MONASH University

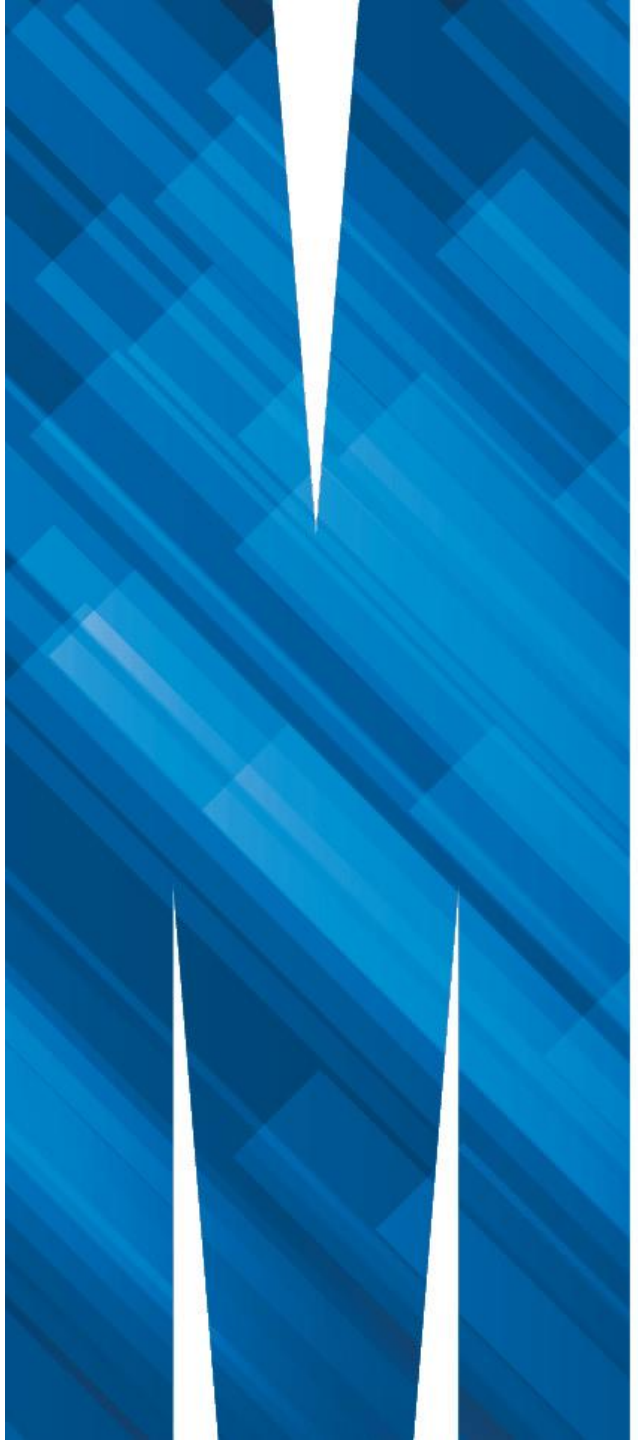
MONASH
INFORMATION
TECHNOLOGY

FIT3003 – Business Intelligence and Data Warehousing

Week 5 – Data Warehousing Architecture: Level of Aggregations

Semester 2, 2022

Developed by:
Dr. Agnes Haryanto
Agnes.Haryanto@monash.edu



Learning Objectives

1. To understand the concepts of granularity in data warehousing architecture.
2. To understand Level of Aggregation and Dimensions.
3. To be able to design lowest level star schema.
4. To understand Star Schemas with No Aggregation.
5. Understanding the relationship between Transactions and Fact Measures.

Agenda

1. Level of Aggregations
    1. Level of Aggregations and Data Warehousing Architecture
    2. Star Schemas with No Aggregation
    3. Understanding the Relationship between Transactions and Fact Measures

Level of Aggregations

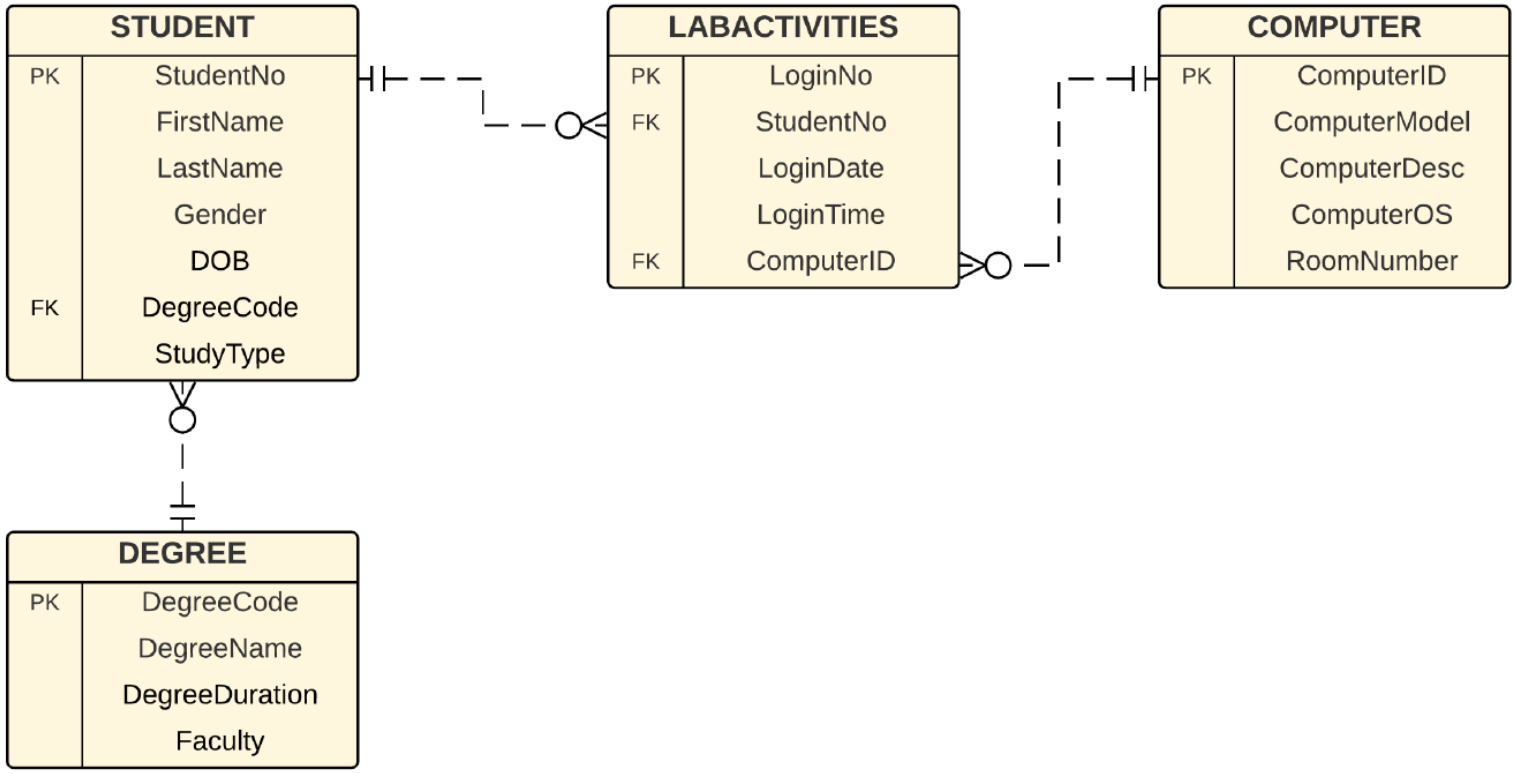
Level of Aggregations

- Aggregate values have different levels of granularity.

  Example:

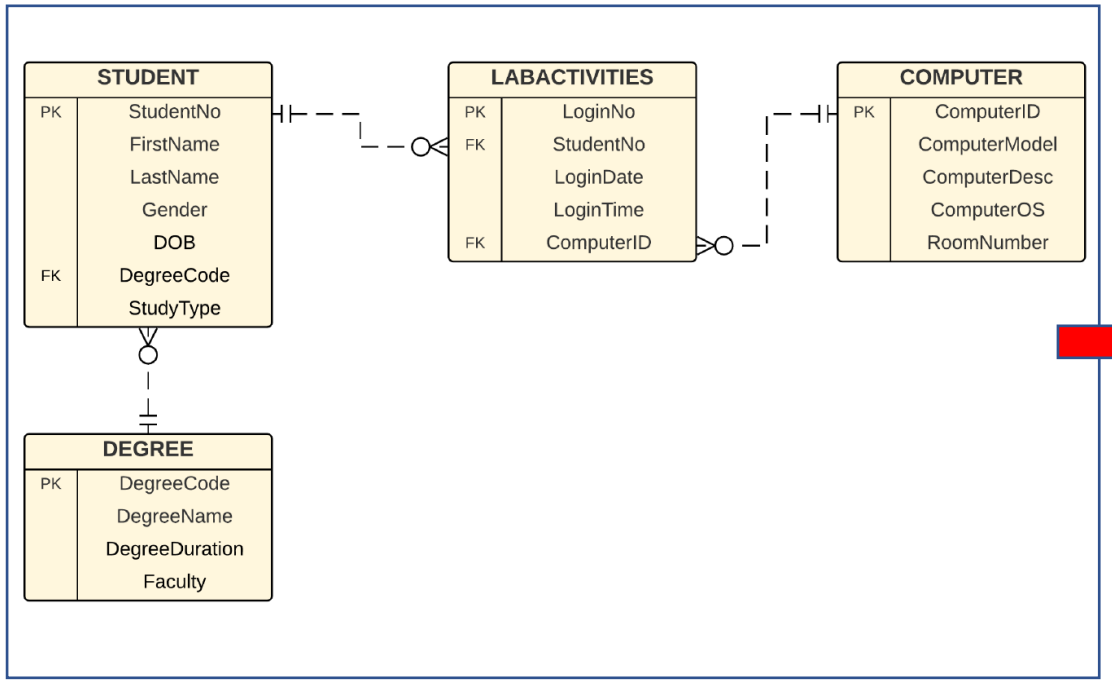
  - Total Sales per Year has lower granularity than Total Sales per Quarter.
  - Number of Logins in the lab per Semester has different level of focus (or granularity) than Number of Logins in the lab per Month

