Mar2008	Camberwell	H6	Harry Potter 6	Rowling	\$10.00	1
Mar2008	Camberwell	DV	Da Vinci Code	Dan Brown	\$27.95	2
Mar2008	Camberwell
Mar2008
...
...
Dec2007	City	C1	CSIRO Diet	CSIRO Team	\$23.00	15
Dec2007	City	H6	Harry Potter 6	Rowling	\$30.95	6
Dec2007	City	DV	Da Vinci Code	Dan Brown	\$27.95	6
Dec2007	City
Dec2007	Chadstone	C1	CSIRO Diet	CSIRO Team	\$23.00	10
Dec2007	Chadstone	H6	Harry Potter 6	Rowling	\$30.95	8
Dec2007	Chadstone	DV	Da Vinci Code	Dan Brown	\$27.95	1
Dec2007	Chadstone
Dec2007	Camberwell	C1	CSIRO Diet	CSIRO Team	\$23.00	18
Dec2007	Camberwell	H6	Harry Potter 6	Rowling	\$30.95	3
Dec2007	Camberwell	DV	Da Vinci Code	Dan Brown	\$27.95	2

```

select
    F.TimeID,
    F.BranchID,
    F.BookID,
    B.BookTitle,
    B.Author,
    P.Price,
    F.Number_of_Books_Sold
from BookSalesFact F, BookDim B, BookPriceDim P
where F.BookID = B.BookID
and B.BookID = P.BookID
and to_date(F.TimeID, 'MonYYYY') >=
    to_date(P.StartDate, 'MonYYYY')
and to_date(F.TimeID, 'MonYYYY') <=
    to_date(P.EndDate, 'MonYYYY');

```

A Bookshop Case Study - Potential Problem on Time Granularity

BookPriceDim Table

BookID	Start Date	End Date	Price	Remarks
C1	Jan2007	July2007	\$45.95	Full Price
C1	Aug2007	Oct2007	\$36.75	20% Discount
C1	Nov2007	15Jan2008	\$23.00	Half Price
C1	16Jan2008	Dec9999	\$45.95	Full Price
H6	Jan2007	Mar2007	\$21.95	Launching
H6	Apr2007	Jan2008	\$30.95	Full Price
H6	Feb2008	Dec9999	\$10.00	End of Product Sale
DV	Jan2007	Dec9999	\$27.95	Full Price
...

The price of Book C1 changes within January instead of 1 February 2008

TimeID = Jan2007

```
and to_date(F.TimeID, 'MonYYYY') >=
to_date(P.StartDate, 'MonYYYY')
and to_date(F.TimeID, 'MonYYYY') <=
to_date(P.EndDate, 'MonYYYY');
```

A Bookshop Case Study - Incorrect Report 3

Report 3 – An Incorrect Report

TimeID	BranchID	BookID	Book Title	Author	Price	Number of Books Sold
Jan2008	City	C1	CSIRO Diet	CSIRO Team	\$23.00	25
Jan2008	City	C1	CSIRO Diet	CSIRO Team	\$45.95	25
Jan2008	City	H6	Harry Potter 6	Rowling	\$30.95	10
Jan2008	City	DV	Da Vinci Code	Dan Brown	\$27.95	7
Jan2008	City
Jan2008	Chadstone	C1	CSIRO Diet	CSIRO Team	\$23.00	30
Jan2008	Chadstone	C1	CSIRO Diet	CSIRO Team	\$45.05	30
Jan2008	Chadstone	H6	Harry Potter 6	Rowling	\$30.95	15
Jan2008	Chadstone	DV	Da Vinci Code	Dan Brown	\$27.95	5
Jan2008	Chadstone
Jan2008	Camberwell	C1	CSIRO Diet	CSIRO Team	\$23.00	20
Jan2008	Camberwell	C1	CSIRO Diet	CSIRO Team	\$45.05	20
Jan2008	Camberwell	H6	Harry Potter 6	Rowling	\$30.95	5
Jan2008	Camberwell	DV	Da Vinci Code	Dan Brown	\$27.95	5
Jan2008	Camberwell
Jan2008
...

A Bookshop Case Study - Solving the Problem of Report 3

- Display "two" prices on the same record in the report.

Report 4 – multiple book prices on one month

TimeID	BranchID	BookID	Book Title	Author	Price	Number of Books Sold
Jan2008	City	C1	CSIRO Diet	CSIRO Team	\$23.00;\$45.95	25
Jan2008	City	H6	Harry Potter 6	Rowling	\$30.95	10
Jan2008	City	DV	Da Vinci Code	Dan Brown	\$27.95	7
Jan2008	City
Jan2008	Chadstone	C1	CSIRO Diet	CSIRO Team	\$23.00;\$45.95	30
Jan2008	Chadstone	H6	Harry Potter 6	Rowling	\$30.95	15
Jan2008	Chadstone	DV	Da Vinci Code	Dan Brown	\$27.95	5
Jan2008	Chadstone
Jan2008	Camberwell	C1	CSIRO Diet	CSIRO Team	\$23.00;\$45.95	20
Jan2008	Camberwell	H6	Harry Potter 6	Rowling	\$30.95	5
Jan2008	Camberwell	DV	Da Vinci Code	Dan Brown	\$27.95	5
Jan2008	Camberwell
Jan2008
...

