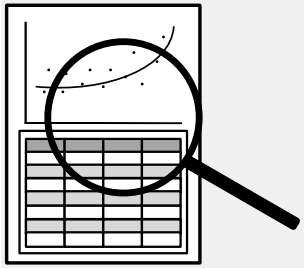
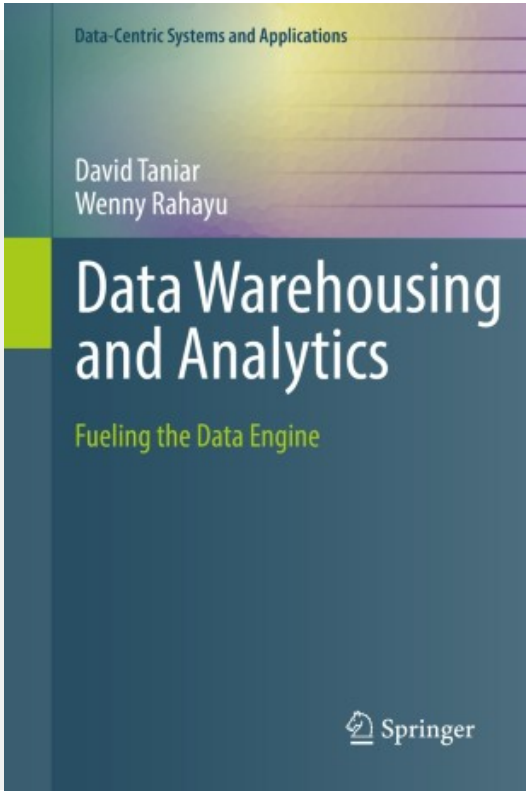