Level of Aggregations

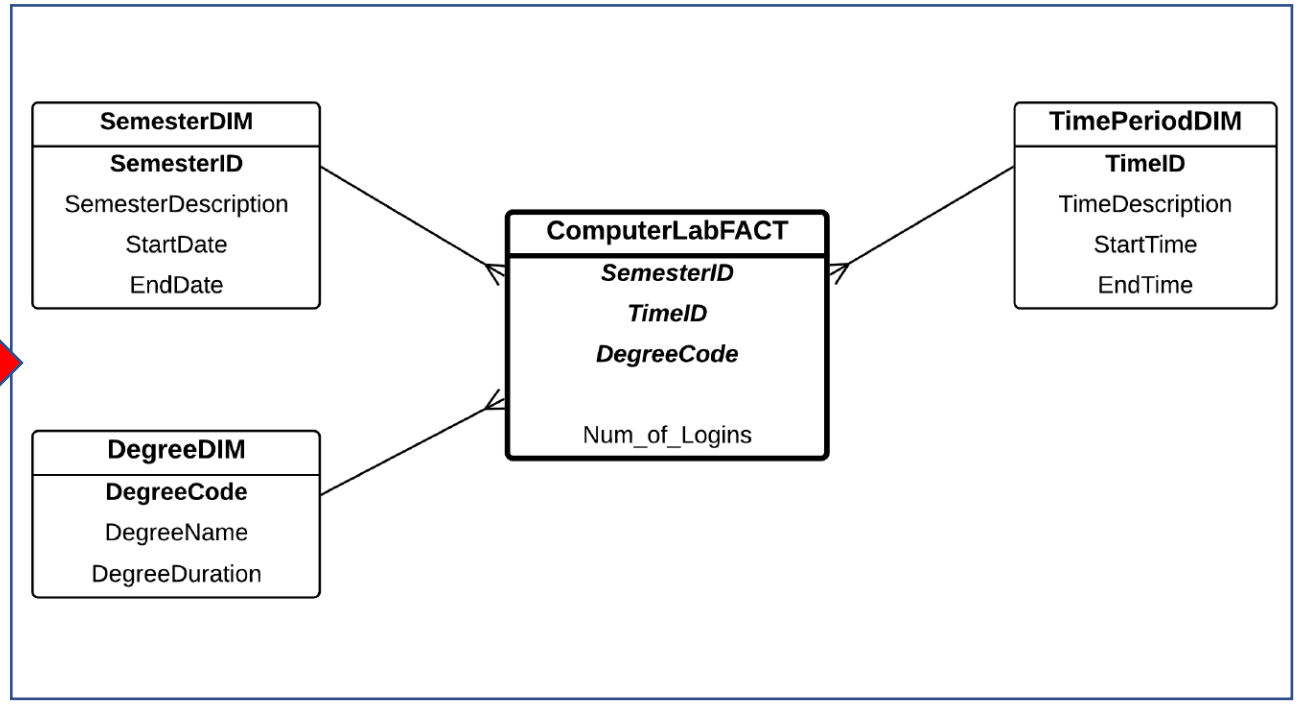


Level of Aggregations

(a) Operational Database

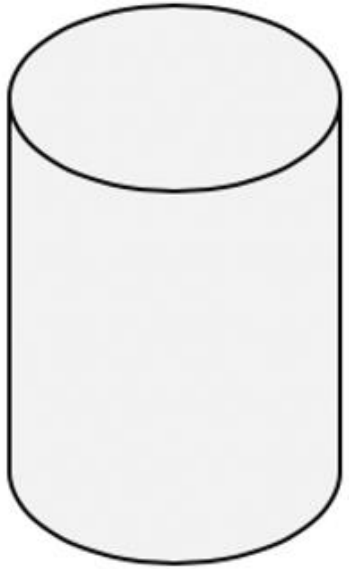


(b) Star Schema



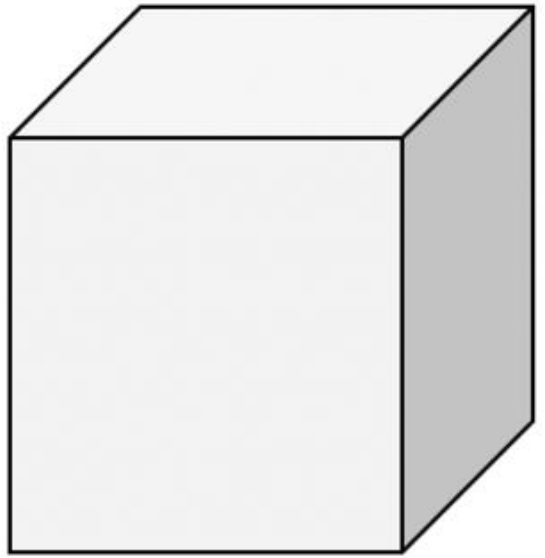
Level of Aggregations

Operational Database



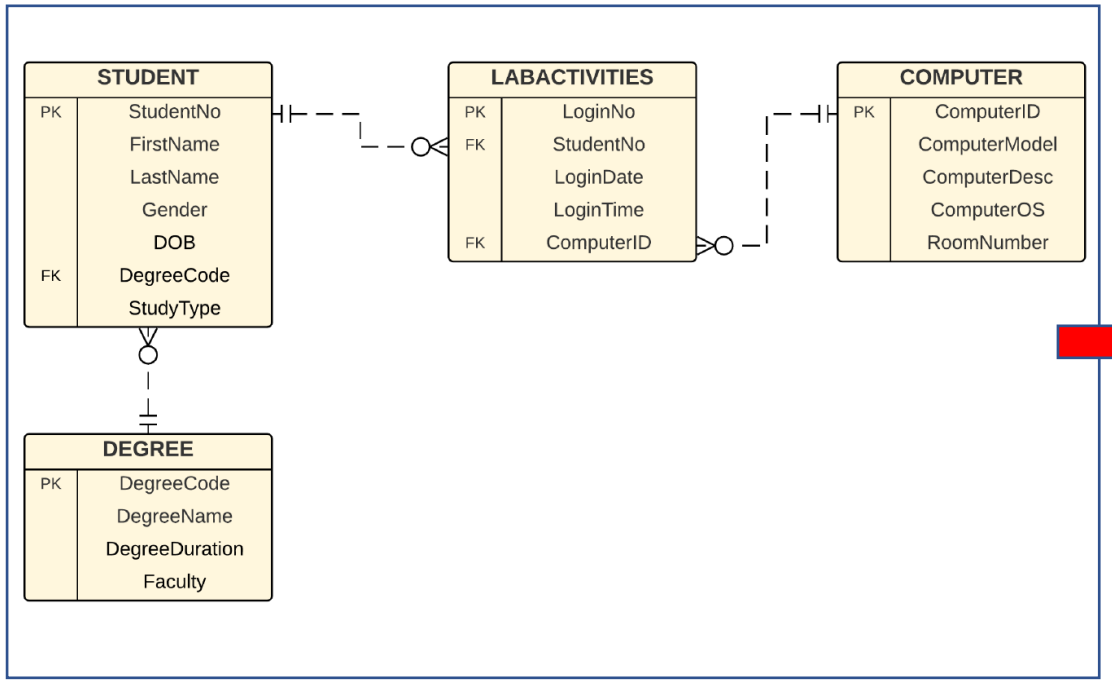
- Extracting
- Cleaning
- Aggregating

Data Warehouse

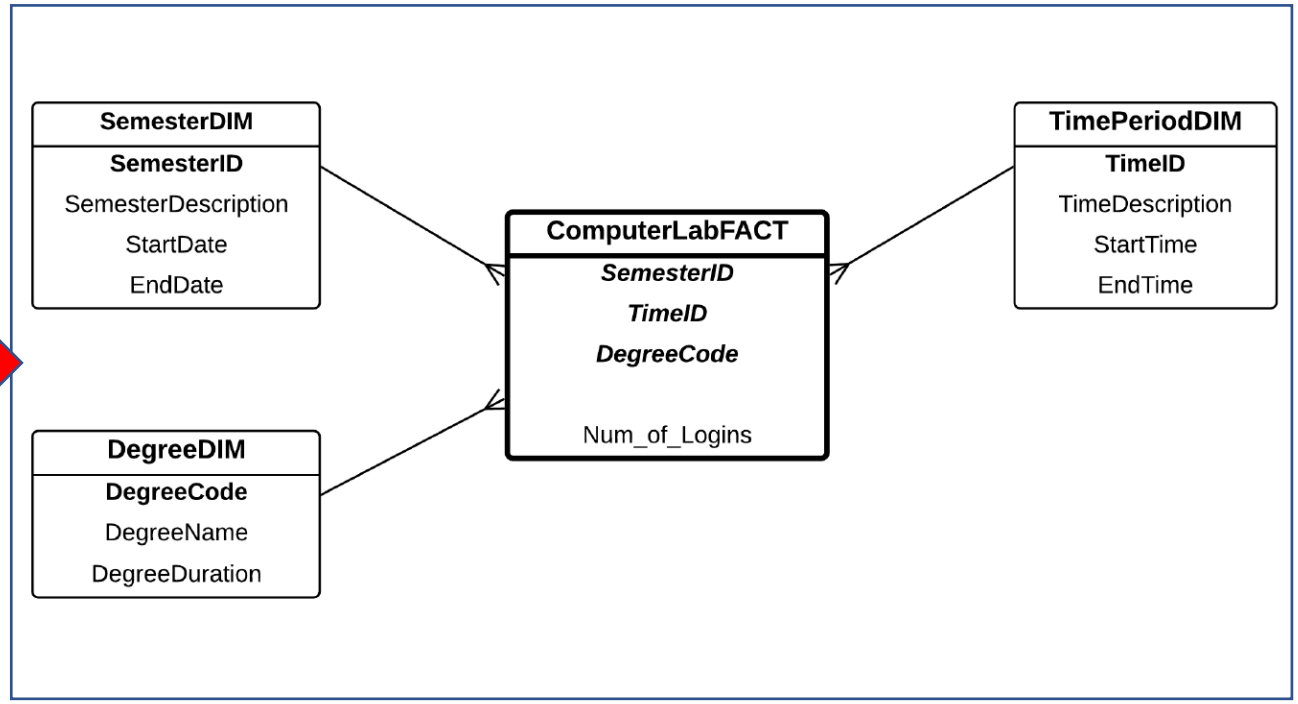


Level of Aggregations

(a) Operational Database

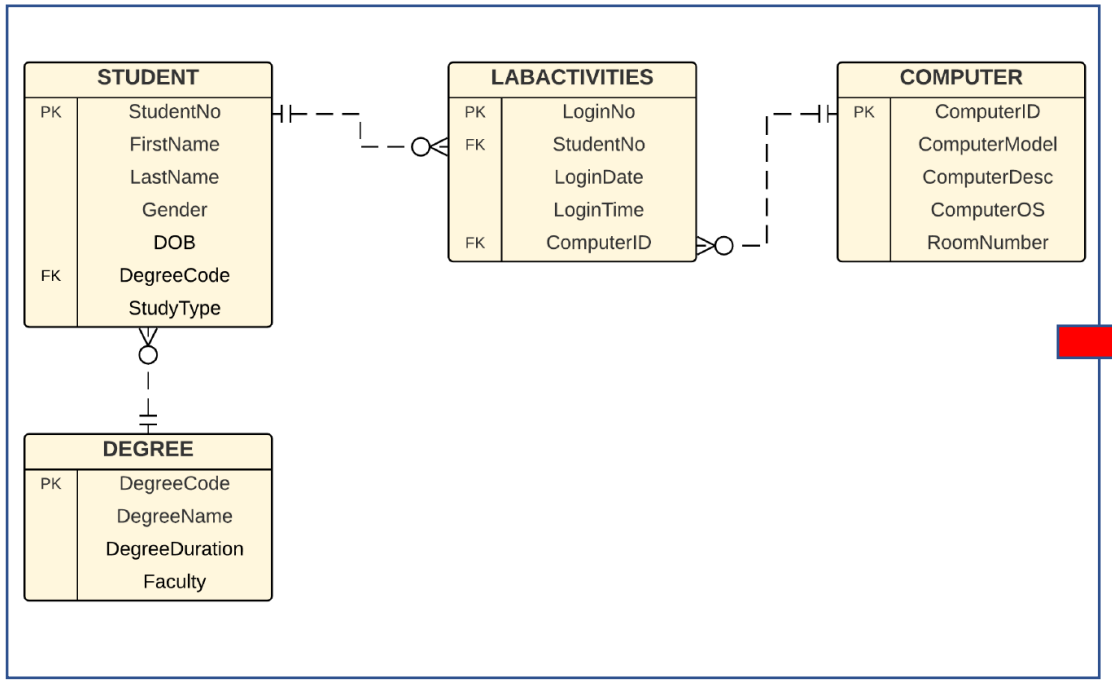


(b) Star Schema

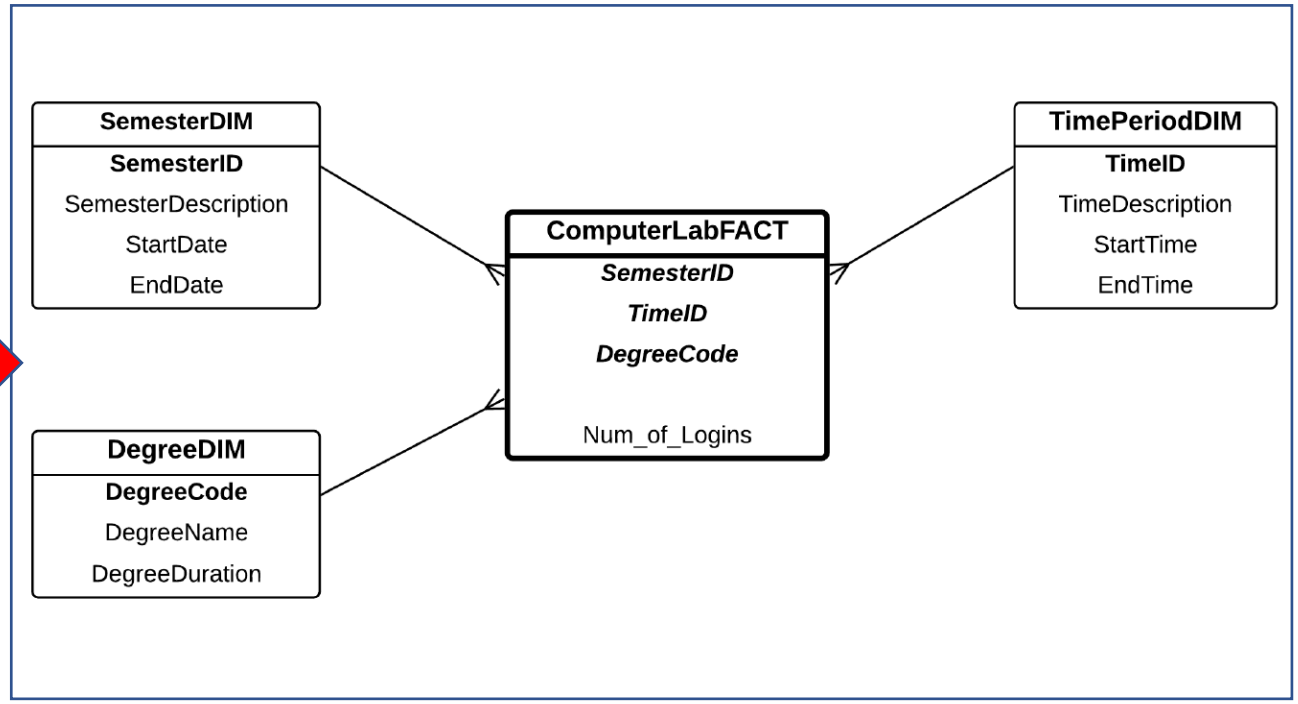


Level of Aggregations

(a) Operational Database



(b) Star Schema

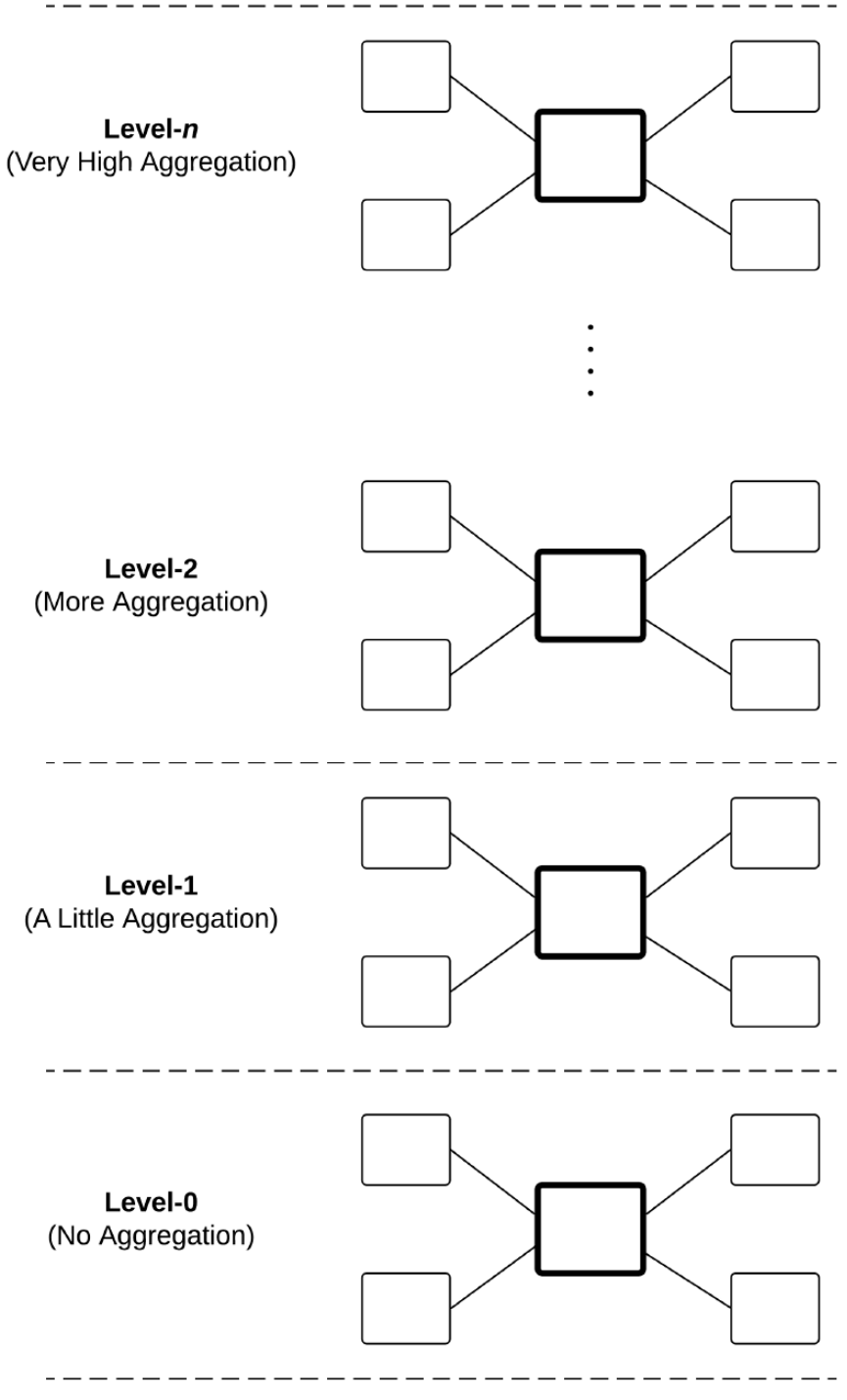


How to find the number of logins per hour at night?

Level of Aggregations

Level of granularity:

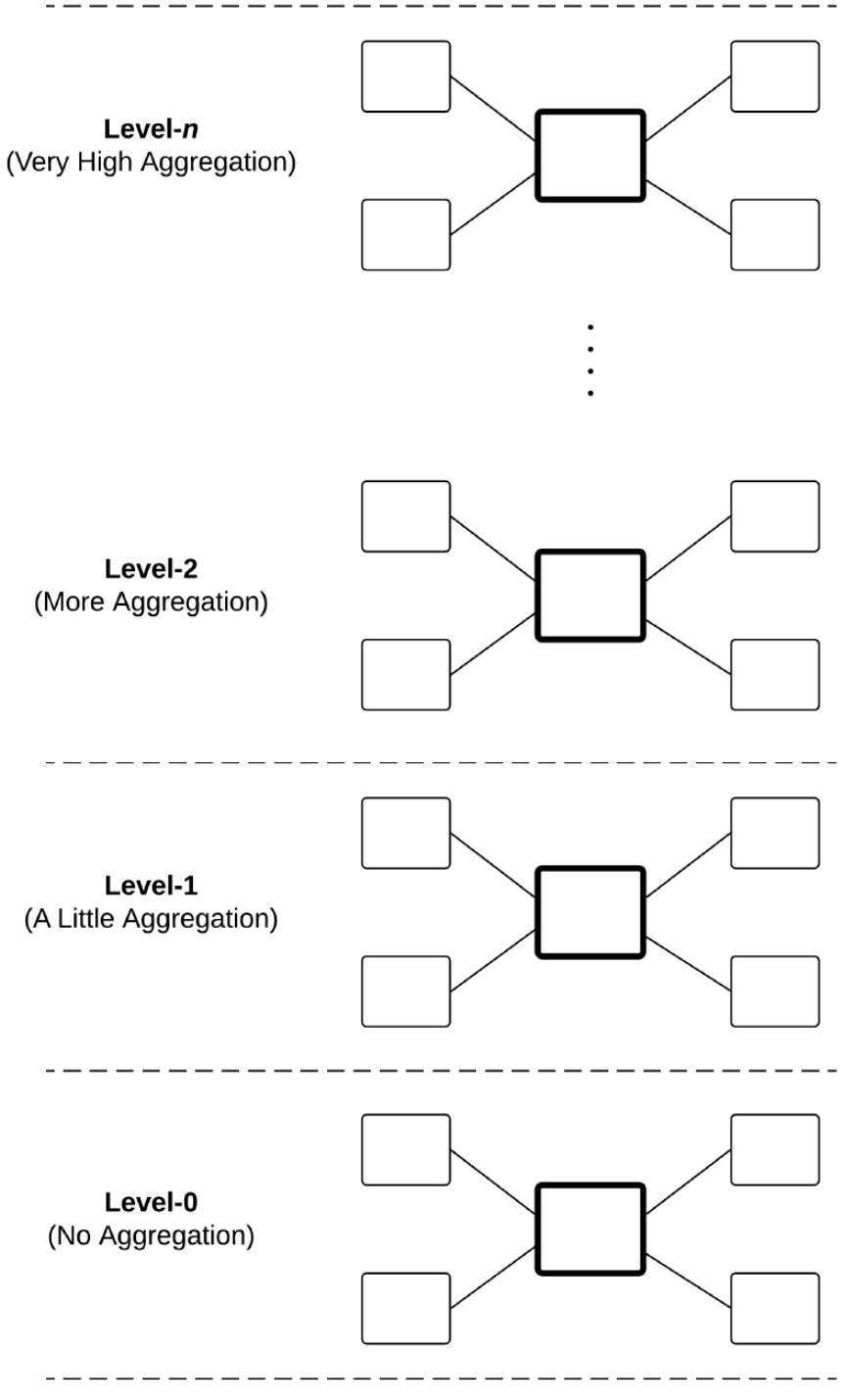
- The highest granularity star schema, which contains the most detail data is Level-0.
- Level-1 star schemas have a lower granularity of the fact measure (e.g. less detail data, as the data is already aggregated).
- Level-2 star schemas are built on top of Level-1 star schemas, and have even a lower granularity of the fact measures.



Level of Aggregations

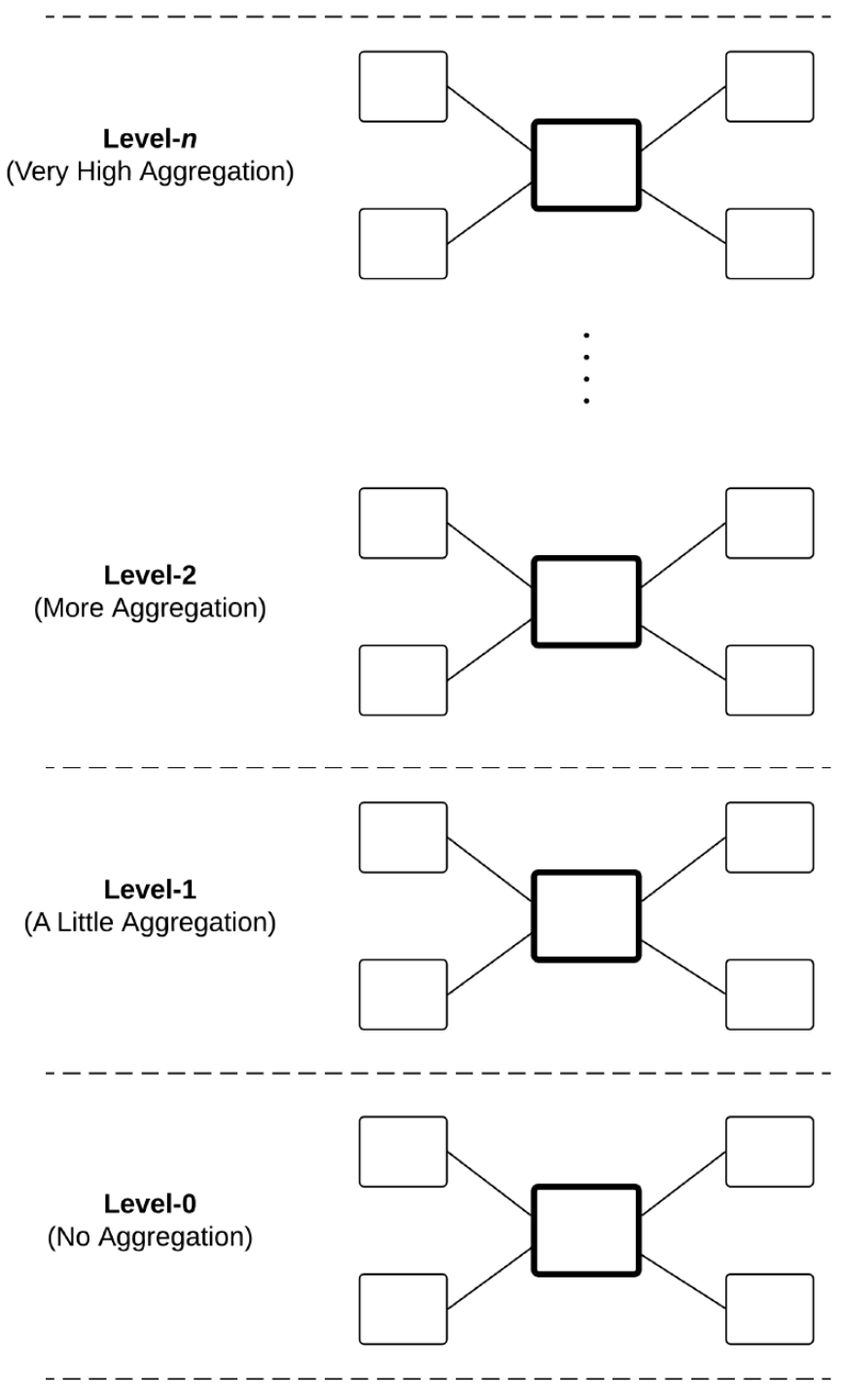
Level of aggregation:

- Level-0 – has the highest level of granularity where no aggregation exists. Level-0 star schema is almost identical to the E/R diagram.
- Level-1 – medium level of aggregation by incorporating ALL domain tables from the operation database but may still incorporate some user-defined groupings.
- Level-2 – high level of aggregation by (i) incorporating selected dimension tables only based on specifications, and (ii) incorporating user-defined grouping for time, distance etc (e.g. yearly/monthly only, short/med/long distance only).