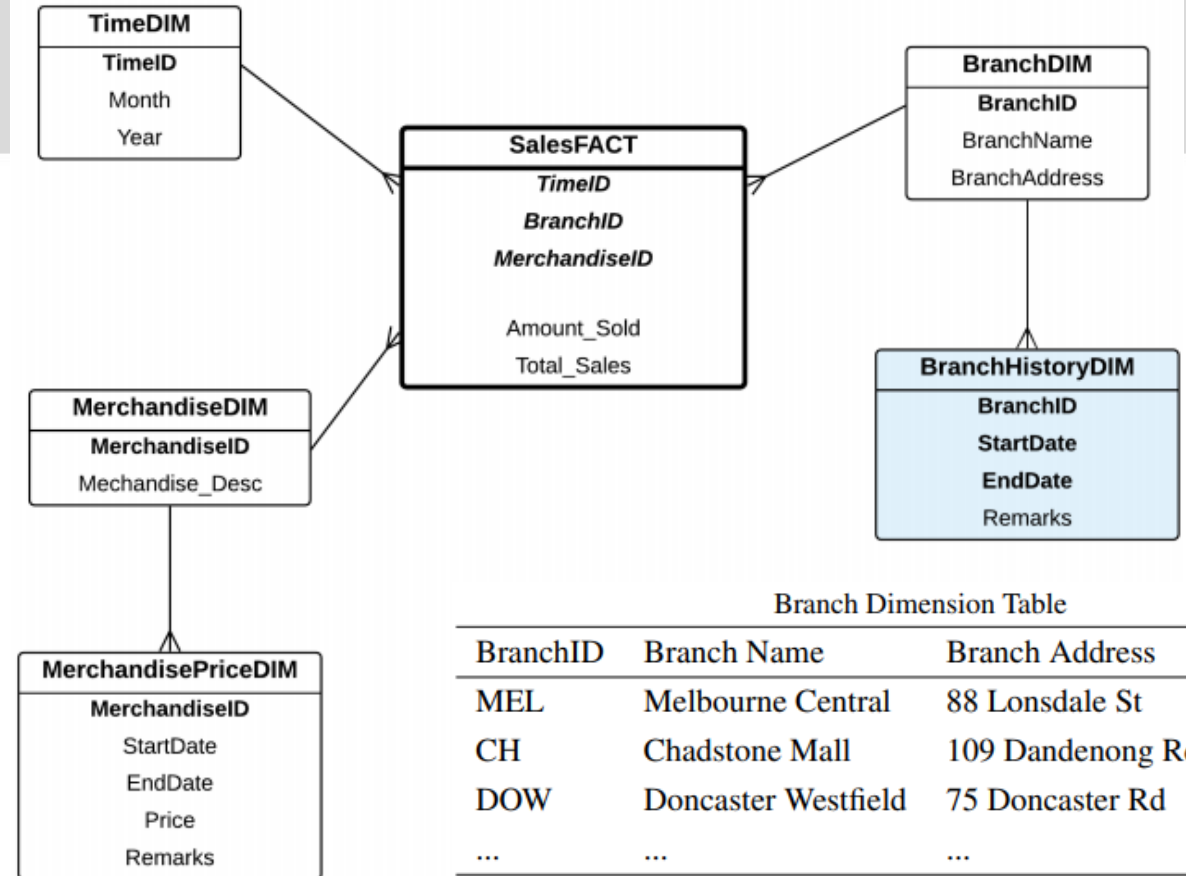
What about *Total_Sales*?

```
select
    F.TimeID,
    F.BranchID,
    F.BookID,
    B.BookTitle,
    B.Author,
    listagg(P.Price, ';') within group (order by P.Price)
    as Price,
    F.Number_of_Books_Sold
from BookSalesFact F, BookDim B, BookPriceDim P
where F.BookID = B.BookID
and B.BookID = P.BookID
and to_date(F.TimeID, 'MonYYYY') >=
    to_date(P.StartDate, 'MonYYYY')
and to_date(F.TimeID, 'MonYYYY') <=
    to_date(P.EndDate, 'MonYYYY')
group by
    F.TimeID,
    F.BranchID,
    F.BookID,
    B.BookTitle,
    B.Author,
    F.Number_of_Books_Sold;
```

Temporal Dimensions

Temporal Dimensions

- Temporal Dimension is a dimension where **the record of the dimension has a specific life span**.
- Case study: a mobile calendar shop
- Temporal attribute
 - MerchandisePriceDIM*
- Temporal dimension:
 - BranchHistoryDIM*



BranchHistoryDim Table

BookID	Start Date	End Date	Remarks	Contact Number
MEL	Jan0000	Dec9999	Main shop	(03) 9859 8070
CH	Oct2007	Mar2008		0411 848 821
CH	Oct2008	Feb2009	Under re-construction	0413 356 665
CH	Oct2009	Feb2010		0412 313 313
DOW	Nov2007	Feb2008		0427 123 456
DOW	Nov2008	Feb2009		0427 123 456
DOW	Oct2009	Feb2010		0427 123 456
...

A Mobile Calendar Shop - A Report

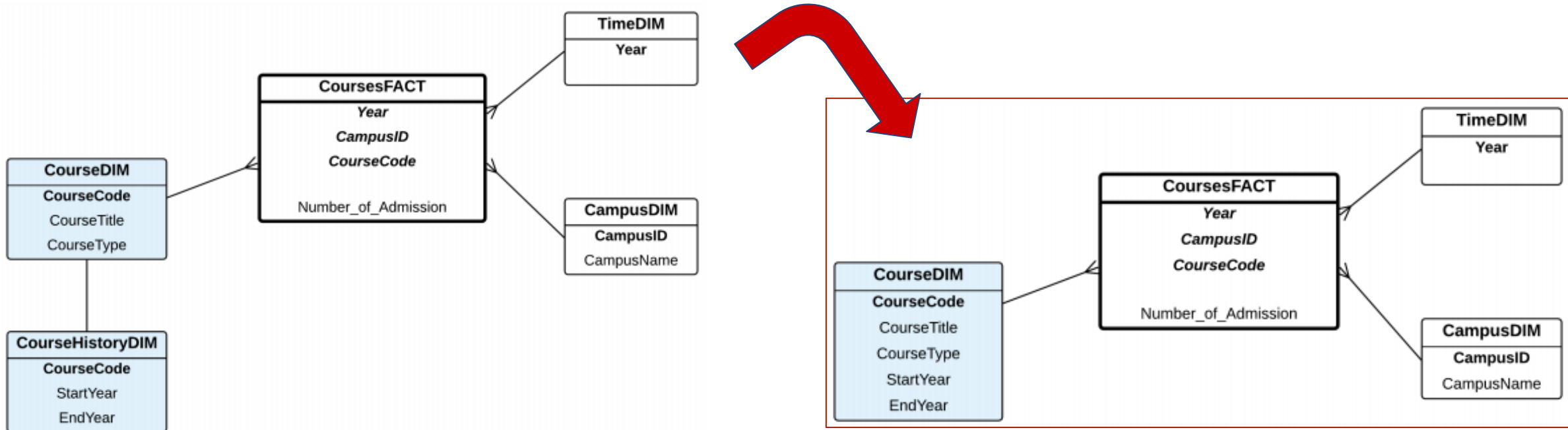
Report – SalesFact joined with Branch Dimension and Branch History Dimension