Chapter 04

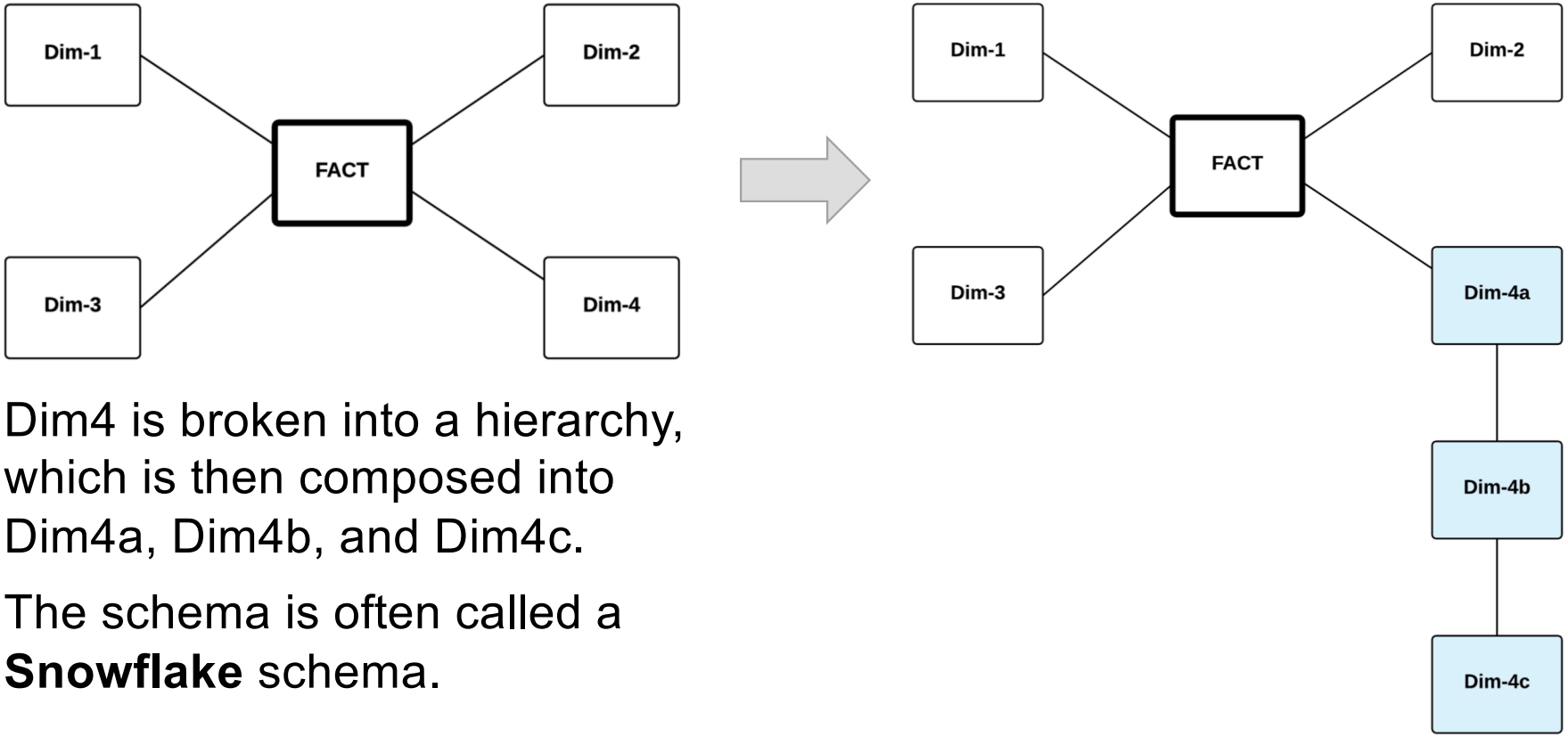
Hierarchy



Outline

- A **Hierarchy** is formed when a dimension is broken down to two or more dimensions in a hierarchical manner.
- This chapter focuses on how and when to design hierarchical dimensions in a star schema, and to compare the differences between the hierarchical version and the non-hierarchical version, which are:
    1) Hierarchy versus Non-Hierarchy Dimensions
    2) Hierarchy versus Multiple Independent Dimensions
    3) Linked Dimensions
    4) Hierarchy Design Considerations

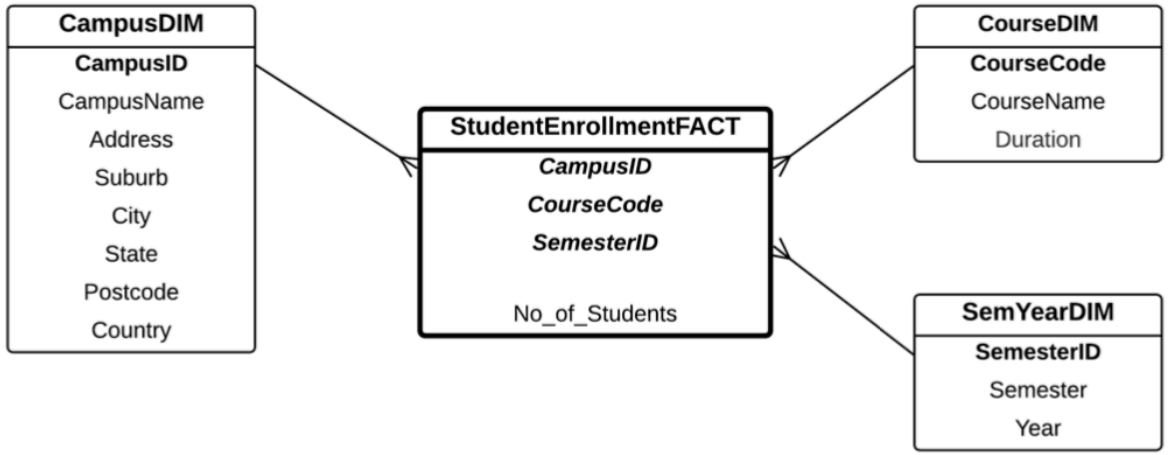
1. Hierarchy versus Non-Hierarchy



- Dim4 is broken into a hierarchy, which is then composed into Dim4a, Dim4b, and Dim4c.
- The schema is often called a **Snowflake** schema.

1. Hierarchy versus Non-Hierarchy

- In the Student Enrolment case study, the fact of number of students enrolment is examined from the course, the campus, and the semester/year point of view.



1.1. Option 1: A Non-hierarchy Version

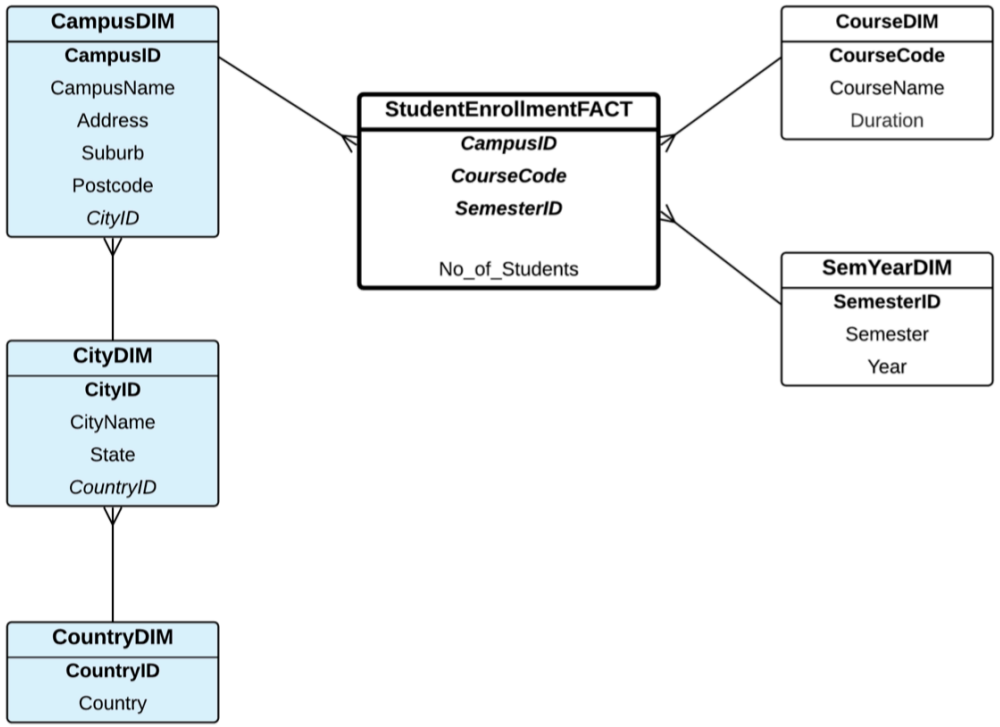
- Campus Dimension

Table 4.1 Campus Dimension table

Campus ID	Campus Name	Address	Suburb	City	State	Postcode	Country
CL	Clayton Campus	Wellington Road	Clayton	Melbourne	Victoria	3800	Australia
CA	Caulfield Campus	Dandenong Road	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 Street	Johannes-burg	Johannes-burg	Johannes-burg	1725	South Africa

