MONASH University

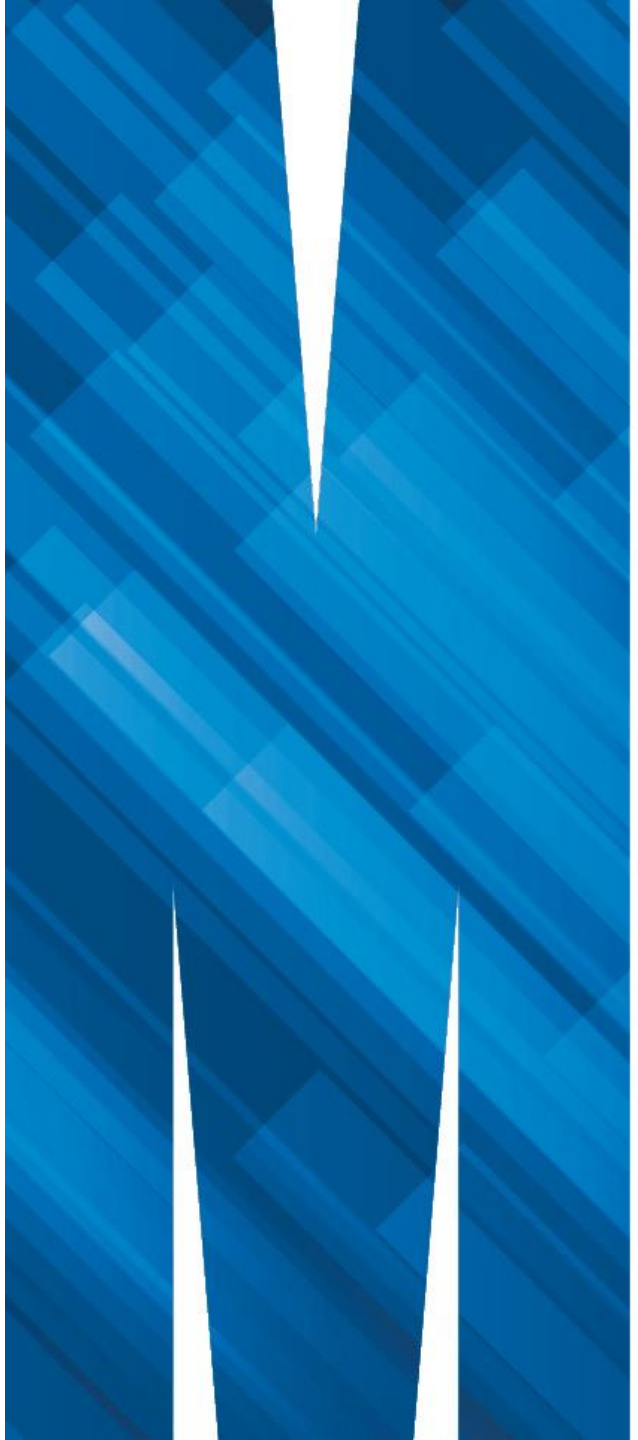
MONASH
INFORMATION
TECHNOLOGY

FIT3003 – Business Intelligence and Data Warehousing

Week 7 – Determinant Dimensions

Semester 2, 2022

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



Learning Objectives

1. To understand the difference between one and split dimensions.
2. To understand the concept of hierarchy.
3. To understand the concept of determinant dimensions.
4. To be able to distinguish the differences between determinant and non-determinant dimensions.

Agenda

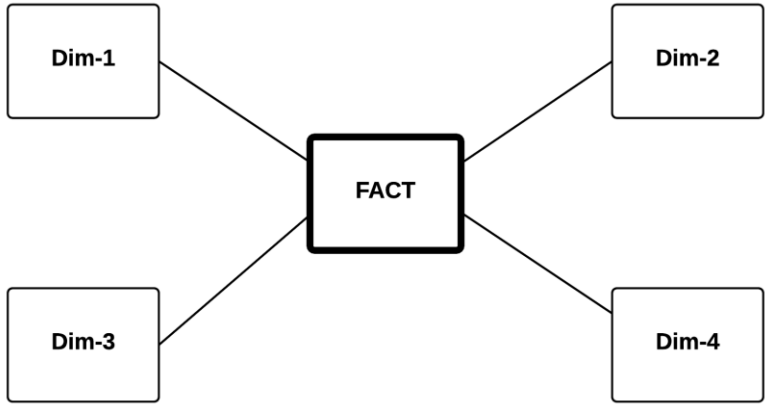
1. Hierarchies
    1. Hierarchy vs. Non-Hierarchy

2. Determinant Dimensions
    1. Determinant vs. Non-Determinant Dimensions
    2. Determinant Dimension vs. Pivoted Fact Table

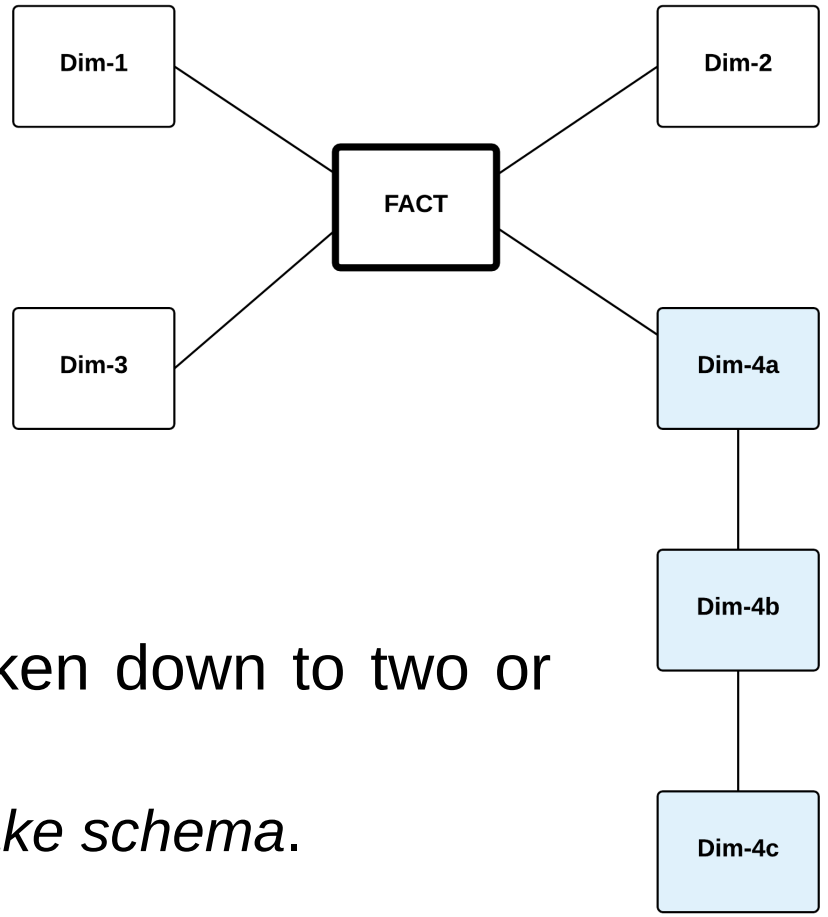
Hierarchies

Hierarchies

(a) A simple Star Schema



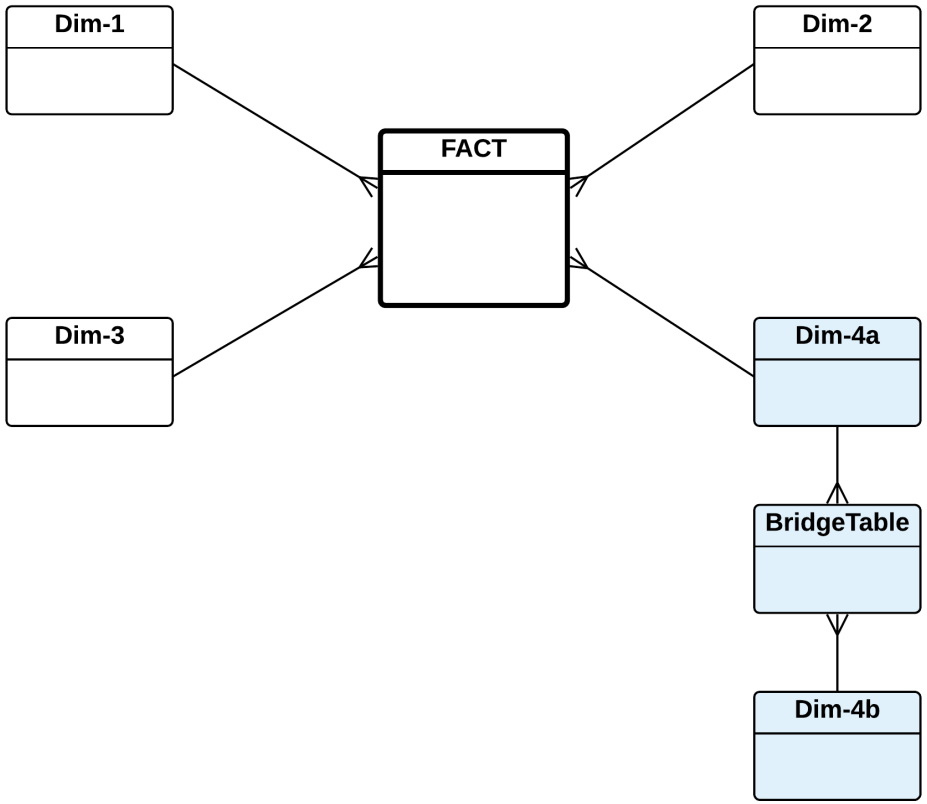
(b) A Snowflake with Hierarchy



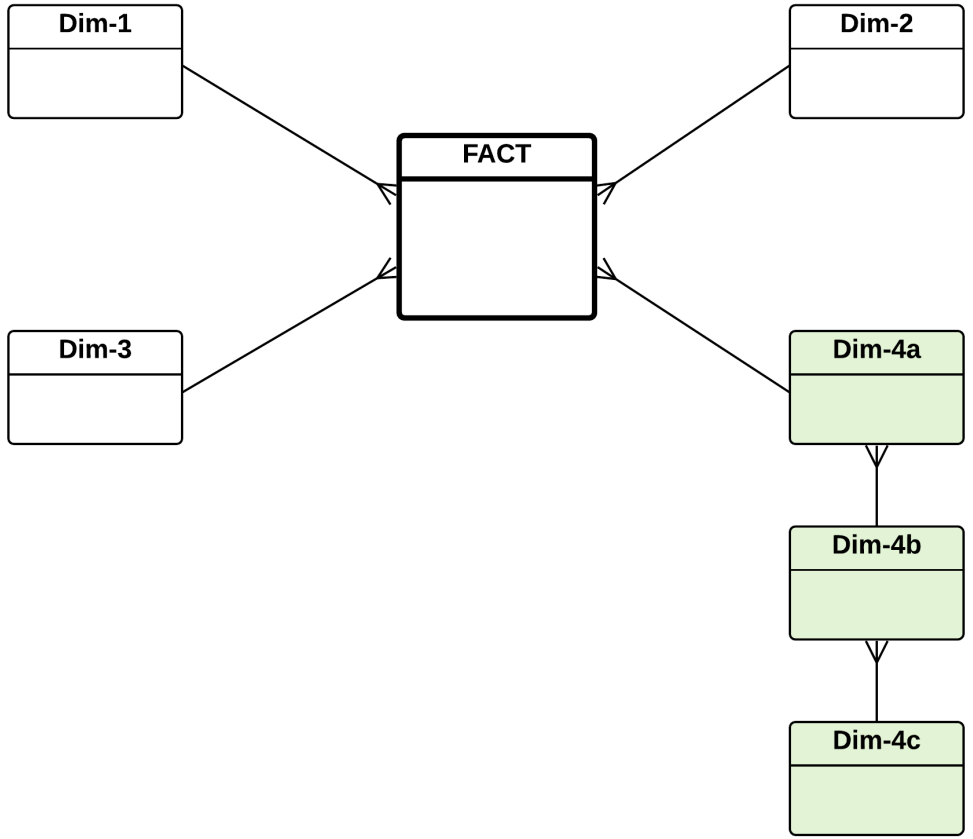
- A *Hierarchy* is formed when a dimension is broken down to two or more dimensions in a hierarchical manner.
  - As a result, the star schema becomes a *snowflake schema*.

