MONASH University

MONASH
INFORMATION
TECHNOLOGY

FIT3003 – Business Intelligence and Data Warehousing

Week 2 – Star Schema

Semester 2, 2022

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

Learning Objectives

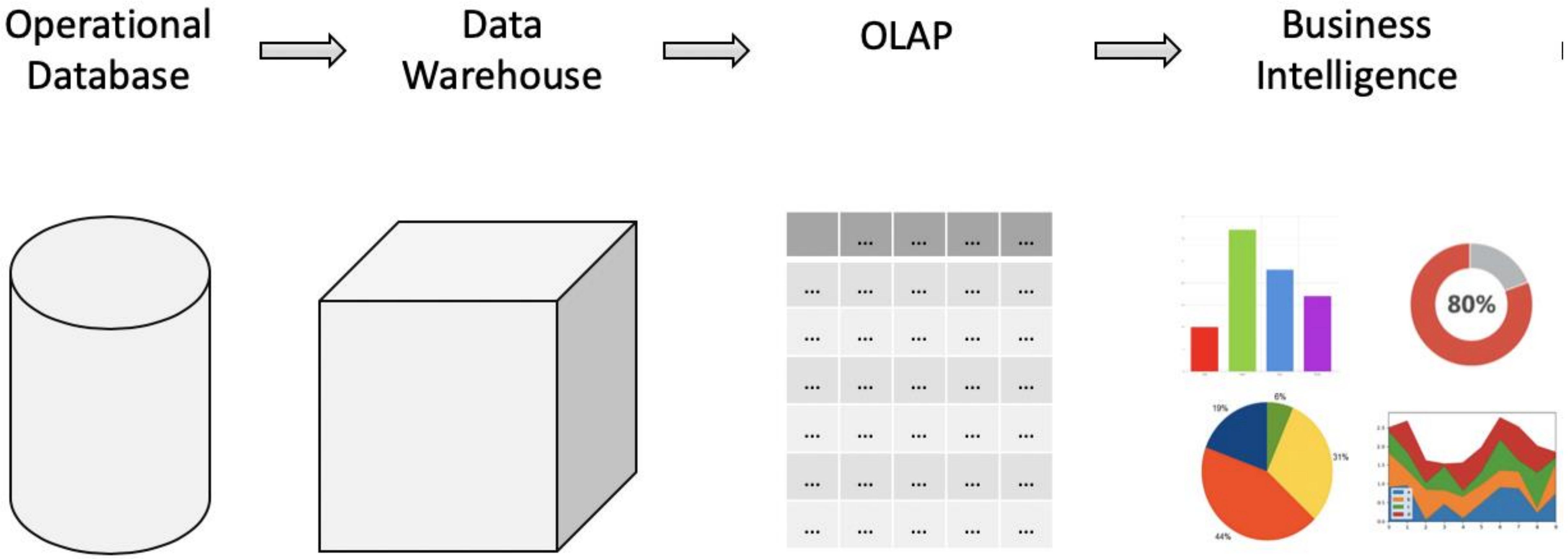
1. understand the concept of Star Schema
2. able to create a data warehousing model using Star Schema
3. understand the concept of Fact and Dimension
4. understand the concept of Two-column Methodology

Agenda

1. Notations and Processes
    1. Star Schema Notation
    2. E/R Diagram Notation
    3. Transformation Process (Case Study)

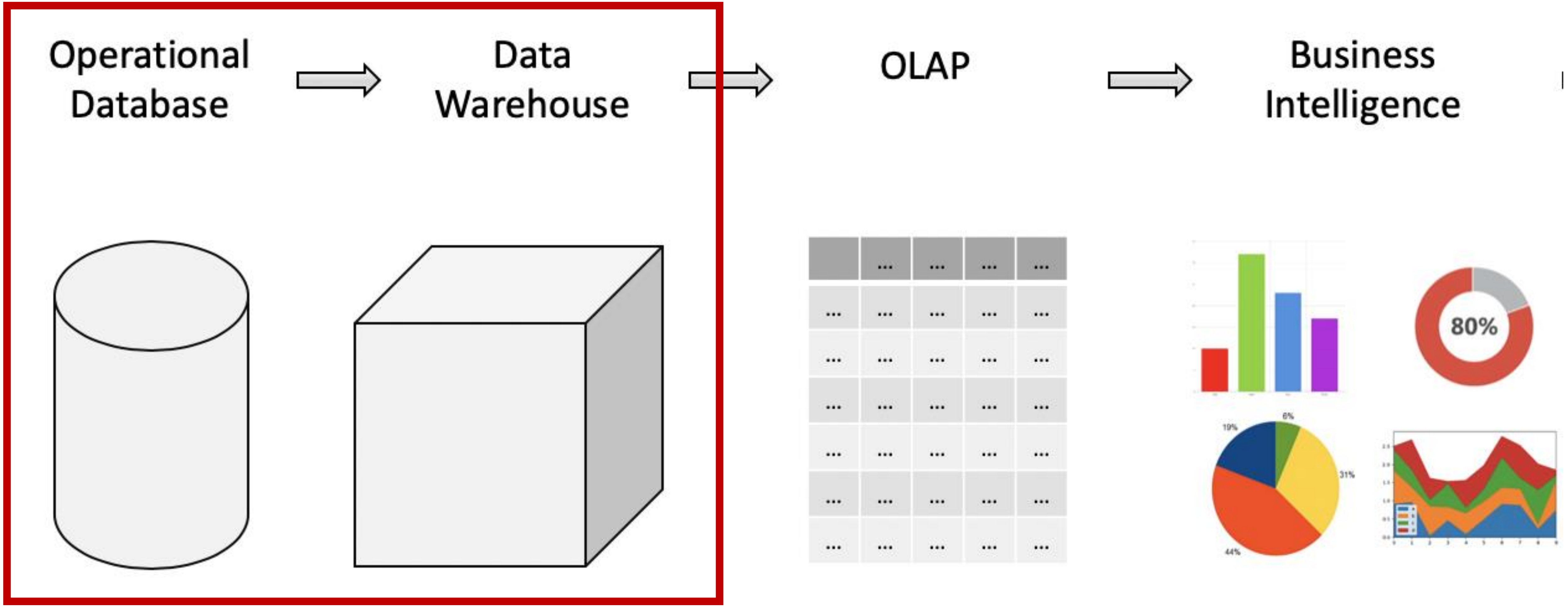
2. Two-Column Table Methodology

Recall – The Big Picture



Operational Database → Data Warehouse → OLAP → Business Intelligence

Recall – The Big Picture



Operational Database → Data Warehouse → OLAP → Business Intelligence

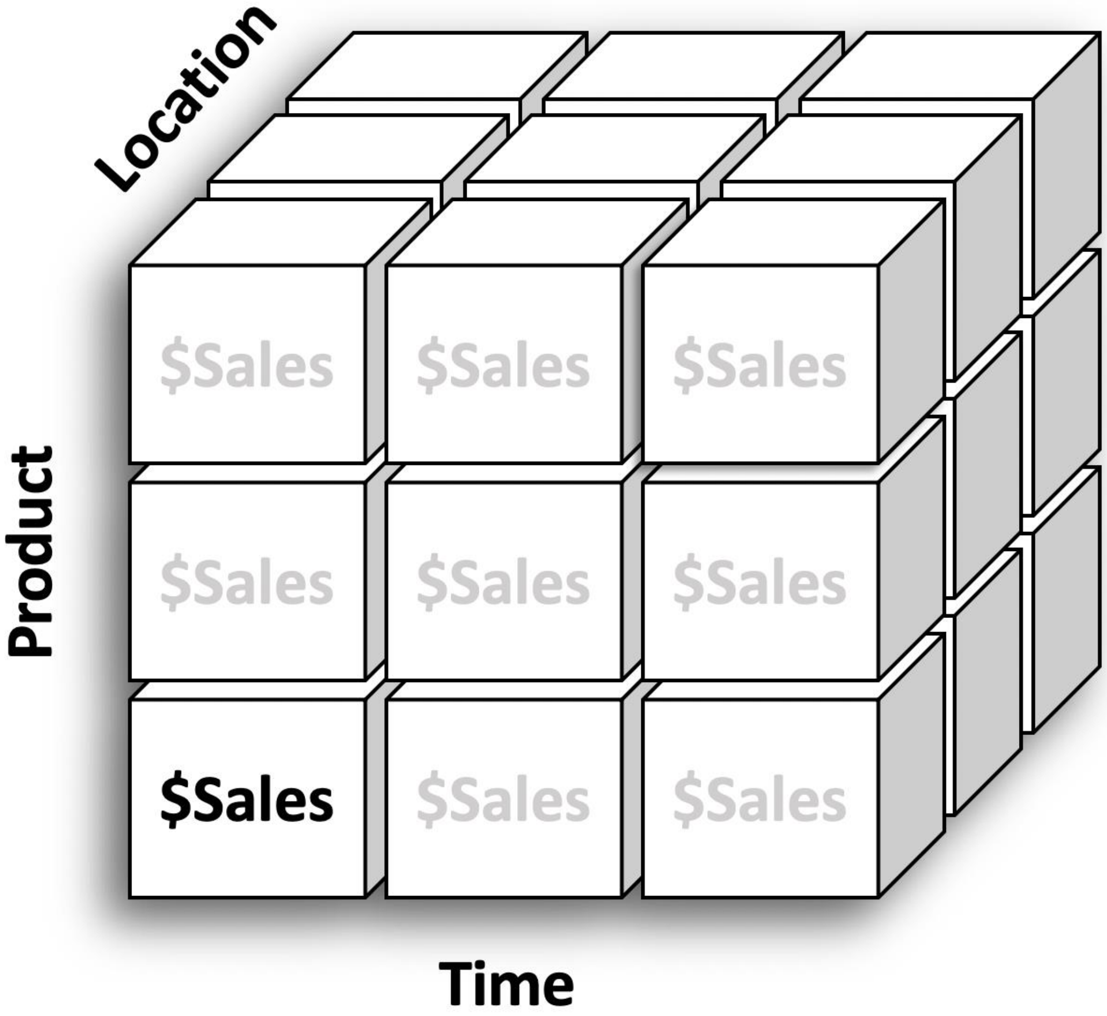
	...	...	...	...
...	...	...	...	...
...	...	...	...	...
...	...	...	...	...
...	...	...	...	...
...	...	...	...	...
...	...	...	...	...

Recall – Data Warehouse

- To address the drawback of operational database, and a need for decision-making support data, **data warehouse is needed**.

- A **data warehouse** is a multi-dimensional view of databases, with aggregates and pre-computed summaries.
  - In many ways, it is basically doing aggregates in advance; that is exactly pre-computation done at the design level, rather than at the query level.

Recall – Data Warehouse



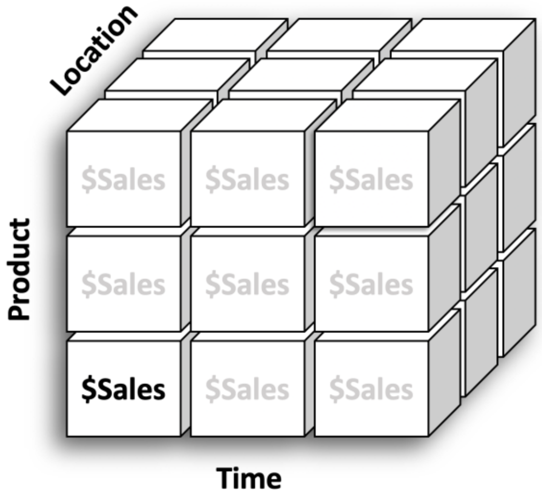
Star Schema

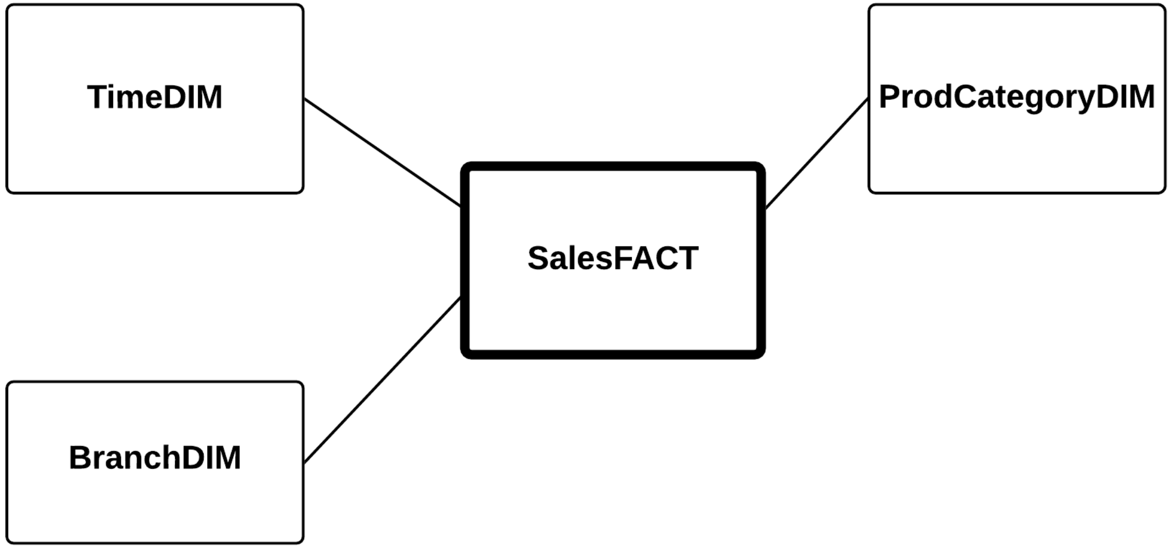
- A Star Schema is a design representation of a multi-dimensional view. It is a data modeling technique used to map multidimensional decision support data into a relational database.