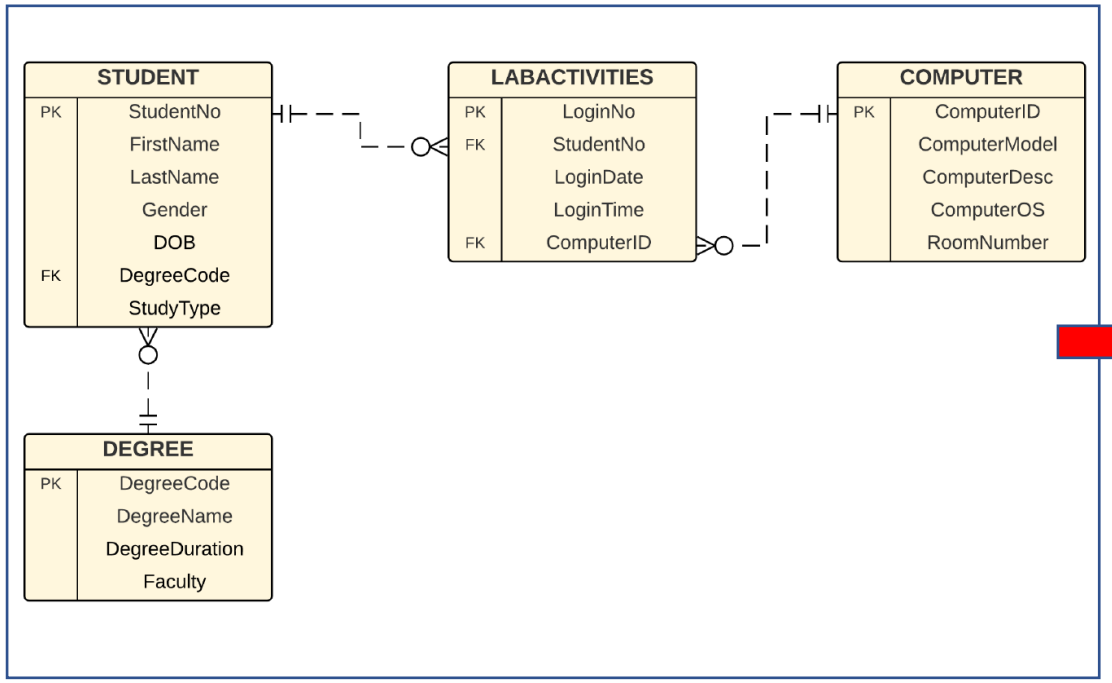
Level of Aggregations

- The lower the level of granularity, the higher the level of aggregations.
- There is no particular rule about determining the level of aggregations (with an exception that Level-0 always means no aggregation).

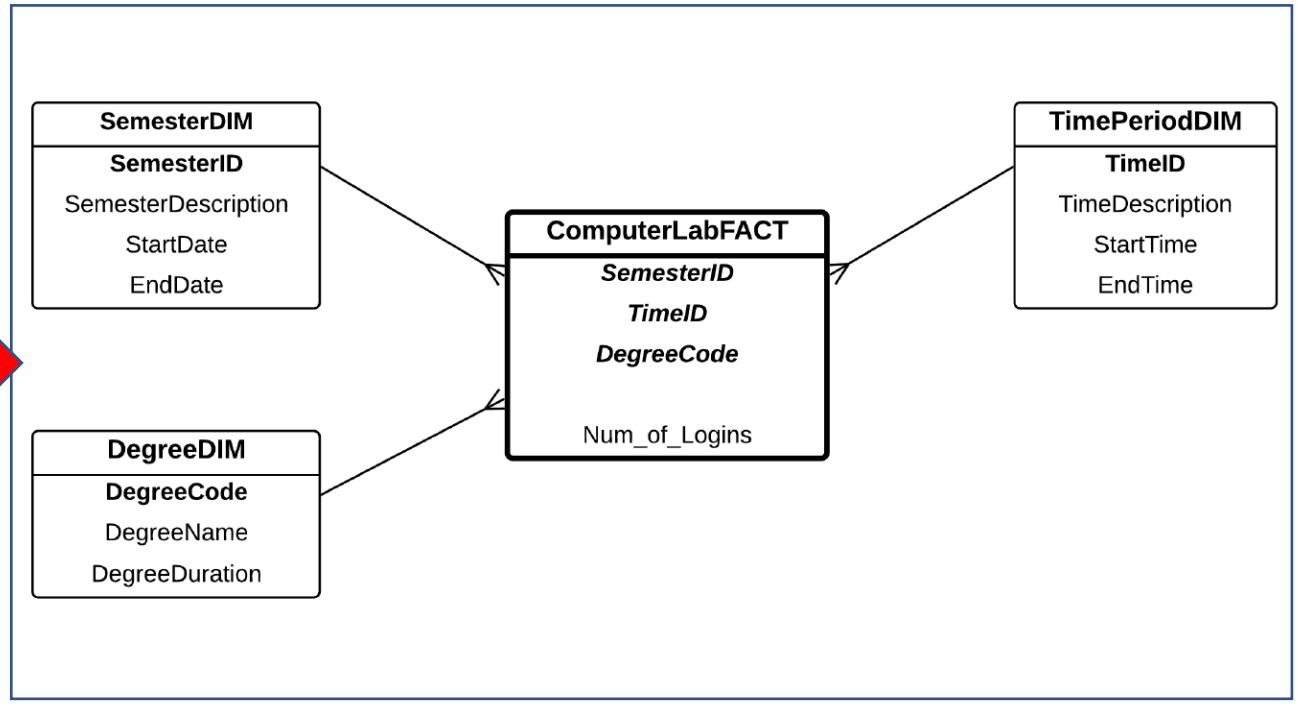


Level of Aggregations

(a) Operational Database

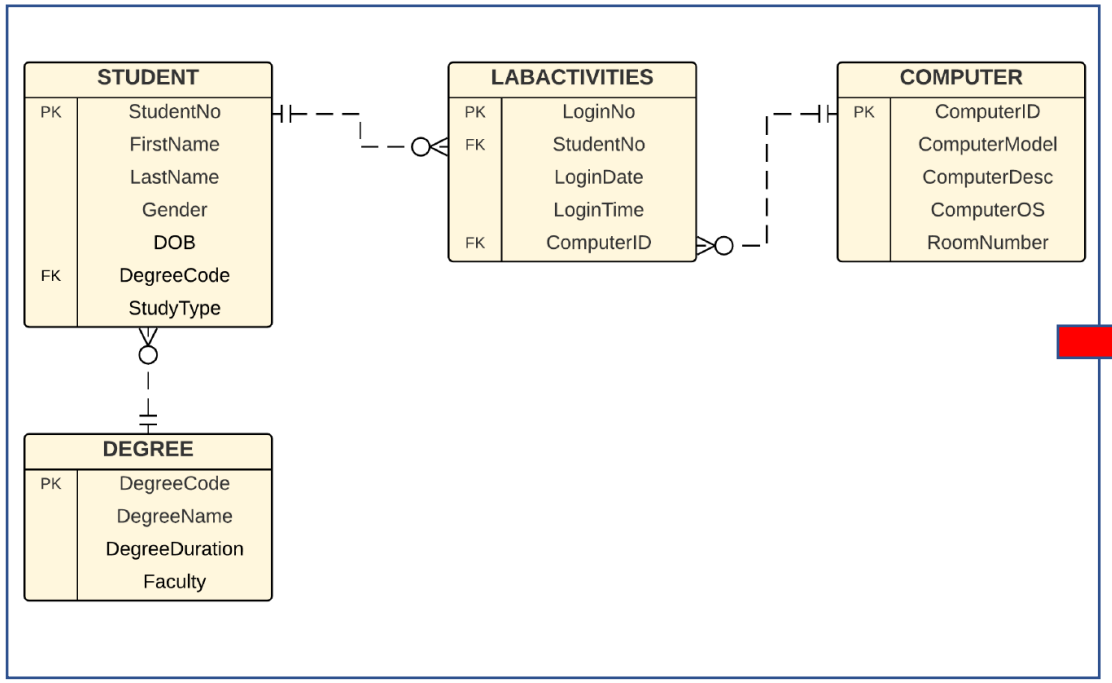


(b) Star Schema

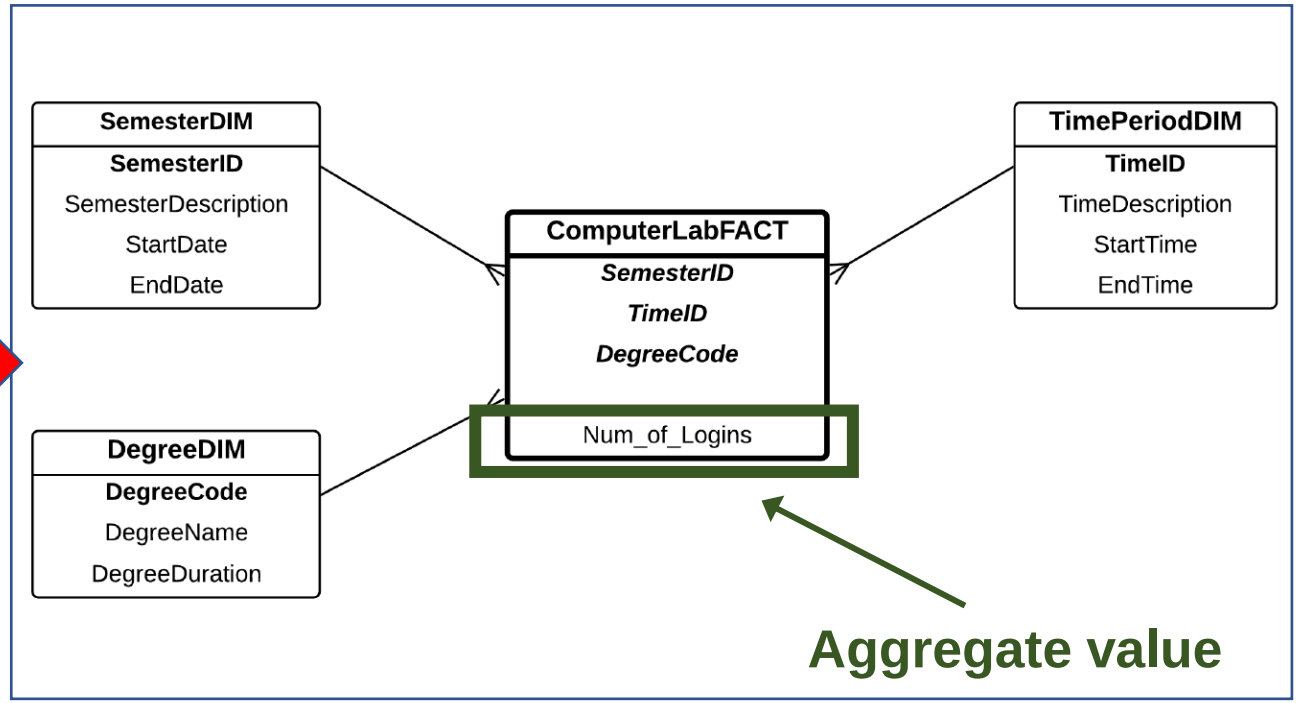


Level of Aggregations

(a) Operational Database

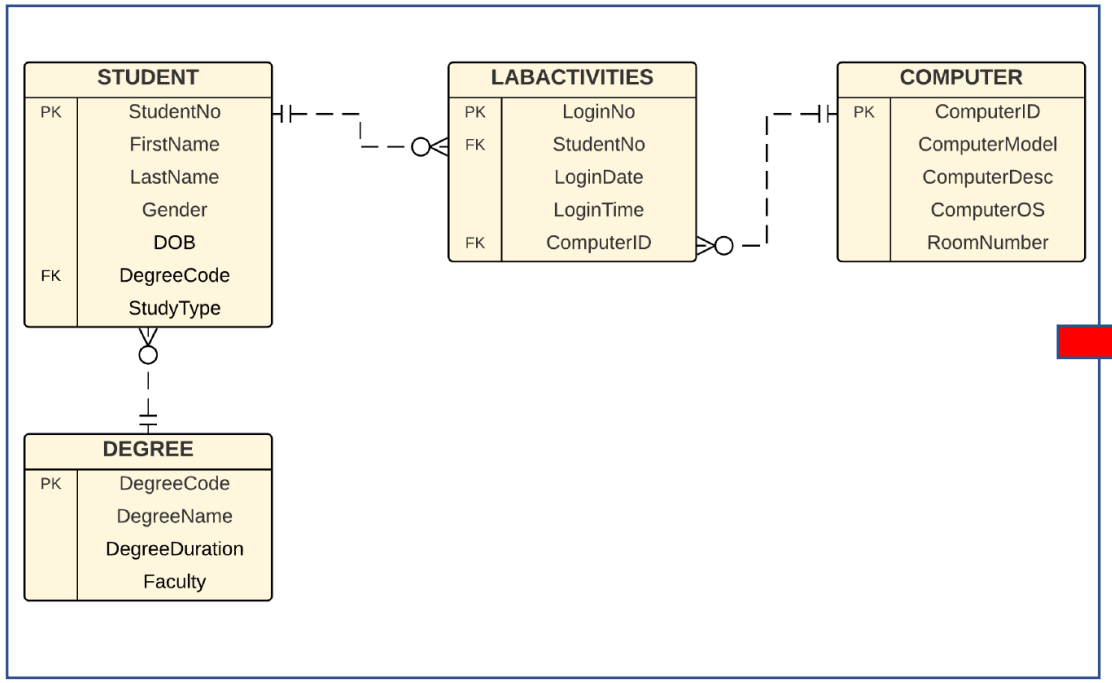


(b) Star Schema

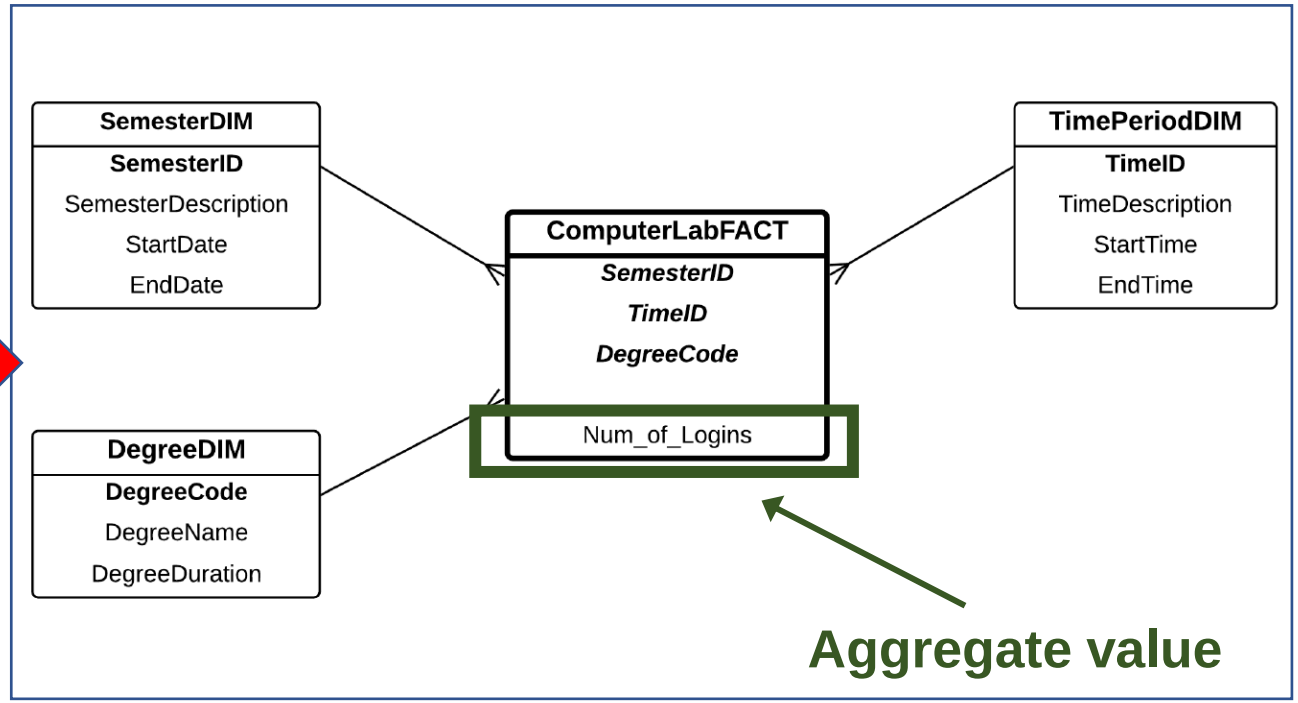


Level of Aggregations

(a) Operational Database



(b) Star Schema



How to find the number of logins per hour at night?