Hierarchies

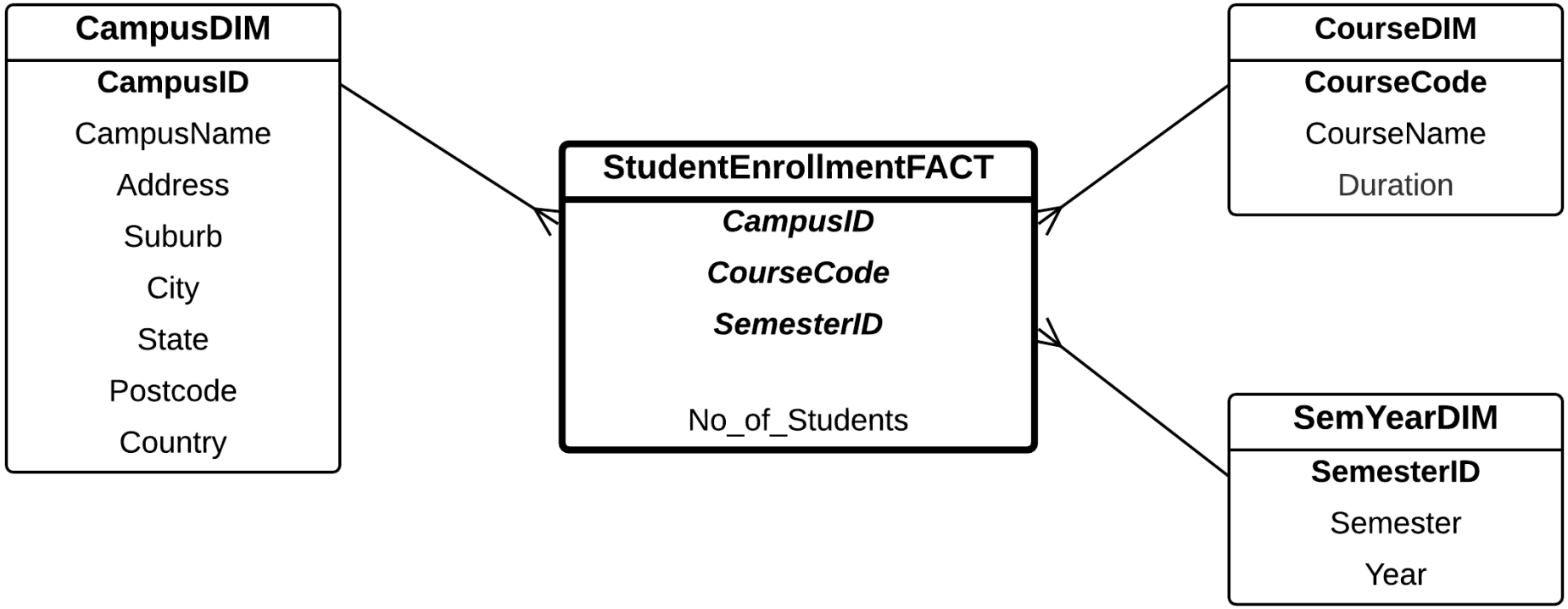
(a) A Snowflake with a Bridge Table



(b) A Snowflake with Hierarchy



Hierarchy vs. Non-Hierarchy

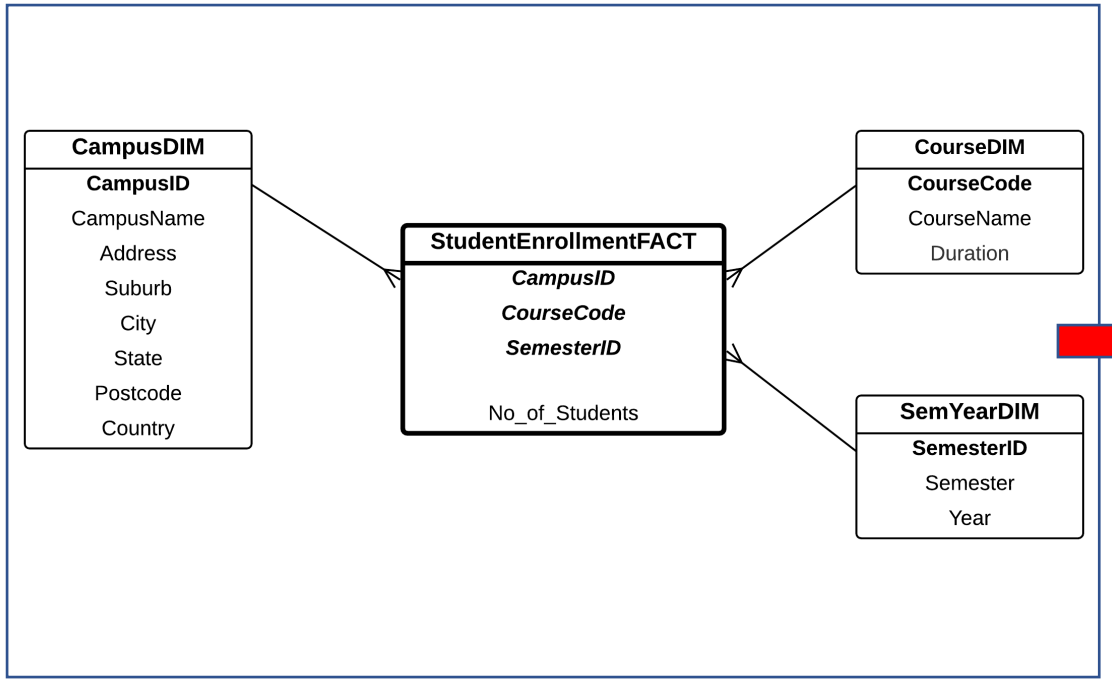


Hierarchy vs. Non-Hierarchy

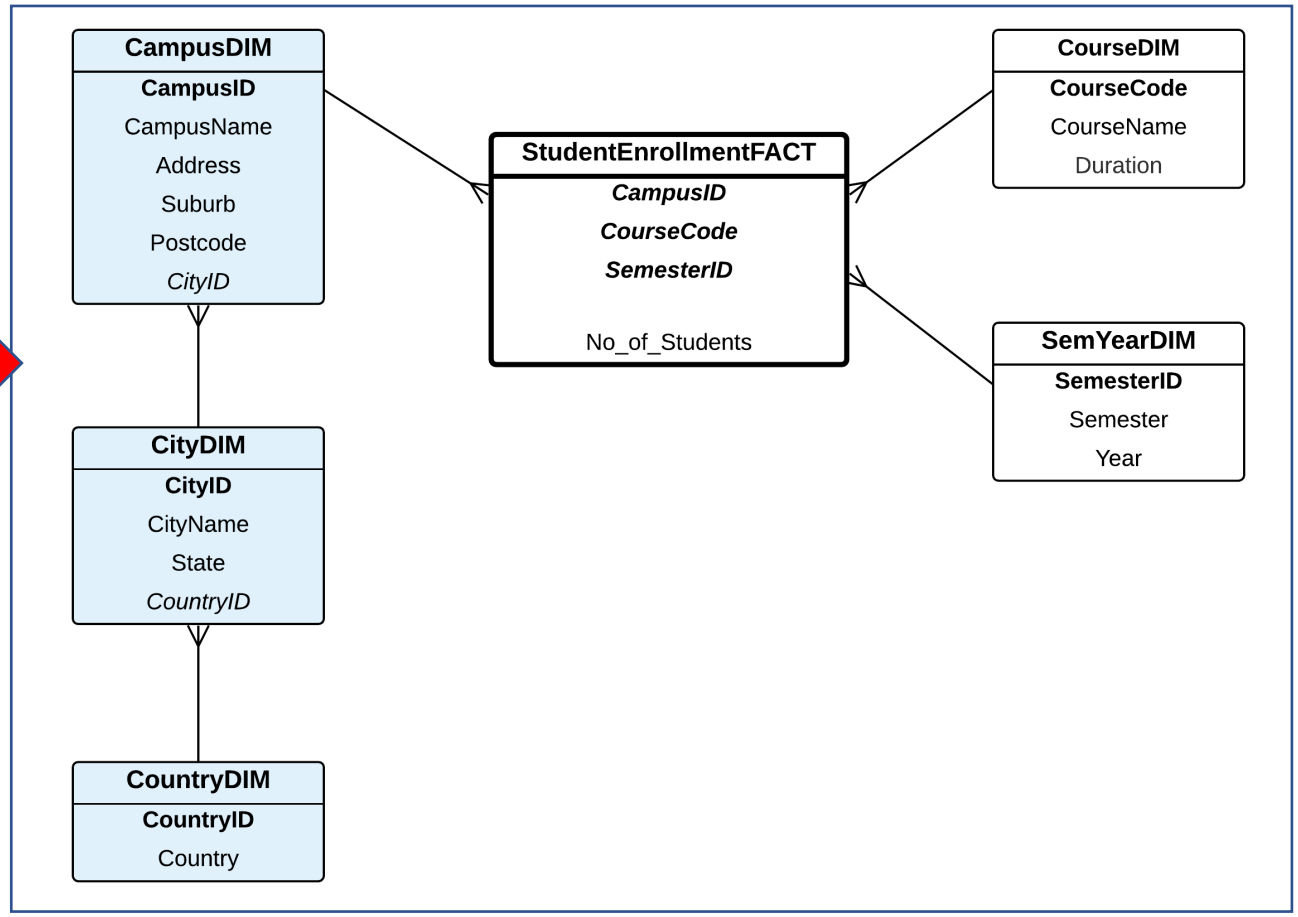
CampusID	CampusName	Address	Suburb	City	State	Postcode	Country
CL	Clayton Campus	Wellington Rd	Clayton	Melbourne	Victoria	3800	Australia
CA	Caulfield Campus	Dandenong Rd	Caulfield East	Melbourne	Victoria	3145	Australia
PA	Parkville Campus	Royal Parade	Parkville	Melbourne	Victoria	3052	Australia
SY	Sydney Campus	Opera Boulevard	Sydney	Sydney	New South Wales	2001	Australia
MUM	Malaysia Campus	Jalan Lagoon	Bandar Sunway	Kuala Lumpur	Selangor	47500	Malaysia
MSA	South Africa Campus	Peter St	Johannesburg	Johannesburg	Johannesburg	1725	South Africa

Hierarchy vs. Non-Hierarchy

Non-Hierarchy



Hierarchy



Hierarchy vs. Non-Hierarchy

(a) Campus Dimension Table

CampusID	CampusName	Address	Suburb	Postcode	CityID
CL	Clayton Campus	Wellington Rd	Clayton	3800	MEL
CA	Caulfield Campus	Dandenong Rd	Caulfield East	3145	MEL
PA	Parkville Campus	Royal Parade	Parkville	3052	MEL
SY	Sydney Campus	Opera Boulevard	Sydney	2001	SYD
MUM	Malaysia Campus	Jalan Lagoon	Bandar Sunway	47500	KUL
MSA	South Africa Campus	Peter St	Johannesburg	1725	JNB