- The reason for the star schema’s development is that existing relational modeling techniques: ER and normalization, did not yield a database structure that served the advanced data analysis requirements well.

Star Schema Components

- There are **Three** main components of the Star Schema:
  1. Facts
  2. Dimensions
  3. Attributes





Star Schema Components

1. **Facts**

   Facts are **numeric measurements** (values) that represent a specific business aspect or activity.

   For example, sales figures are numeric measurements that represent product and/or service sales.

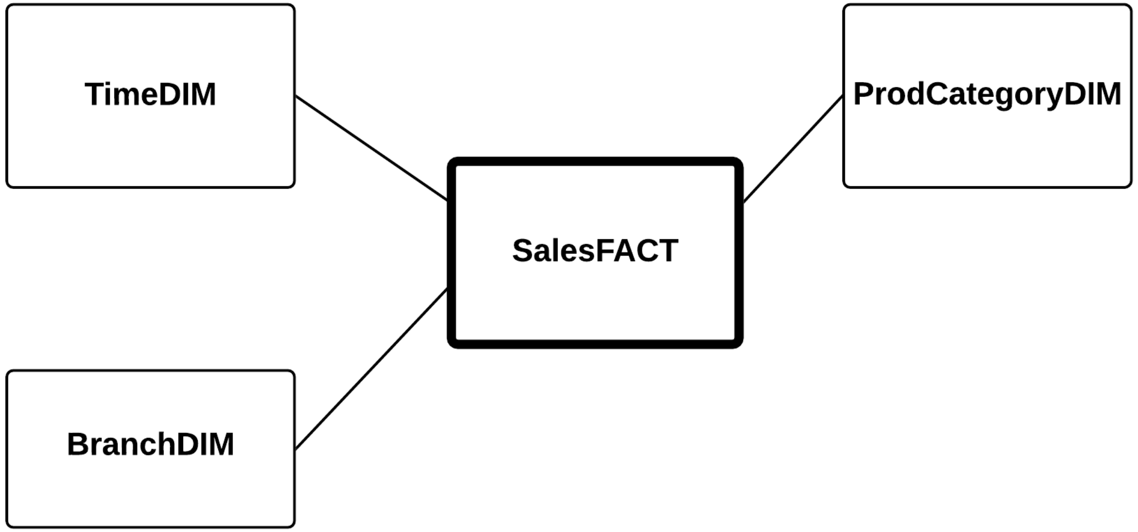
2. **Dimensions**

   Dimensions are qualifying characteristics that provide **additional perspectives** to a given fact.

   For example, sales might be *viewed* from specific dimension(s), such as sales location, sales period, sales product, etc.

Star Schema Notation

- A Sales Star Schema
  - **Fact**:
    - Sales
  - **Dimensions**:
    - Time
    - Product
    - Branch

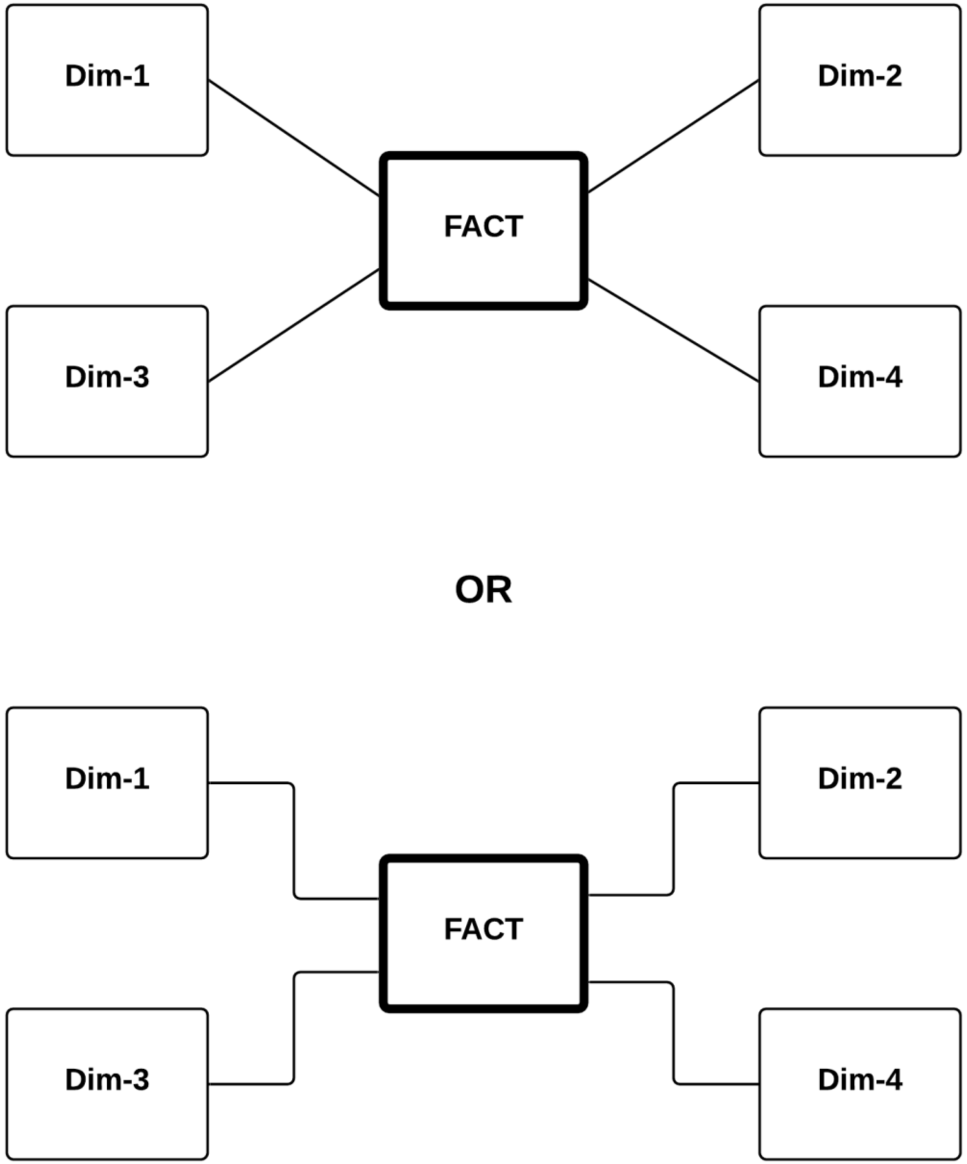


- Notation-wise, the Fact uses a bolder line, to differentiate between Fact from Dimensions.

Star Schema Notation

- A Sales Star Schema
  - **Fact**:
    - Sales
  - **Dimensions**:
    - Time
    - Product
    - Branch

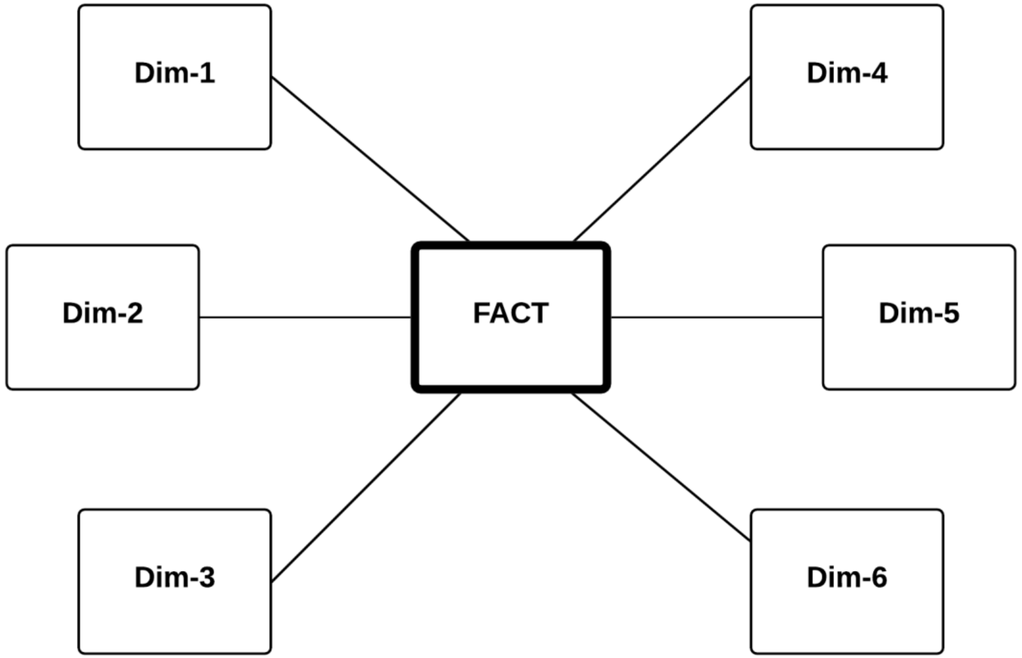
- The lines that represent a relationship between the fact and dimensions can be straight lines or bended lines.



Star Schema Notation

- Using the star schema notation, the number of dimensions can be unlimited.

- If there is more dimensions, then we just add more dimensions linked to the Fact.



Star Schema Components

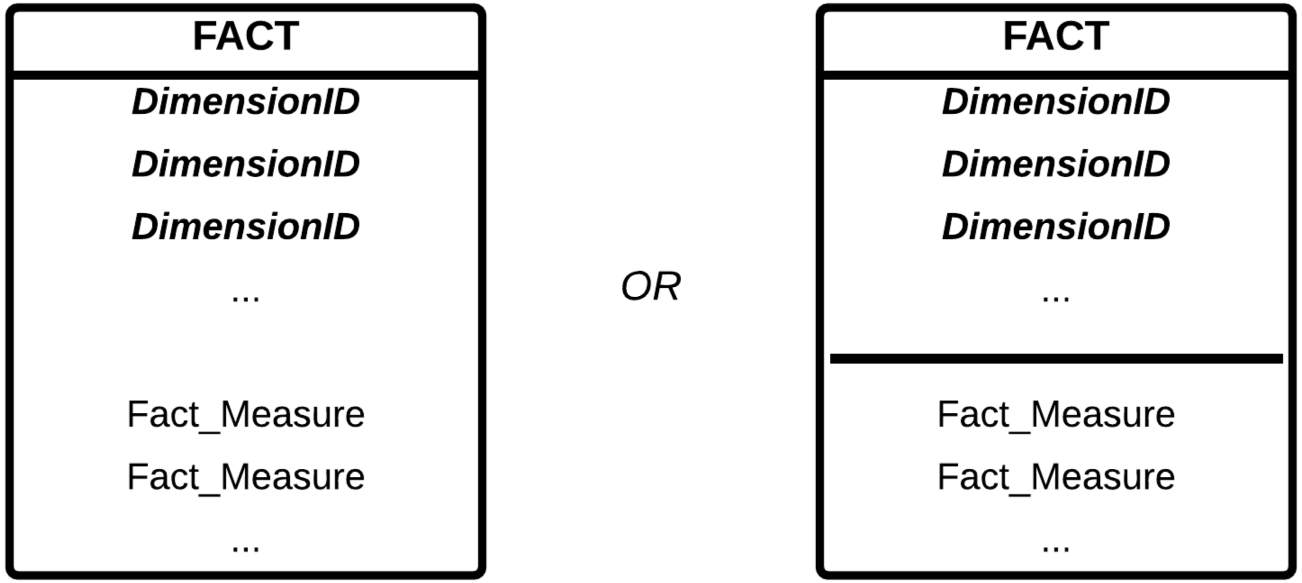
3. **Attributes**

   Each dimension table contains attributes.

   For example:

Product dimension:	Prod Type, Description.
Location dimension:	Region, State, City.
Time dimension:	Year, Month.

Star Schema Notation



FACT

DimensionID

DimensionID

DimensionID

...

Fact_Measure

Fact_Measure

...

OR

FACT

DimensionID

DimensionID

DimensionID

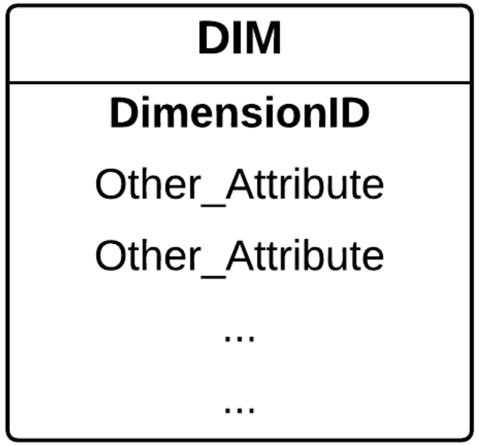
...

Fact_Measure

Fact_Measure

...