- We must create another star schema with a lower level of aggregation

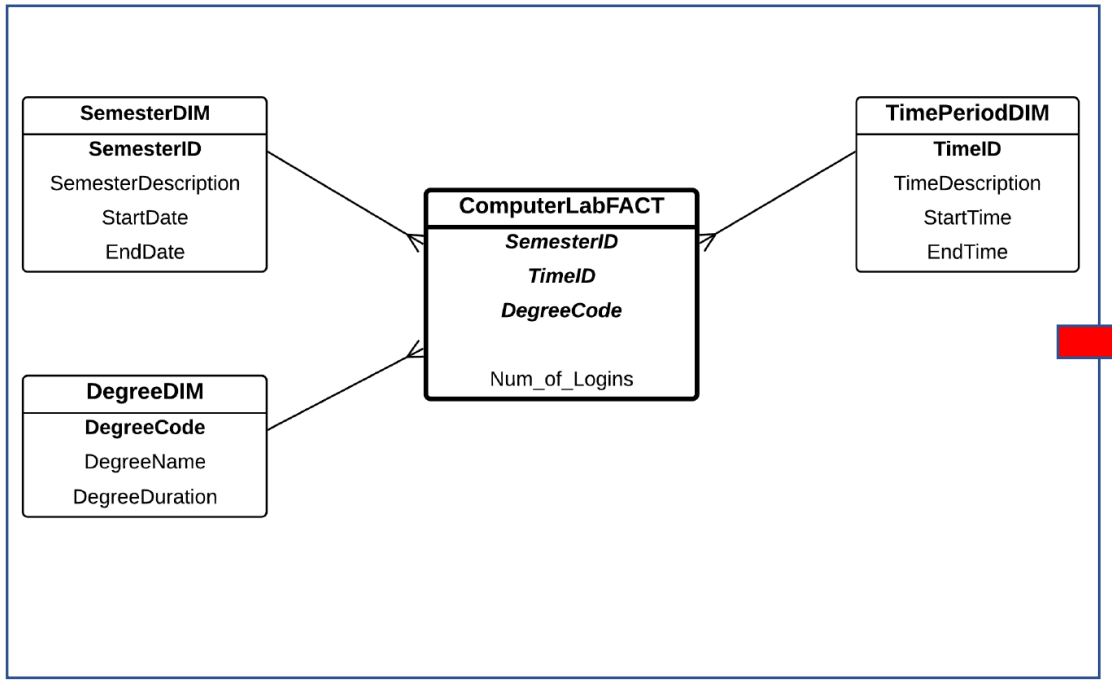
MONASH University

Level of Aggregations

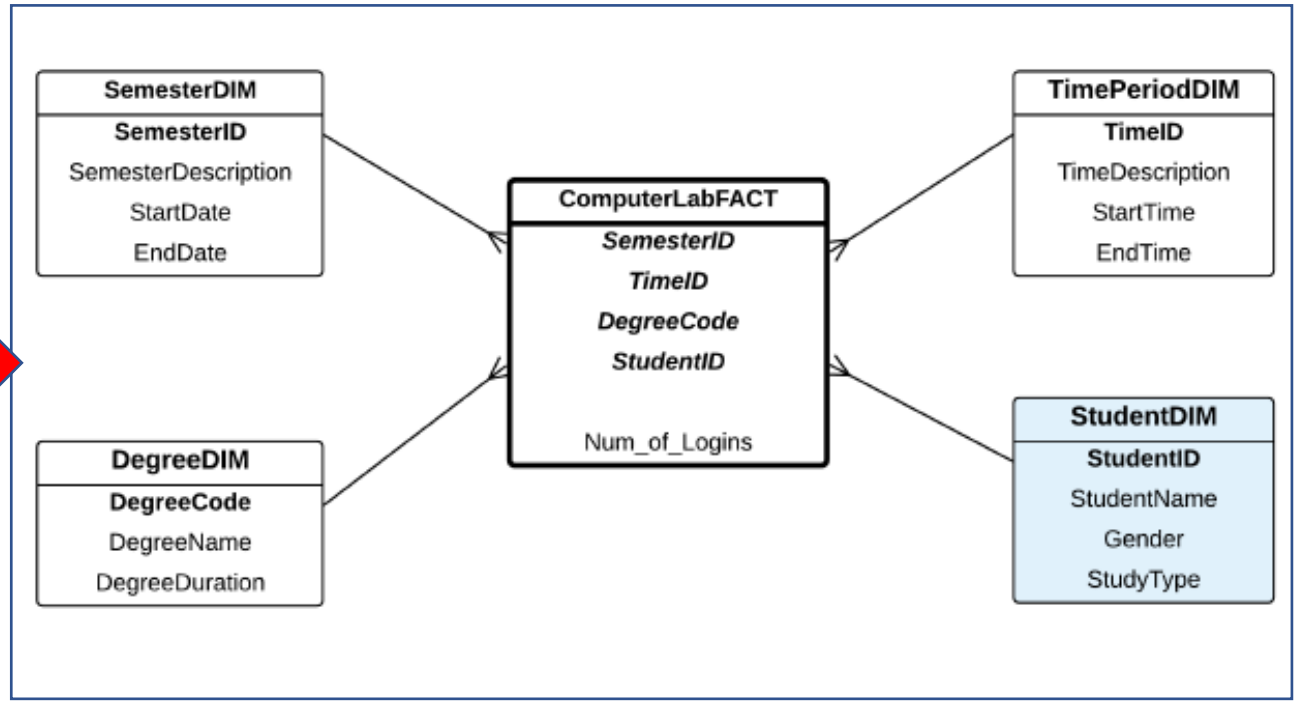
- Two ways to lower down the level of aggregations of a star schema:
    1. *Add a new dimension*. When we add a new dimension, each value in the fact measure will literally be broken down more details on each record of the new dimension.
    2. *Replace an existing dimension with a higher granularity dimension*. The values of the fact measures will also be broken down more details because the fact measure has a lower detail dimension.

Level of Aggregations

(a) The Computer Lab Usage Star Schema

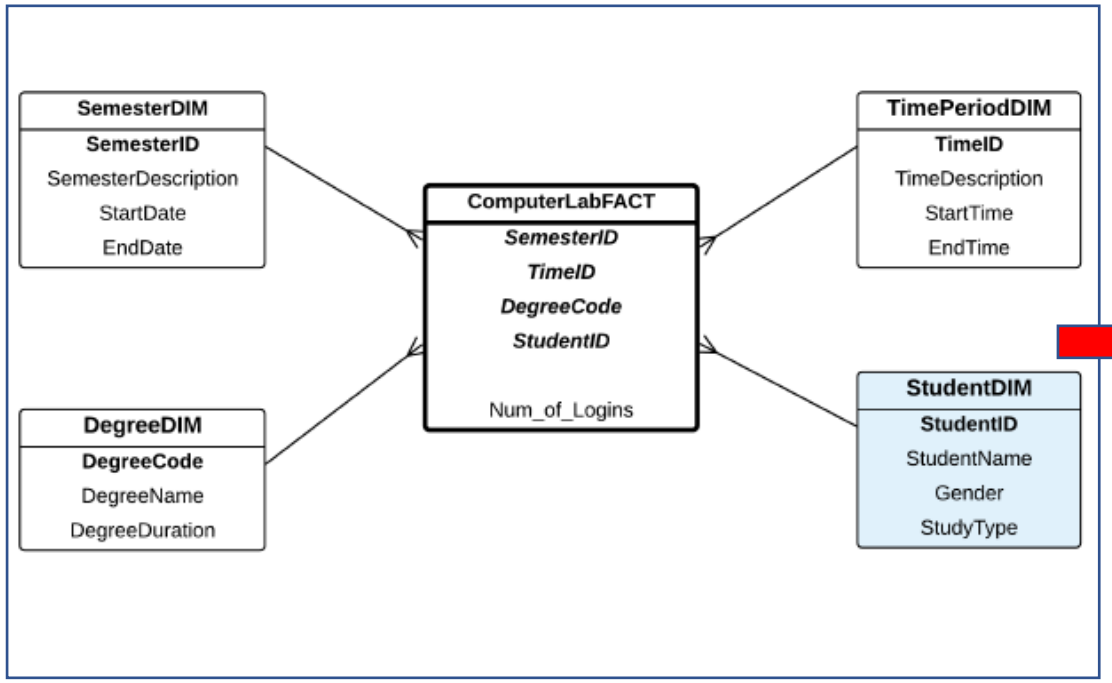


(b) A Lower Level of Aggregation

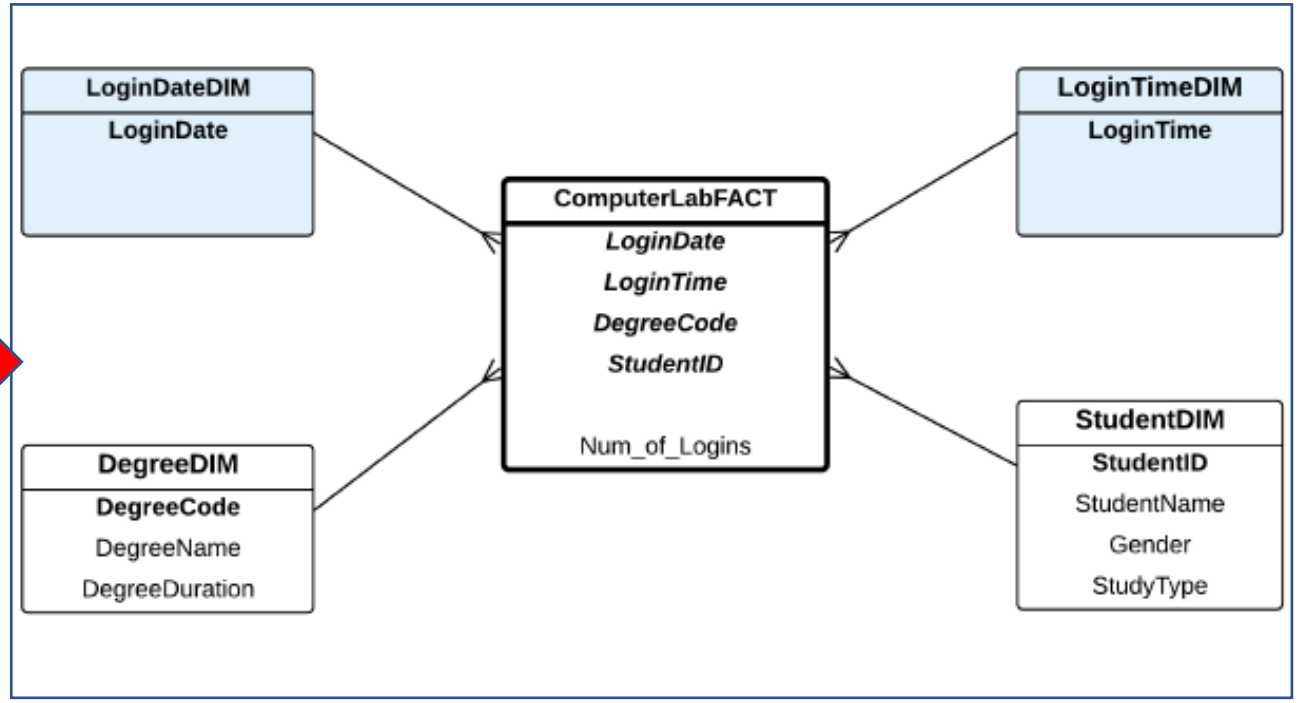


Level of Aggregations

(a) Level-1 Star Schema



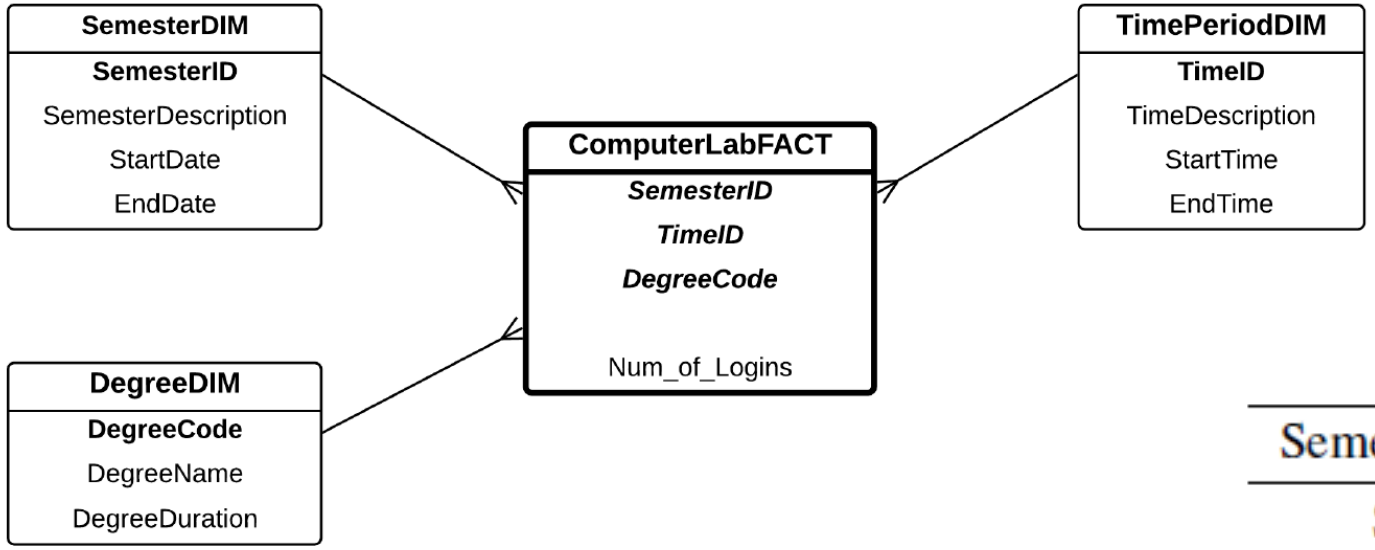
(b) Level-0 Star Schema



Level of Aggregations

- Level-0 star schema provides the most detailed information about the data warehouse.
- The upper levels provide some levels of aggregated information, which has a higher level of aggregation.
  - Level (x+1) star schema is more aggregated than level x star schema.
- There is no particular guideline on how many levels we should have in the data warehousing architecture. It can be less or more than three, depending on how many level of aggregation is needed.
- There is no particular rule on what kind of aggregation a star schema needs on a particular level (except that Level-0 must not have any aggregation).

Level of Aggregations

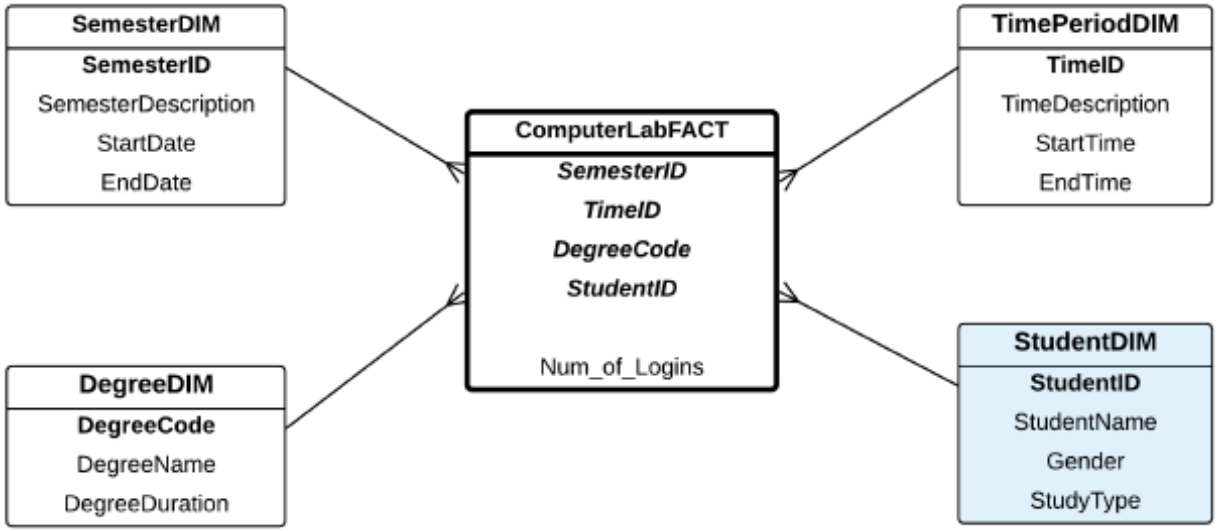


(a) Level-2 Star Schema

(b) Level-2 Fact Table

SemesterID	TimeID	DegreeCode	Num of Logins
S1	3	BIT	**1500**
S1	3	BEng	1250
S1	3	BSc	788
S1	3	MBA	980
...	...	...	...

Level of Aggregations

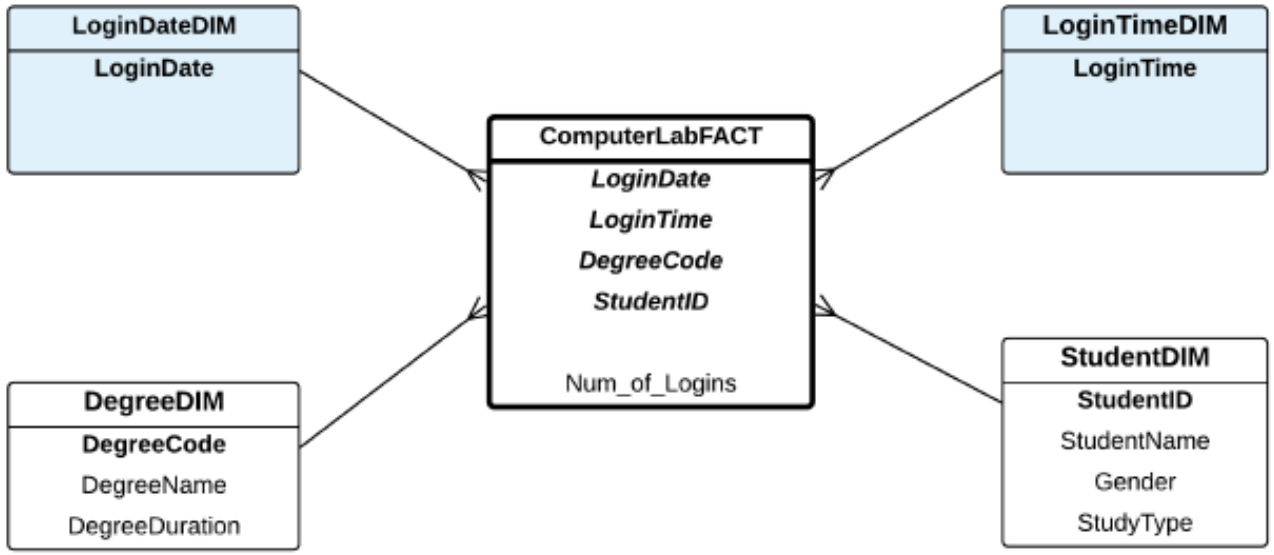


(a) Level-1 Star Schema

(b) Level-1 Fact Table

SemesterID	TimeID	DegreeCode	StudentID	Num of Logins
S1	3	BIT	21002	**120**
S1	3	BIT	21023	90
S1	3	BIT	21025	55
S1	3	BIT	21066	37
S1	3	BIT	...	...
S1	3	BIT	...	...
...	...	...	...	...