(b) City Dimension Table

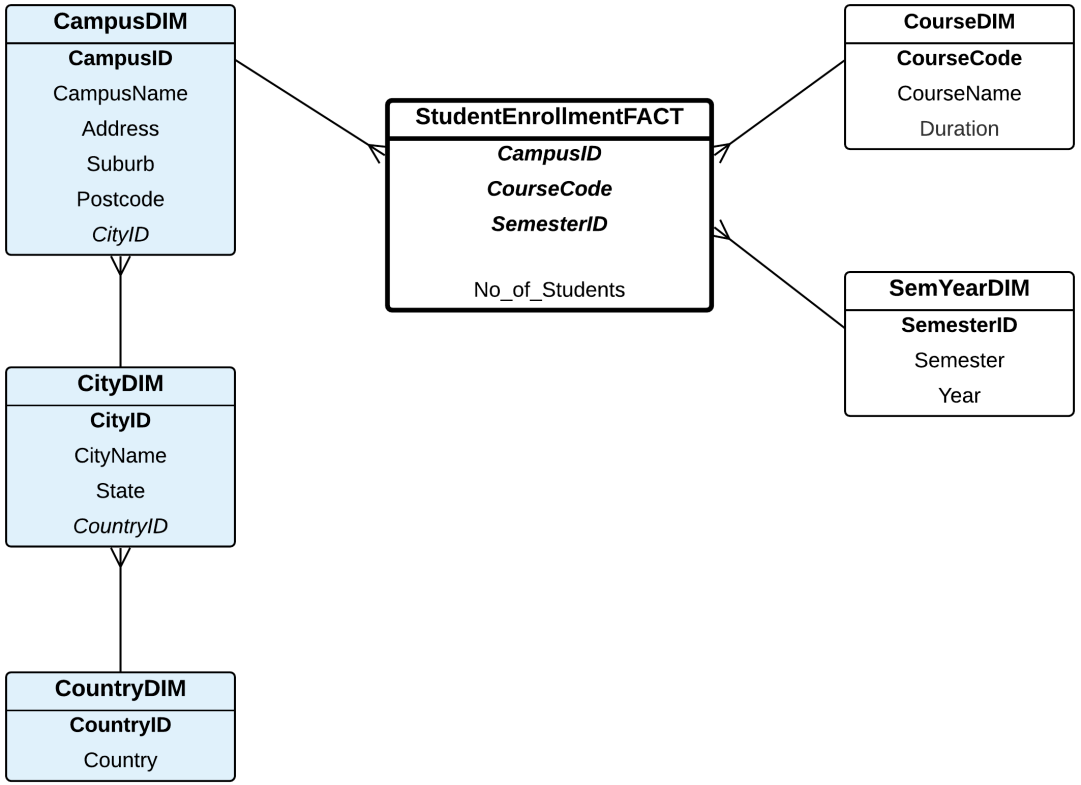
CityID	City Name	State	CountryID
MEL	Melbourne	Victoria	AU
SYD	Sydney	New South Wales	AU
KUL	Kuala Lumpur	Selangor	MA
JNB	Johannesburg	Johannesburg	SA

(c) Country Dimension Table

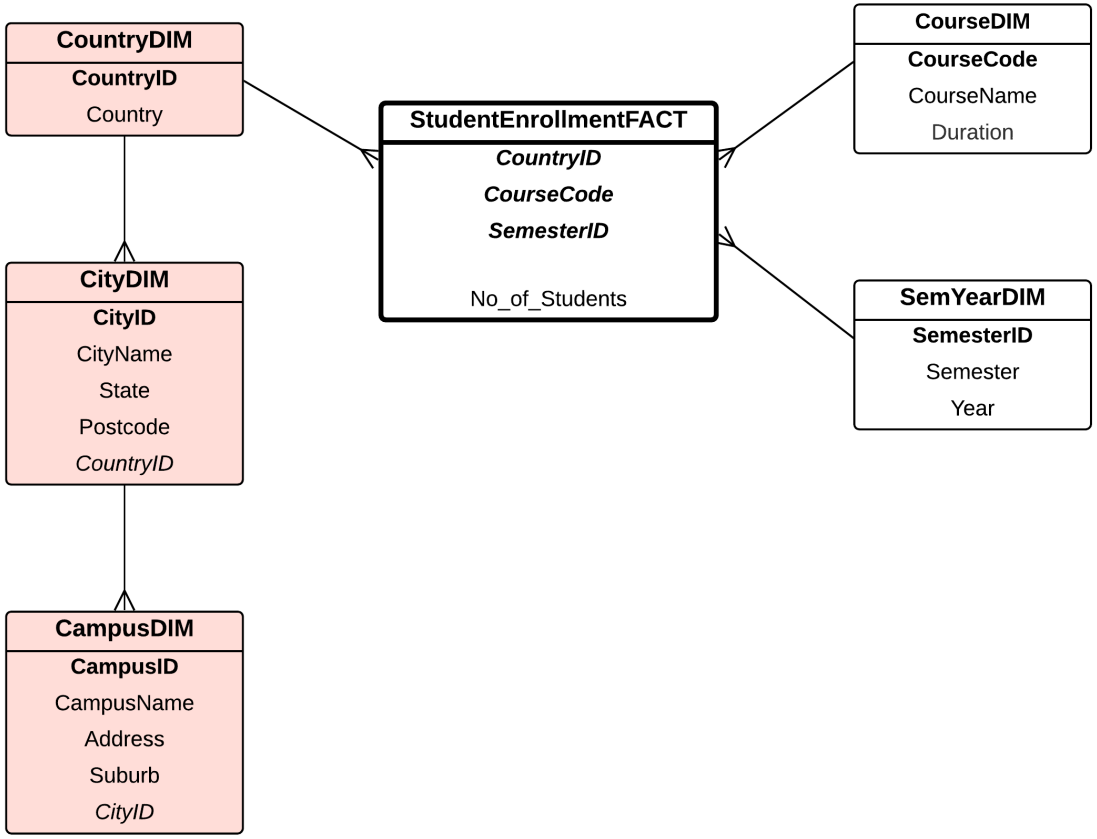
CountryID	Country Name
AU	Australia
MA	Malaysia
SA	South Africa

Hierarchy vs. Non-Hierarchy

(a) Correct Hierarchy



(b) A Wrong Hierarchy



Hierarchy vs. Non-Hierarchy

1. **One table vs. many tables**

   The consequence is in the join query processing when producing reports.

2. **Normalized vs. Un-normalized**

   With the hierarchy option, the tables are normalized, which follows the relational model. In contrast, the non-hierarchy option, the table (e.g. the Campus Dimension table) is unnormalized, which is basically 1NF with visible replication on information, which is prone to anomalies (e.g. insert, update, and delete anomalies).

3. **Drilling down and rolling up**

   The hierarchy model is not about drilling down information exploration.

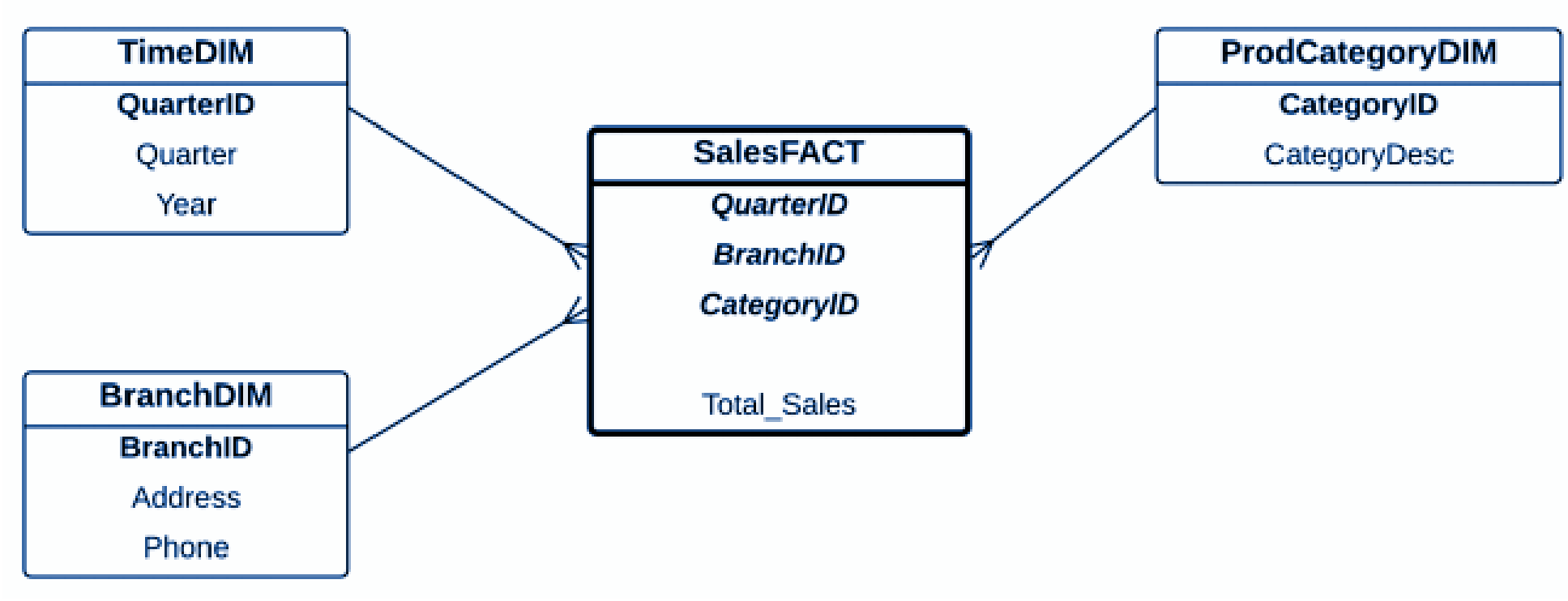
Hierarchy vs. Non-Hierarchy – Summary

- From the **query point of view**, both versions need two queries.
- From the **query processing point of view**, the non-hierarchy version uses one join operation only, because it only needs to join the fact and one dimension.
- From the **conceptual point of view**, the hierarchy model does not actually offer a better roll up or drill down features.

Determinant Dimensions

MONASH University

Dimensions and Fact Tables - Recap



A Star Schema for the Sales Case Study

Querying Fact Table

Example #1:

```
select
  T.Year, P.ProductName,
  sum(Total_Sales) as TotalSales
from
  SalesFact S,
  TimeDim T,
  ProductDim P,
  LocationDim L
where S.TimeID = T.TimeID
and S.ProductNo = P.ProductNo
and S.LocationID = L.LocationID
and T.Year = 2019 and T.Year = 2020
and P.ProductName in ('Shoes', 'Jeans')
and L.Country = 'Australia'
group by cube(T.Year, P.ProductName);
```

Year	Product Name	Total Sales
2019	Shoes	1359800
2019	Jeans	941330
2019		2301130
2020	Shoes	2861450
2020	Jeans	1827480
2020		4688930
	Shoes	4221250
	Jeans	2768810
		6990060

Querying Fact Table (cont.)