(*a*) Fact



DIM

DimensionID

Other_Attribute

Other_Attribute

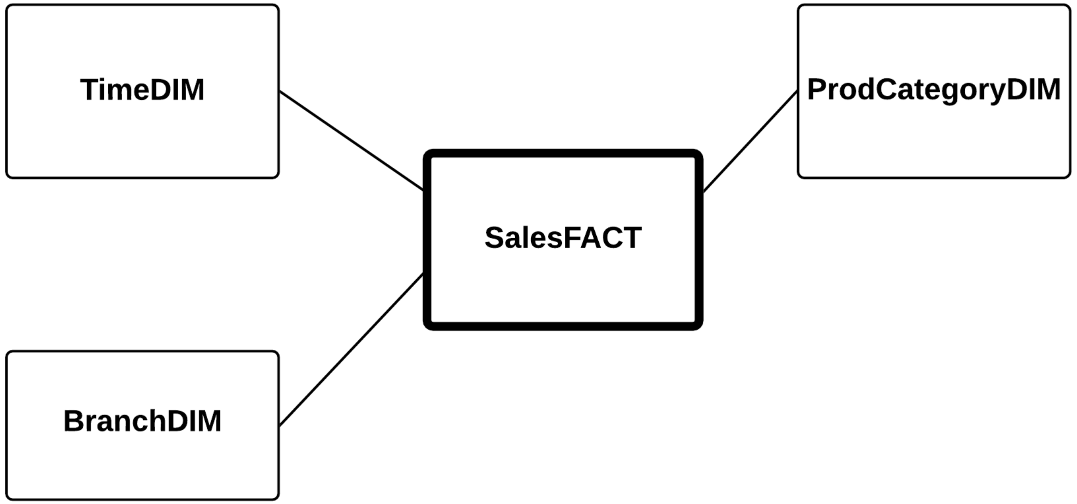
...

...

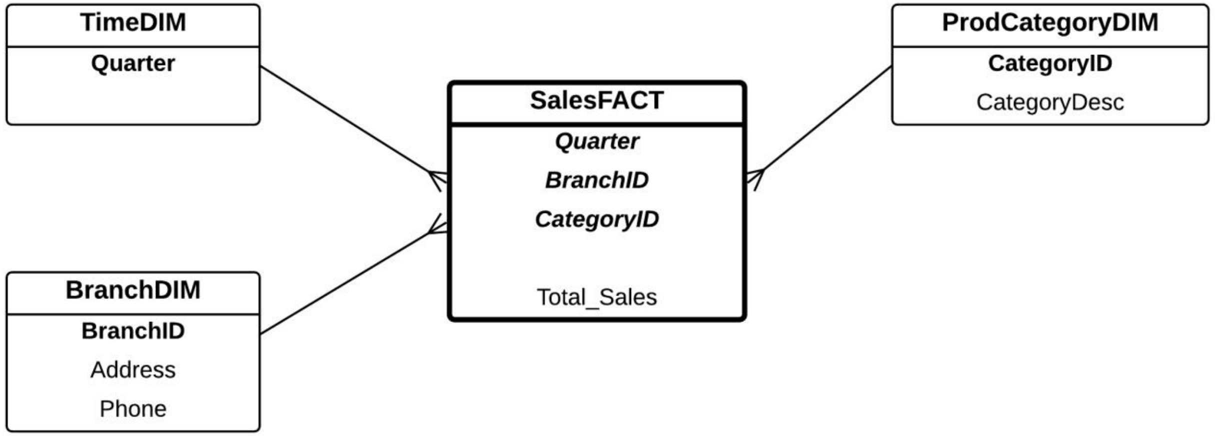
(*b*) Dimension

Star Schema Notation

- Sales Star Schema



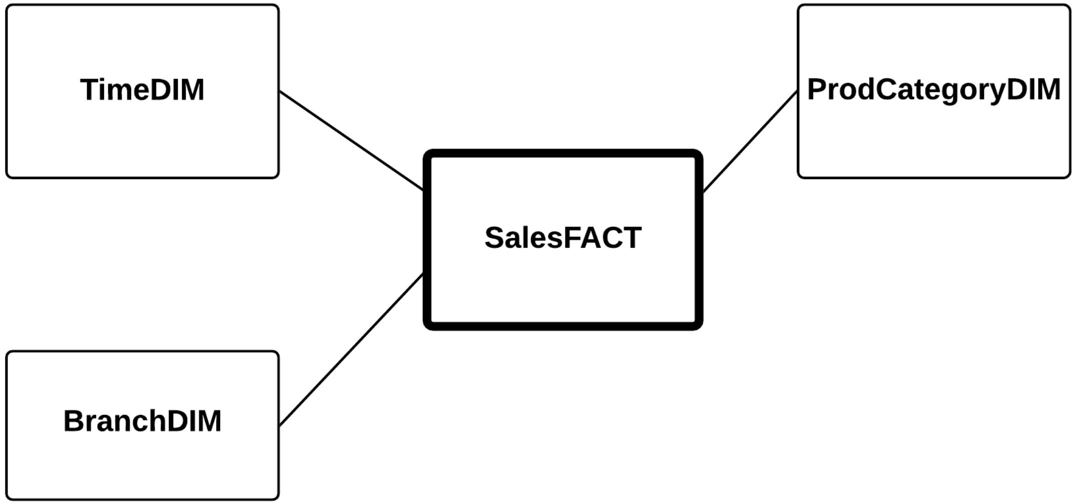
(a) Outline star schema for Sales



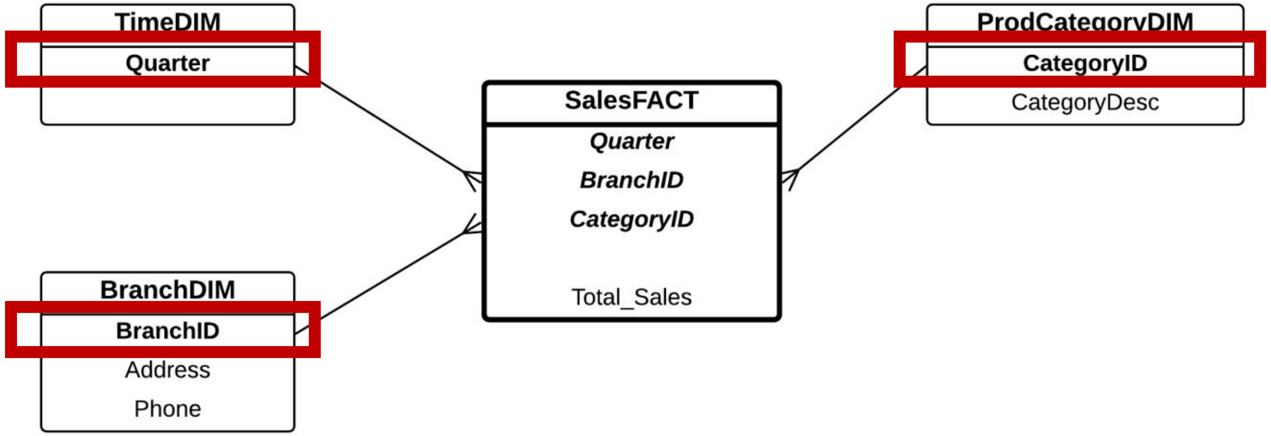
(b) Sales Star Schema complete with the Attributes

Star Schema Notation

- Sales Star Schema



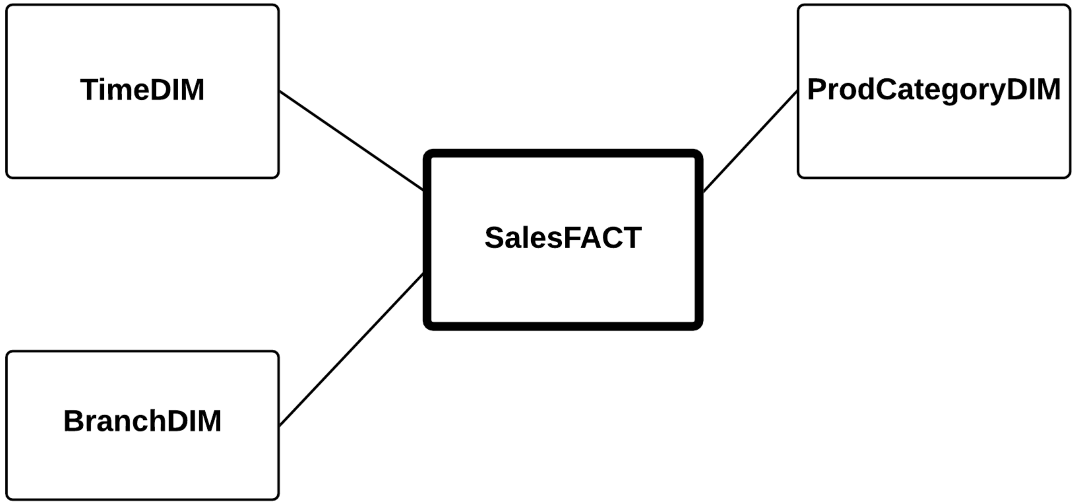
(a) Outline star schema for Sales



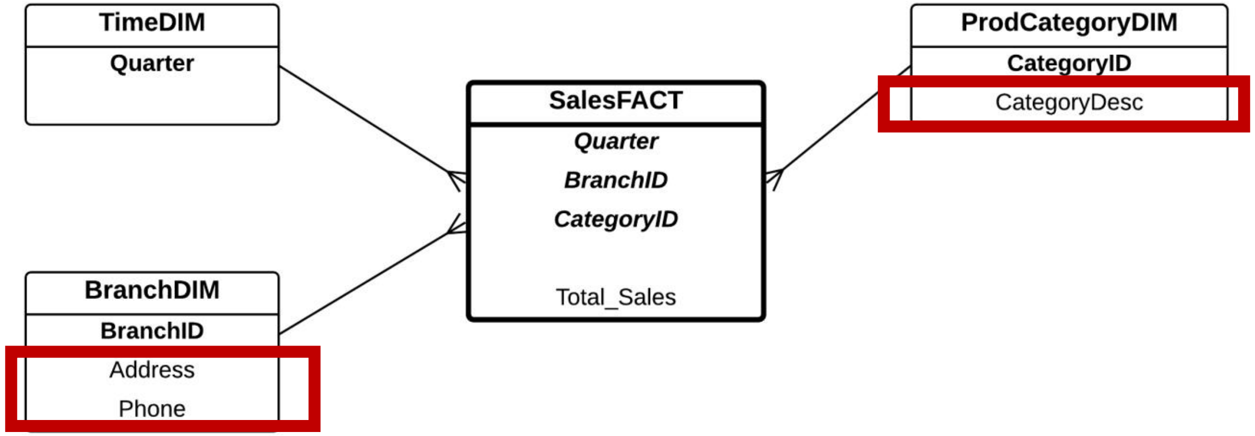
(b) Sales Star Schema complete with the Attributes

Star Schema Notation

- Sales Star Schema



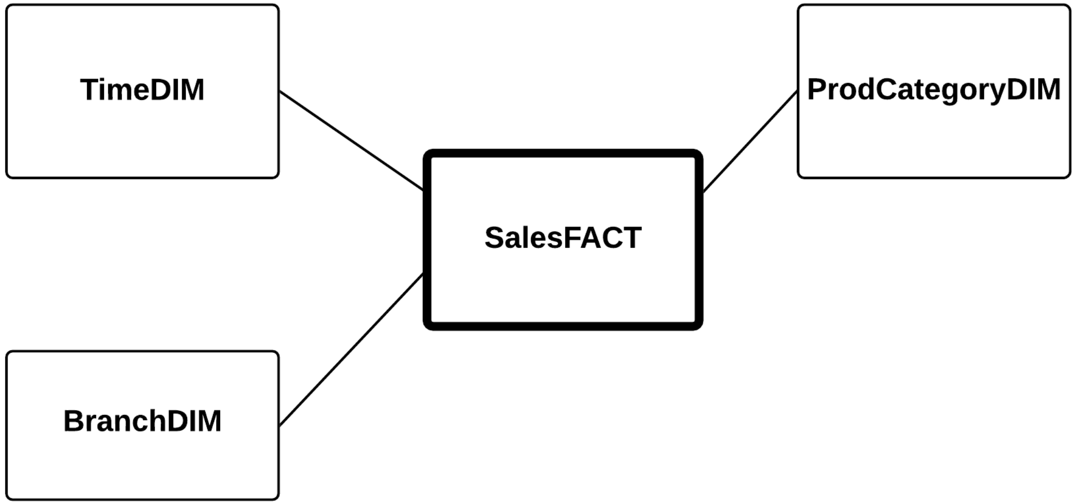
(a) Outline star schema for Sales



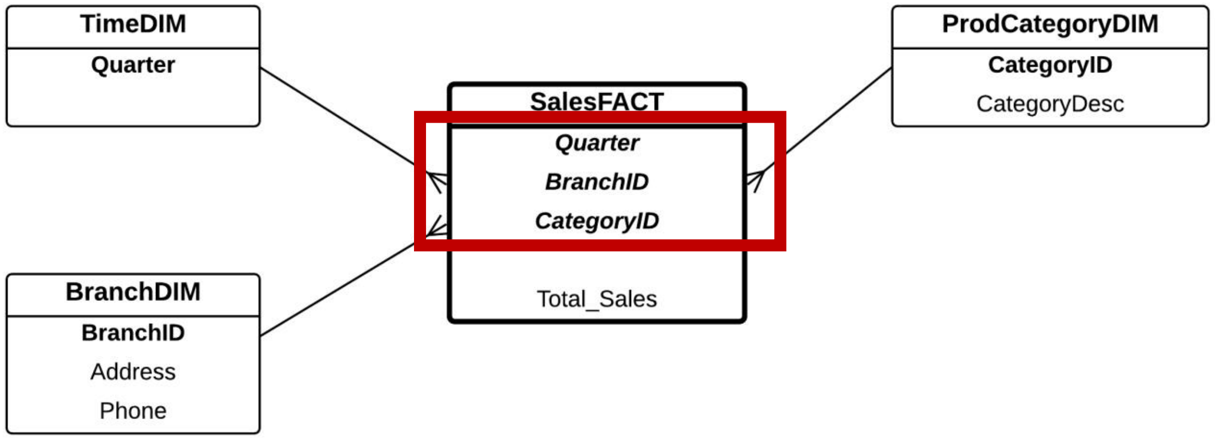
(b) Sales Star Schema complete with the Attributes