TimeID	BranchID	Branch Name	Branch Address	Start Date	End Date	Remarks	ContactNo	Merchand	Amount	TotSales
Oct2007	MEL	Melbourne Central	88 Lonsdale St	Jan0000	Dec9999	Main shop
Oct2007	CH	Chadstone Mall	109 Dandenong Rd	Oct2007	Mar2008	
Nov2007	MEL	Melbourne Central	88 Lonsdale St	Jan0000	Dec9999	Main shop
Nov2007	CH	Chadstone Mall	109 Dandenong Rd	Oct2007	Mar2008	
Nov2007	DOW	Doncaster Westfield	75 Doncaster Rd	Nov2007	Feb2008	
Dec2007	MEL	Melbourne Central	88 Lonsdale St	Jan0000	Dec9999	Main shop
Dec2007	CH	Chadstone Mall	109 Dandenong Rd	Oct2007	Mar2008	
Dec2007	DOW	Doncaster Westfield	75 Doncaster Rd	Nov2007	Feb2008	
Jan2008	MEL	Melbourne Central	88 Lonsdale St	Jan0000	Dec9999	Main shop
Jan2008	CH	Chadstone Mall	109 Dandenong Rd	Oct2007	Mar2008	
Jan2008	DOW	Doncaster Westfield	75 Doncaster Rd	Nov2007	Feb2008	
Feb2008	MEL	Melbourne Central	88 Lonsdale St	Jan0000	Dec9999	Main shop
Feb2008	CH	Chadstone Mall	109 Dandenong Rd	Oct2007	Mar2008	
Feb2008	DOW	Doncaster Westfield	75 Doncaster Rd	Nov2007	Feb2008	
Mar2008	MEL	Melbourne Central	88 Lonsdale St	Jan0000	Dec9999	Main shop
Mar2008	CH	Chadstone Mall	109 Dandenong Rd	Oct2007	Mar2008	
Apr2008	MEL	Melbourne Central	88 Lonsdale St	Jan0000	Dec9999	Main shop
May2008	MEL	Melbourne Central	88 Lonsdale St	Jan0000	Dec9999	Main shop
...
Oct2008	MEL	Melbourne Central	88 Lonsdale St	Jan0000	Dec9999	Main shop
Oct2008	CH	Chadstone Mall	109 Dandenong Rd	Oct2008	Feb2009	Under re-construction
...



Another Temporal Dimension

- Course History as a Temporal Dimension – Using a Bridge Table
- Because the relationship cardinality between Course and Course History is a 1-1 relationship, both dimensions can be combined into one dimension, called the Course dimension.



Slowly Changing Dimensions (SCD)

Slowly Changing Dimensions (SCD)

- Slowly changing dimensions
 - Dimensions where the records of these dimensions **change slowly over a period of time**
 - E.g. Book dimension
 - It has price information, and it is common that the price of a book changes "slowly" over time
- Different from attributes/records that change 'rapidly', e.g. share price, location of a taxi
 - These are related to real-time data warehousing (stream data warehousing)

Slowly Changing Dimensions (SCD) - Type 0 and Type 1

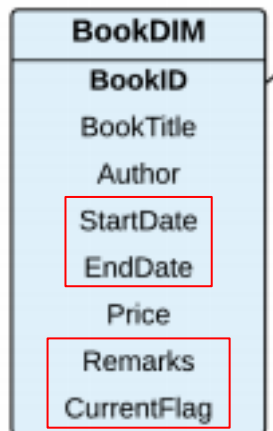
- Both types do not actually record the history of changes in the dimension
- Type 0
 - The dimension stores the "Original or Initial" value of the records, when the data warehousing is built
 - E.g. full price of books will be recorded in the book dimension
- Type 1
 - It only records the latest value of the record
 - E.g. the latest price of books will be recorded in the book dimension

Book Dimension Table (SCD Type 0)			
BookID	Book Title	Author	Price
C1	CSIRO Diet	CSIRO Team	\$45.95
H6	Harry Potter 6	Rowling	\$30.95
DV	Da Vinci Code	Dan Brown	\$27.95
...

Book Dimension Table (SCD Type 1)			
BookID	Book Title	Author	Price
C1	CSIRO Diet	CSIRO Team	\$45.95
H6	Harry Potter 6	Rowling	\$10.00
DV	Da Vinci Code	Dan Brown	\$27.95
...