Example #2:

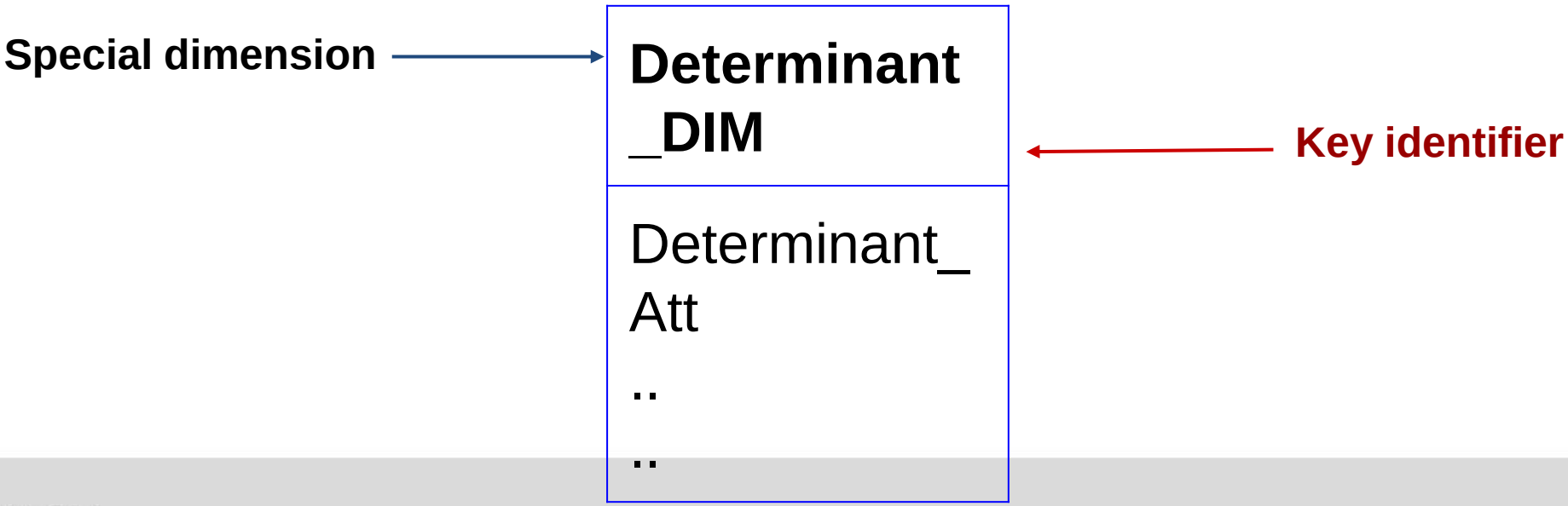
```
select
  T.Year, P.ProductName,
  sum(Total_Sales) as TotalSales
from
  SalesFact S,
  TimeDim T,
  ProductDim P,
where S.TimeID = T.TimeID
and S.ProductNo = P.ProductNo
and T.Year = 2019 and T.Year = 2020
and P.ProductName in ('Shoes', 'Jeans')
group by cube(T.Year, P.ProductName);
```

Determinant Dimensions

A star schema may have a **special dimension**

all data retrieval from this star schema must use this dimension

else, the retrieved data will become meaningless.



Introducing a Determinant Dimension – Petrol Station Case Study

Sample operational database of a company about petrol prices from all petrols stations in Victoria:

Petrol station	Company	Address	Fuel Type	Price	Date	Day of Week	Suburb
Caltex Star Mart Box Hill	Caltex	793-797 Whitehorse Road, Box Hill	Unleaded	120.9	1/08/16	Monday	Box Hill
Caltex Star Mart Box Hill	Caltex	793-797 Whitehorse Road, Box Hill	Diesel	118.9	1/08/16	Monday	Box Hill
Caltex Star Mart Box Hill	Caltex	793-797 Whitehorse Road, Box Hill	LPG	58.9	1/08/16	Monday	Box Hill
Caltex Star Mart Box Hill	Caltex	793-797 Whitehorse Road, Box Hill	U95	132.9	1/08/16	Monday	Box Hill
Caltex Star Mart Box Hill	Caltex	793-797 Whitehorse Road, Box Hill	Premium Unl	139.9	1/08/16	Monday	Box Hill
7-Eleven Clayton	7-Eleven	187-191 Clayton Road, Clayton	unleaded	122.7	1/08/16	Monday	Clayton
7-Eleven Clayton	7-Eleven	187-191 Clayton Road, Clayton	LTG	56.7	1/08/16	Monday	Clayton
7-Eleven Clayton	7-Eleven	187-191 Clayton Road, Clayton	diesel	115.9	1/08/16	Monday	Clayton
United Fitzroy	United	390 Nicholson Street, Fitzroy	Unleaded	119.9	1/08/16	Monday	Fitzroy
United Fitzroy	United	390 Nicholson Street, Fitzroy	Diesel	109.9	1/08/16	Monday	Fitzroy
United Fitzroy	United	390 Nicholson Street, Fitzroy	LPG	59.7	1/08/16	Monday	Fitzroy
7-Eleven Hawthorn	7-Eleven	747-755 Toorak Road, Hawthorn East	Unleaded	121.9	1/08/16	Monday	Hawthorn East

Petrol Station - Star Schema Version-1

The requirements for the data warehouse are to answer the questions related to
(a) average petrol price,
(b) minimum petrol price, and
(c) maximum petrol price

Chosen Dimensions:
(i) day of week,
(ii) suburb, and
(iii) petrol company.

Category	Fact Measures			
Day of week	**Total Petrol Price**	**Num of Petrol Station**	**Min Petrol Price**	**Max Petrol Price**
Monday				
Tuesday				
...				

Two-column Table for Petrol Prices based on Day of Week

Petrol Station - Star Schema Version-1 (cont.)

Minimum price for which petrol type?

Day of week	Total Petrol Price	Num of Petrol Station	Min Petrol Price	Max Petrol Price
Monday				
Tuesday				
...				

Day of week	Total Petrol Price	Num of Petrol Station	Min Petrol Price						Max Petrol Price
			Unleaded	P95	P98	E10	Diesel	LPG	
Monday			119.9c	132.9c	139.9c	114.9c	115.9c	55.7c	
Tuesday			118.9c	131.9c	136.9c	114.9c	115.9c	56.9c	
...									

Petrol Station - Star Schema Version-1 (cont.)

Day of Week	Total Petrol Price						Num of Petrol Station	Min Petrol Price						Max Petrol Price					
	U	95	98	E	D	LPG		U	95	98	E	D	LPG	U	95	98	E	D	LPG
Monday																			
Tuesday																			

- Not all petrol stations sell all types of petrol, e.g. Premium 98.
- Calculating the average price for Premium 98 on Monday **cannot be**:

```
Total_Petrol_Price_P98 / Number_of_Petrol_Station
```

- The number of petrol stations is different for different petrol type!

Petrol Station - Star Schema Version-1 (cont.)

- Solution: Split the Number of Petrol Station column into six columns – one for each petrol type.
- **Question**: How many columns in total?

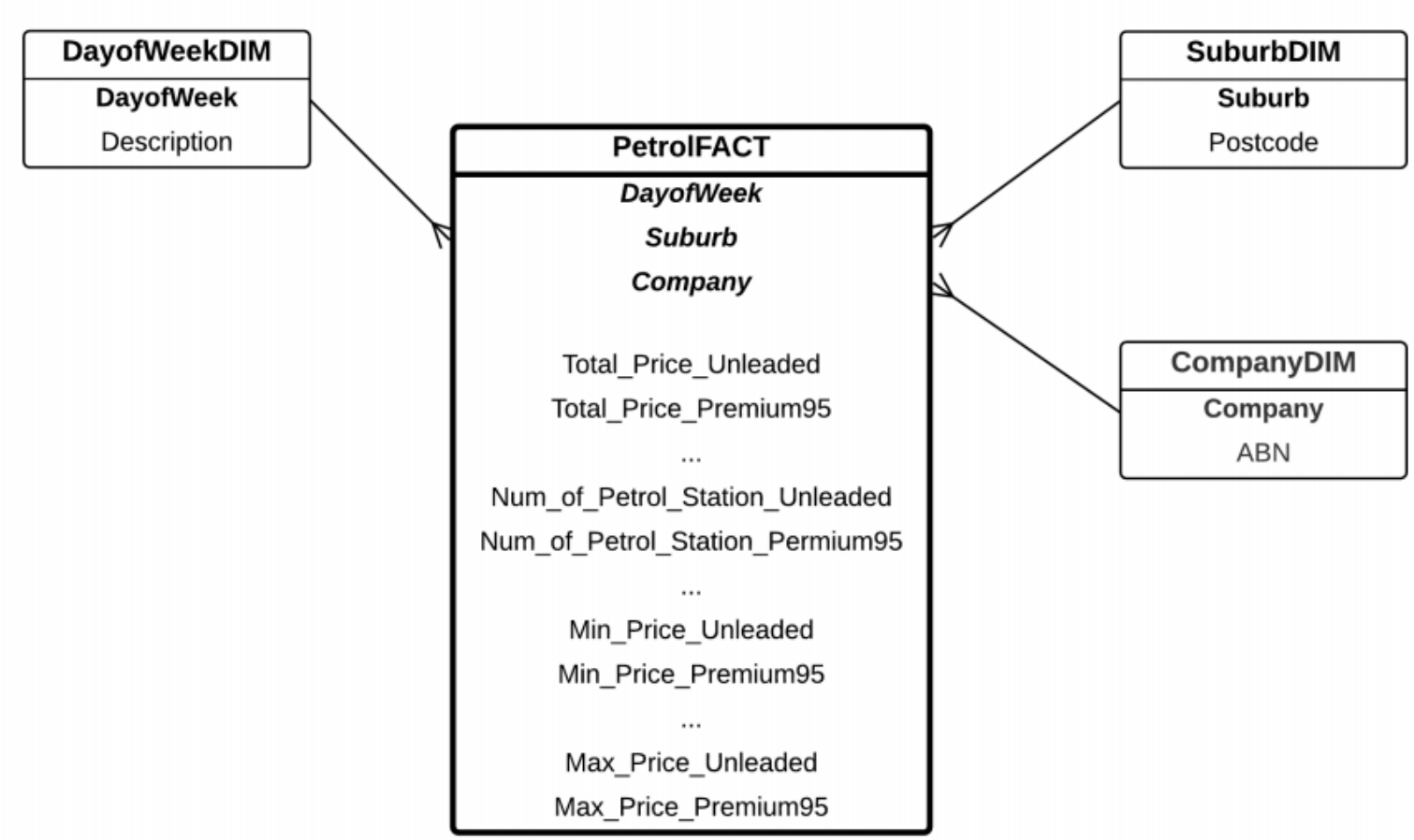
Suburb	Total Petrol Price						Num of Petrol Station						Min Petrol Price						Max Petrol Price					
	U	95	98	E10	D	L	U	95	98	E10	D	L	U	95	98	E10	D	L	U	95	98	E10	D	L
Box Hill																								
Clayton																								