Star Schema Notation

- Sales Star Schema



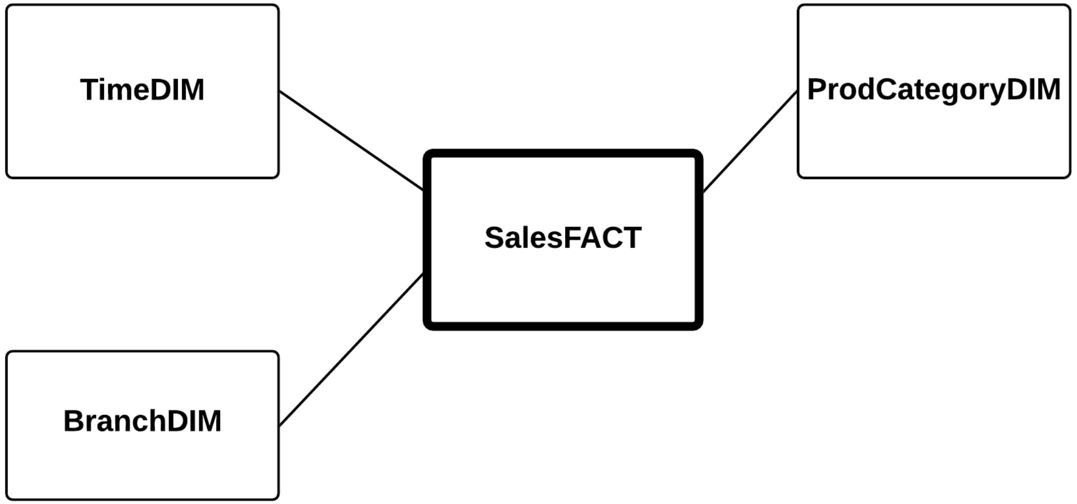
(a) Outline star schema for Sales



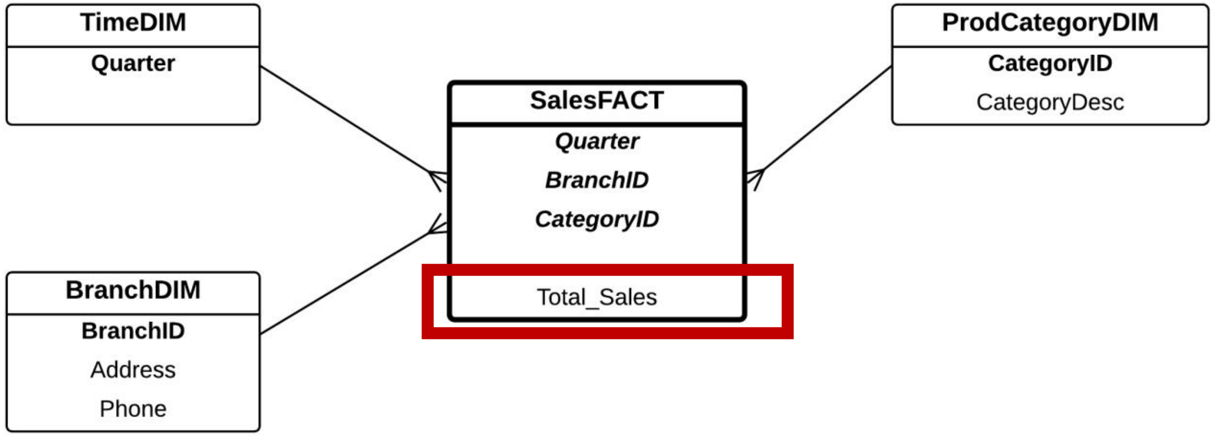
(b) Sales Star Schema complete with the Attributes

Star Schema Notation

- Sales Star Schema



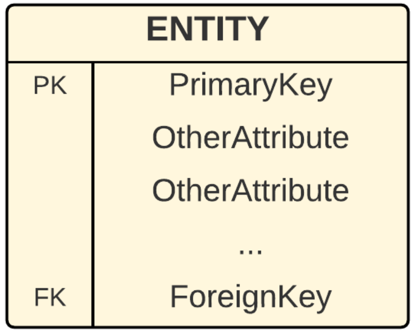
(a) Outline star schema for Sales



(b) Sales Star Schema complete with the Attributes

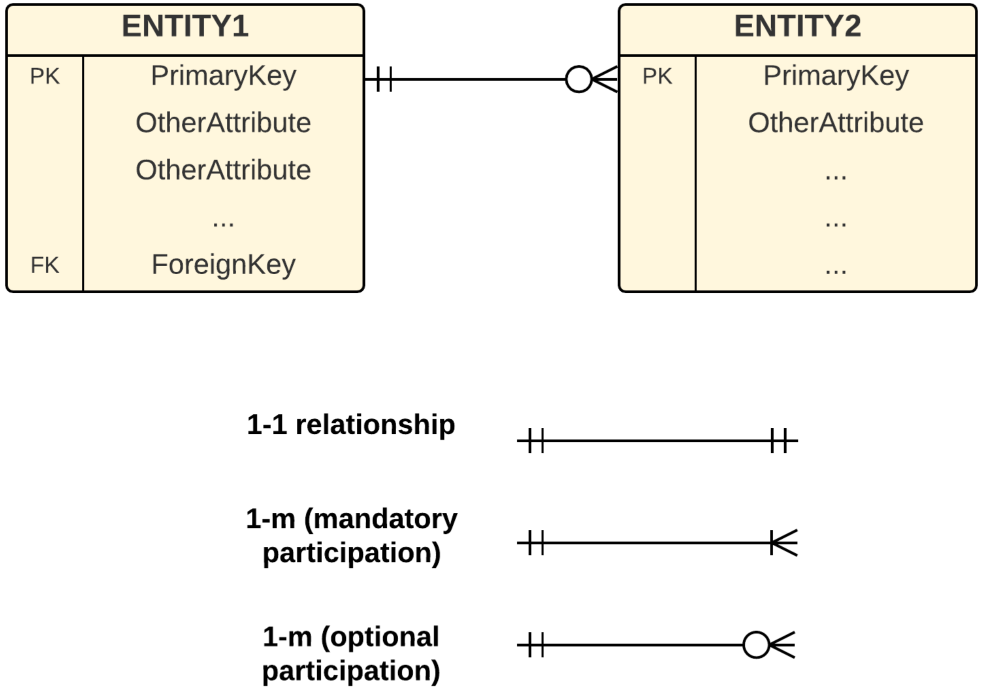
E/R Diagram Notation

E/R Diagram Notation



ENTITY	
PK	PrimaryKey
	OtherAttribute
	OtherAttribute
	...
FK	ForeignKey

(a) An Entity in E/R Diagram

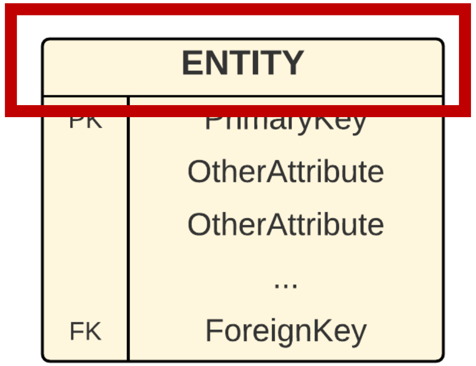


ENTITY1	
PK	PrimaryKey
	OtherAttribute
	OtherAttribute
	...
FK	ForeignKey

ENTITY2	
PK	PrimaryKey
	OtherAttribute
	...
	...
	...

(b) Relationships in E/R Diagram

E/R Diagram Notation



ENTITY

PK PrimaryKey

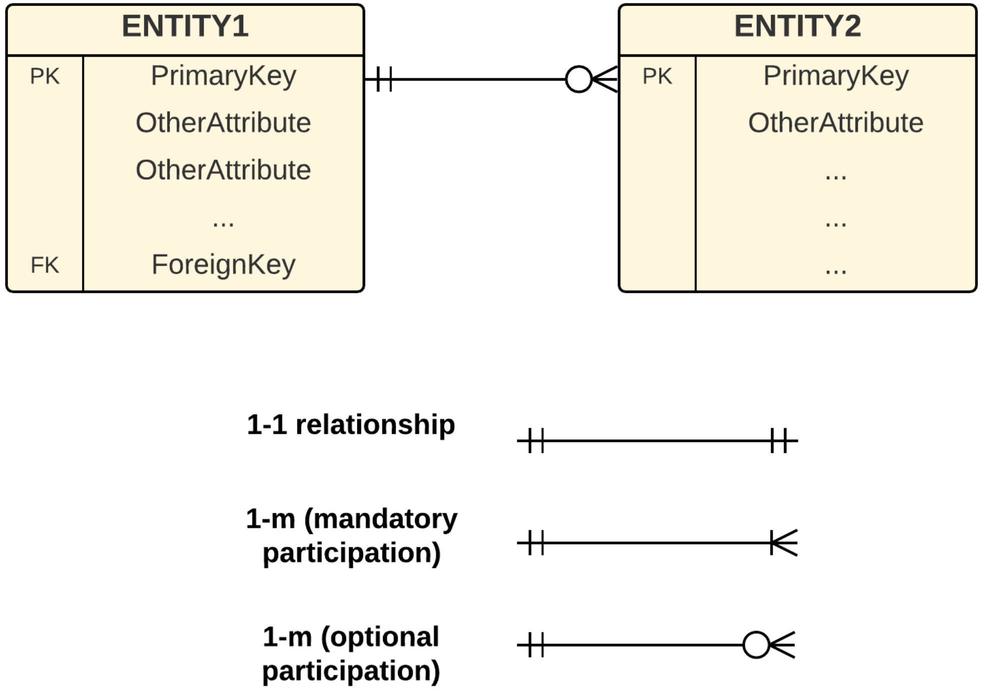
OtherAttribute

OtherAttribute

...

FK ForeignKey

(a) An Entity in E/R Diagram



ENTITY1

PK PrimaryKey

OtherAttribute

OtherAttribute

...

FK ForeignKey

ENTITY2

PK PrimaryKey

OtherAttribute

...

...

...

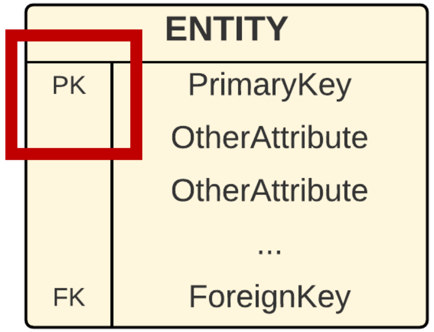
1-1 relationship

1-m (mandatory participation)

1-m (optional participation)

(b) Relationships in E/R Diagram

E/R Diagram Notation



ENTITY

PK PrimaryKey

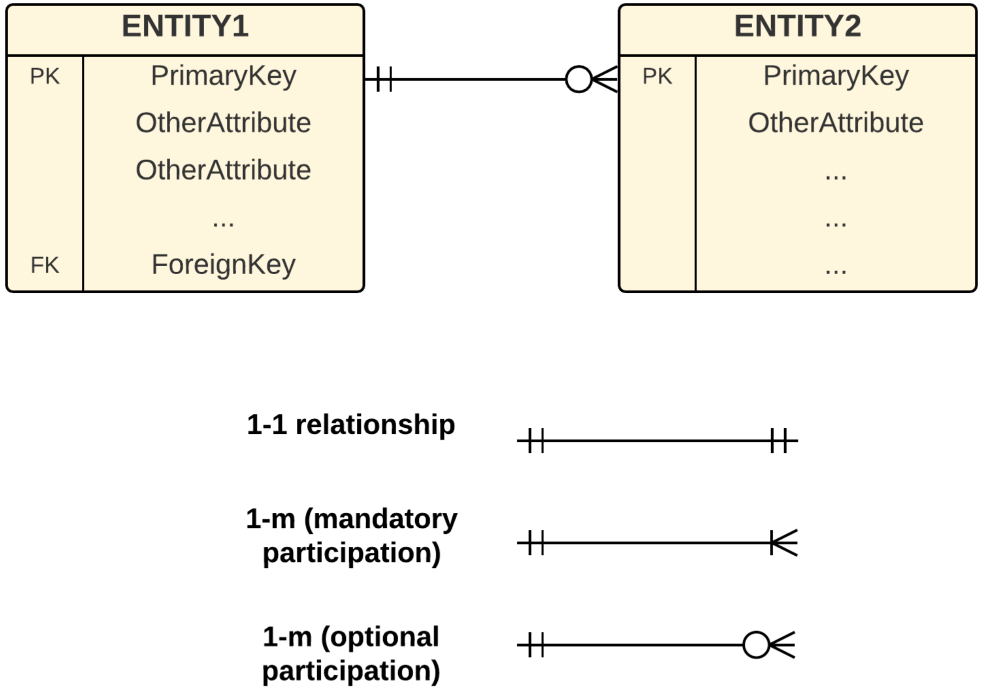
OtherAttribute

OtherAttribute

...