1.2. Option 2: A Hierarchy Version

- Another option to present the schema is by using a hierarchy on the Campus Dimension
- The *CityID* attribute in Campus dimension is linked to the *CityID* attribute in the City dimension; through a *many*-1 relationship.
- The same is applied to the City and Country hierarchy, the *CountryID* attribute in City dimension is linked to the *CountryID* attribute in the Country dimension, through a *many*-1 relationship.



1.2. Option 2: A Hierarchy Version

Table 4.2 Campus Dimension table

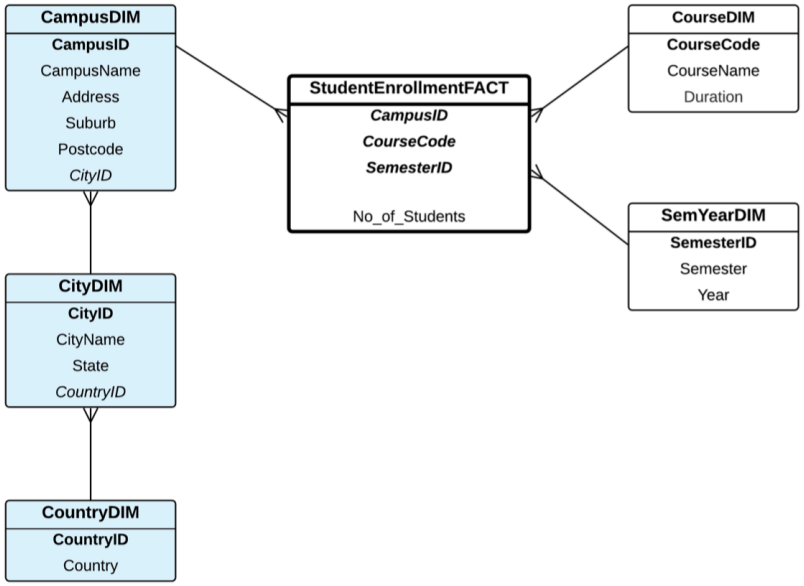
Campus ID	Campus name	Address	Suburb	Postcode	CityID
CL	Clayton Campus	Wellington Road	Clayton	3800	MEL
CA	Caulfield Campus	Dandenong Road	Caulfield East	3145	MEL
PA	Parkville Campus	Royal Parade	Parkville	3052	MEL
SY	Sydney Campus	Opera Boulevard	Sydney	2001	SYD
MUM	Malaysia Campus	Jalan Lagoon	Bandar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

Table 4.3 City Dimension table

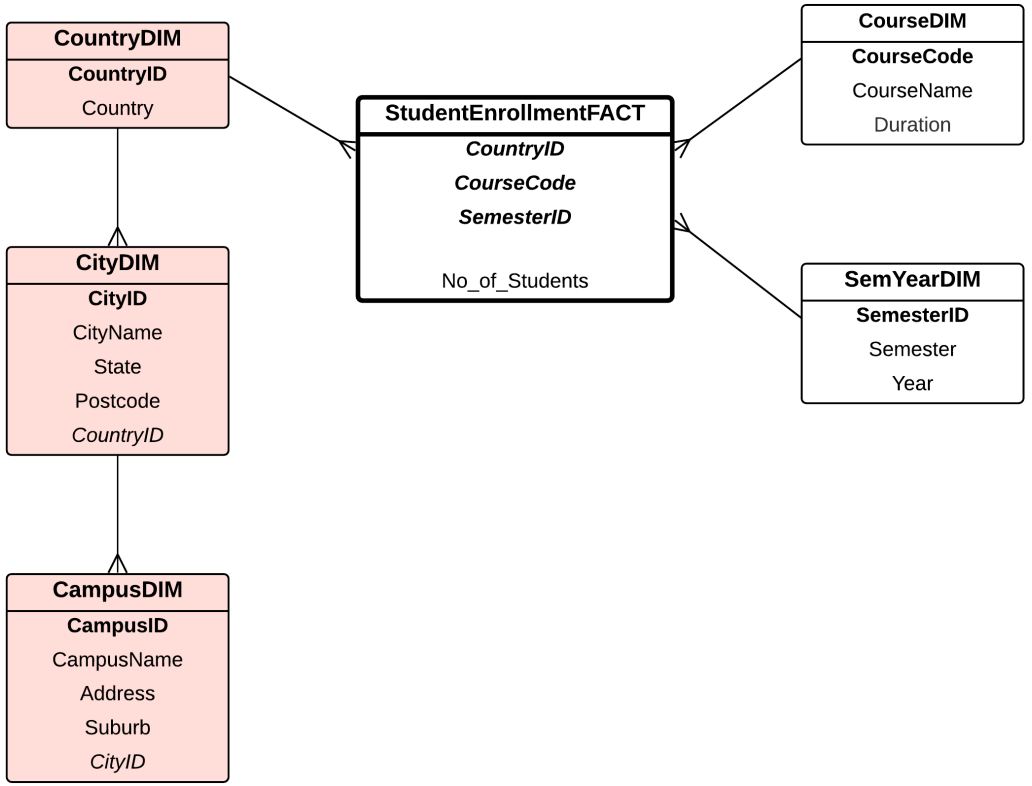
City ID	City name	State	Country ID
MEL	Melbourne	Victoria	AU
SYD	Sydney	New South Wales	AU
KUL	Kuala Lumpur	Selangor	MA
JNB	Johannesburg	Johannesburg	SA