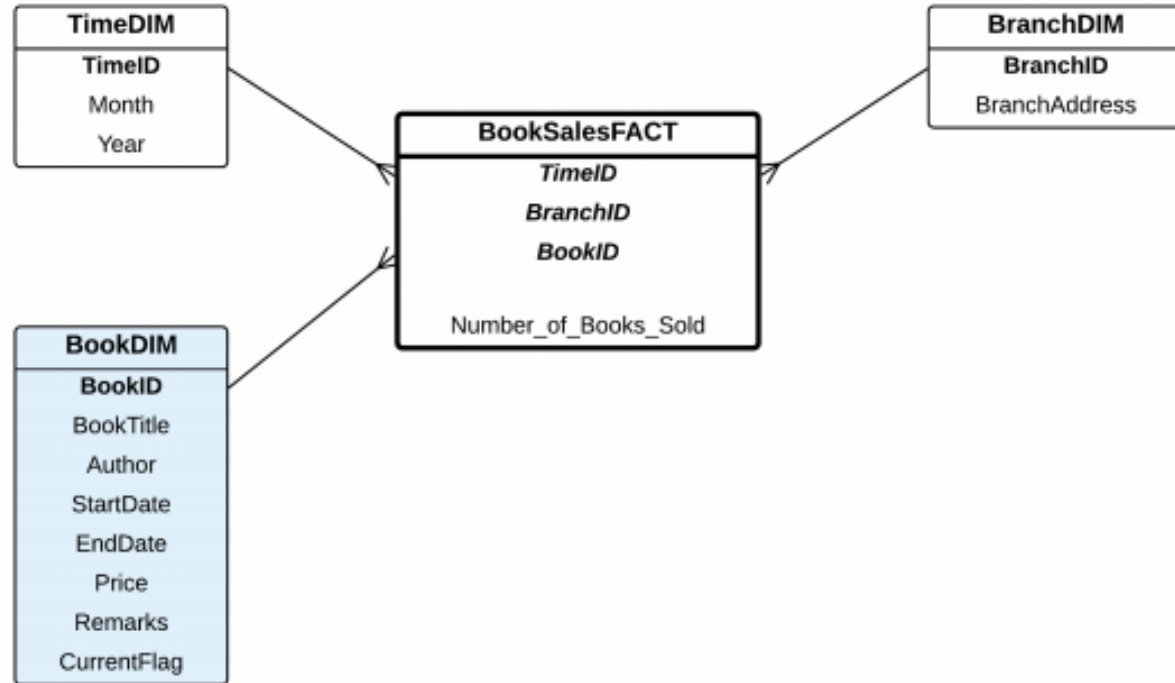
Slowly Changing Dimensions (SCD) - Type 2

- Type 2
 - Keeps track of the history from the main dimension
 - E.g. when the price of a book is changed, “another book” with the same details is created, with the new Book ID, and the new price



Book Dimension Table (SCD Type 2)							
BookID	Book Title	Author	Start Date	End Date	Price	Remarks	Current Flag
C1_1	CSIRO Diet	CSIRO Team	Jan2007	July2007	\$45.95	Full Price	N
C1_2	CSIRO Diet	CSIRO Team	Aug2007	Oct2007	\$36.75	20% Discount	N
C1_3	CSIRO Diet	CSIRO Team	Nov2007	Jan2008	\$23.00	Half Price	N
C1_4	CSIRO Diet	CSIRO Team	Feb2008	Dec9999	\$45.95	Full Price	Y
H6_1	Harry Potter 6	Rowling	Jan2007	Mar2007	\$21.95	Launching	N
H6_2	Harry Potter 6	Rowling	Apr2007	Jan2008	\$30.95	Full Price	N
H6_3	Harry Potter 6	Rowling	Feb2008	Dec9999	\$10.00	End of Product Sale	Y
DV_1	Da Vinci Code	Dan Brown	Jan2007	Dec9999	\$27.95	Full Price	Y
...

Slowly Changing Dimensions (SCD) Type 2 (cont.)



Report 3 (SCD Type 2)

TimeID	BranchID	BookID	Book Title	Author	Price	Number of Books Sold
Mar2008	City	C1_4	CSIRO Diet	CSIRO Team	\$45.95	5
Mar2008	City	H6_3	Harry Potter 6	Rowling	\$10.00	15
Mar2008	City	DV_1	Da Vinci Code	Dan Brown	\$27.95	23
Mar2008	City
Mar2008	Chadstone	C1_4	CSIRO Diet	CSIRO Team	\$45.95	15
Mar2008	Chadstone	H6_3	Harry Potter 6	Rowling	\$10.00	3
Mar2008	Chadstone	DV_1	Da Vinci Code	Dan Brown	\$27.95	2
Mar2008	Chadstone
Mar2008	Camberwell	C1_4	CSIRO Diet	CSIRO Team	\$45.95	1
Mar2008	Camberwell	H6_3	Harry Potter 6	Rowling	\$10.00	1
Mar2008	Camberwell	DV_1	Da Vinci Code	Dan Brown	\$27.95	2
Mar2008	Camberwell
Mar2008
...
...
Dec2007	City	C1_3	CSIRO Diet	CSIRO Team	\$23.00	15
Dec2007	City	H6_2	Harry Potter 6	Rowling	\$30.95	6
Dec2007	City	DV_1	Da Vinci Code	Dan Brown	\$27.95	6
Dec2007	City
Dec2007	Chadstone	C1_3	CSIRO Diet	CSIRO Team	\$23.00	10
Dec2007	Chadstone	H6_2	Harry Potter 6	Rowling	\$30.95	8
Dec2007	Chadstone	DV_1	Da Vinci Code	Dan Brown	\$27.95	1
Dec2007	Chadstone
Dec2007	Camberwell	C1_3	CSIRO Diet	CSIRO Team	\$23.00	18
Dec2007	Camberwell	H6_2	Harry Potter 6	Rowling	\$30.95	3
Dec2007	Camberwell	DV_1	Da Vinci Code	Dan Brown	\$27.95	2
Dec2007	Camberwell
Dec2007
...