FK ForeignKey

(a) An Entity in E/R Diagram



ENTITY1

PK PrimaryKey

OtherAttribute

OtherAttribute

...

FK ForeignKey

ENTITY2

PK PrimaryKey

OtherAttribute

...

...

...

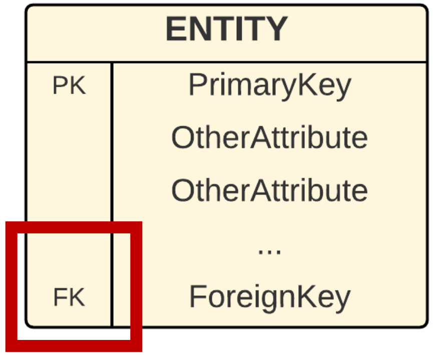
1-1 relationship

1-m (mandatory participation)

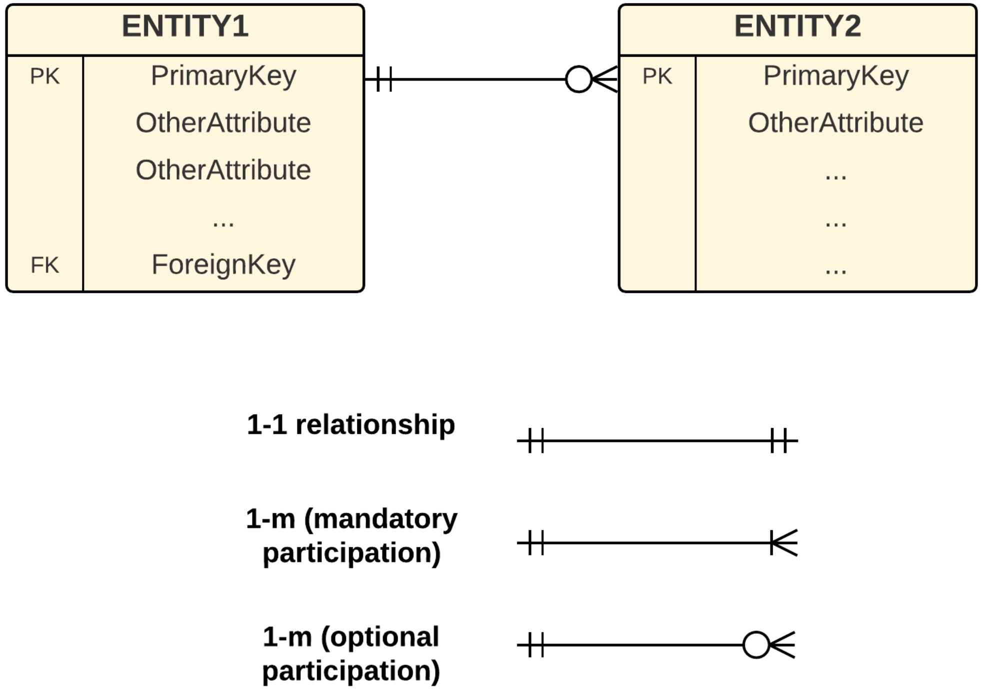
1-m (optional participation)

(b) Relationships in E/R Diagram

E/R Diagram Notation

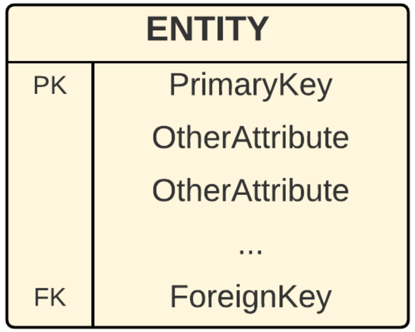


(a) An Entity in E/R Diagram



(b) Relationships in E/R Diagram

E/R Diagram Notation



ENTITY

PK PrimaryKey

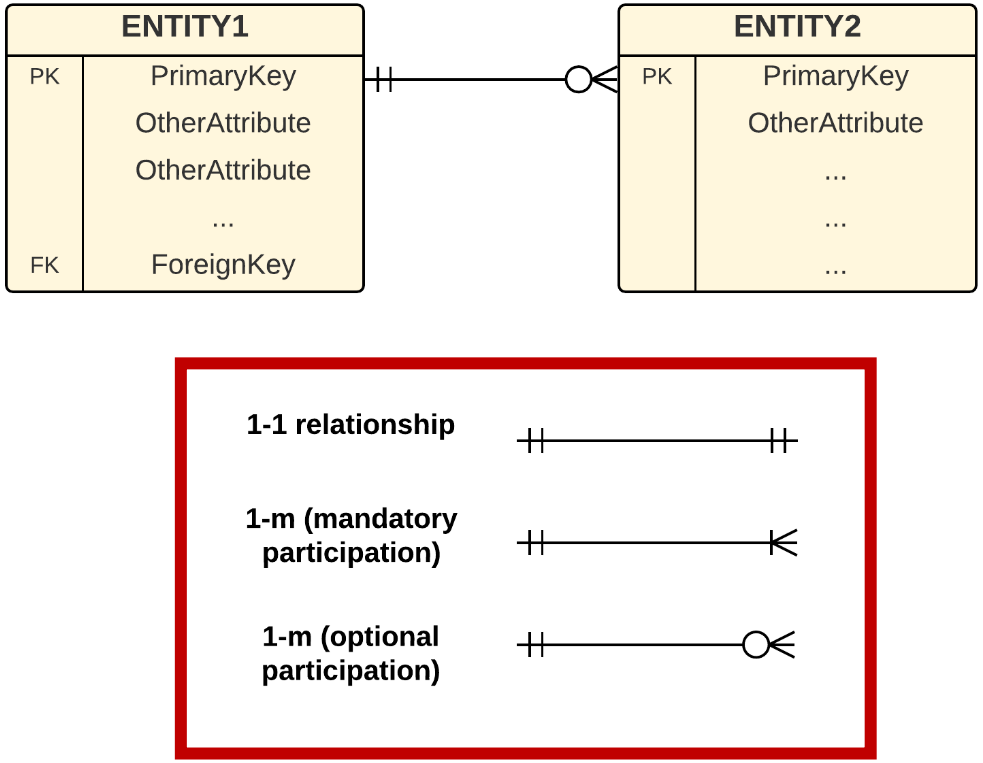
OtherAttribute

OtherAttribute

...

FK ForeignKey

(a) An Entity in E/R Diagram



ENTITY1

PK PrimaryKey

OtherAttribute

OtherAttribute

...

FK ForeignKey

ENTITY2

PK PrimaryKey

OtherAttribute

...

...

...

1-1 relationship

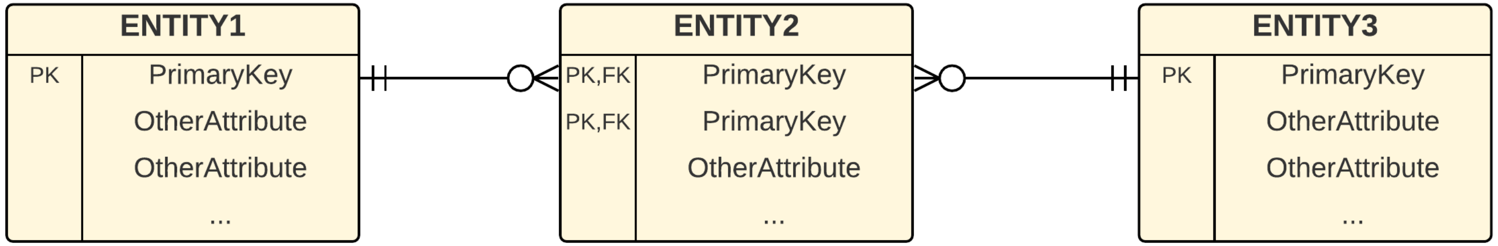
1-m (mandatory participation)

1-m (optional participation)

(b) Relationships in E/R Diagram

E/R Diagram Notation

Associative Relationship (m-m)

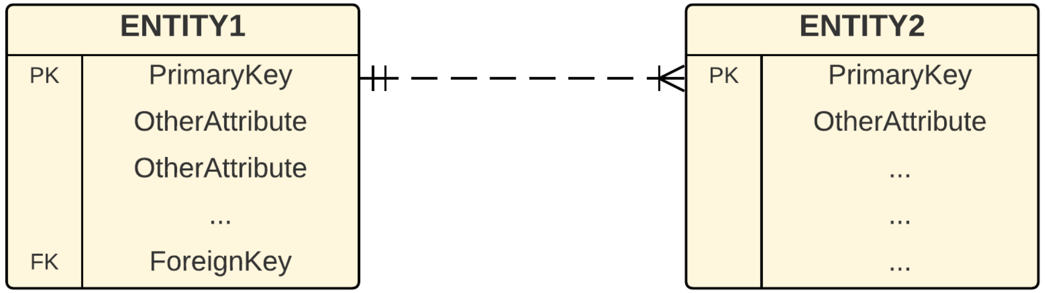


	ENTITY1
PK	PrimaryKey
	OtherAttribute
	OtherAttribute
	...

	ENTITY2
PK,FK	PrimaryKey
PK,FK	PrimaryKey
	OtherAttribute
	...

	ENTITY3
PK	PrimaryKey
	OtherAttribute
	OtherAttribute
	...

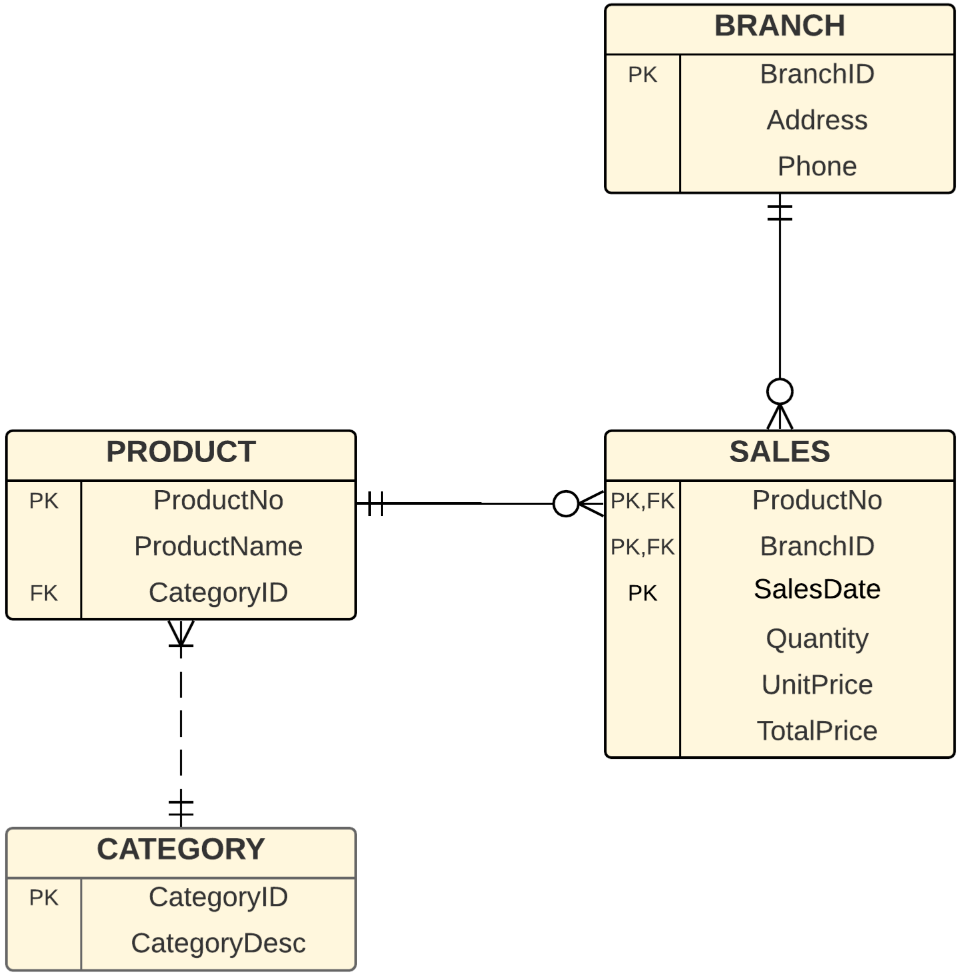
Non-Associative Relationship (1-m)