Table 4.4 Country Dimension table

Country ID	Country name
AU	Australia
MA	Malaysia
SA	South Africa



1.3. Wrong Hierarchy



1. Hierarchy versus Non-Hierarchy

1) One tables vs. many tables
   - Without hierarchy, there is only one table.
   - With hierarchy, there are three tables

2) 3NF vs. lower than 3NF
   - With the hierarchy option, the tables are in 3NF.
   - With the non-hierarchy option, the table (e.g. the Campus Dimension table) is in 2NF.

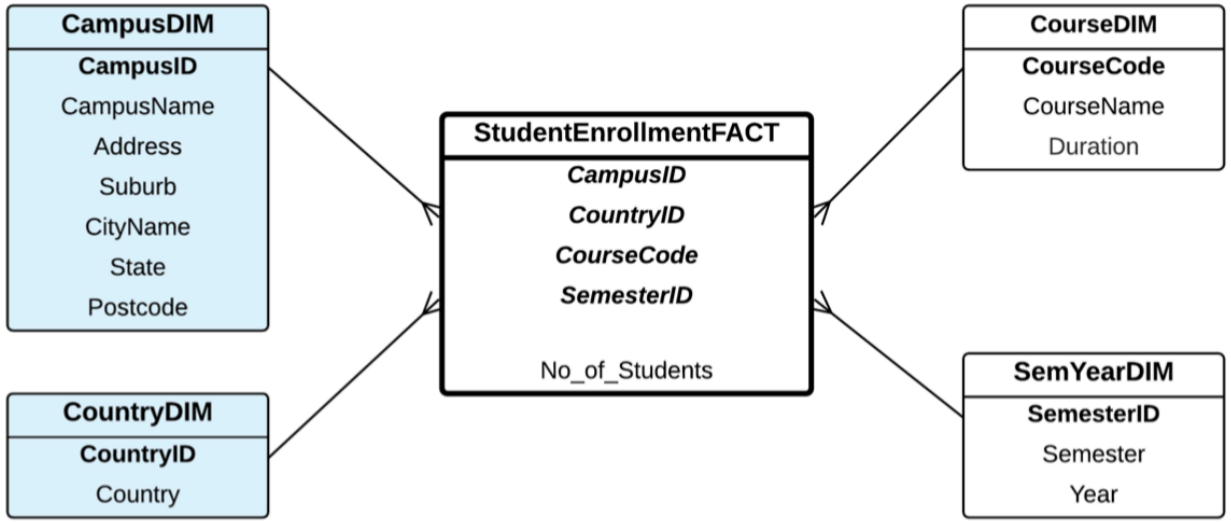
3)Drilling down and rolling up
   - The hierarchy model is not about drilling down information exploration.
   - Rolling up is exploring information from the detail to a more general.

1. Hierarchy versus Non-Hierarchy

- 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.

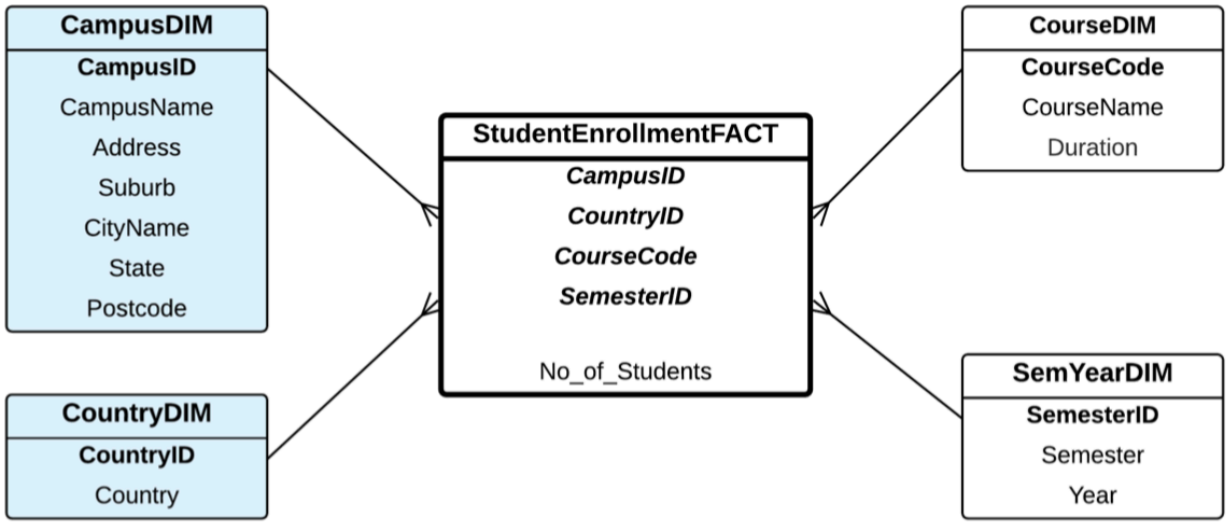
2. Hierarchy versus Multiple Independent Dimensions