Level of Aggregations



(a) Level-0 Star Schema

(b) Level-0 Fact Table

Login Date	Login Time	DegreeCode	StudentID	Num of Logins
4-Apr	19:00	BIT	21002	1
6-Apr	21:20	BIT	21002	1
8-Apr	02:30	BIT	21002	1
3-May	18:55	BIT	21002	1
7-May	19:30	BIT	21002	1
8-May	19:45	BIT	21002	1
...	...	...	...	...

Level of Aggregations

(1:01:48) You cannot compare Lv 1a and 1b, because they are lowered down from lv2 with different dims. e.g., we lower to 1a using SEMDIM whereas we lowered to 1b using LoginTIMEDIM

(a) The New Level-3 Fact Table

SemesterID	TimeID	DegreeCode	Num of Logins
S1	3	BIT	**1500**
S1	3	BEng	1250
S1	3	BSc	788
S1	3	MBA	980
...	...	...	...

(b) The New Level-2 Fact Table

SemesterID	TimeID	DegreeCode	StudentID	Num of Logins
S1	3	BIT	21002	**120**
S1	3	BIT	21023	90
S1	3	BIT	21025	55
S1	3	BIT	21066	37
S1	3	BIT	...	...
S1	3	BIT	...	...
...	...	...	...	...

(c) The New Level-1a Fact Table

MonthID	TimeID	DegreeCode	StudentID	Num of Logins
Jan	3	BIT	21002	**10**
Feb	3	BIT	21002	**5**
Mar	3	BIT	21002	**7**
Apr	3	BIT	21002	**10**
May	3	BIT	21002	**25**
Jun	3	BIT	21002	**20**
Jul	3	BIT	21002	**34**
...	...	...	...	...

(d) The New Level-1b Fact Table

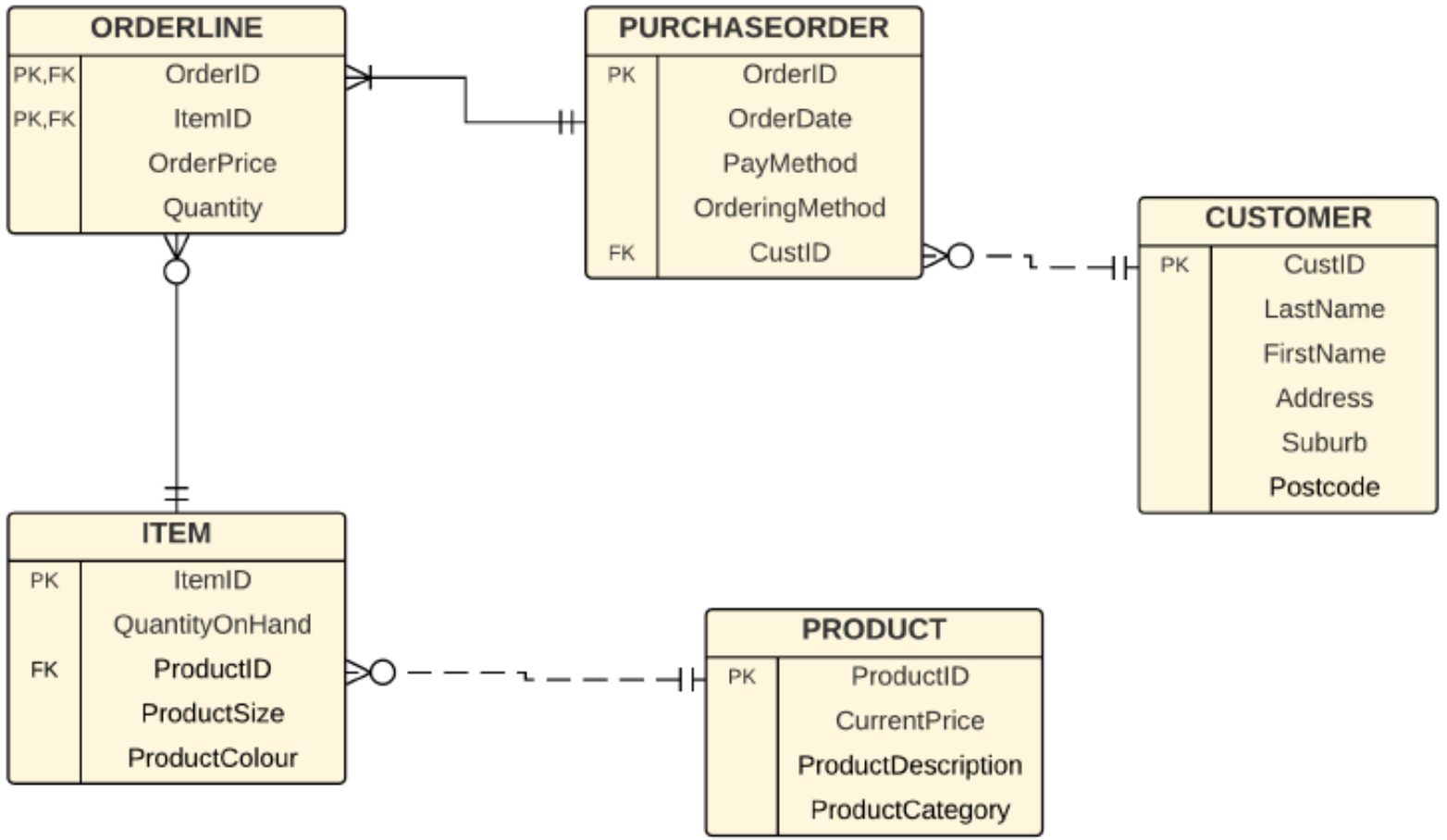
MonthID	HourID	DegreeCode	StudentID	Num of Logins
S1	6:00-6:59pm	BIT	21002	**25**
S1	7:00-7:59pm	BIT	21002	**10**
S1	8:00-8:59pm	BIT	21002	**5**
S1	9:00-9:59pm	BIT	21002	**5**
S1	10:00-10:59pm	BIT	21002	**3**
...	...	...	...	...

(e) The New Level-0 Fact Table

Login Date	Login Time	DegreeCode	StudentID	Num of Logins
4-Apr	19:00	BIT	21002	**1**
6-Apr	21:20	BIT	21002	**1**
8-Apr	02:30	BIT	21002	**1**
3-May	18:55	BIT	21002	**1**
7-May	19:30	BIT	21002	**1**
8-May	19:45	BIT	21002	**1**
...	...	...	...	...