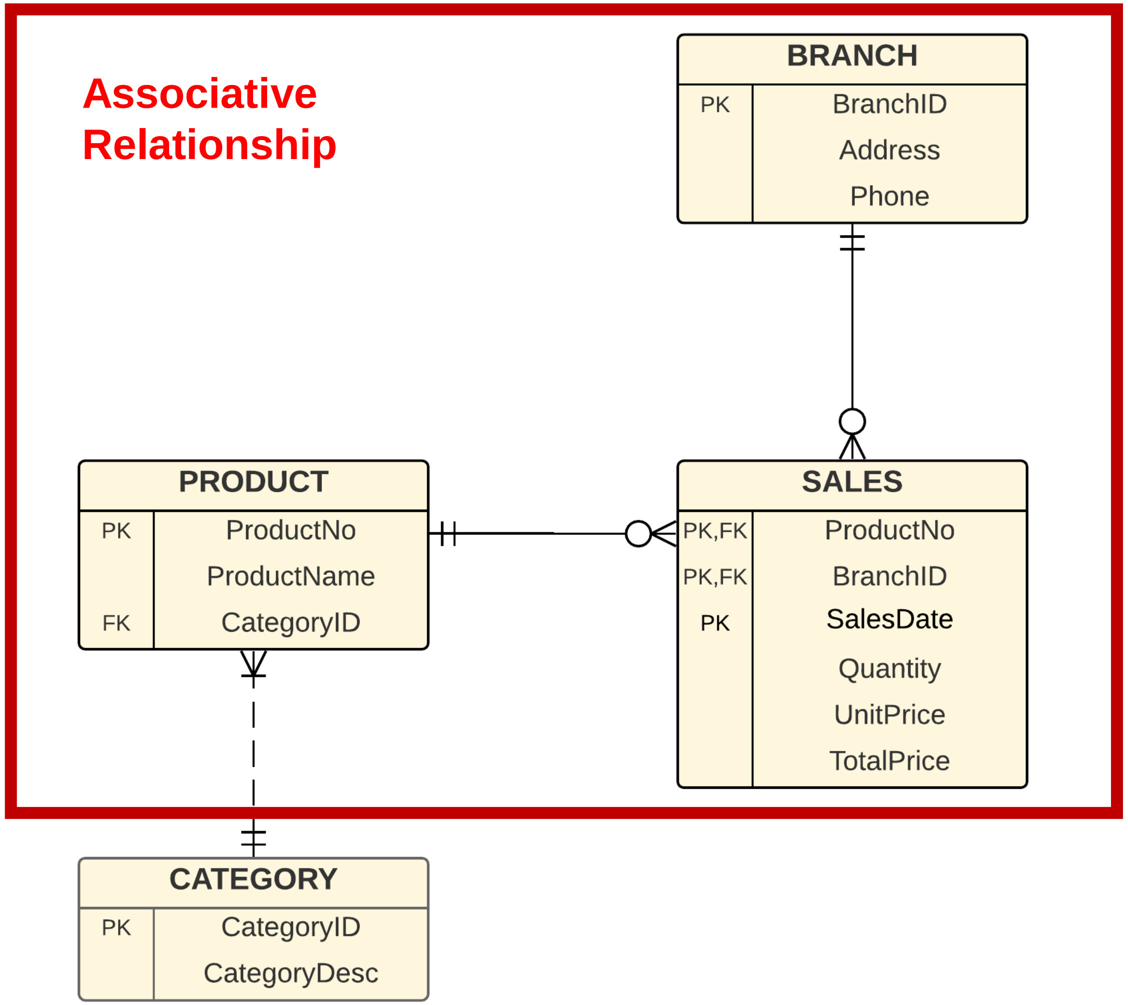
	ENTITY1
PK	PrimaryKey
	OtherAttribute
	OtherAttribute
	...
FK	ForeignKey

	ENTITY2
PK	PrimaryKey
	OtherAttribute
	...
	...
	...

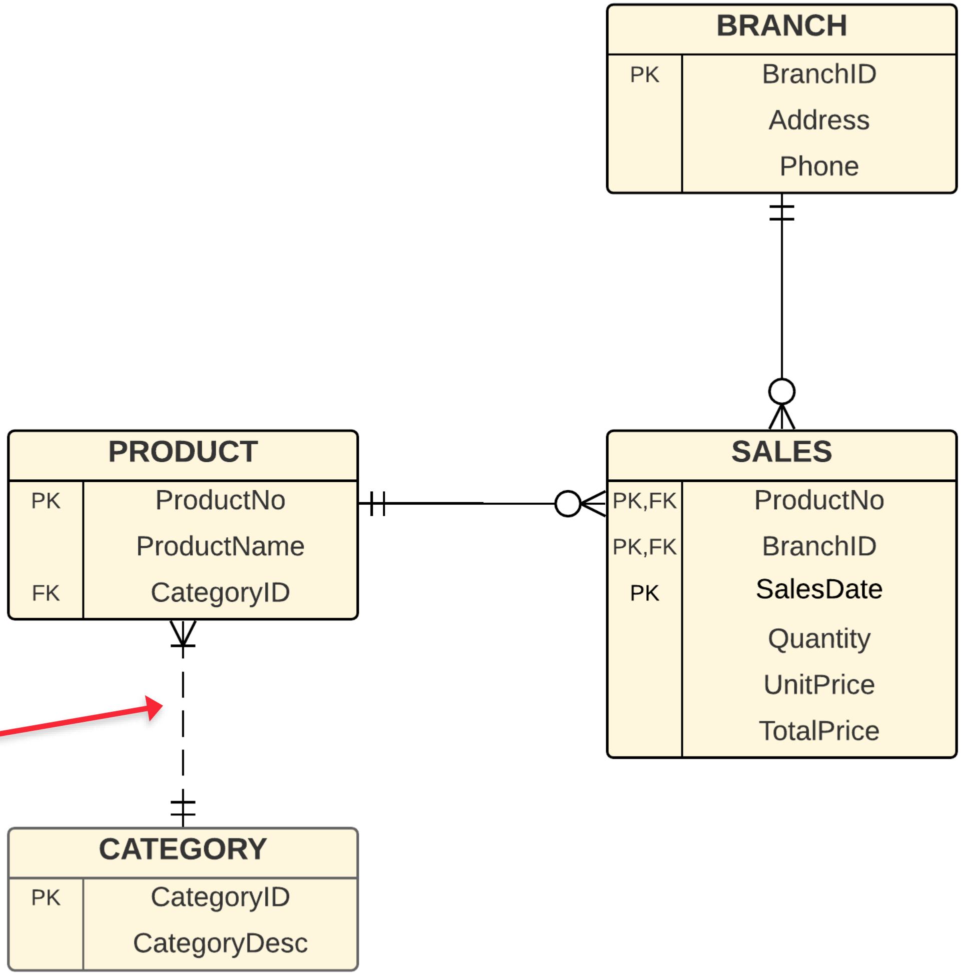
E/R Diagram Notation



E/R Diagram Notation



E/R Diagram Notation



BRANCH	
PK	BranchID
	Address
	Phone

PRODUCT	
PK	ProductNo
	ProductName
FK	CategoryID

SALES	
PK,FK	ProductNo
PK,FK	BranchID
PK	SalesDate
	Quantity
	UnitPrice
	TotalPrice

CATEGORY	
PK	CategoryID
	CategoryDesc

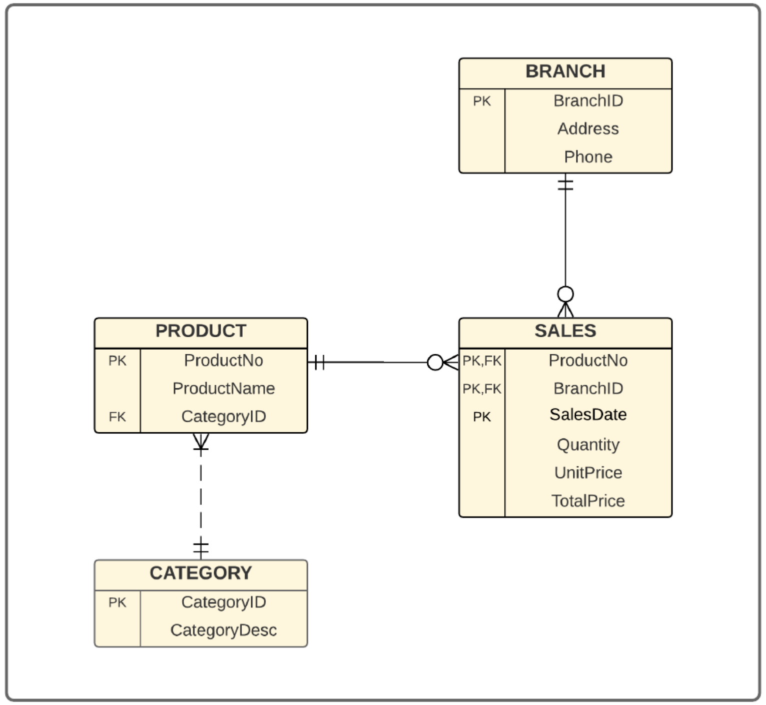
Non-Associative Relationship

Transformation Process

MONASH
University

Transformation Process

Operational Database (E/R Diagram)



BRANCH	
PK	BranchID
	Address
	Phone

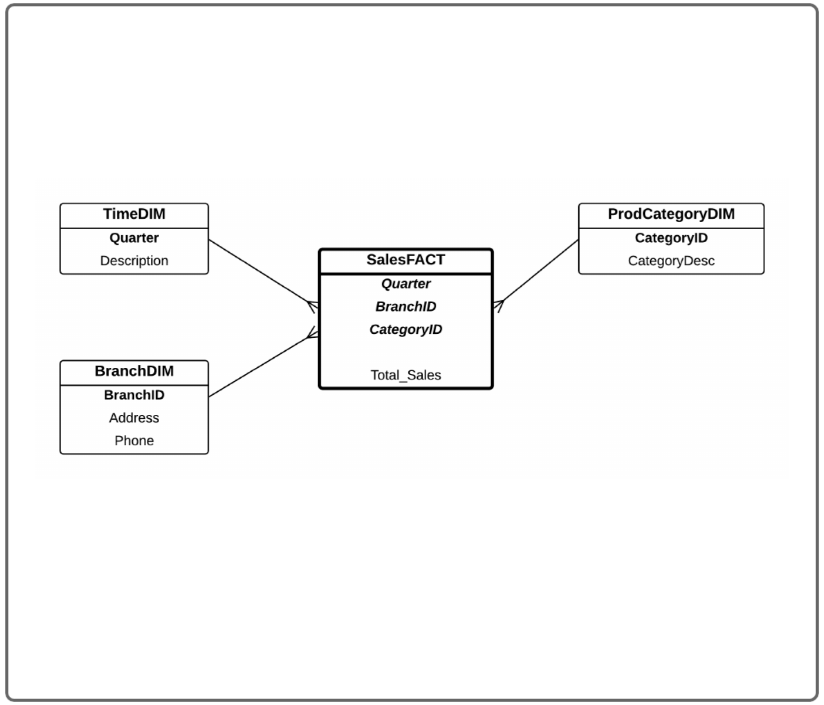
PRODUCT	
PK	ProductNo
	ProductName
FK	CategoryID

SALES	
PK,FK	ProductNo
PK,FK	BranchID
PK	SalesDate
	Quantity
	UnitPrice
	TotalPrice

CATEGORY	
PK	CategoryID
	CategoryDesc

Transformation (ETL)

Data Warehouse (Star Schema)



TimeDIM
Quarter
Description

BranchDIM
BranchID
Address
Phone

SalesFACT
Quarter
BranchID
CategoryID
Total_Sales

ProdCategoryDIM
CategoryID
CategoryDesc

Transformation Process
Case Study #1

MONASH University

Case Study #1: International College

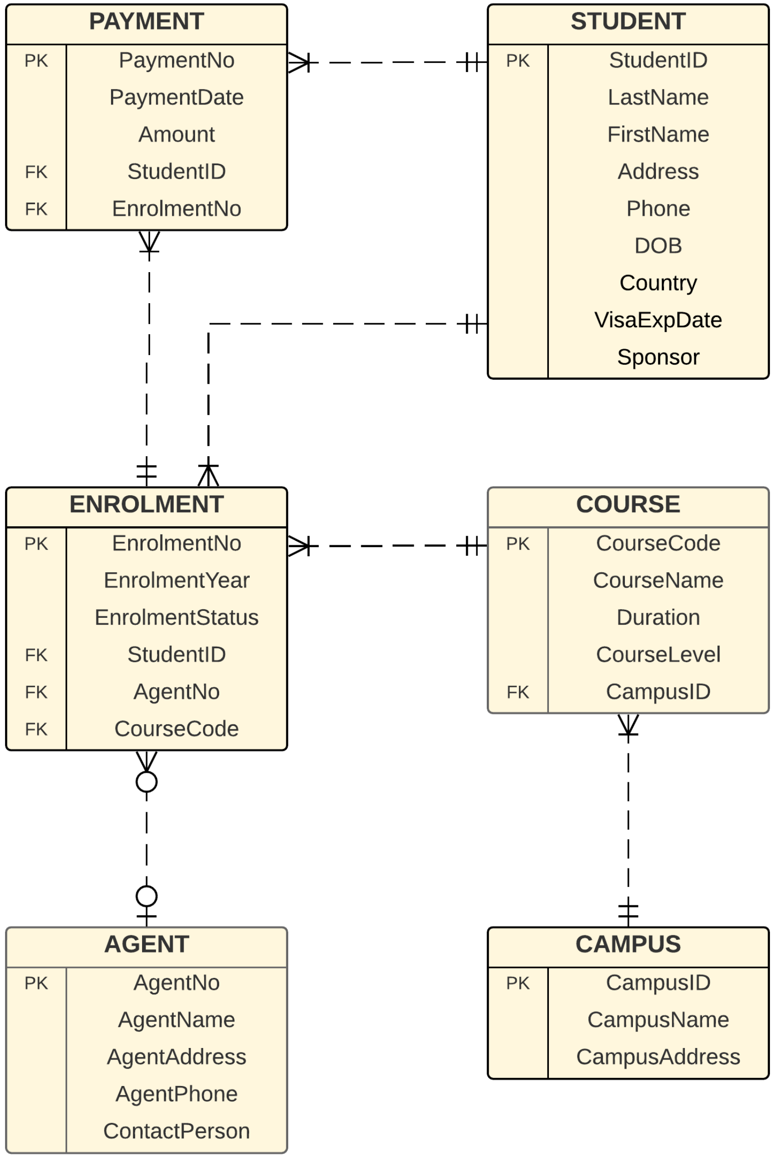
The admission office handles enrolment, payment, and marketing campaigns to international students, often through educational agents located overseas. This admission office has an operational system that maintains all the details of international students enrolled in the College. Payment details are also handled by this office. Basically, the operational system has the following features:

- Every student details are kept in the database. This includes the courses that the students enroll.
- As the College is a multi-campus university, some courses are offered in a different campus. The admission office handles international students of all campuses.
- Some international students coming to the College are handled by an educational agent. This is particularly common for the first course that a student enrolls in. Subsequent courses are not normally handled by an agent, because the students themselves deal directly with the College.
- International students pay tuition fees several times (normally once every semester) for each course they are doing.