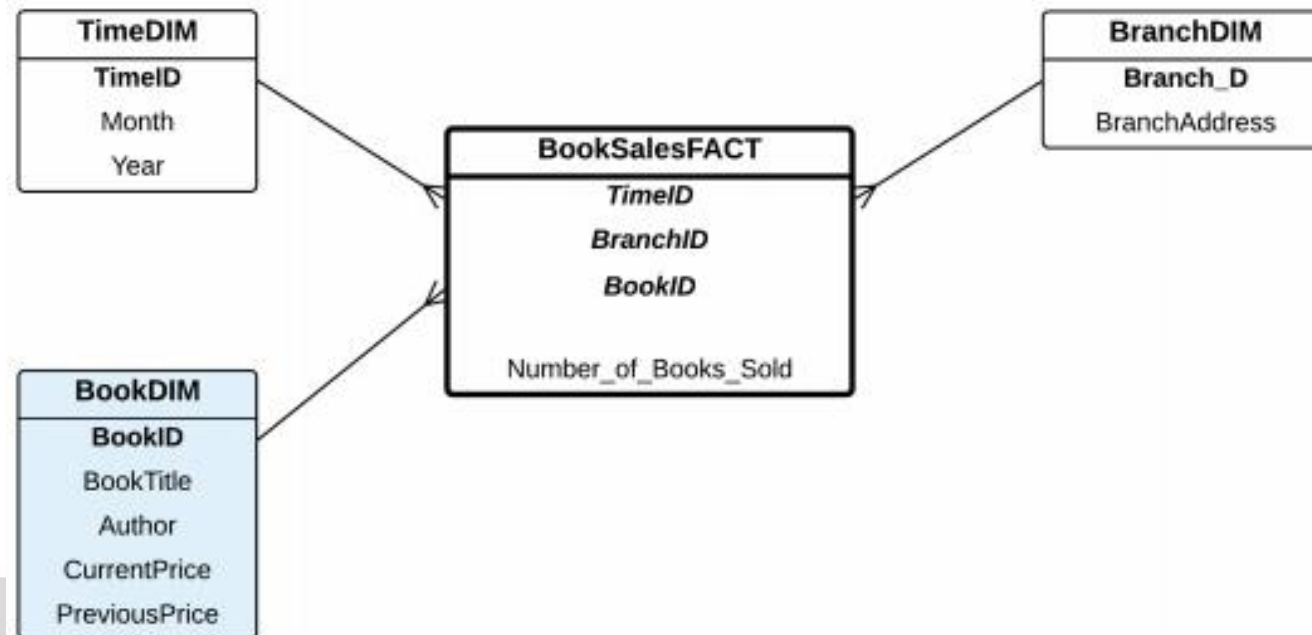
Slowly Changing Dimensions (SCD) - Type 3

- Type 3
 - Simplification of Type 2
 - Maintains only the **current** and the **previous values**, not the entire history
 - E.g. only the last two prices of the book are recorded
 - Rationale
 - Assume that analyses of complete history is not necessary
 - Most analyses will be done on the current price, and at most one past price, e.g. for comparison with the trend

Slowly Changing Dimensions (SCD) - Type 3 (cont.)

Book Dimension Table (SCD Type 3)

BookID	Book Title	Author	Current Price	Previous Price
C1	CSIRO Diet	CSIRO Team	\$45.95	\$23.00
H6	Harry Potter 6	Rowling	\$10.00	\$30.95
DV	Da Vinci Code	Dan Brown	\$27.95	Null
...



Slowly Changing Dimensions (SCD) - Type 4

- Type 4
 - Create a new dimension to maintain the history of attribute value change
 - E.g. *BookPriceDIM* in the Bookshop case study
 - Main advantage:
 - Do not need to have a different *BookID* for the same book. Additionally, the entire history of changes is kept.

Book Dimension Table		
BookID	Book Title	Author
C1	CSIRO Diet	CSIRO Team
H6	Harry Potter 6	Rowling
DV	Da Vinci Code	Dan Brown
...

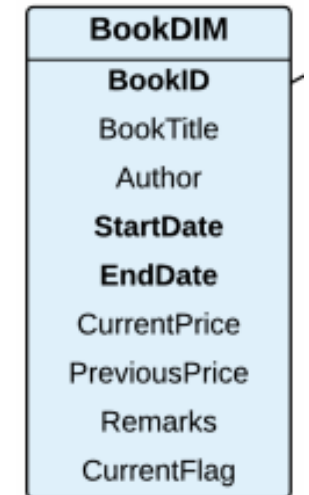
Book Price Dimension Table				
BookID	Start Date	End Date	Price	Remarks
C1	Jan2007	July2007	\$45.95	Full Price
C1	Aug2007	Oct2007	\$36.75	20% Discount
C1	Nov2007	Jan2008	\$23.00	Half Price
C1	Feb2008	Dec9999	\$45.95	Full Price
H6	Jan2007	Mar2007	\$21.95	Launching
H6	Apr2007	Jan2008	\$30.95	Full Price
H6	Feb2008	Dec9999	\$10.00	End of Product Sale
DV	Jan2007	Dec9999	\$27.95	Full Price
...

Slowly Changing Dimensions (SCD) - Type 6

- Type 6
 - Combination of Type 2 and Type 3
 - A separate identifier for the same book is not needed (Type 3), but the entire history is kept (Type 2)

Book Dimension (SCD Type6)