- CampusDIM and CountryDIM are independent dimensions, connected to the Fact



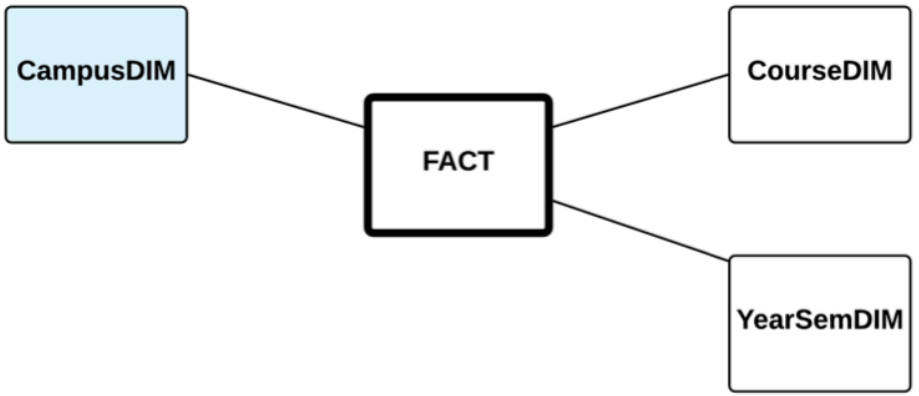
2. Hierarchy versus Multiple Independent Dimensions

- A few possibilities:
  - **Separate Dimensions**
  - Combined Dimension
  - Hierarchy Dimensions



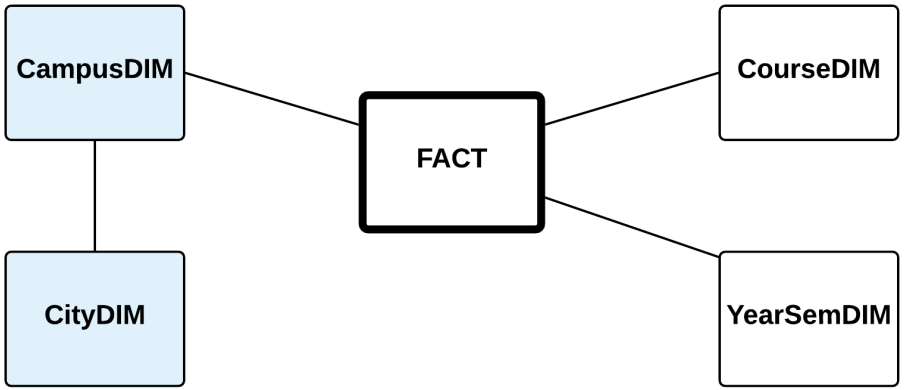
2. Hierarchy versus Multiple Independent Dimensions

- A few possibilities:
  - Separate Dimensions
  - **Combined Dimension**
  - Hierarchy Dimensions



2. Hierarchy versus Multiple Independent Dimensions

- A few possibilities:
  - Separate Dimensions
  - Combined Dimension
  - **Hierarchy Dimensions**



2.1. Separate vs Combined Dimension

- Separate Dimension Model – Campus Dimension Table

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

- Separate Dimension Model – Country Dimension Table

CountryID	Country Name
AU	Australia
MA	Malaysia
SA	South Africa

- Separate Dimension Model – Fact Table

CampusID	CountryID	CourseCode	SemesterID	No of Students
CL	AU	A3001	202001	450
CL	AU	A3002	202001	150
CL	AU	A3003	202001	200
CL	AU	...	...	...
CA	AU	B5001	202001	115
CA	AU	B5002	202001	160
CA	AU	...	...	...
PA	AU	C6001	202001	75
PA	AU	C6002	202001	50
PA	AU	...	...	...
SY	AU	A3001	202001	40
SY	AU	B5002	202001	35
SY	AU	...	...	...
MUM	MA	A3001	202001	150
MUM	MA	B5001	202001	80
MUM	MA	B5002	202001	100
MUM	MA	...	...	...
MSA	SA	A3002	202001	25
MSA	SA	A3002	202001	20
MSA	SA	...	...	...