Case Study #1: International College

The College now requires a data warehouse for analysis purposes. The analysis is needed for identifying at least the following questions:

1. How many students come from certain countries?
2. What is the total income for certain postgraduate courses?
3. How many students are handled by certain agents?
4. How the number of enrolment of courses fluctuates across the year?



	PAYMENT
PK	PaymentNo
	PaymentDate
	Amount
FK	StudentID
FK	EnrolmentNo

	STUDENT
PK	StudentID
	LastName
	FirstName
	Address
	Phone
	DOB
	Country
	VisaExpDate
	Sponsor

	ENROLMENT
PK	EnrolmentNo
	EnrolmentYear
	EnrolmentStatus
FK	StudentID
FK	AgentNo
FK	CourseCode

	COURSE
PK	CourseCode
	CourseName
	Duration
	CourseLevel
FK	CampusID

	AGENT
PK	AgentNo
	AgentName
	AgentAddress
	AgentPhone
	ContactPerson

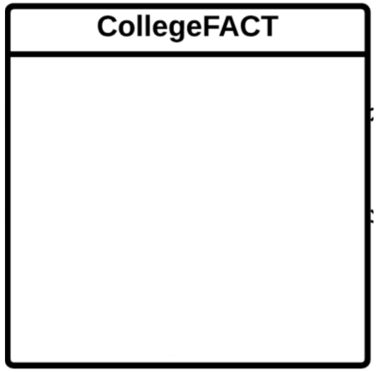
	CAMPUS
PK	CampusID
	CampusName
	CampusAddress

Case Study #1: International College

- College Star Schema
  - **Fact**:
    - Number of Students
    - Total Income
  - **Dimensions**:
    - Country
    - Agent
    - Course
    - Year

Case Study #1: International College

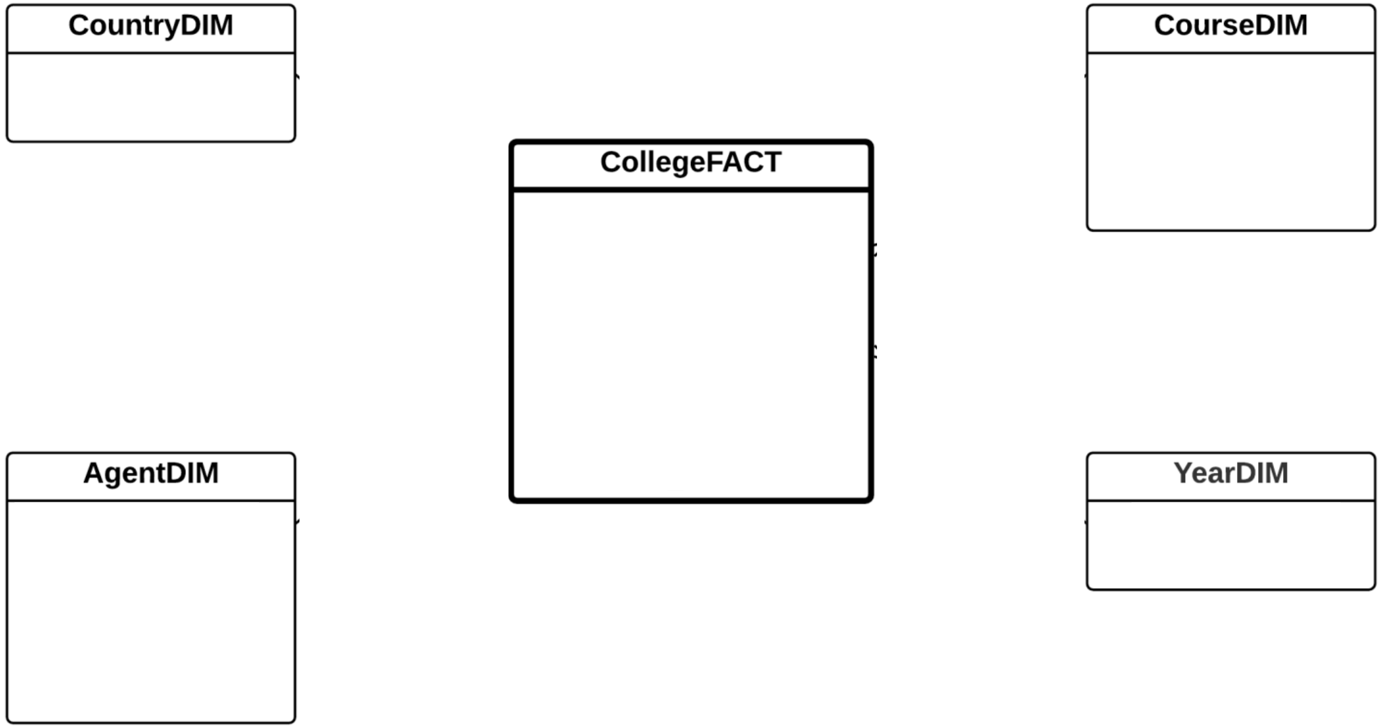
- College Star Schema
  - **Fact**:
    - Number of Students
    - Total Income
  - **Dimensions**:
    - Country
    - Agent
    - Course
    - Year



CollegeFACT

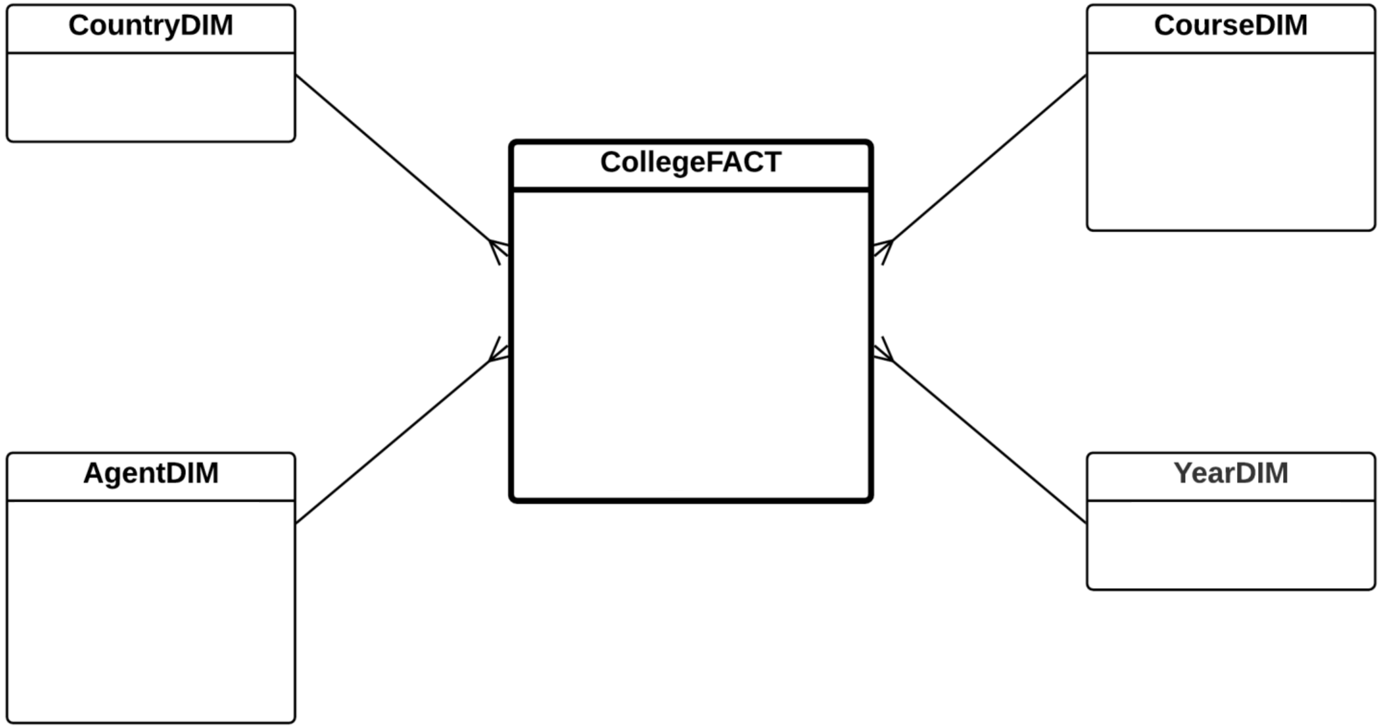
Case Study #1: International College

- College Star Schema
  - **Fact**:
    - Number of Students
    - Total Income
  - **Dimensions**:
    - Country
    - Agent
    - Course
    - Year



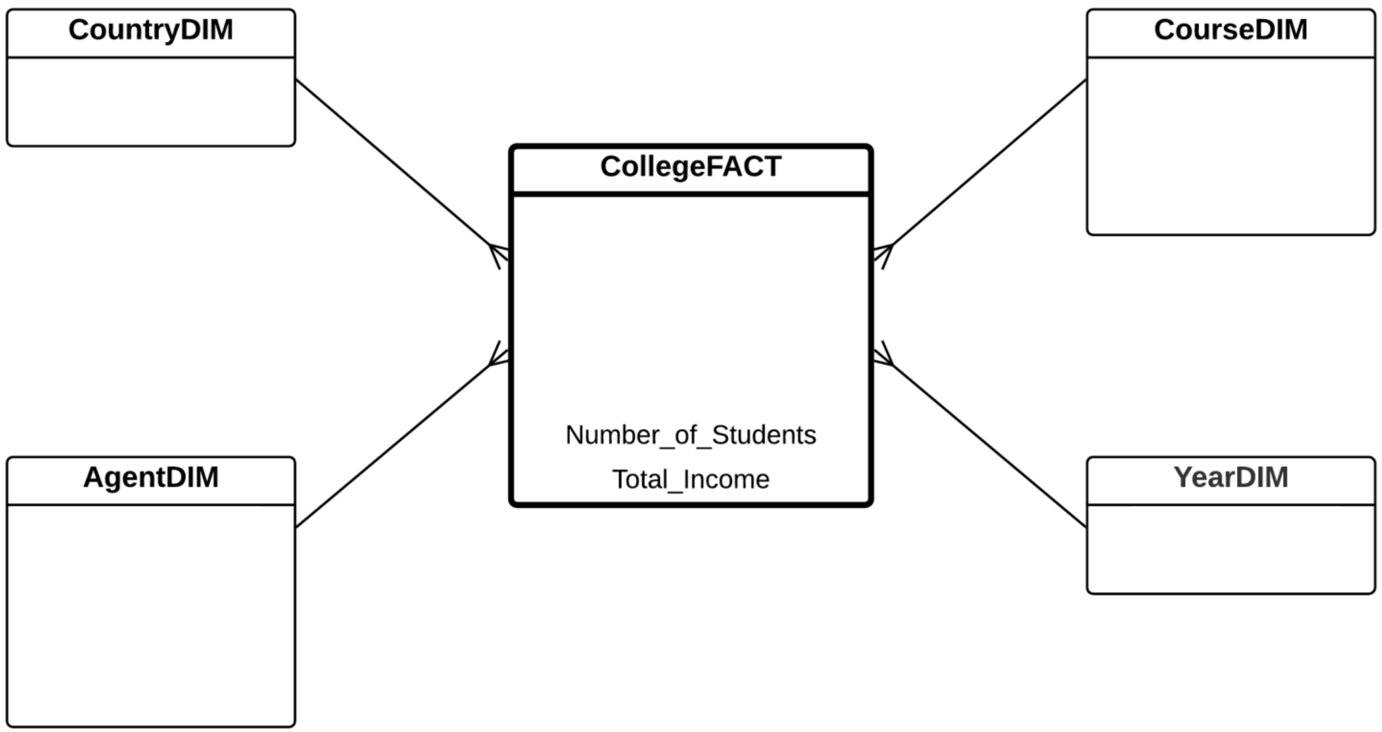
Case Study #1: International College

- College Star Schema
  - **Fact**:
    - Number of Students
    - Total Income
  - **Dimensions**:
    - Country
    - Agent
    - Course
    - Year