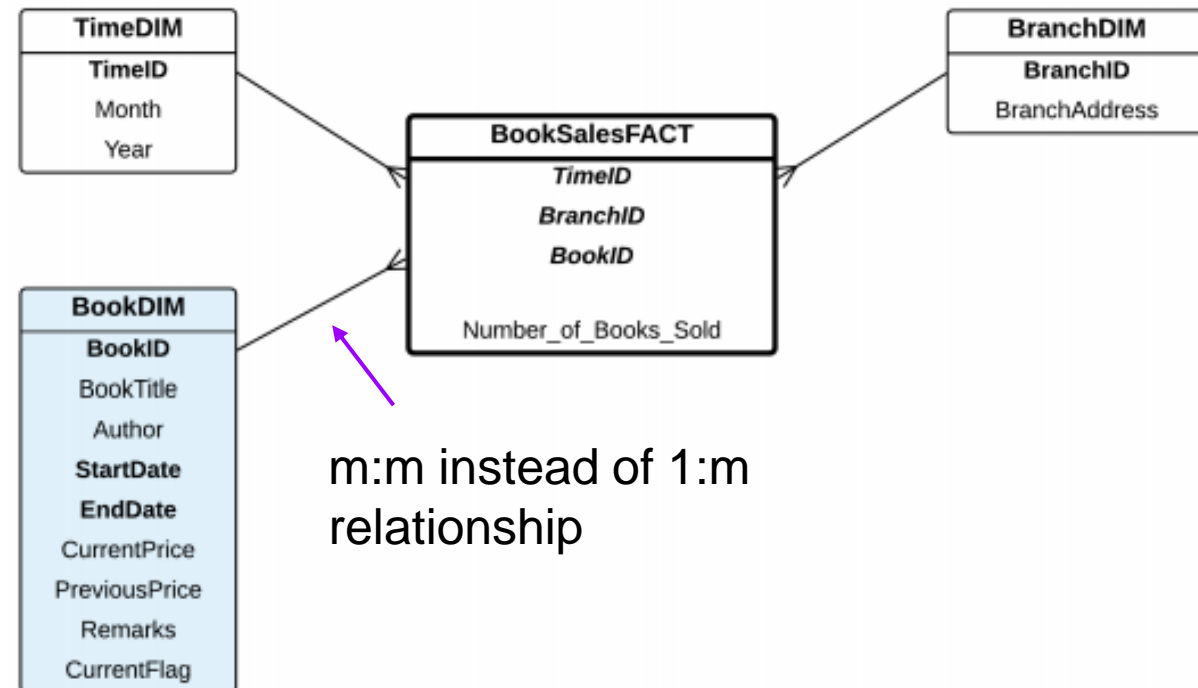
BookID	Book Title	Author	Start Date	End Date	Current Price	Previous Price	Remarks	Current Flag
C1	CSIRO Diet	CSIRO Team	Jan2007	July2007	\$45.95	Null	Full Price	N
C1	CSIRO Diet	CSIRO Team	Aug2007	Oct2007	\$36.75	\$45.95	20% Discount	N
C1	CSIRO Diet	CSIRO Team	Nov2007	Jan2008	\$23.00	\$36.75	Half Price	N
C1	CSIRO Diet	CSIRO Team	Feb2008	Dec9999	\$45.95	\$23.00	Full Price	Y
H6	Harry Potter 6	Rowling	Jan2007	Mar2007	\$21.95	Null	Launching	N
H6	Harry Potter 6	Rowling	Apr2007	Jan2008	\$30.95	\$21.95	Full Price	N
H6	Harry Potter 6	Rowling	Feb2008	Dec9999	\$10.00	\$30.95	End of Product Sale	Y
DV	Da Vinci Code	Dan Brown	Jan2007	Dec9999	\$27.95	Null	Full Price	Y
...



Composite key
{BookID, StartDate,
EndDate}

Slowly Changing Dimensions (SCD) - Type 6 (cont.)

- Type 6
 - **m:m relationship** between *BookDIM* and *BookSalesFACT*
- Solution #1:
 - Add new surrogate key to Book Dimension table (Type 2)
 - Consequently, the Fact table will have this surrogate key from the Book Dimension as a reference.
 - The surrogate key can also simply be a concatenation between *BookID*, *StartDate*, and *EndDate*, or, *BookID* with a sequence number



Slowly Changing Dimensions (SCD) - Type 6 (cont.)

- Solution #2:
 - Include the *StartDate* and *EndDate* in the fact table
 - Messy because the fact table already has TimeID
- Solution #3:
 - Add an associative table (or a bridge table) between Book Dimension and the fact table
 - The associative table will have a composite key comprising key identifier from Book dimension and from the fact
 - Messy as well
- **Conclusion:**
 - Solution #1 is the best: change to Type 2
 - So Type 6 is rather unnecessary!

Implementation of SCD in SQL

Implementation of SCD in SQL - Type 0

Book Dimension Table (SCD Type 0)			
BookID	Book Title	Author	Price
C1	CSIRO Diet	CSIRO Team	\$45.95
H6	Harry Potter 6	Rowling	\$30.95
DV	Da Vinci Code	Dan Brown	\$27.95
...

Book (BookID, BookTitle, Author)

BookPriceHistory (BookID, StartDate, EndDate, Price, Remarks)

```
-- Non-temporal dimension
create table SCD0
as
select distinct B.BookID, B.BookTitle, B.Author,
               H.Price as OriginalPrice
from Book B, BookPriceHistory H
where B.BookID = H.BookID and H.Remarks = 'Full Price';
```


Implementation of SCD in SQL - Type 1

```
-- Latest price
create table SCD1
as
select T.BookID, T.BookTitle, T.Author, T.Price as CurrentPrice
from (
    select B.BookID, B.BookTitle, B.Author,
        to_date(H.StartDate, 'MonYYYY'), H.Price,
        rank() over( partition by B.BookID order by
            to_date(H.StartDate, 'MonYYYY') desc) as Rank
    from Book B, BookPriceHistory H where B.BookID = H.BookID) T
where T.Rank = 1;
```

Book Dimension Table (SCD Type 1)			
BookID	Book Title	Author	Price
C1	CSIRO Diet	CSIRO Team	\$45.95
H6	Harry Potter 6	Rowling	\$10.00
DV	Da Vinci Code	Dan Brown	\$27.95
...