2.1. Separate vs Combined Dimension

- Combined Dimension Model – Campus Dimension Table

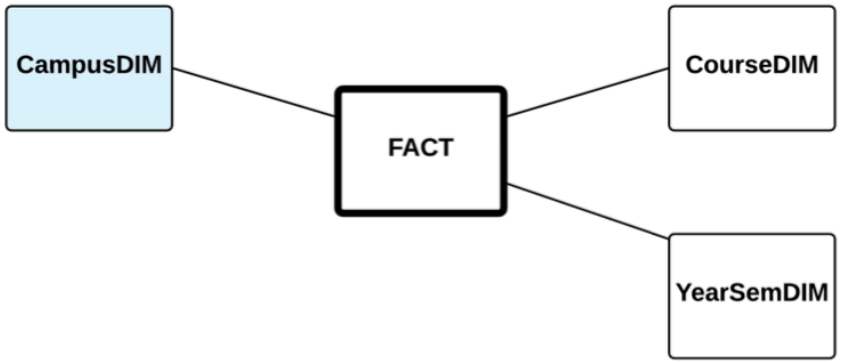
CampusID	CampusName	Address	Suburb	CityName	State	Postcode	Country
CL	Clayton Campus	Wellington Rd	Clayton	Melbourne	Victoria	3800	Australia
CA	Caulfield Campus	Dandenong Rd	Caulfield East	Melbourne	Victoria	3145	Australia
PA	Parkville Campus	Royal Parade	Parkville	Melbourne	Victoria	3052	Australia
SY	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

- Combined Dimension Model – Fact Table

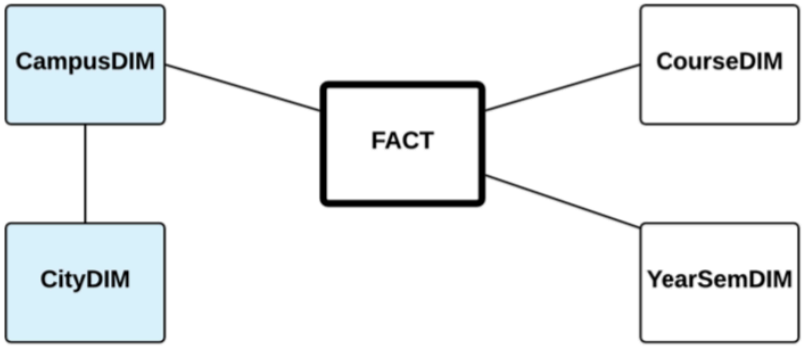
CampusID	CourseCode	SemesterID	No of Students
CL	A3001	202001	450
CL	A3002	202001	150
CL	A3003	202001	200
CL	...	...	...
CA	B5001	202001	115
CA	B5002	202001	160
CA	...	...	...
PA	C6001	202001	75
PA	C6002	202001	50
PA	...	...	...
SY	A3001	202001	40
SY	B5002	202001	35
SY	...	...	...
MUM	A3001	202001	150
MUM	B5001	202001	80
MUM	B5002	202001	100
MUM	...	...	...
MSA	A3002	202001	25
MSA	A3002	202001	20
MSA	...	...	...

2.2. Combined Dimension vs. Hierarchy

- Combined Dimensions



- Hierarchy Dimensions



2.2. Combined Dimension vs. Hierarchy

- Hierarchy Model – Campus Dimension Table

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

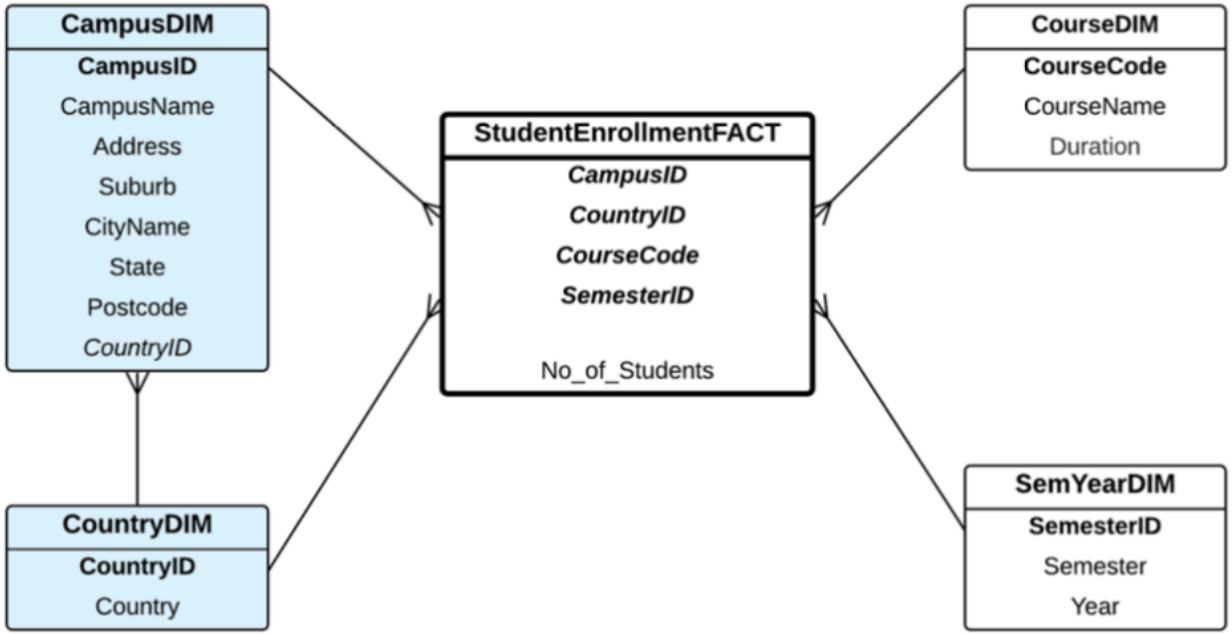
- Hierarchy Model – Country Dimension Table