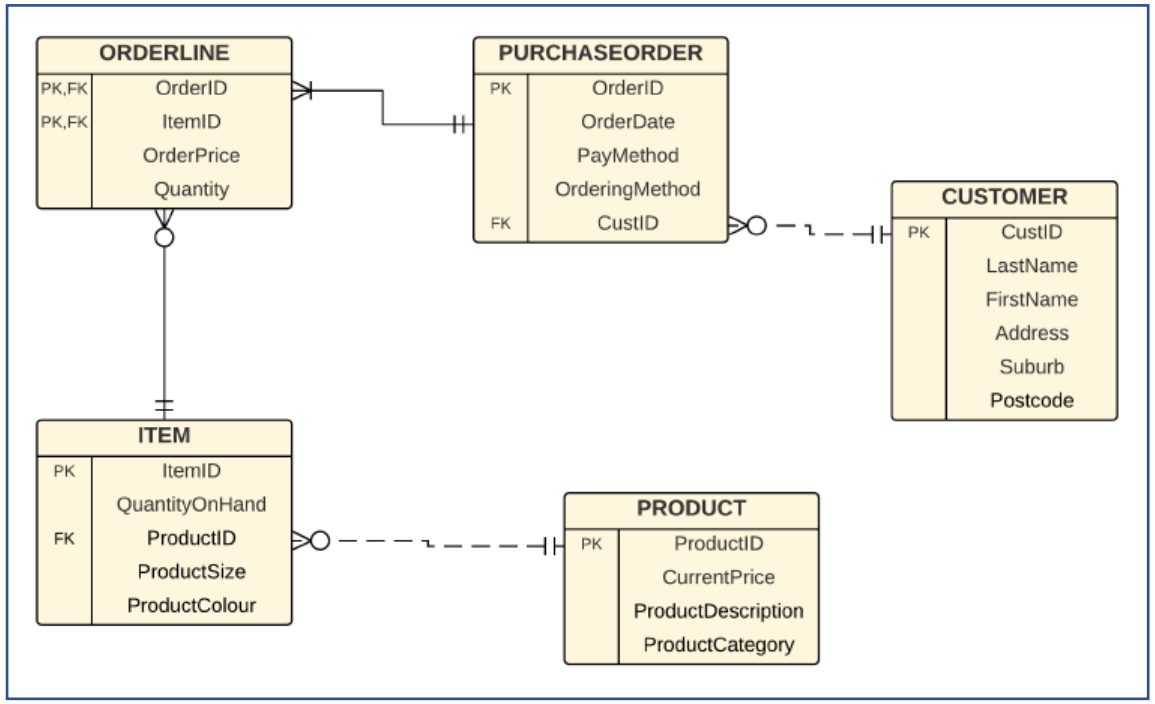
Star Schemas with No Aggregation

Star Schemas with No Aggregation

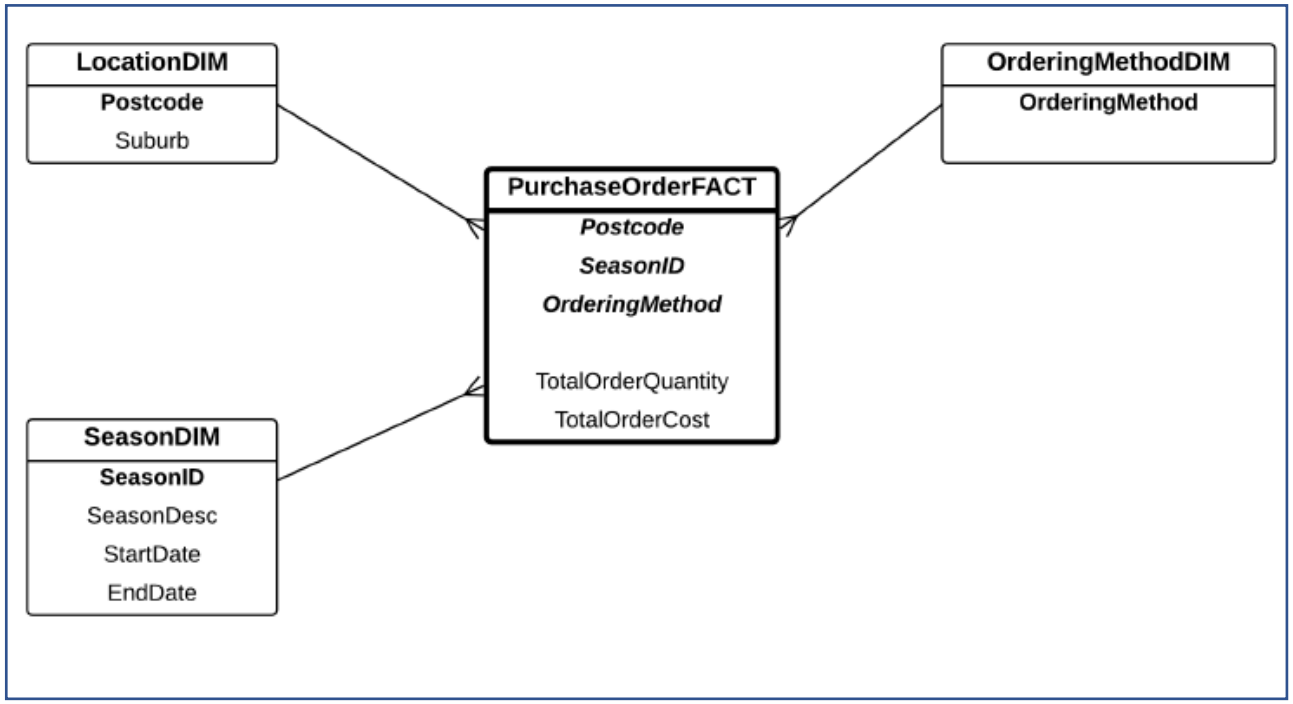


Star Schemas with No Aggregation

(a) Purchase Order E/R Diagram

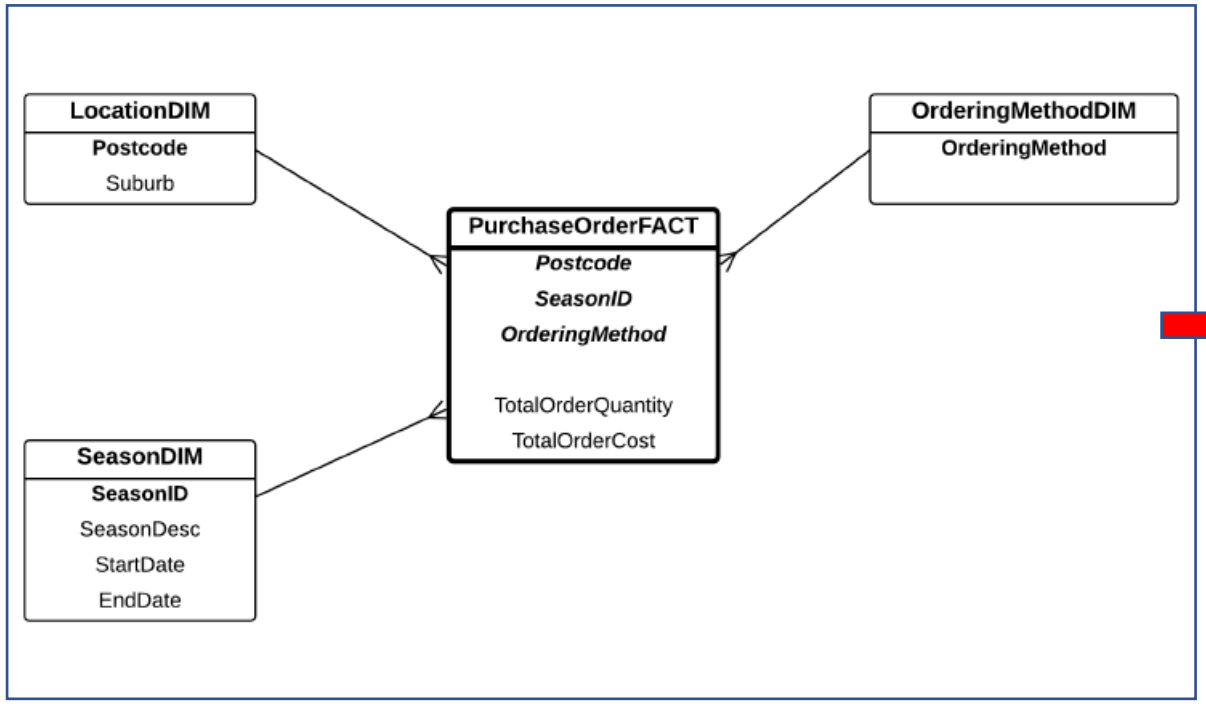


(b) Purchase Order Star Schema – High Level of Aggregation

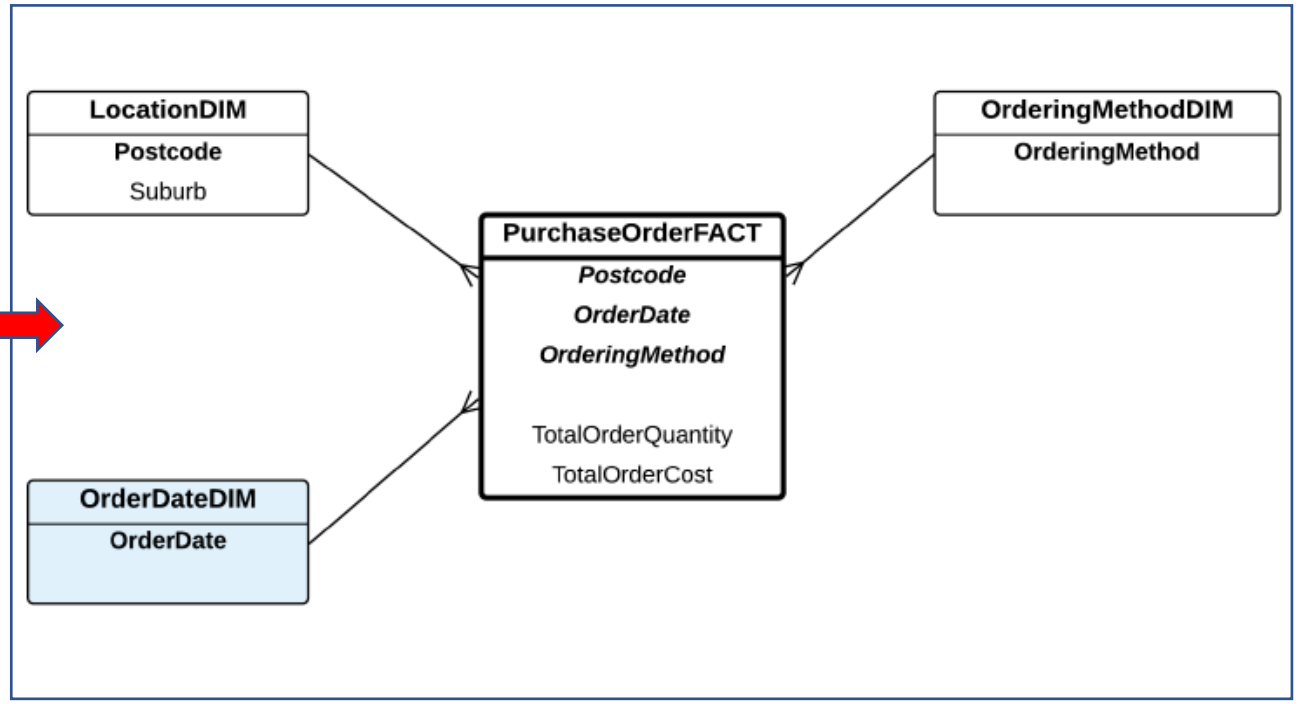


Star Schemas with No Aggregation

(b) Purchase Order Star Schema – High Level of Aggregation

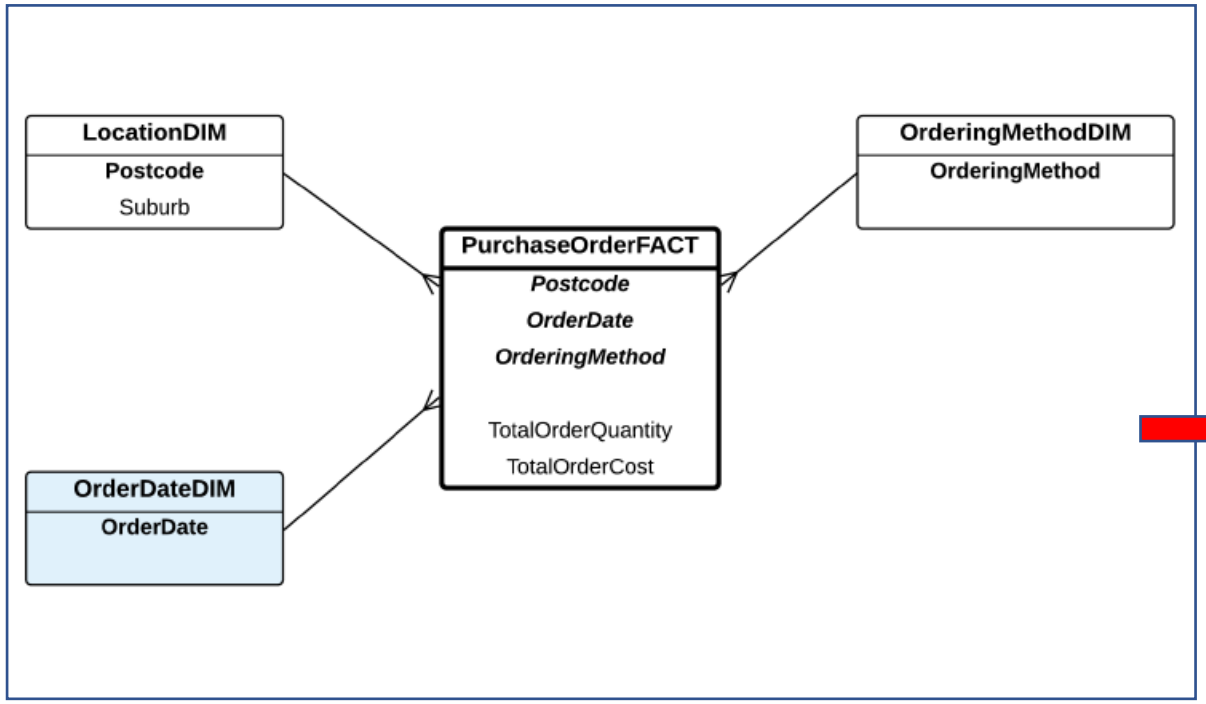


(c) Purchase Order Star Schema – with Order Date Dimension

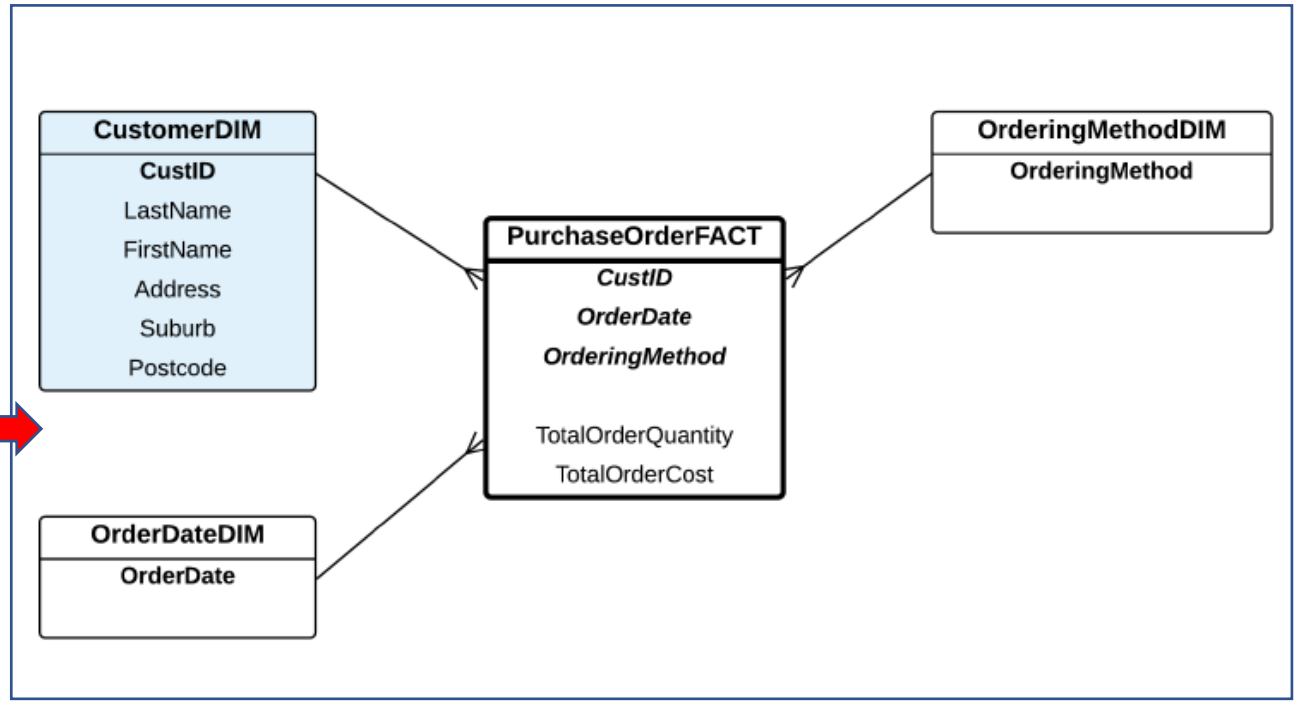


Star Schemas with No Aggregation

(c) Purchase Order Star Schema – with Order Date Dimension

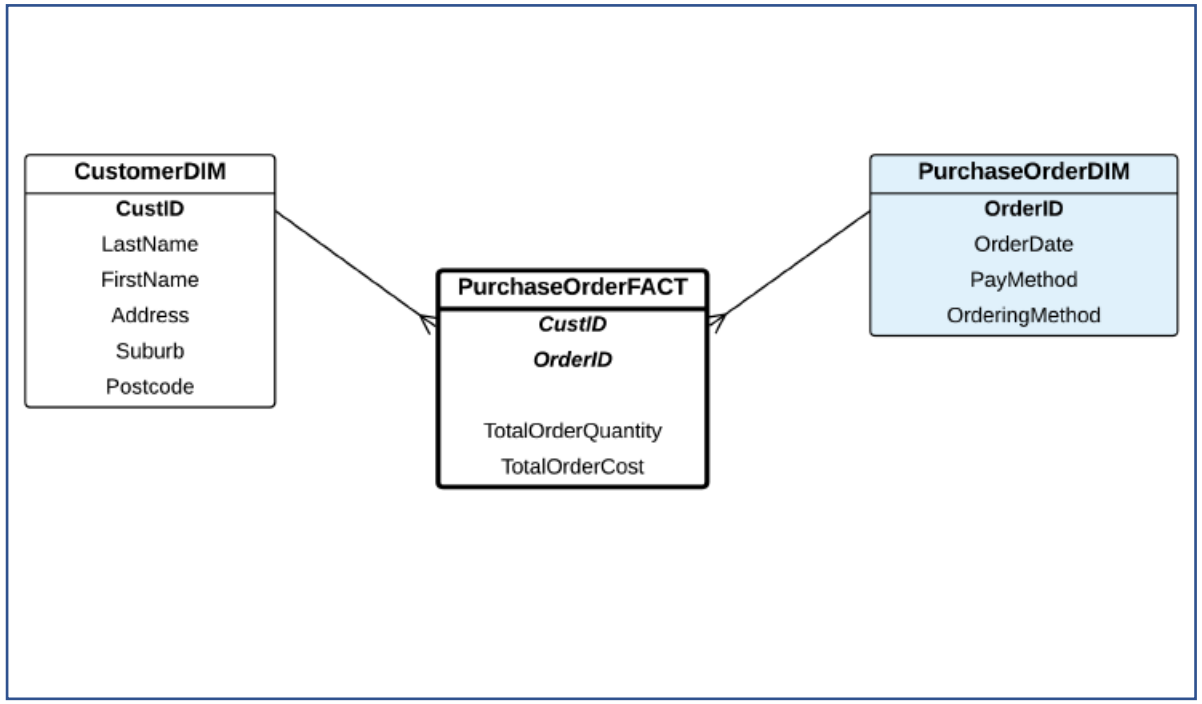


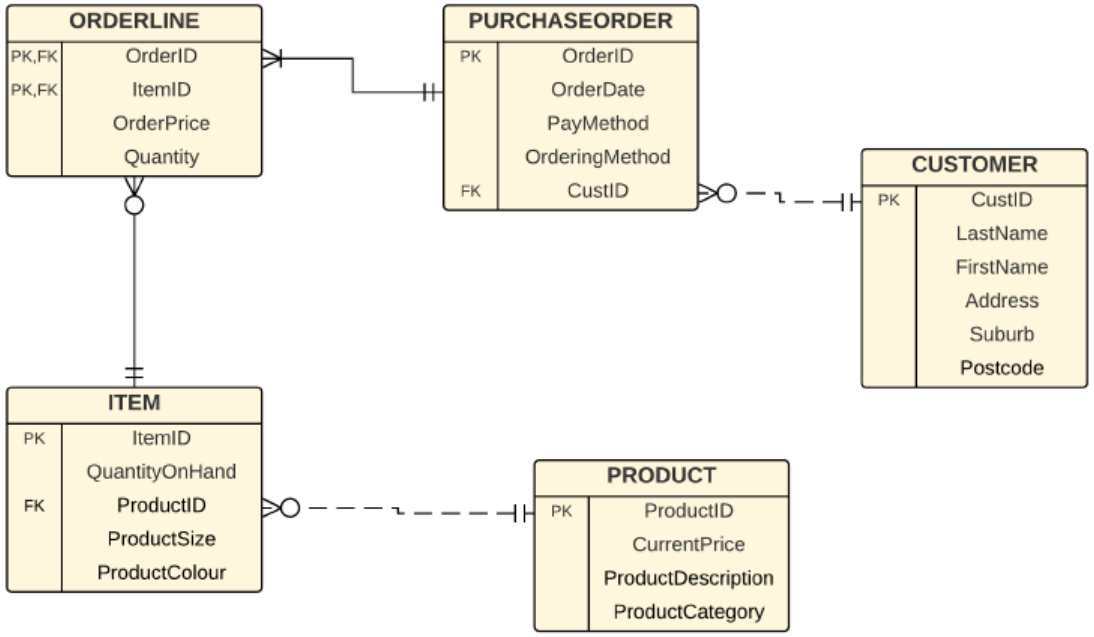
(d) Purchase Order Star Schema – with Customer Dimension



Star Schemas with No Aggregation

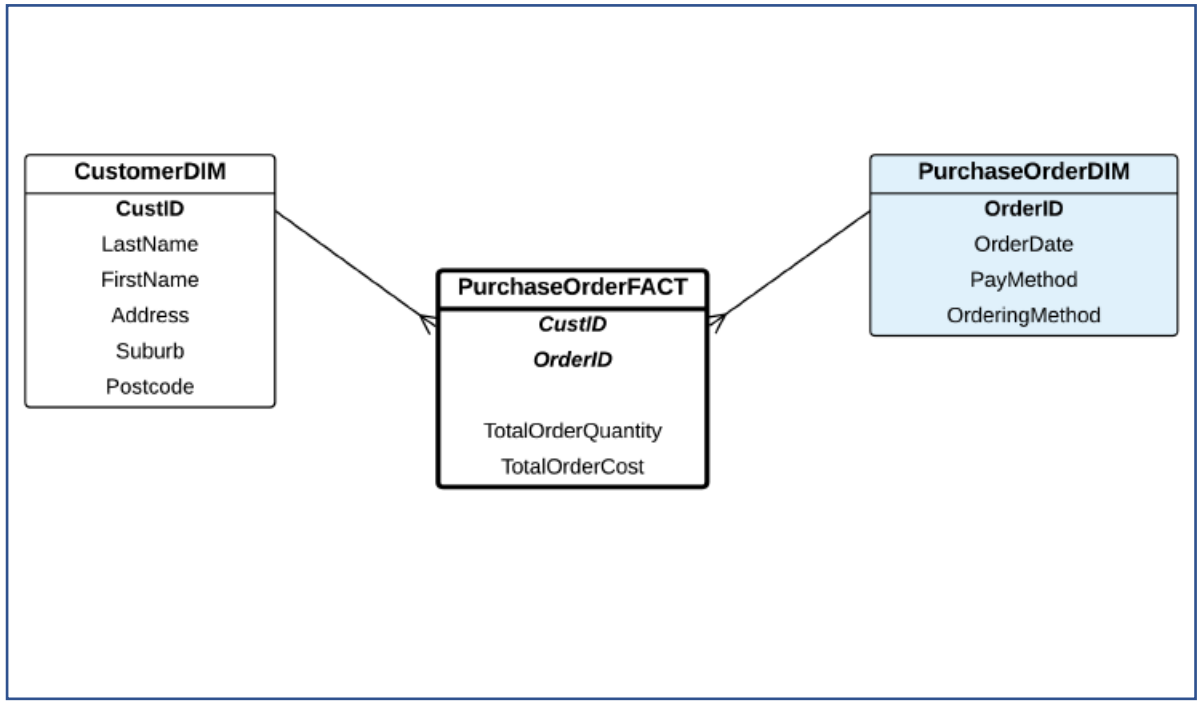
(e) Purchase Order Star Schema – with Purchase Order Dimension