Two-column Table for Petrol Prices based on Suburb

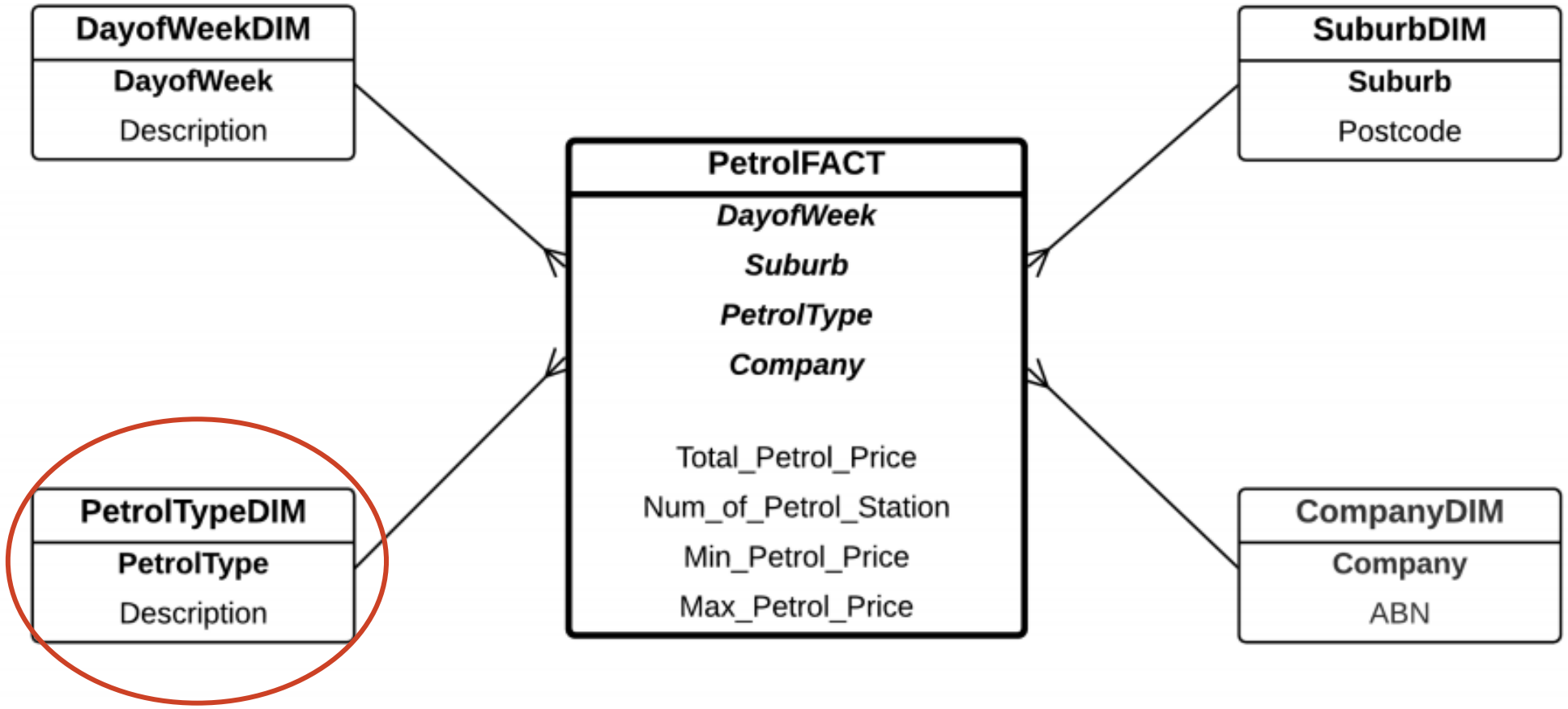
Company	Total Petrol Price						Num of Petrol Station						Min Petrol Price						Max Petrol Price					
	U	95	98	E	D	L	U	95	98	E	D	L	U	95	98	E	D	L	U	95	98	E	D	L
Caltex																								
7-Eleven																								

Two-column Table for Petrol Prices based on Company

Petrol Station - Star Schema Version-1 (cont.)



Petrol Station - Star Schema Version-2

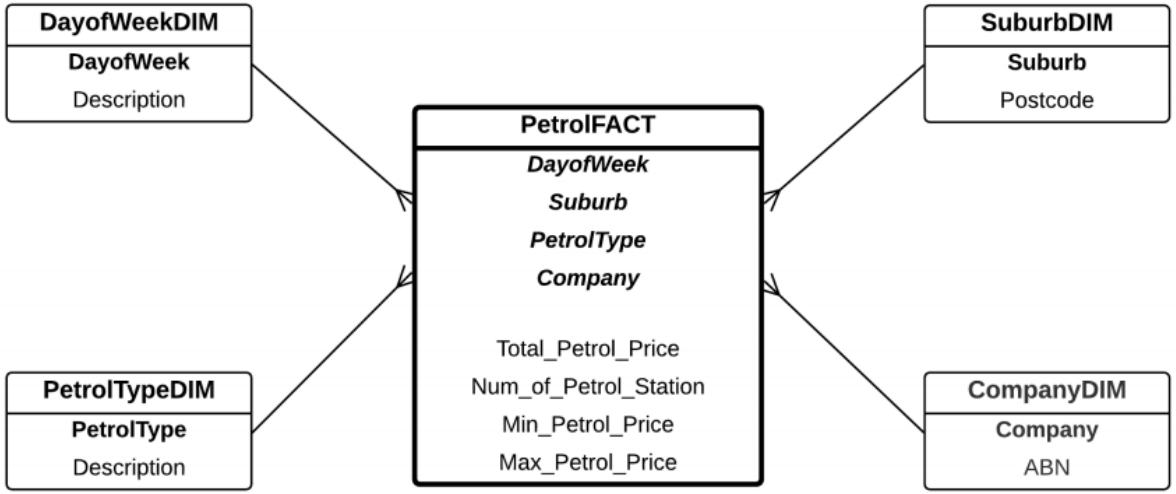


Petrol Station - Star Schema Version-2

```
select DayofWeek, Min_Petrol_Price
from PetrolFact
where DayofWeek = 'Monday';
```

Day of Week	Min Petrol Price
Monday	117.9 cents
Monday	56.7 cents
Monday	109.9cents
Monday	125.9 cents
Monday	133.9 cents
Monday	114.9 cents

Min. price for which petrol type??



Petrol Station - Star Schema Version-2

```
select      DayofWeek,
            Min_Petrol_Price
from PetrolFact
where DayofWeek = 'Monday'
and PetrolType = 'Unleaded';
```

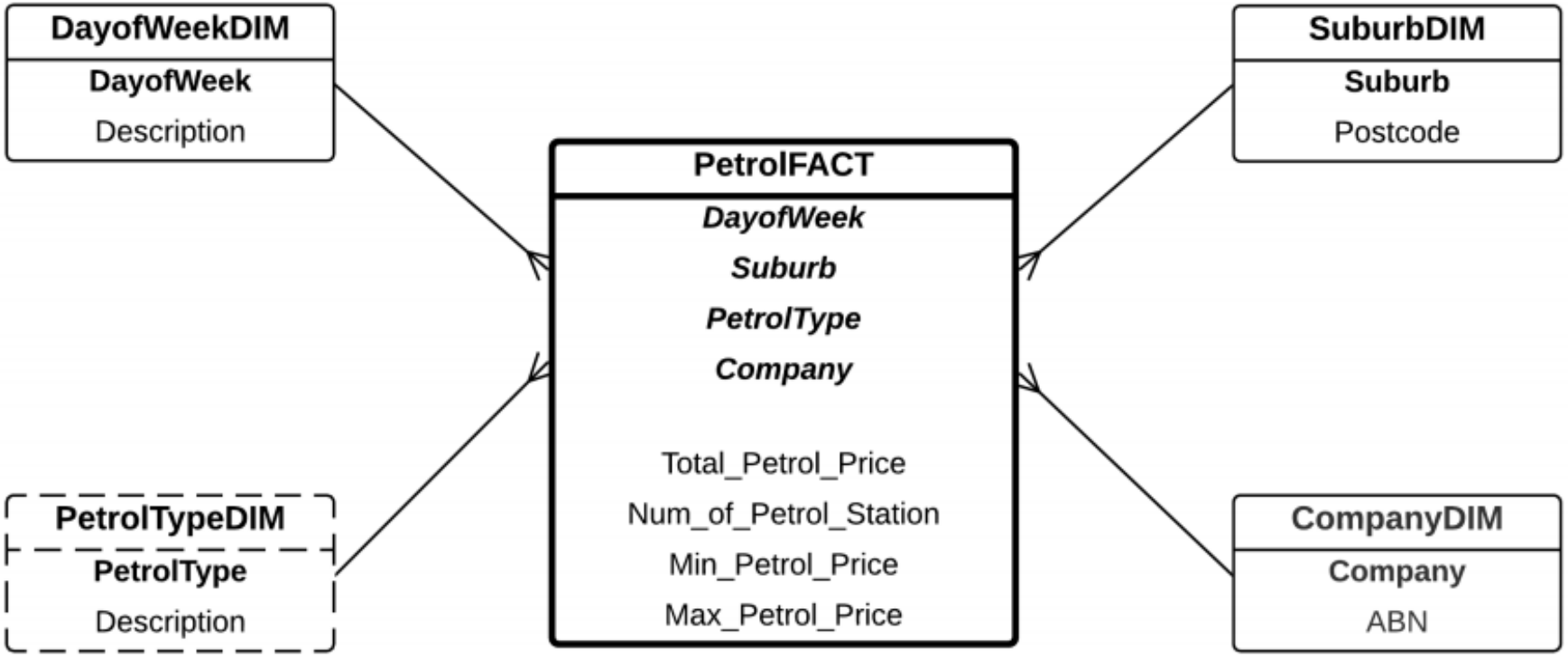
Day of Week	Min Petrol Price
Monday	117.9 cents

```
select      DayofWeek, PetrolType,
            Min_Petrol_Price
from PetrolFact
where DayofWeek = 'Monday';
```

Day of Week	PetrolType	Min Petrol Price
Monday	Unleaded	117.9 cents
Monday	LPG	56.7 cents
Monday	Diesel	109.9cents
Monday	E10	125.9 cents
Monday	Premium 95	133.9 cents
Monday	Premium 98	114.9 cents

Petrol Station - Star Schema Version-2

Determinant Dimension uses dotted box in the star schema.



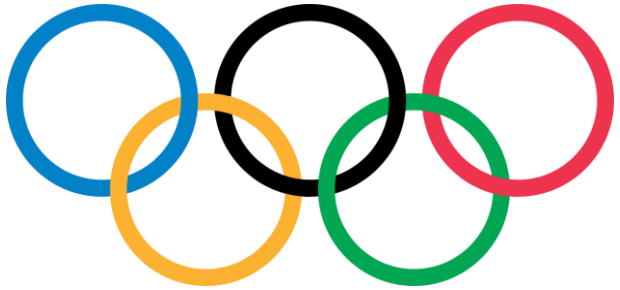
The new star schema with a Determinant Dimension for Petrol Case Study

Determinant vs. Non-Determinant Dimensions

Determinant vs. Non-Determinant Dimensions

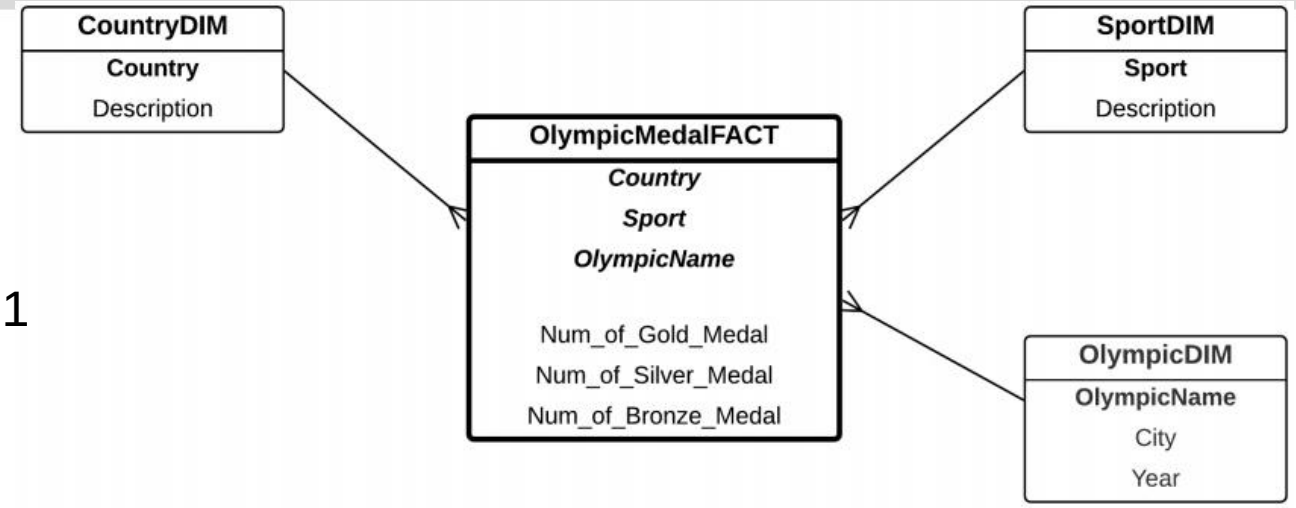
The Olympic Games Case Study

- The Olympic Games committee maintains an operational database that stores all **matches**, **games**, as well as the **medal winner**s of the Olympic Games **over the years**
- Let's build a data warehouse to analyse the medal counts, by each country, sport, and at which Olympic Games