CountryID	Country Name
AU	Australia
MA	Malaysia
SA	South Africa

- Hierarchy Model – Fact Table

CampusID	CourseCode	SemesterID	No of Students
CL	A3001	202001	450
CL	A3002	202001	150
CL	A3003	202001	200
CL	...	...	...
CA	B5001	202001	115
CA	B5002	202001	160
CA	...	...	...
PA	C6001	202001	75
PA	C6002	202001	50
PA	...	...	...
SY	A3001	202001	40
SY	B5002	202001	35
SY	...	...	...
MUM	A3001	202001	150
MUM	B5001	202001	80
MUM	B5002	202001	100
MUM	...	...	...
MSA	A3002	202001	25
MSA	A3002	202001	20
MSA	...	...	...

3. Linked Dimensions



3. Linked Dimensions

- Linked Dimension Model – Campus Dimension Table

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

- Linked Dimension Model – Country Dimension Table

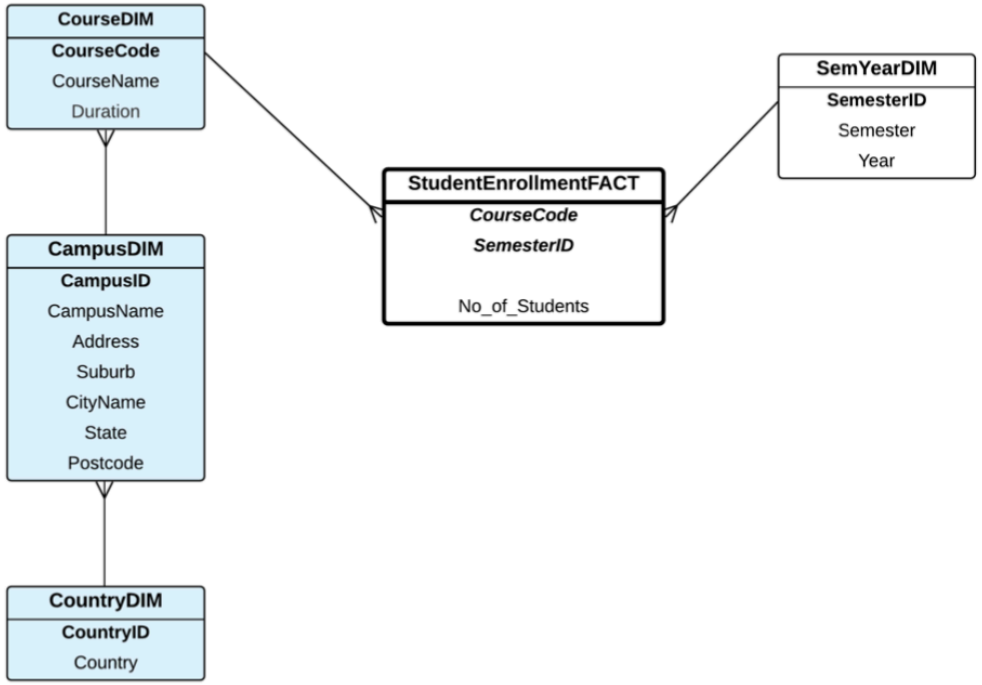
CountryID	Country Name
AU	Australia
MA	Malaysia
SA	South Africa

- Linked Dimension Model – Fact Table

CampusID	CountryID	CourseCode	SemesterID	No of Students
CL	AU	A3001	202001	450
CL	AU	A3002	202001	150
CL	AU	A3003	202001	200
CL	AU	...	...	...
CA	AU	B5001	202001	115
CA	AU	B5002	202001	160
CA	AU	...	...	...
PA	AU	C6001	202001	75
PA	AU	C6002	202001	50
PA	AU	...	...	...
SY	AU	A3001	202001	40
SY	AU	B5002	202001	35
SY	AU	...	...	...
MUM	MA	A3001	202001	150
MUM	MA	B5001	202001	80
MUM	MA	B5002	202001	100
MUM	MA	...	...	...
MSA	SA	A3002	202001	25
MSA	SA	A3002	202001	20
MSA	SA	...	...	...