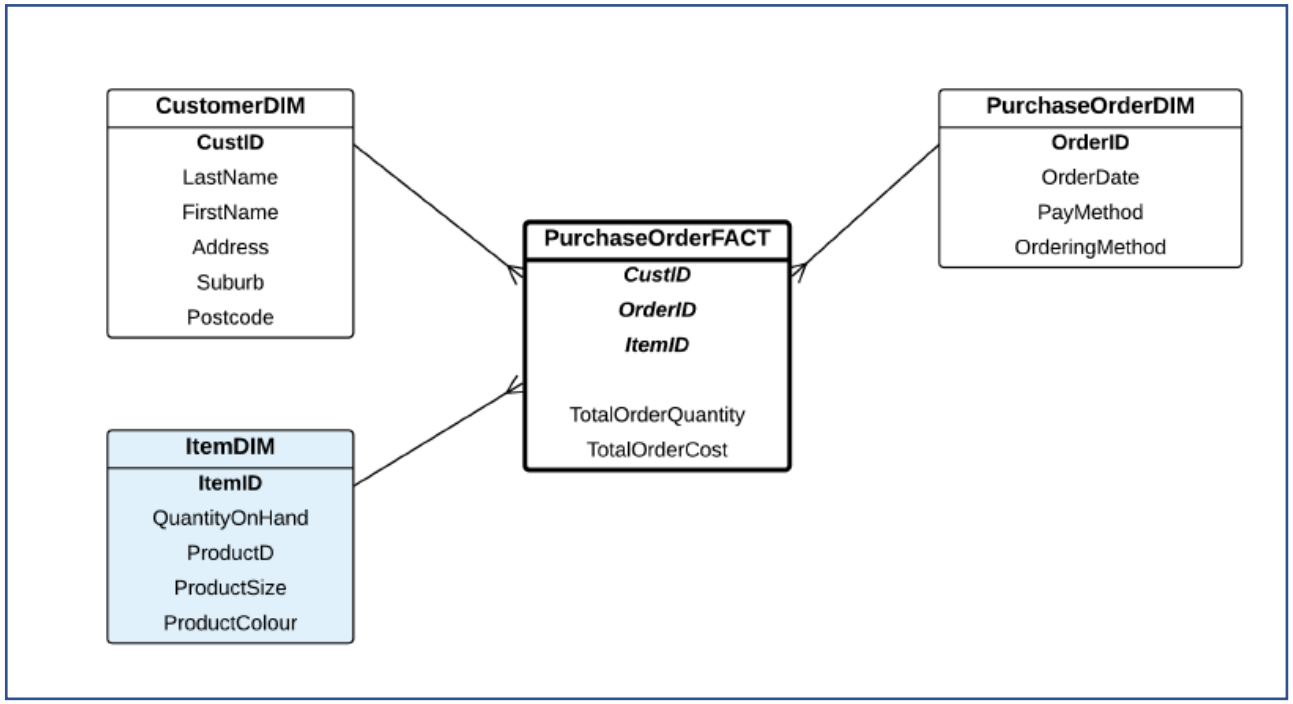


Star Schemas with No Aggregation

(e) Purchase Order Star Schema – with Purchase Order Dimension

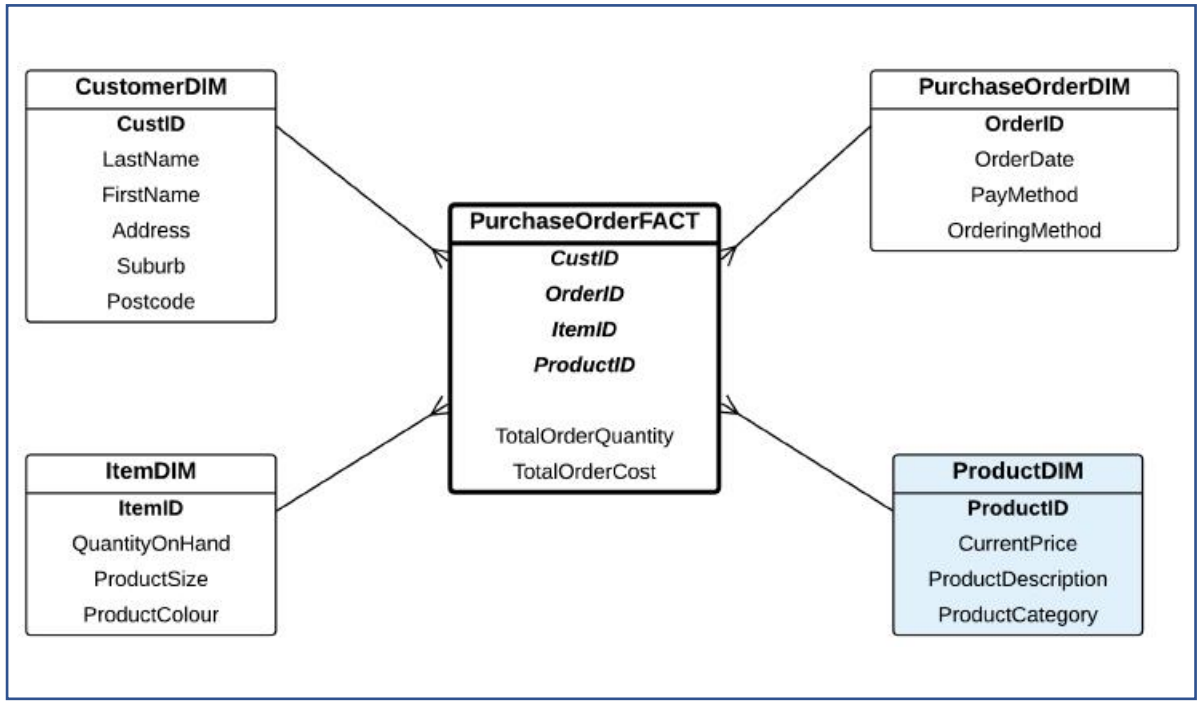


(f) Purchase Order Star Schema – with Item Dimension

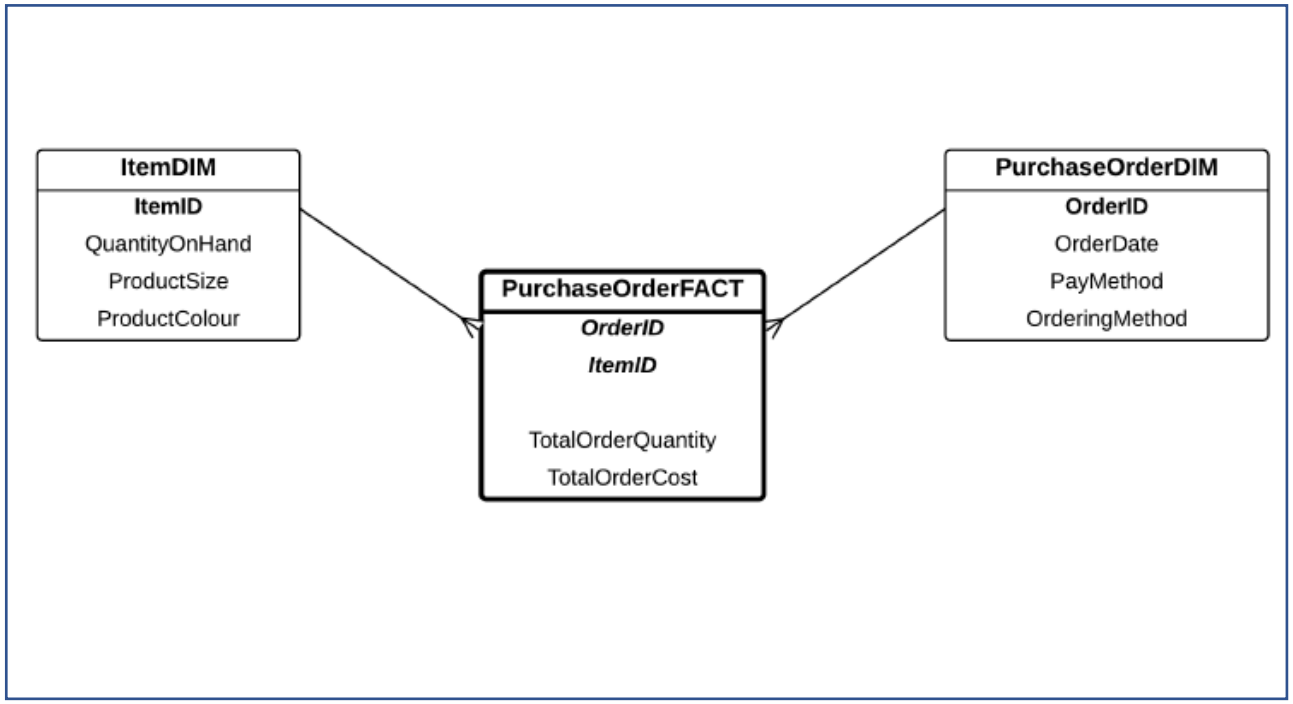


Star Schemas with No Aggregation

(g) Purchase Order Star Schema – with Product Dimension



(h) Purchase Order Star Schema – the minimum requirement

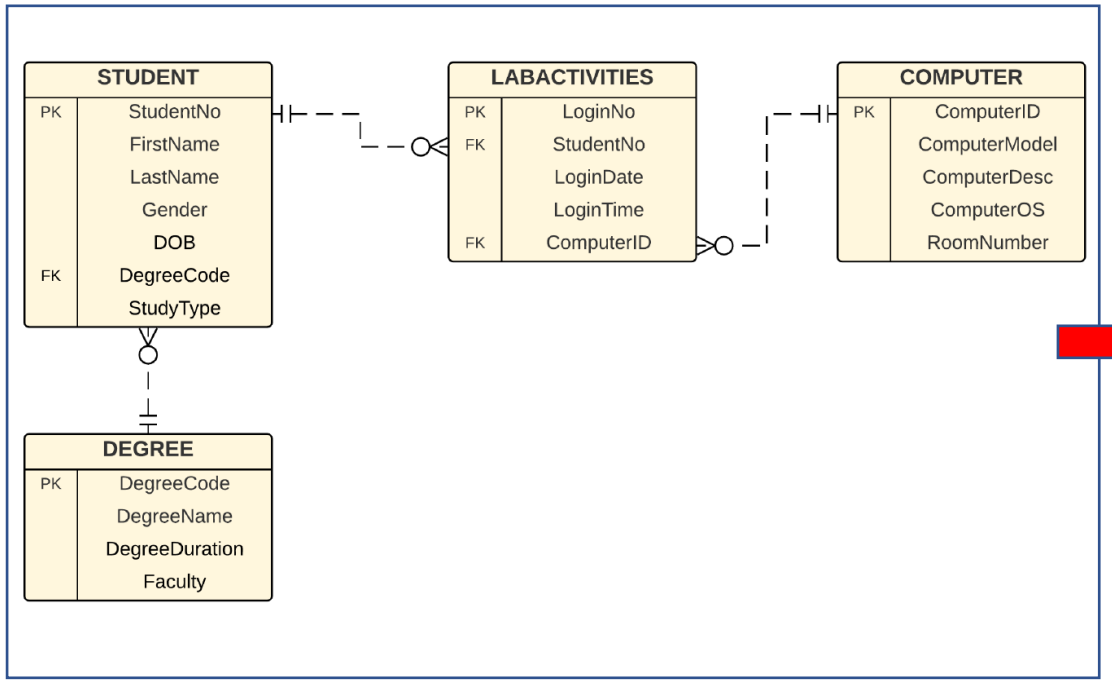


Understanding the Relationship between Transactions and Fact Measures

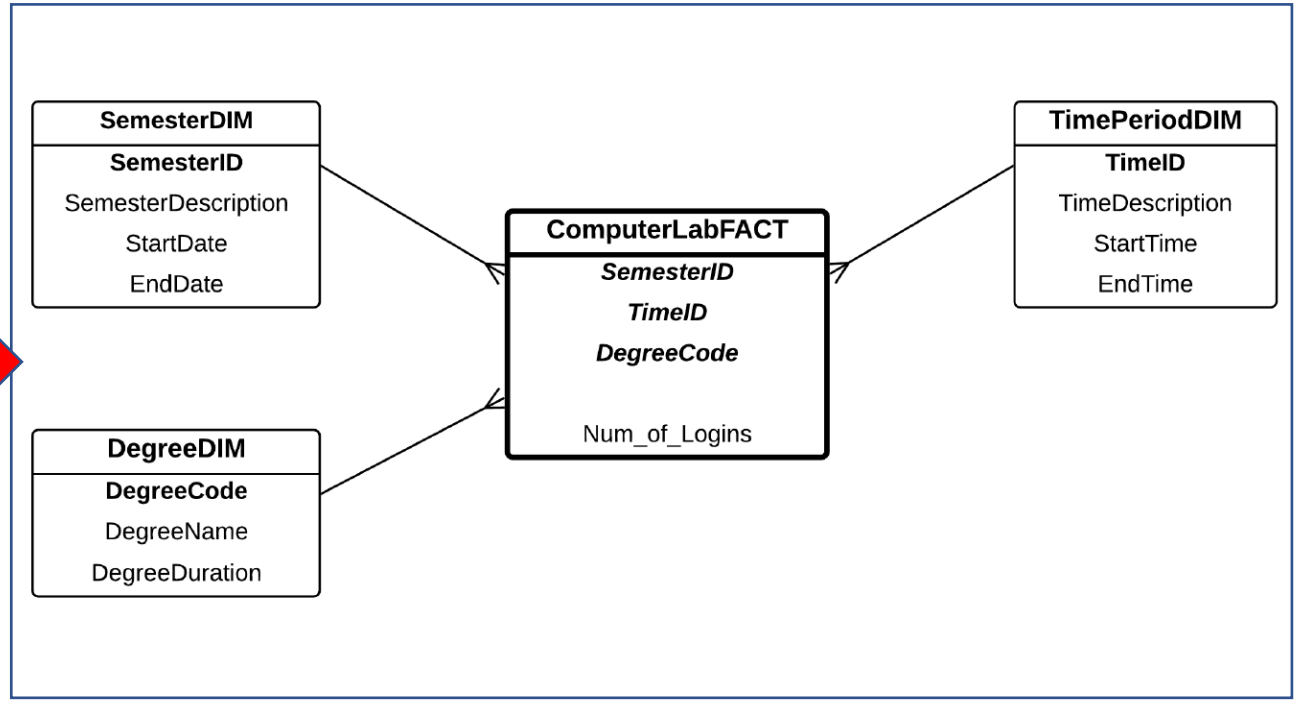
MONASH University

Relationship between Transactions and Fact Measures

(a) Operational Database

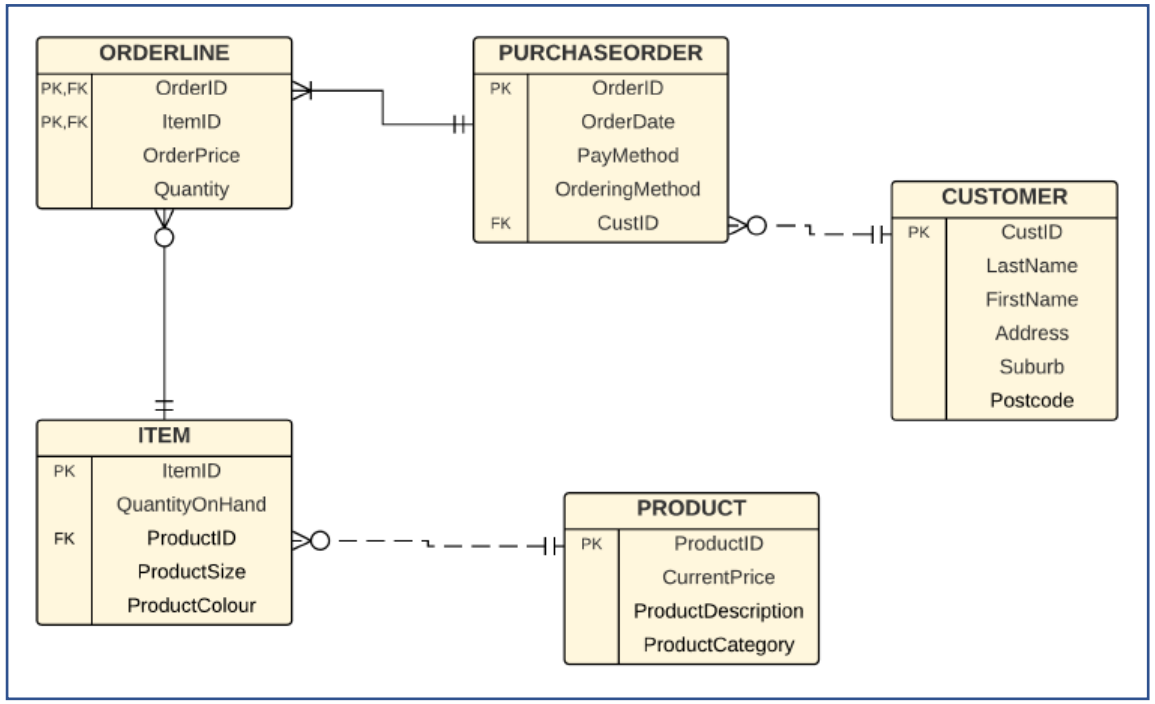


(b) Star Schema

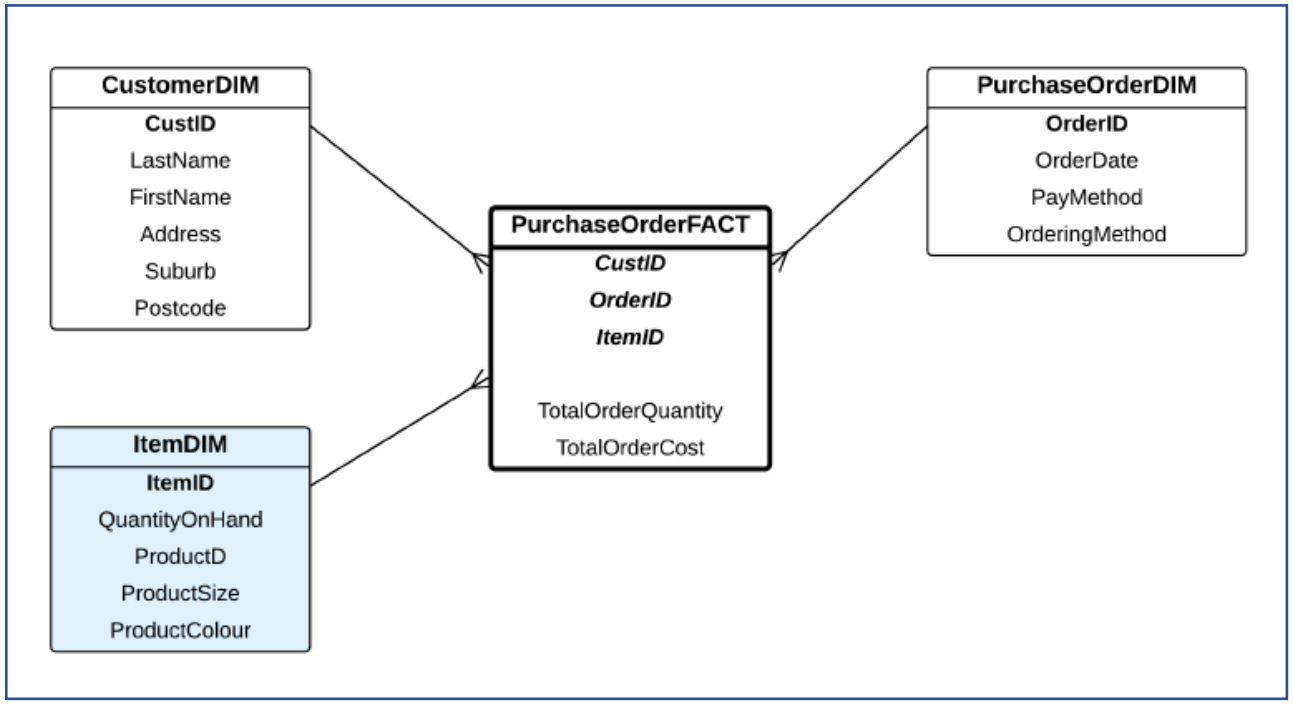


Relationship between Transactions and Fact Measures

(a) Purchase Order E/R Diagram

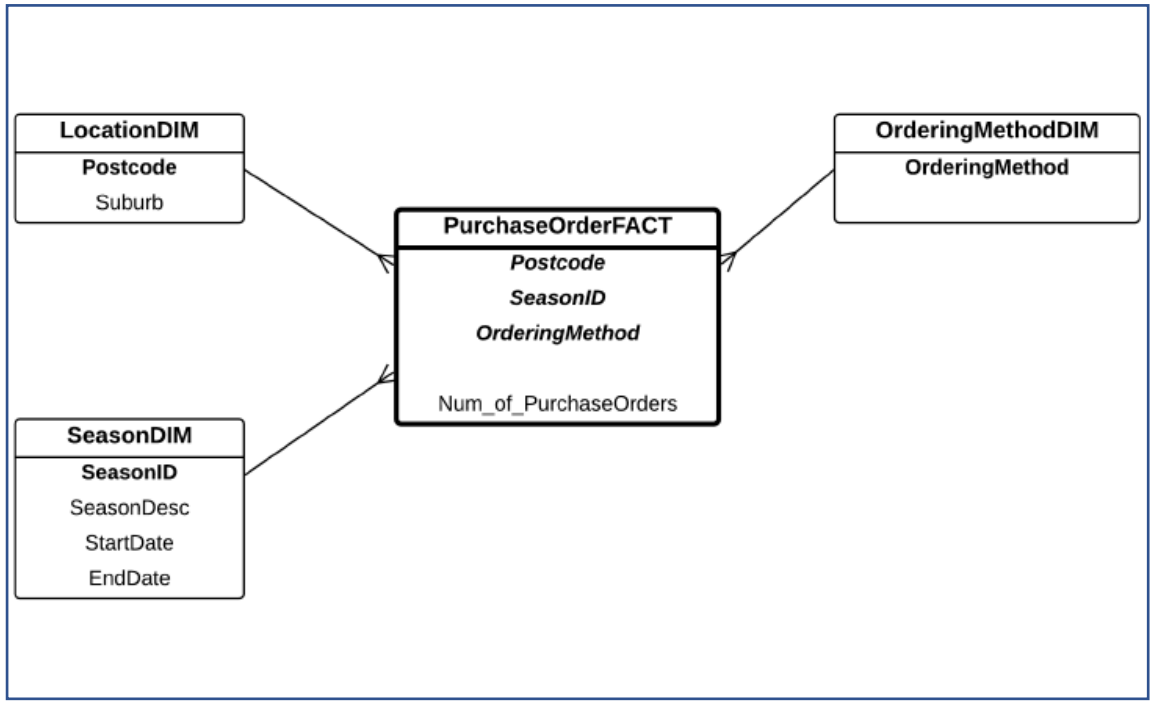


(b) Purchase Order Star Schema – with Item Dimension

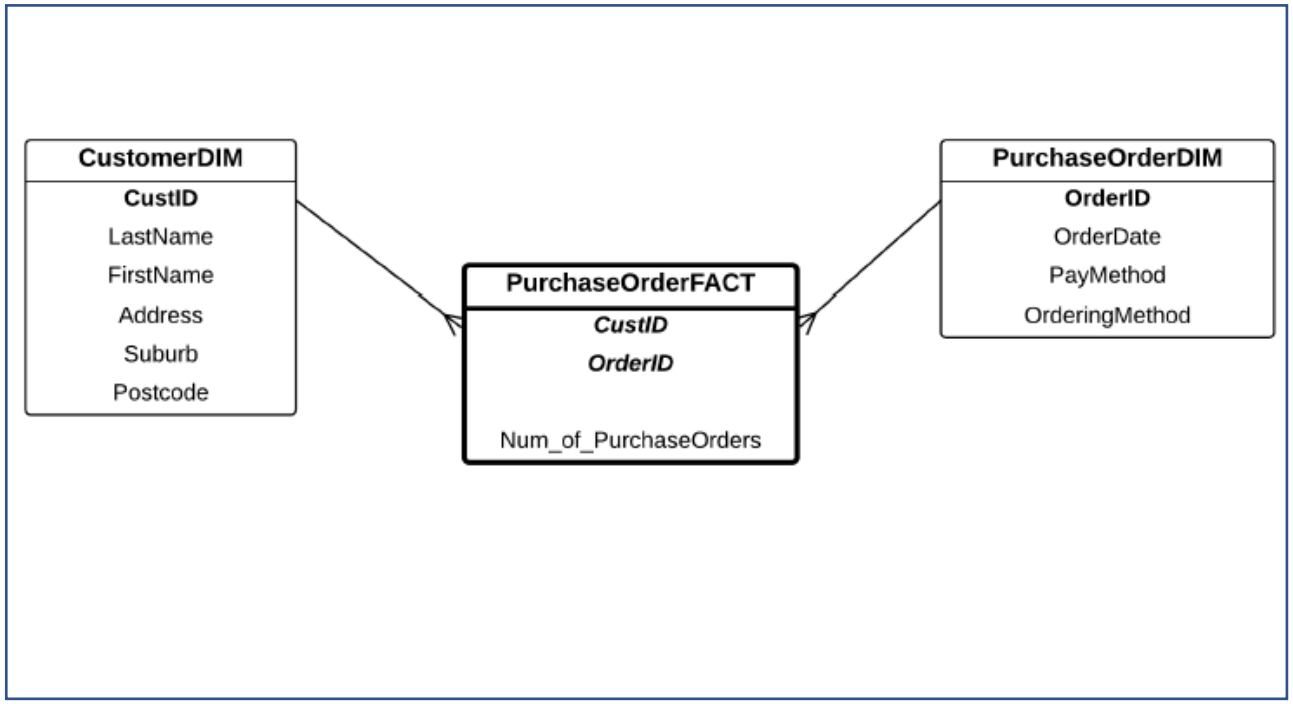


Relationship between Transactions and Fact Measures

(c) Purchase Order Star Schema – with Number of Purchase Orders as the fact measure