Implementation of SCD in SQL - Type 2

```
-- New BookID for every changed price
create table SCD2
as
select B.BookID || '_' ||
       rank() over(partition by B.BookID order by
                    to_date(H.StartDate, 'MonYYYY') asc) as BookID,
       B.BookTitle, B.Author, H.StartDate, H.EndDate, H.Price, H.Remarks,
       case H.EndDate
         When 'Dec9999' then 'Y'
         else 'N'
       end as CurrentFlag
from Book B, BookPriceHistory H
where B.BookID = H.BookID;
```

Book Dimension Table (SCD Type 2)							
BookID	Book Title	Author	Start Date	End Date	Price	Remarks	Current Flag
C1.1	CSIRO Diet	CSIRO Team	Jan2007	July2007	\$45.95	Full Price	N
C1.2	CSIRO Diet	CSIRO Team	Aug2007	Oct2007	\$36.75	20% Discount	N
C1.3	CSIRO Diet	CSIRO Team	Nov2007	Jan2008	\$23.00	Half Price	N
C1.4	CSIRO Diet	CSIRO Team	Feb2008	Dec9999	\$45.95	Full Price	Y
H6.1	Harry Potter 6	Rowling	Jan2007	Mar2007	\$21.95	Launching	N
H6.2	Harry Potter 6	Rowling	Apr2007	Jan2008	\$30.95	Full Price	N
H6.3	Harry Potter 6	Rowling	Feb2008	Dec9999	\$10.00	End of Product Sale	Y
DV.1	Da Vinci Code	Dan Brown	Jan2007	Dec9999	\$27.95	Full Price	Y
...

Implementation of SCD in SQL - Type 3

-- Include current and previous price using outer join

create table SCD3

as

select T1.BookID, T1.BookTitle, T1.Author, **T1.CurrentPrice**, **T2.CurrentPrice as PreviousPrice**
from

(select T.BookID, T.BookTitle, T.Author, **T.Price as CurrentPrice** from (
select B.BookID, B.BookTitle, B.Author, to_date(H.StartDate, 'MonYYYY'), H.Price,
rank() over(partition by B.BookID order by to_date(H.StartDate, 'MonYYYY')
desc) as Rank
from Book B, BookPriceHistory H
where B.BookID = H.BookID) T

where **T.Rank = 1**) **T1**,

(select T.BookID, T.BookTitle, T.Author, **T.Price as CurrentPrice** from (
select B.BookID, B.BookTitle, B.Author, to_date(H.StartDate, 'MonYYYY'), H.Price,
rank() over(partition by B.BookID order by to_date(H.StartDate, 'MonYYYY')
desc) as Rank
from Book B, BookPriceHistory H
where B.BookID = H.BookID) T

where **T.Rank = 2**) **T2**

where **T1.BookID = T2.BookID(+)**; -- Use outer join as some books do not have previous price

Book Dimension Table (SCD Type 3)				
BookID	Book Title	Author	Current Price	Previous Price
C1	CSIRO Diet	CSIRO Team	\$45.95	\$23.00
H6	Harry Potter 6	Rowling	\$10.00	\$30.95
DV	Da Vinci Code	Dan Brown	\$27.95	Null
...

Implementation of SCD in SQL - Type 4

- Type 4: Same as the Temporal Data Warehousing implementation

```
create table SCD4 as  
select * from BookPriceHistory;
```

BOOKPRICEHISTORY	
PK,FK	BookID
PK	StartDate
PK	EndDate
	Price
	Remarks



Book Price Dimension Table				
BookID	Start Date	End Date	Price	Remarks
C1	Jan2007	July2007	\$45.95	Full Price
C1	Aug2007	Oct2007	\$36.75	20% Discount
C1	Nov2007	Jan2008	\$23.00	Half Price
C1	Feb2008	Dec9999	\$45.95	Full Price
H6	Jan2007	Mar2007	\$21.95	Launching
H6	Apr2007	Jan2008	\$30.95	Full Price
H6	Feb2008	Dec9999	\$10.00	End of Product Sale
DV	Jan2007	Dec9999	\$27.95	Full Price
...

Implementation of SCD in SQL - Type 6

- Type 6: Combination of SCD2 (use different Book ID) and SCD3 (current and previous values)
- SCD Type 2 identifier (Book ID) is different from the original Book ID used by SCD Type 3
 - Cannot simply use an equi-join between SCD Type 2 and SCD Type 3
 - Instead, check if the Book ID of SCD Type 3 **is part of** the Book ID of the SCD Type 2, using the like operator in SQL:

```
create table SCD6 as
select SCD3.BookID, SCD3.BookTitle, SCD3.Author,
       SCD2.StartDate, SCD2.EndDate, SCD3.CurrentPrice,
       SCD3.PreviousPrice, SCD2.Remarks, SCD2.CurrentFlag
from SCD2, SCD3
where SCD2.BookID like SCD3.BookID||'_%';
```

Creating Fact Tables

Creating the Fact Tables

Book Dimension Table (SCD Type 0)

BookID	Book Title	Author	Price
C1	CSIRO Diet	CSIRO Team	\$45.95
H6	Harry Potter 6	Rowling	\$30.95
DV	Da Vinci Code	Dan Brown	\$27.95
...

Book Dimension Table (SCD Type 1)

BookID	Book Title	Author	Price
C1	CSIRO Diet	CSIRO Team	\$45.95
H6	Harry Potter 6	Rowling	\$10.00
DV	Da Vinci Code	Dan Brown	\$27.95
...

Book Dimension Table (SCD Type 3)

BookID	Book Title	Author	Current Price	Previous Price
C1	CSIRO Diet	CSIRO Team	\$45.95	\$23.00
H6	Harry Potter 6	Rowling	\$10.00	\$30.95
DV	Da Vinci Code	Dan Brown	\$27.95	Null
...

- Book dimension table in SCD type 0, 1, 3 and 4 have the same number of records with same BookID
- Fact table is not affected
- BUT, when creating report, book price might not be accurate based on the Time ID
 - Except SCD Type 4

Book Dimension Table

BookID	Book Title	Author
C1	CSIRO Diet	CSIRO Team
H6	Harry Potter 6	Rowling
DV	Da Vinci Code	Dan Brown
...