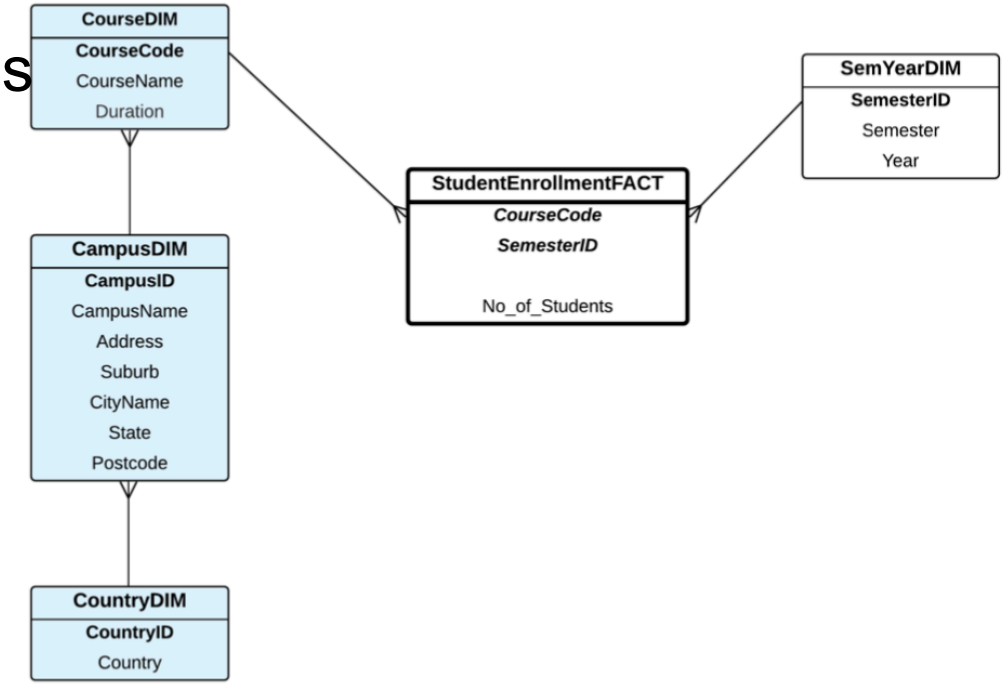
4. Hierarchy Design Considerations

- Hierarchy = many-1 relationship
- Assume that each course is offered in one campus
- Is this Course-Campus-Country hierarchy proper?
- Course and Campus is not a good hierarchy



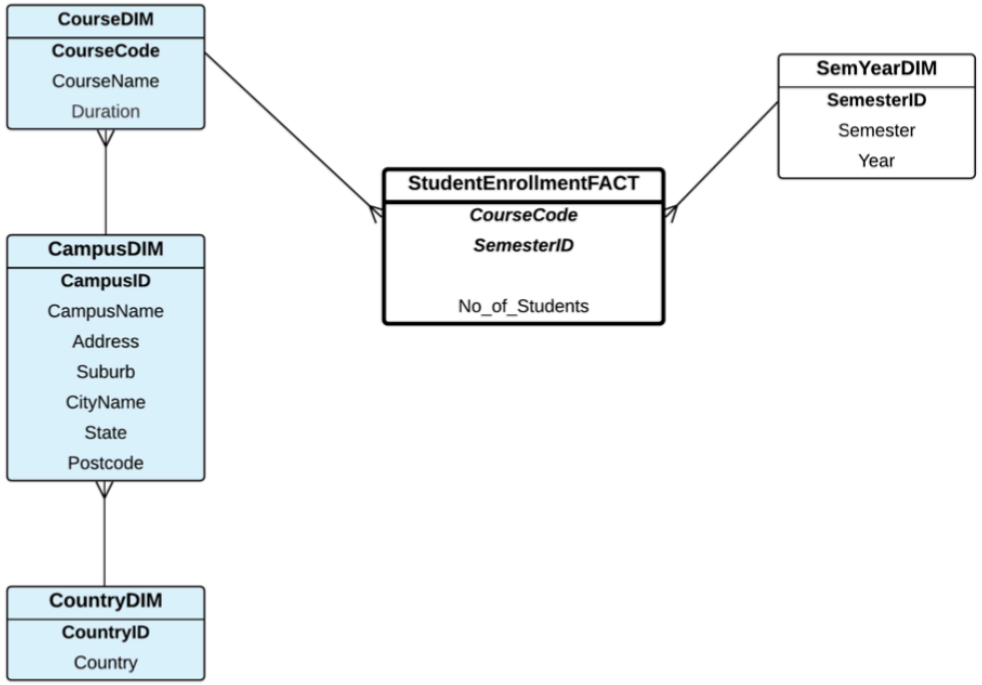
4. Hierarchy Design Considerations

- Campus and Country dimensions have a similarity, which is the spatial context
- Course is totally a different thing
- The Course-Campus-Country hierarchy is not appropriate, although they have many-1 relationships



4. Hierarchy Design Considerations

- Data access to the data warehouse is always through the Fact (e.g. fact measure)
- Fact measures are the focus of data retrieval
- All dimensions should be linked directly to the fact whenever possible



Summary

- A dimension hierarchy is connecting two or more dimensions in a hierarchical manner, using a many-1 relationship. As a result, the dimensions in a hierarchy are normalized, in 3NF, using the context of Relational Database Design.
- Compare and contrasts five different models:
  1) Separate Dimension model
  2) Combined Dimension model
  3) Hierarchy model
  4) Linked Dimension model
  5) Hierarchy Design Considerations