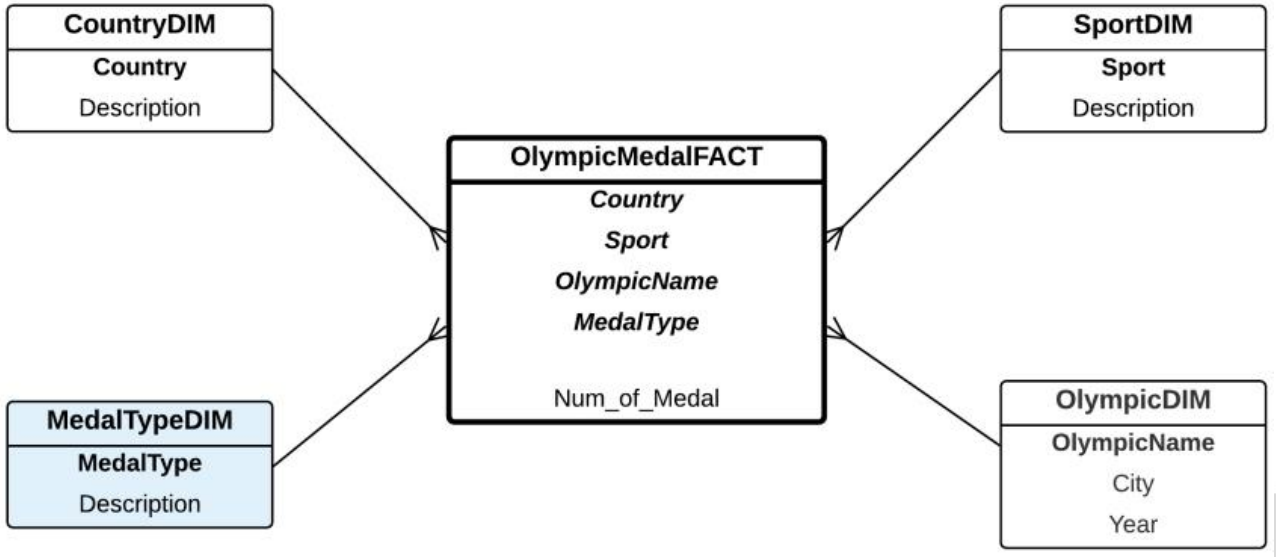


The Olympic Games Case Study - The Star Schemas

Version-1



Version-2



The questions are:

1. What is the difference between these two versions of star schema?
2. Is Medal Type Dimension a Determinant Dimension?

The Olympic Games Case Study - The Star Schema Version-1

Two-column tables (pivot tables) for each dimension category in star schema version-1

Country	Num of Gold	Num of Silver	Num of Bronze
USA	733	602	488
China	199	143	133
Australia	167	170	189

Sport	Num of Gold	Num of Silver	Num of Bronze
Swimming 100m Butterfly Men	20	20	20
Swimming 400m Freestyle Women	20	20	20
Swimming 4x100m Medley Relay Men	20	20	20

Olympic Name	Num of Gold	Num of Silver	Num of Bronze
London 2012	302	304	356
Beijing 2008	302	303	353
Athens 2004	301	301	327

The Olympic Games Case Study - The Star Schema Version-2

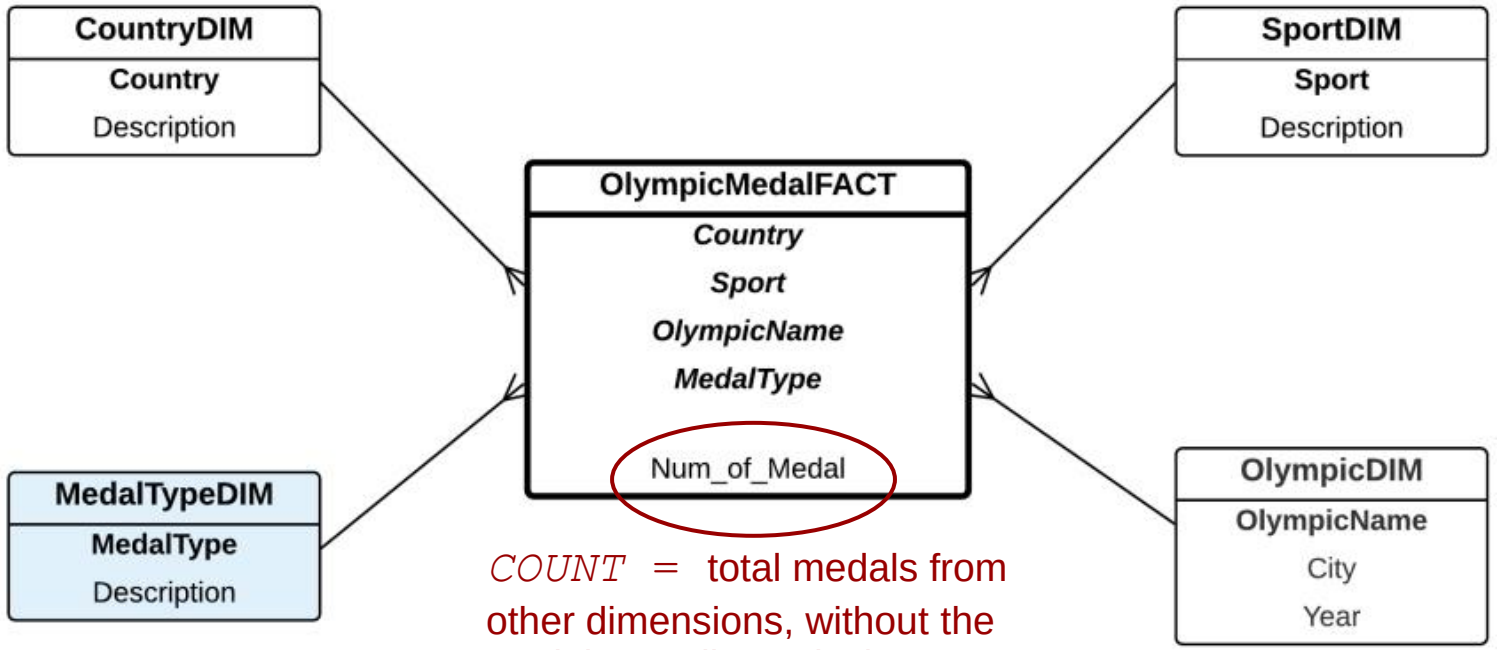
Country	Num of Medals
USA	1823
China	475
Australia	526

Sport	Num of Medals
London 2012	962
Beijing 2008	958
Athens 2004	929

Sport	Num of Medals
Swimming 100m Butterfly Men	60
Swimming 400m Freestyle Women	60
Swimming 4x100m Medley Relay Men	60

Medal Type	Num of Medals
Gold	4115
Silver	4095
Bronze	4474

Determinant vs. Non-Determinant Dimensions



Version-2

- Is MedalTypeDIM a determinant dimension?
  - To retrieve the data from version-2 star schema, must we have the information from Medal Type Dimension?

The Olympic Games Case Study - The Star Schema Version-1 vs Version-2

Fact Table for Version-1

Country	Sport	Olympic Name	Num of Gold	Num of Silver	Num of Bronze
USA	Swimming	London 2012	16	9	6
China	Swimming	London 2012	5	1	4
Australia	Swimming	London 2012	1	6	3

Fact Table for Version-2

Country	Sport	Olympic Name	Medal Type	Num of Medals
USA	Swimming	London 2012	Gold	16
USA	Swimming	London 2012	Silver	9
USA	Swimming	London 2012	Bronze	6
China	Swimming	London 2012	Gold	5
China	Swimming	London 2012	Silver	1
China	Swimming	London 2012	Bronze	4
Australia	Swimming	London 2012	Gold	1
Australia	Swimming	London 2012	Silver	6
Australia	Swimming	London 2012	Bronze	3

The Olympic Games Case Study - The Star Schema Version-1 vs Version-2

Differences	Version 1	Version 2 (with MedalTypeDIM)
Storage	Lower with only three records	Nine records
Modelling	The schema looks more complex and crowded; but the storage cost is lower	More concise and compact with less number of measures in the fact, easier to understand
Query Processing	Less join processing between dimension tables and the fact table.	Requires additional join with MedalTypeDIM

Determinant Dimensions vs. Pivoted Fact Table

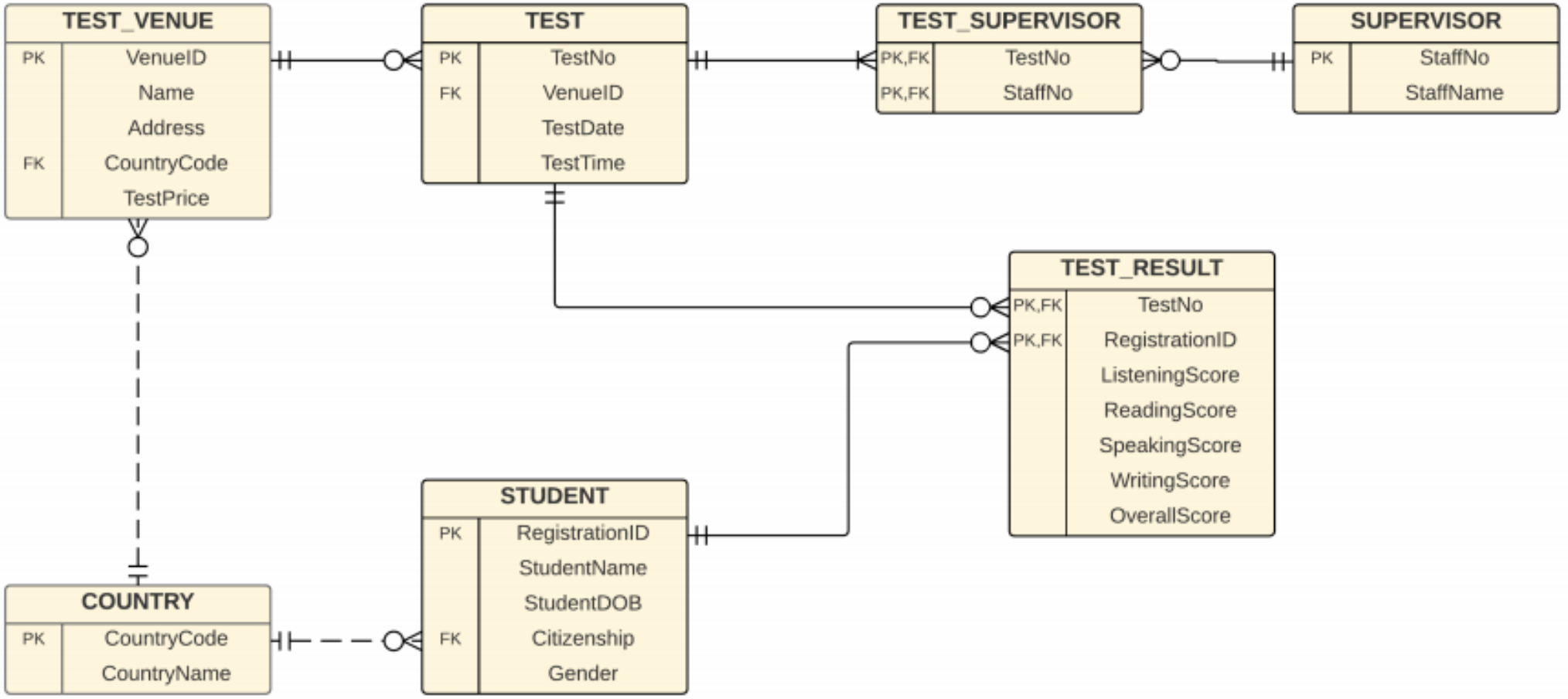
Determinant Dimensions vs. Pivoted Fact Table