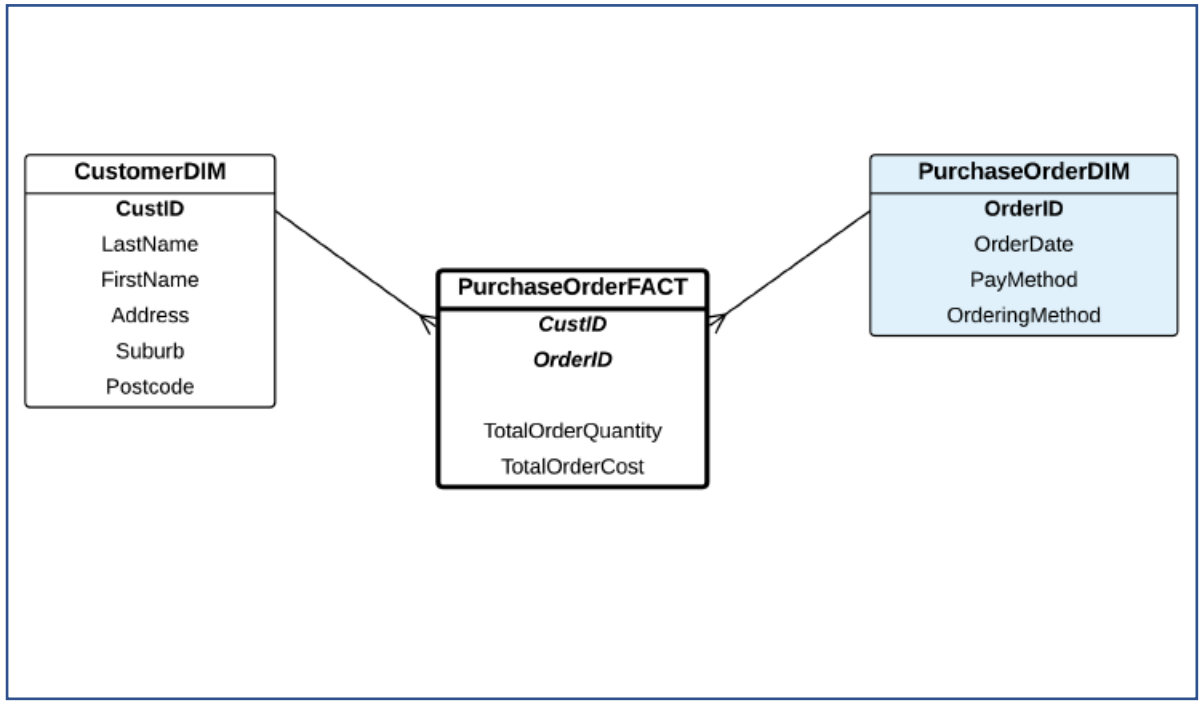


(d) Purchase Order Star Schema – Level-0

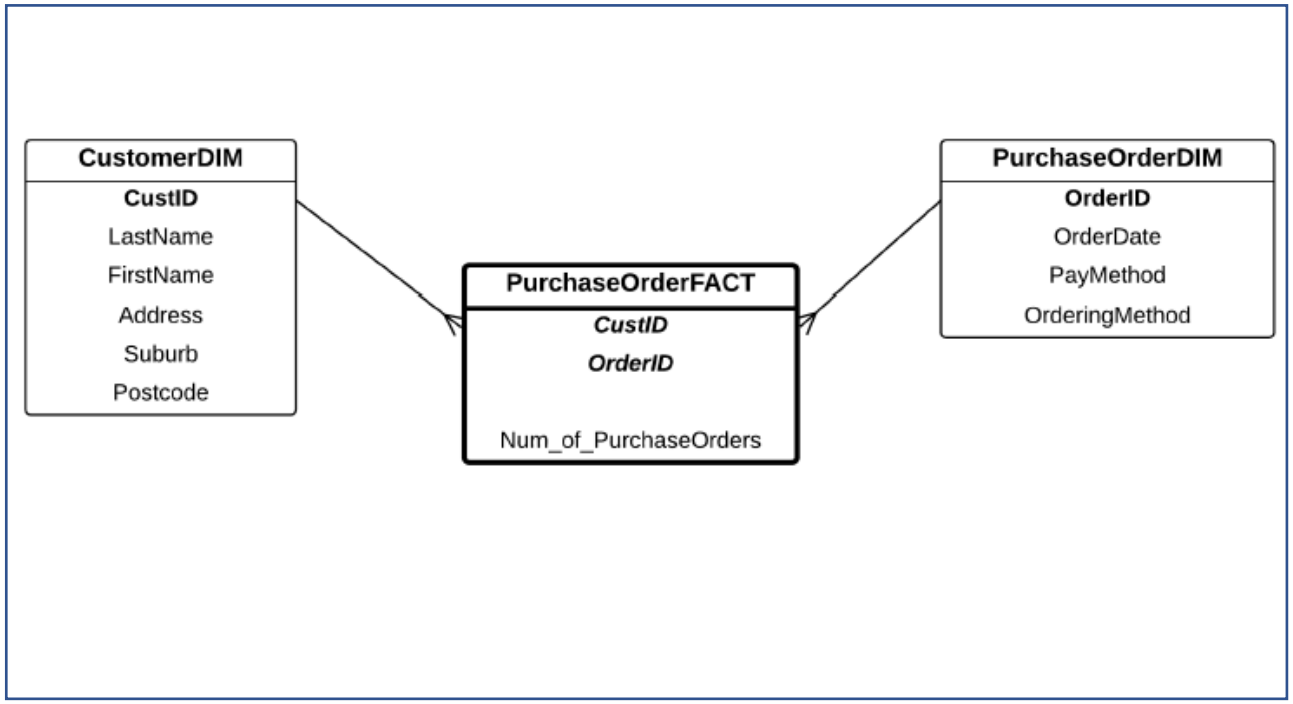


Relationship between Transactions and Fact Measures

(e) Purchase Order Star Schema – with Purchase Order Dimension

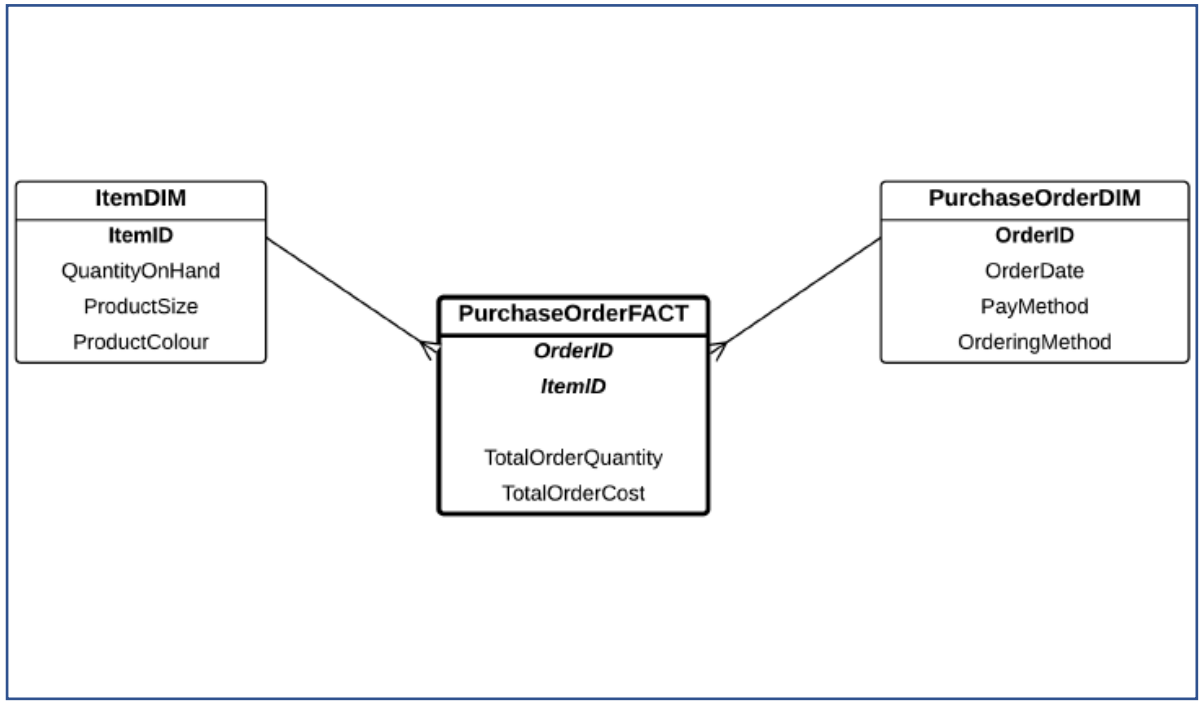


(d) Purchase Order Star Schema – Level-0

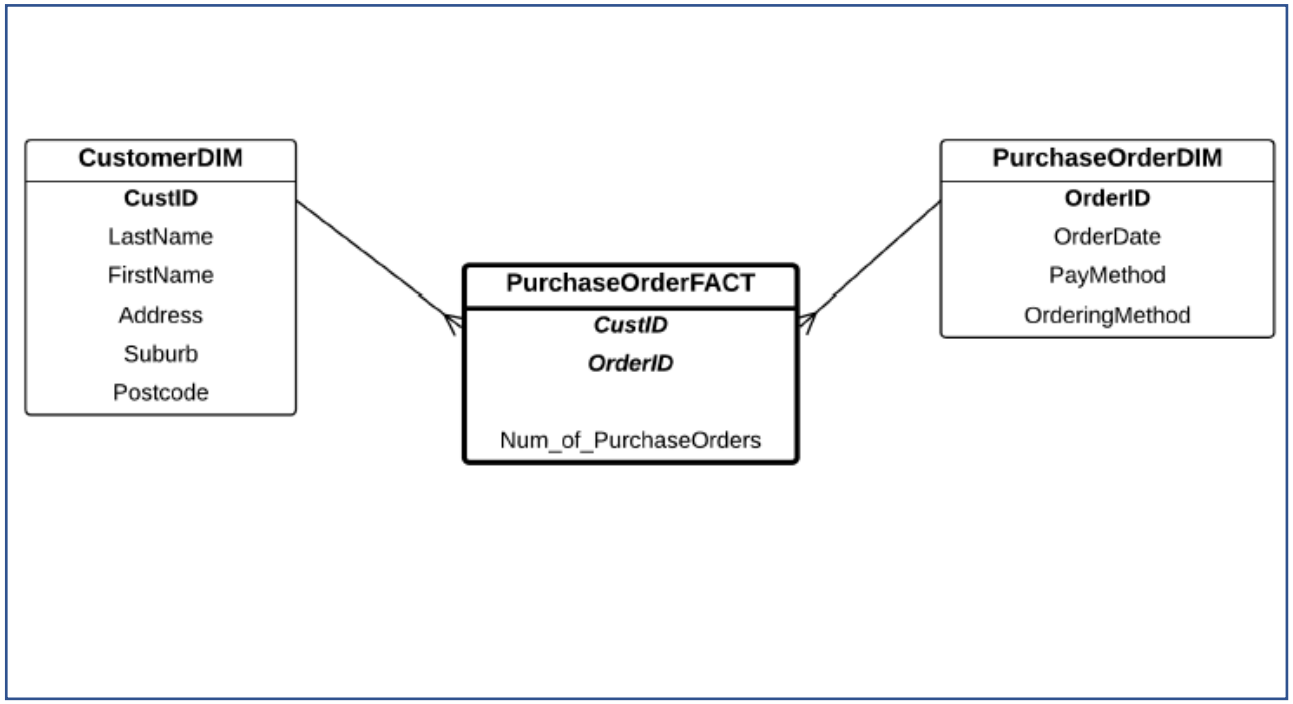


Relationship between Transactions and Fact Measures

(f) Purchase Order Star Schema – the minimum requirement

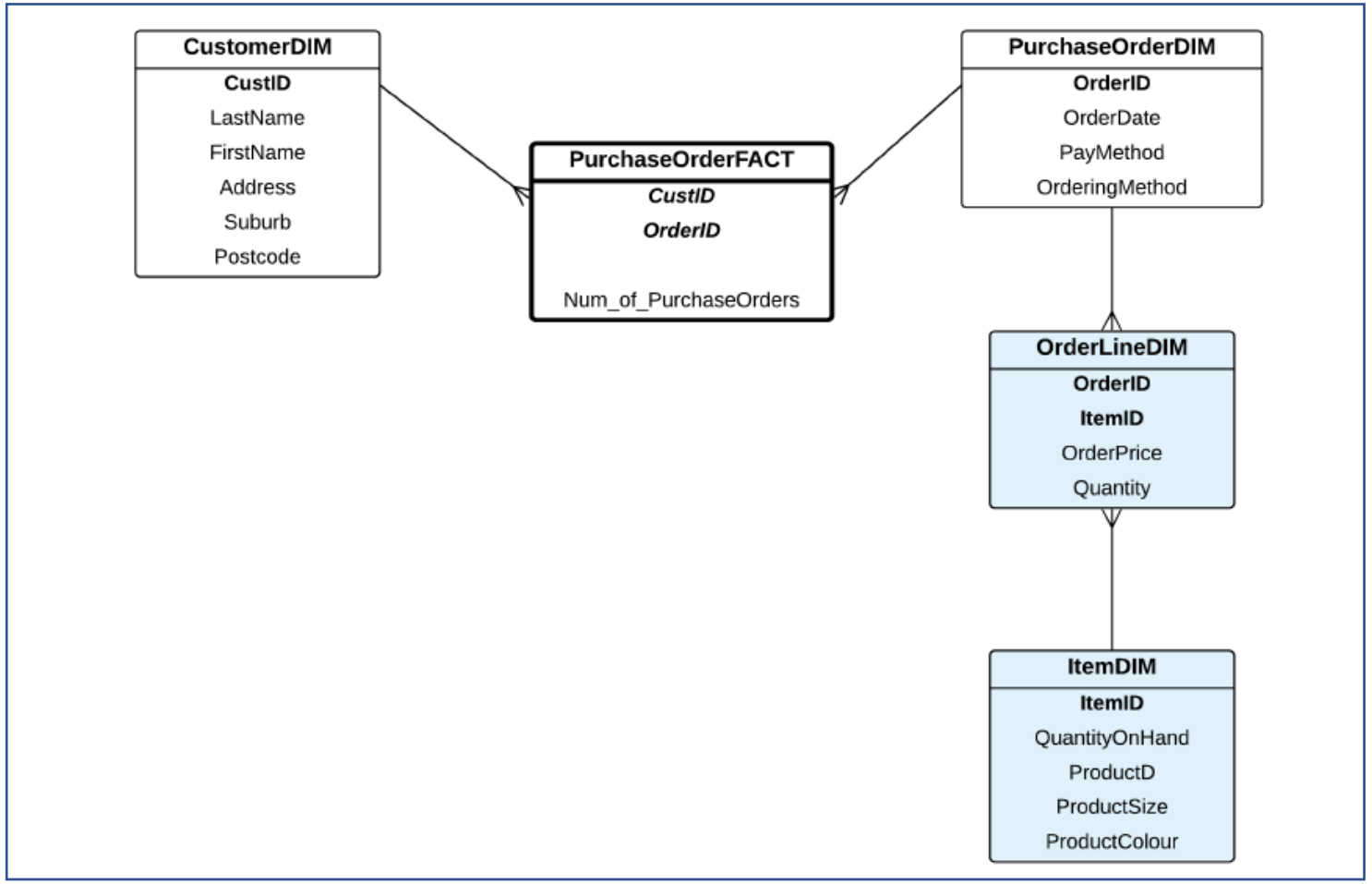


(d) Purchase Order Star Schema – Level-0



Relationship between Transactions and Fact Measures

(e) Purchase Order Star Schema – Level-0 with a Bridge to Item Dimension



Summary

MONASH University

Summary

- The higher the level of aggregation, the more aggregation in the fact measure.
- A data warehouse is built primarily used for drilling down some interesting data for business decision.
- It is common when we design a data warehouse, we start from a high level of aggregation, where fact measures contain aggregated values. Lowering down the level of aggregation can be done by changing the granularity of the dimension, or by simply adding new dimensions.
- Determining whether a star schema is in Level-0 or not can be tricky. Not having an aggregated fact measure does not always mean that the star schema is on Level-0. Hence, it is important to understand the concept of transaction recorded in the E/R diagram of the operational database.