SCD Type 4

Book Price Dimension Table

BookID	Start Date	End Date	Price	Remarks
C1	Jan2007	July2007	\$45.95	Full Price
C1	Aug2007	Oct2007	\$36.75	20% Discount
C1	Nov2007	Jan2008	\$23.00	Half Price
C1	Feb2008	Dec9999	\$45.95	Full Price
H6	Jan2007	Mar2007	\$21.95	Launching
H6	Apr2007	Jan2008	\$30.95	Full Price
H6	Feb2008	Dec9999	\$10.00	End of Product Sale
DV	Jan2007	Dec9999	\$27.95	Full Price
...

Creating Fact Tables - SCD Type 2

Book Dimension Table (SCD Type 2)							
BookID	Book Title	Author	Start Date	End Date	Price	Remarks	Current Flag
C1.1	CSIRO Diet	CSIRO Team	Jan2007	July2007	\$45.95	Full Price	N
C1.2	CSIRO Diet	CSIRO Team	Aug2007	Oct2007	\$36.75	20% Discount	N
C1.3	CSIRO Diet	CSIRO Team	Nov2007	Jan2008	\$23.00	Half Price	N
C1.4	CSIRO Diet	CSIRO Team	Feb2008	Dec9999	\$45.95	Full Price	Y
H6.1	Harry Potter 6	Rowling	Jan2007	Mar2007	\$21.95	Launching	N
H6.2	Harry Potter 6	Rowling	Apr2007	Jan2008	\$30.95	Full Price	N
H6.3	Harry Potter 6	Rowling	Feb2008	Dec9999	\$10.00	End of Product Sale	Y
DV.1	Da Vinci Code	Dan Brown	Jan2007	Dec9999	\$27.95	Full Price	Y
...

- New row being added when price change
 - Fact table is impacted, must be updated to contain the correct Book ID

```
create table BookSalesFactWithSCD2 as
select to_char(T.Transaction_Date, 'MonYYYY') as TimeID,
       BK.BookID, BR.BranchID,
       sum(T.Quantity) as Number_Of_Books_Sold
from BookTransaction T, SCD2 BK, Branch BR
where T.BranchID = BR.BranchID
      and BK.BookID LIKE T.BookID||'_%'
      and to_date(BK.StartDate, 'MonYYYY') <= T.Transaction_Date
      and T.Transaction_Date <= to_date(BK.EndDate, 'MonYYYY')
group by to_char(T.Transaction_Date, 'MonYYYY'), BK.BookID, BR.BranchID;
```


Creating Fact Tables - SCD Type 2 (cont.)

Fact Table (SCD Type 2)			
TimeID	BranchID	BookID	Number of Books Sold
Mar2008	City	C1_4	5
Mar2008	City	H6_3	15
Mar2008	City	DV_1	23
Mar2008	City
Mar2008	Chadstone	C1_4	15
Mar2008	Chadstone	H6_3	3
Mar2008	Chadstone	DV_1	2
Mar2008	Chadstone
Mar2008	Camberwell	C1_4	1
Mar2008	Camberwell	H6_3	1
Mar2008	Camberwell	DV_1	2
Mar2008	Camberwell
Mar2008
...
...
Dec2007	City	C1_3	15
Dec2007	City	H6_2	6

Creating Fact Tables - SCD Type 6

- SCD Type 6 is similar to SCD Type 2, where it contains the same number of records
- The only difference is that in SCD Type 6, Book ID does not change; the original Book ID is used

```
create table BookSalesFactWithSCD6 as
Select to_char(T.Transaction_Date, 'MonYYYY') as TimeID,
       BK.BookID, BR.BranchID, sum(T.Quantity) as Number_Of_Books_Sold
from BookTransaction T, SCD6 BK, Branch BR
where T.BranchID = BR.BranchID and BK.BookID = T.BookID
      and to_date(BK.StartDate, 'MonYYYY') <= T.Transaction_Date
      and T.Transaction_Date <= to_date(BK.EndDate, 'MonYYYY')
group by to_char(T.Transaction_Date, 'MonYYYY'), BK.BookID,
BR.BranchID;
```

Creating Fact Tables - SCD Type 6 (cont.)

- When we join SCD Type 2 and the Fact Table, or when we join SCD Type 6 and the Fact Table
 - Correct book price will be shown
 - The reason is that both SCD Type 2 and Type 6 maintain the complete history of Book Prices; they keep the *StartDate* and *EndDate* of each Book Price

Fact Table (SCD Type 6)			
TimeID	BranchID	BookID	Number of Books Sold
Mar2008	City	C1	5
Mar2008	City	H6	15
Mar2008	City	DV	23
Mar2008	City
Mar2008	Chadstone	C1	15
Mar2008	Chadstone	H6	3
Mar2008	Chadstone	DV	2
Mar2008	Chadstone
Mar2008	Camberwell	C1	1
Mar2008	Camberwell	H6	1
Mar2008	Camberwell	DV	2
Mar2008	Camberwell
Mar2008
...
...
Dec2007	City	C1	15
Dec2007	City	H6	6
Dec2007	City	DV	6
Dec2007	City
Dec2007	Chadstone	C1	15

Creating Fact Tables - Conclusion

- SCD Type 2, 4 and 6 will ensure correct price is shown
 - Reason: complete price history is kept
- **Differences**
 - SCD Type 4 uses two tables (i.e. Book Dimension and Book Price Dimension)
 - SCD Type 2 and SCD Type 6 uses one table only, where the Book Price history is maintained in the Book Dimension

Summary

- A temporal data warehousing uses the concept of the **Bridge Table** (or a Weak Entity), where the history is maintained in a bridge table.
- Maintaining the **history of certain attributes** is important in order to make associative analysis more accurate when analysing the reports produced by the fact and dimensions.
- However, certain degree of caution when joining the fact table and the temporal dimension, especially when the **level of granularity** of time between the fact and the temporal dimension is not the same.
- Temporal data warehousing is also known as **Slowly Changing Dimensions(SCD)**.
 - Different types will server different purposes of the data warehousing.