PTE ACADEMIC™
Official Test Centre
PEARSON

The PTE Academic Test Case Study

- PTE Test consists of four main components: Listening, Reading, Speaking, and Writing
- There are many test venues (all major cities in Australia, and overseas as well), each country may have a different price setting
- A student taking a PTE test will have one score for each of these four components, as well as one overall score
- The score is a numerical score, ranging from 10 to 90
- The results will normally be released five days after the actual test date
- A student may take a PTE Academic Test several times, in order to improve the score
- A student is identified by his/her RegistrationID, which stays the same student stays the same

ERD for PTE Academic Test



The PTE Academic Test Case Study - Sample Records for 8 Students

Table Test Venue

Venue ID	Name	Address	Price	Country Code
1	RMIT English Worldwide	Level 10 235-251 Bourke Street Melbourne Victoria 3000	300AUD 300AUD	AU AU
2	Pearson Professional Centers San Francisco	Suite 200 201 Filbert Street San Francisco, California	200USD	US
3	PLT at Cliftons Sydney	Level 12 60 Margaret Street Sydney New South Wales 2000	300AUD	AU

Table Test

Test No	Test Date	Test Time	Venue ID
1	01/AUG/17	10:00am	1
2	01/AUG/17	11:00am	1
3	10/AUG/17	10:00am	2
4	10/AUG/17	11:00am	2
5	17/AUG/17	10:00am	3
6	17/AUG/17	11:00am	3
7	17/AUG/17	10:00am	1
8	25/AUG/17	10:00am	1

Table Student

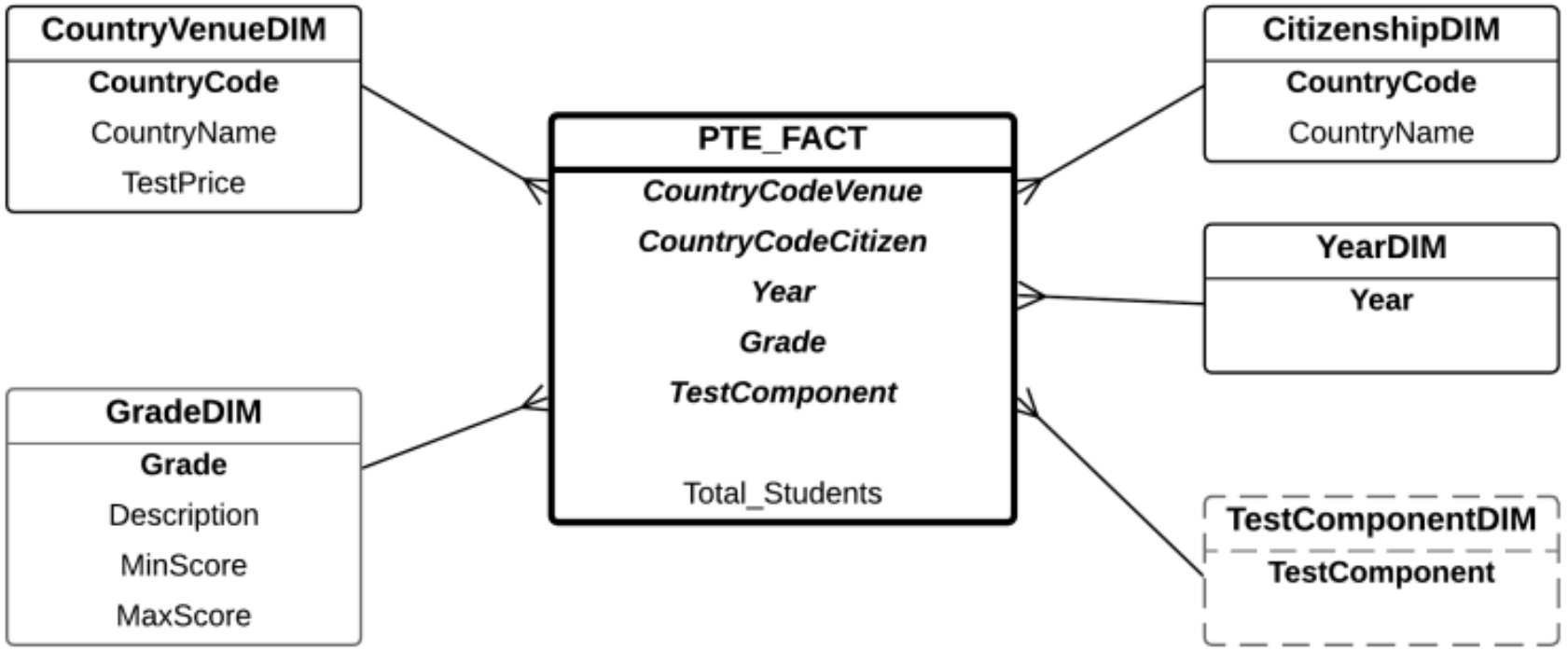
Registration ID	Student Name	DOB	Gender	Citizenship
318678895	Ying Chang	27/AUG/89	F	CN
312124682	Ryou Kato	23/MAR/94	M	JP
117609895	Susi Bambang	19/FEB/91	F	ID
891186588	Su-yeon Park	07/JUL/94	F	KR
387160595	Hong Lei	16/DEC/90	M	CN
312905800	Tim Yeoh	08/OCT/91	M	SG
311765711	Berta Ferrari	27/DEC/87	F	IT
392837402	Wenbo Lin	27/AUG/92	M	CN

The PTE Academic Test Case Study - Sample Records for 8 Students

Table Test Result

Test No	Registration ID	Listening Score	Reading Score	Speaking Score	Writing Score	Overall Score
1	318678895	69	77	78	70	**74**
1	312124682	55	57	60	75	60
2	117609895	68	68	67	69	68
2	891186588	62	63	57	68	61
3	387160595	67	72	78	66	72
3	312905800	66	73	71	67	70
4	311765711	57	55	59	60	57
7	117609895	75	68	78	78	75
7	891186588	69	70	76	75	72
8	318678895	80	79	78	78	**78**
8	312124682	55	45	47	50	47

The PTE Academic Test Case Study - Star Schema with Determinant Dim.



The PTE Academic Test - Star Schema with Determinant Dim. (cont.)

To implement the star schema in SQL:

```
create table CountryVenueDim
as
select distinct C.CountryCode, C.CountryName, T.TestPrice
from Test_Venue T, Country C
where T.CountryCode = C.CountryCode;

create table CitizenshipDim
as
select distinct C.CountryCode as Citizenship, C.CountryName
from Student S, Country C
where S.Citizenship = C.CountryCode;

create table YearDim
as
select distinct to_char(TestDate, 'YYYY') as Year from Test;
```

The PTE Academic Test - Star Schema with Determinant Dim. (cont.)

```
create table GradeDim (
    Grade varchar2(3),
    Description varchar2(20),
    MinScore number,
    MaxScore number
);
insert into GradeDim values ('4.5', 'Functional', 30, 35);
insert into GradeDim values ('5', 'Vocational', 36, 49);
insert into GradeDim values ('6', 'Competent', 50, 64);
insert into GradeDim values ('7', 'Proficient', 65, 78);
insert into GradeDim values ('8-9', 'Superior', 79, 90);
```

PTE band-scale

Proficiency Level	Grade	Test Component	Test Score
Functional	4.5	Overall score	30-35
Vocational	5	Listening	36-49
		Reading	
		Writing	
		Speaking	
Competent	6	Listening	50-64
		Reading	
		Writing	
		Speaking	
Proficient	7	Listening	65-78
		Reading	
		Writing	
		Speaking	
Superior	8-9	Listening	79-90
		Reading	
		Writing	
		Speaking	

The PTE Academic Test - Star Schema with Determinant Dim. (cont.)

```
create table TestComponentDim (
    TestComponent varchar2(20)
);

insert into TestComponentDim values
('Listening');
insert into TestComponentDim values
('Reading');
insert into TestComponentDim values
('Writing');
insert into TestComponentDim values
('Speaking');
insert into TestComponentDim values
('Overall');
```

Create temporary fact table:

```
create table TempFact
as
select TV.CountryCode, S.Citizenship,
to_char(T.TestDate, 'YYYY') as Year,
TR.ListeningScore, TR.ReadingScore,
TR.WritingScore, TR.SpeakingScore,
TR.OverallScore, TR.RegistrationID
from Test_Venue TV, Test T, Student S,
Test_Result TR
where TV.VenueID = T.VenueID
and T.TestNo = TR.TestNo
and TR.RegistrationID = S.RegistrationID;
```

The PTE Academic Test - Star Schema with Determinant Dim. (cont.)

```
alter table TempFact
add (
      GradeOverall varchar2(3),
      GradeListening varchar2(3),
      GradeReading varchar2(3),
      GradeWriting varchar2(3),
      GradeSpeaking varchar2(3)
);
```

```
update TempFact
set GradeOverall = (
case
   when OverallScore >= 30 and
      OverallScore <= 35 then '4.5'
   when OverallScore >= 36 and
      OverallScore <= 49 then '5'
   when OverallScore >= 50 and
      OverallScore <= 64 then '6'
   when OverallScore >= 65 and
      OverallScore <= 78 then '7'
   when OverallScore >= 79 and
      OverallScore <= 90 then '8-9'
end);
```

Repeat for GradeListening, GradeReading, GradeWriting, and GradeSpeaking

PTE band-scale

Proficiency Level	Grade	Test Component	Test Score
Functional	4.5	Overall score	30-35
Vocational	5	Listening	36-49
		Reading	
		Writing	
		Speaking	
Competent	6	Listening	50-64
		Reading	
		Writing	
		Speaking	
Proficient	7	Listening	65-78
		Reading	
		Writing	
		Speaking	
Superior	8-9	Listening	79-90
		Reading	
		Writing	
		Speaking	

The PTE Academic Test - Star Schema with Determinant Dim. (cont.)

- The TempFact table has the correct grades for each student.
- The problem now is to breakdown the grades for a student into multiple test components
  - Create another temporary fact for each of the test components, eg. a temporary fact called 'OverallFact' for the Overall Score, and so on so forth

Repeat for
GradeListening, GradeReading, GradeWriting, and GradeSpeaking

```
  create table OverallFact
as
select   CountryCode, Citizenship, Year,
         GradeOverall As Grade, 'Overall' as TestComponent,
         count(RegistrationID) as Total_Students_Overall
from TempFact
group by CountryCode, Citizenship, Year, GradeOverall, 'Overall';
```

The PTE Academic Test - Star Schema with Determinant Dim. (cont.)

- There were eleven students taking the test which are now grouped into nine records in Table OverallFact

OverallFact Table

Country Code	Citizenship	Year	Grade	Test Component	Total Students Overall
AU	JP	2017	6	Overall	1
AU	KR	2017	6	Overall	1
AU	ID	2017	7	Overall	2
US	IT	2017	6	Overall	1
AU	JP	2017	5	Overall	1
US	CN	2017	7	Overall	1
US	SG	2017	7	Overall	1
AU	KR	2017	7	Overall	1
AU	CN	2017	7	Overall	2

ListeningFact Table

Country Code	Citizenship	Year	Grade	Test Component	Total Students Listening
AU	CN	2017	7	Listening	1
AU	ID	2017	7	Listening	2
US	IT	2017	6	Listening	1
AU	KR	2017	6	Listening	1
US	CN	2017	7	Listening	1
US	SG	2017	7	Listening	1
AU	JP	2017	6	Listening	2
AU	CN	2017	8-9	Listening	1
AU	KR	2017	7	Listening	1

The PTE Academic Test - Star Schema with Determinant Dim. (cont.)

- There are also three other fact tables: ReadingFact, WritingFact, and SpeakingFact, with identical table structures
- These five fact tables can now be 'combined' to form one final fact table using a *union* operator:

```
create table FinalFact as
select CountryCode, Citizenship, Year, Grade, TestComponent,
Total_Students_Overall as Total_Students
from OverallFact
union select * from ListeningFact
union select * from ReadingFact
union select * from WritingFact
union select * from SpeakingFact;
```

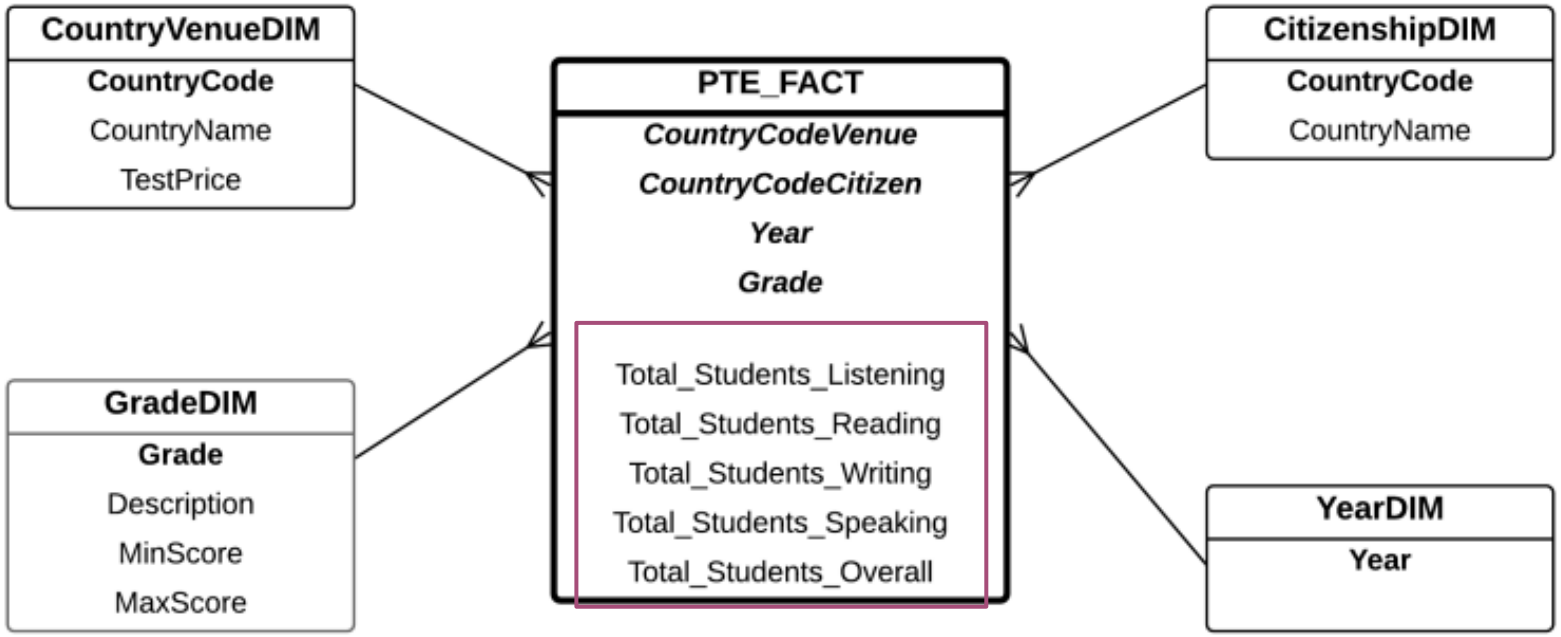
The PTE Academic Test - Star Schema with Determinant Dim. (cont.)

- The final fact table contains 45 records
- Conclusion:
  - When the input record in the operational database contains the scores of all the test components as one record, it will be **challenging to break down the records into multiple records** in the final fact table, because there is a Test Component dimension.

Final Fact Table

Country Code	Citizenship	Year	Grade	Test Component	Total Students
AU	CN	2017	7	Listening	1
AU	CN	2017	7	Overall	2
AU	CN	2017	7	Reading	1
AU	CN	2017	7	Speaking	2
AU	CN	2017	7	Writing	2
AU	CN	2017	8-9	Listening	1
AU	CN	2017	8-9	Reading	1
...	...	...	...	...	...
...	...	...	...	...	...
US	SG	2017	7	Listening	1
US	SG	2017	7	Overall	1
US	SG	2017	7	Reading	1
US	SG	2017	7	Speaking	1
US	SG	2017	7	Writing	1

The PTE Academic Test - Star Schema without Determinant Dim.



- Fact measure for all the test components
- This means that the fact table is a Pivoted Fact Table
  - Shifts the Test Component into the fact measure;
  - The Determinant Dimension - TestComponentDim is then removed from the star schema

The PTE Academic Test - Star Schema without Determinant Dim. (cont.)

- How many records do we have in fact table with five fact measures?

Answer:
60 records in the final fact table
= 2 countries of test venues * 6 countries of citizenship * 1 year * 5 grades

- However, since our test result data is incomplete, we will have problems with the final fact, because in the non-determinant dimension version, **we need to keep track the zeroes**
  - Even if the data is the test result is a lot, but still does not cover all possible combinations from the dimension!

The PTE Academic Test - Star Schema without Determinant Dim. (cont.)

Table Test Result

Test No	Registration ID	Listening Score	Reading Score	Speaking Score	Writing Score	Overall Score
1	318678895	69	77	78	70	**74**
1	312124682	55	57	60	75	60
2	117609895	68	68	67	69	68
2	891186588	62	63	57	68	61
3	387160595	67	72	78	66	72
3	312905800	66	73	71	67	70
4	311765711	57	55	59	60	57
7	117609895	75	68	78	78	75
7	891186588	69	70	76	75	72
8	318678895	80	79	78	78	**78**
8	312124682	55	45	47	50	47

PTE band-scale

Proficiency Level	Grade	Test Component	Test Score
Functional	4.5	Overall score	30-35
Vocational	5	Listening	36-49
		Reading	
		Writing	
		Speaking	
Competent	6	Listening	50-64
		Reading	
		Writing	
		Speaking	
Proficient	7	Listening	65-78
		Reading	
		Writing	
		Speaking	
Superior	8-9	Listening	79-90
		Reading	
		Writing	
		Speaking	

- 'Superior' grade is only available for Listening and Reading scores
- Using the normal fact table, there will not be entry in the fact table for Superior-Speaking, and for Superior-Writing;
- There will not be zero value in TotalStudents, where the Grade column is 'Superior' and the Test Component column is either 'Speaking' or 'Writing' → less than 60 rows of records in the fact table

The PTE Academic Test - Star Schema without Determinant Dim. (cont.)

Solution: Step1

Get all possible combinations from all dimensions using Cartesian product between all dimensions:

```
create table AllDimensions
as
select CO.CountryCode, CI.Citizenship, Y.Year, G.Grade
from   CountryVenueDim CO,CitizenshipDim CI,
       YearDim Y, GradeDim G;
```

The PTE Academic Test - Star Schema without Determinant Dim. (cont.)

Solution: Step 2

Re-create five temporary fact tables, one for each test component, using Outer Join operation between AllDimensions table and each of the temporary fact tables created in the previous section (see slide #51).

```
create table OverallFactNew as
select  A.CountryCode, A.Citizenship,
          A.Year, A.Grade,
          nvl(O.Total_Students_Overall, 0)
          as Total_Students_Overall
from AllDimensions A, OverallFact O
where A.CountryCode = O.CountryCode(+)
and A.Citizenship = O.Citizenship(+)
and A.Year = O.Year(+)
and A.Grade = O.Grade(+);
```

```
create table ListeningFactNew as
select    A.CountryCode, A.Citizenship,
          A.Year, A.Grade,
          nvl(O.Total_Students_Listening, 0)
          as Total_Students_Listening
from AllDimensions A, ListeningFact O where
A.CountryCode = O.CountryCode(+)
and A.Citizenship = O.Citizenship(+)
and A.Year = O.Year(+)
and A.Grade = O.Grade(+);
create table ReadingFactNew as select ...;
create table WritingFactNew as select ...;
create table SpeakingFactNew as select ...;
```

The PTE Academic Test - Star Schema without Determinant Dim. (cont.)

Solution: Step 3

Create the final fact table using the join operation

```
create table FinalFact2 as
select O.CountryCode, O.Citizenship, O.Year, O.Grade, O.Total_Students_Overall, L.Total_Students_Listening,
            R.Total_Students_Reading, W.Total_Students_Writing, S.Total_Students_Speaking

from     OverallFactNew O, ListeningFactNew L, ReadingFactNew R, WritingFactNew W, SpeakingFactNew S

where O.CountryCode = L.CountryCode
and L.CountryCode = R.CountryCode and R.CountryCode = W.CountryCode
and W.CountryCode = S.CountryCode and O.Citizenship = L.Citizenship
and L.Citizenship = R.Citizenship and R.Citizenship = W.Citizenship and W.Citizenship = S.Citizenship
and O.Year = L.Year and L.Year = R.Year and R.Year = W.Year and W.Year = S.Year
and O.Grade = L.Grade and L.Grade = R.Grade and R.Grade = W.Grade and W.Grade = S.Grade;
```

The PTE Academic Test - Star Schema without Determinant Dim. (cont.)

Solution: Step 4

Because our sample data is small, there are many records with zeroes values in the Total Students columns. Delete records where Total Students in all of the five test components are equal to zero.

```
delete from FinalFact2
where Total_Students_Overall = 0
and Total_Students_Listening = 0
and Total_Students_Reading = 0
and Total_Students_Writing = 0
and Total_Students_Speaking = 0;
```

Final Fact Table

Country Code	Citizenship	Year	Grade	Test Students Overall	Total Students Listening	Total Students Reading	Total Students Writing	Total Students Speaking
AU	KR	2017	6	1	1	1	0	1
US	IT	2017	6	1	1	1	1	1
AU	CN	2017	7	2	1	1	2	2
AU	JP	2017	6	1	2	1	1	1
AU	ID	2017	7	2	2	2	2	2
AU	KR	2017	7	1	1	1	2	1
US	CN	2017	7	1	1	1	1	1
AU	JP	2017	5	1	0	1	0	1
US	SG	2017	7	1	1	1	1	1
AU	CN	2017	8-9	0	1	1	0	0
AU	JP	2017	7	0	0	0	1	0

Determinant Dimension - Conclusion

A Determinant Dimension (or a Determinant Attribute) must be used in retrieving the fact, in order to make the retrieved data more meaningful.

In other words, it is critical to use the Determinant Attribute in the query, either as a filtering mechanism (in the where clause) or in the display (in the select clause) in the SQL command.