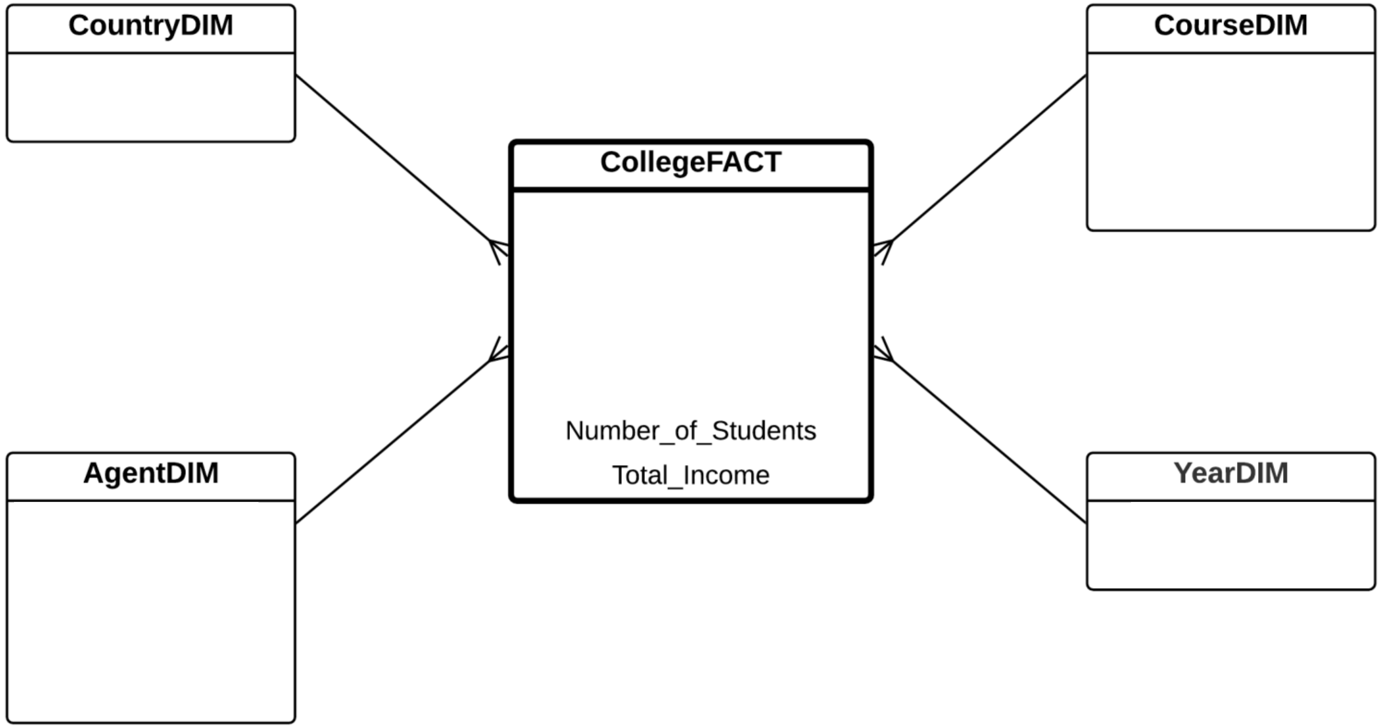


Case Study #1: International College

- College Star Schema
  - **Fact**:
    - Number of Students
    - Total Income
  - **Dimensions**:
    - Country
    - Agent
    - Course
    - Year

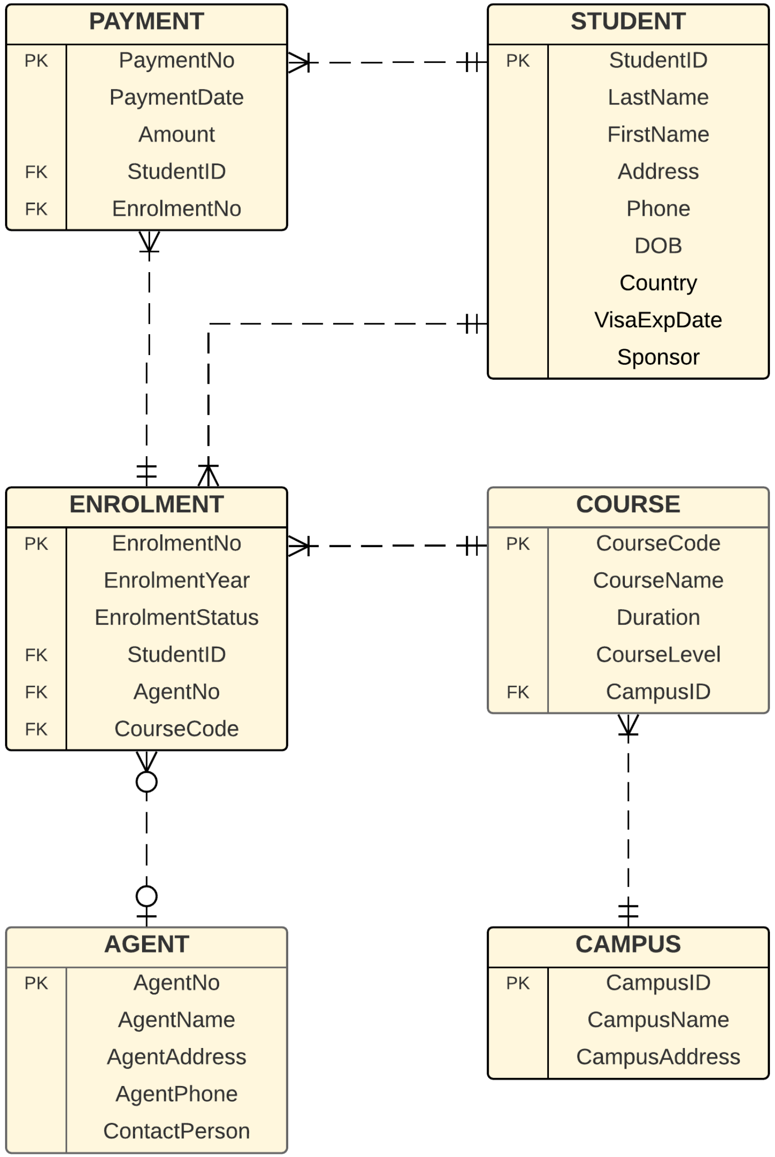


Case Study #1: International College



CollegeFACT

Number_of_Students
Total_Income



PAYMENT

PK	PaymentNo
	PaymentDate
	Amount
FK	StudentID
FK	EnrolmentNo

STUDENT

PK	StudentID
	LastName
	FirstName
	Address
	Phone
	DOB
	Country
	VisaExpDate
	Sponsor

ENROLMENT

PK	EnrolmentNo
	EnrolmentYear
	EnrolmentStatus
FK	StudentID
FK	AgentNo
FK	CourseCode

COURSE

PK	CourseCode
	CourseName
	Duration
	CourseLevel
FK	CampusID

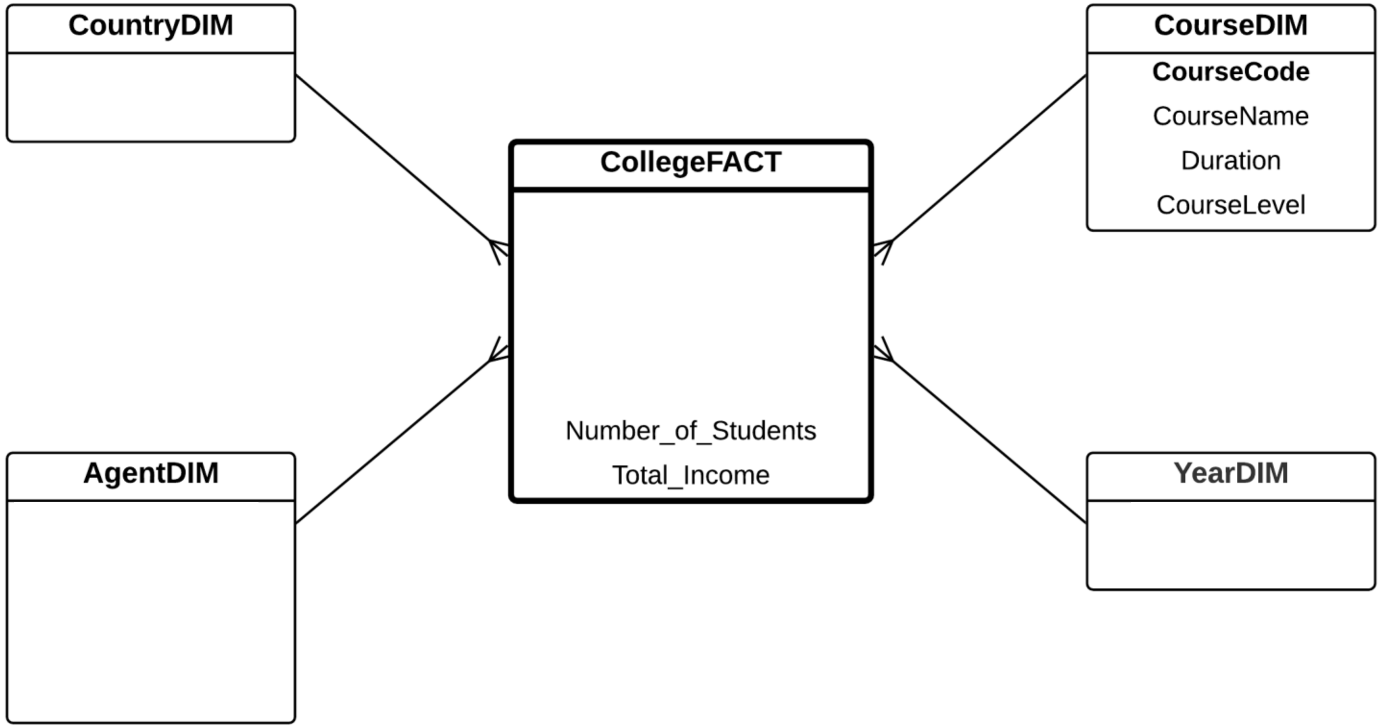
AGENT

PK	AgentNo
	AgentName
	AgentAddress
	AgentPhone
	ContactPerson

CAMPUS

PK	CampusID
	CampusName
	CampusAddress

Case Study #1: International College



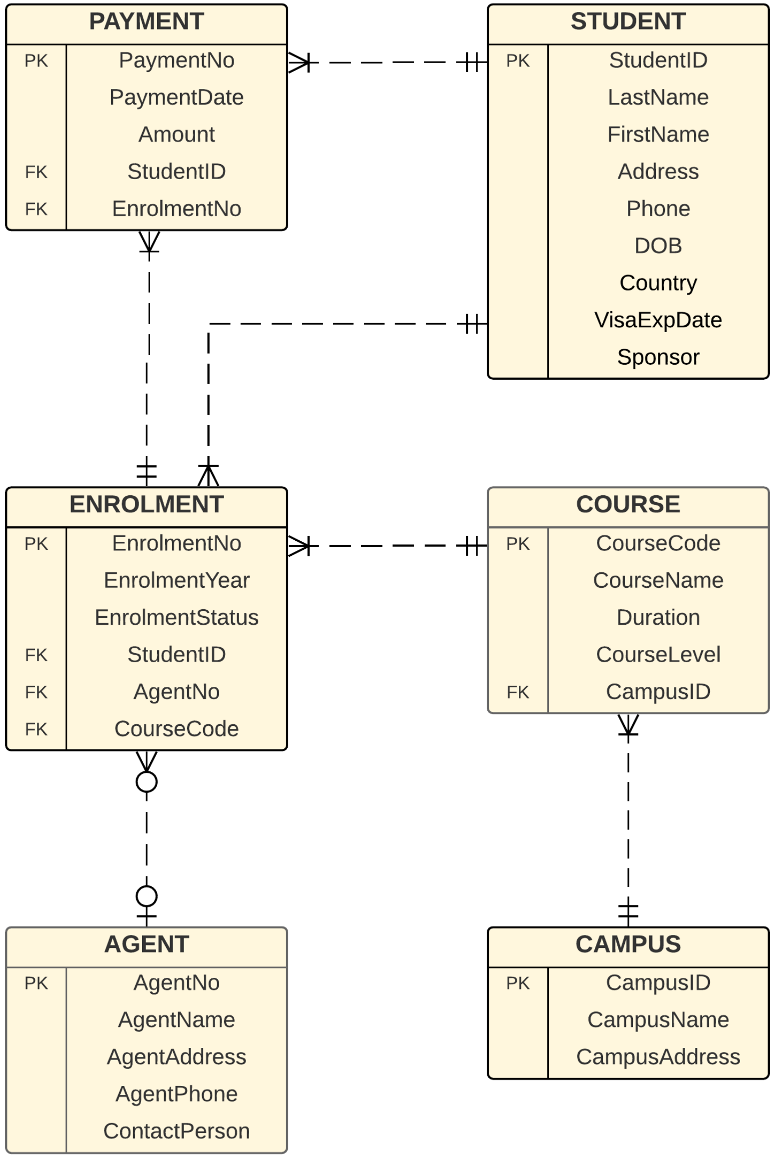
CountryDIM

CourseDIM
CourseCode
CourseName
Duration
CourseLevel

CollegeFACT
Number_of_Students
Total_Income

AgentDIM

YearDIM



PAYMENT
PK PaymentNo
PaymentDate
Amount
FK StudentID
FK EnrolmentNo

STUDENT
PK StudentID
LastName
FirstName
Address
Phone
DOB
Country
VisaExpDate
Sponsor

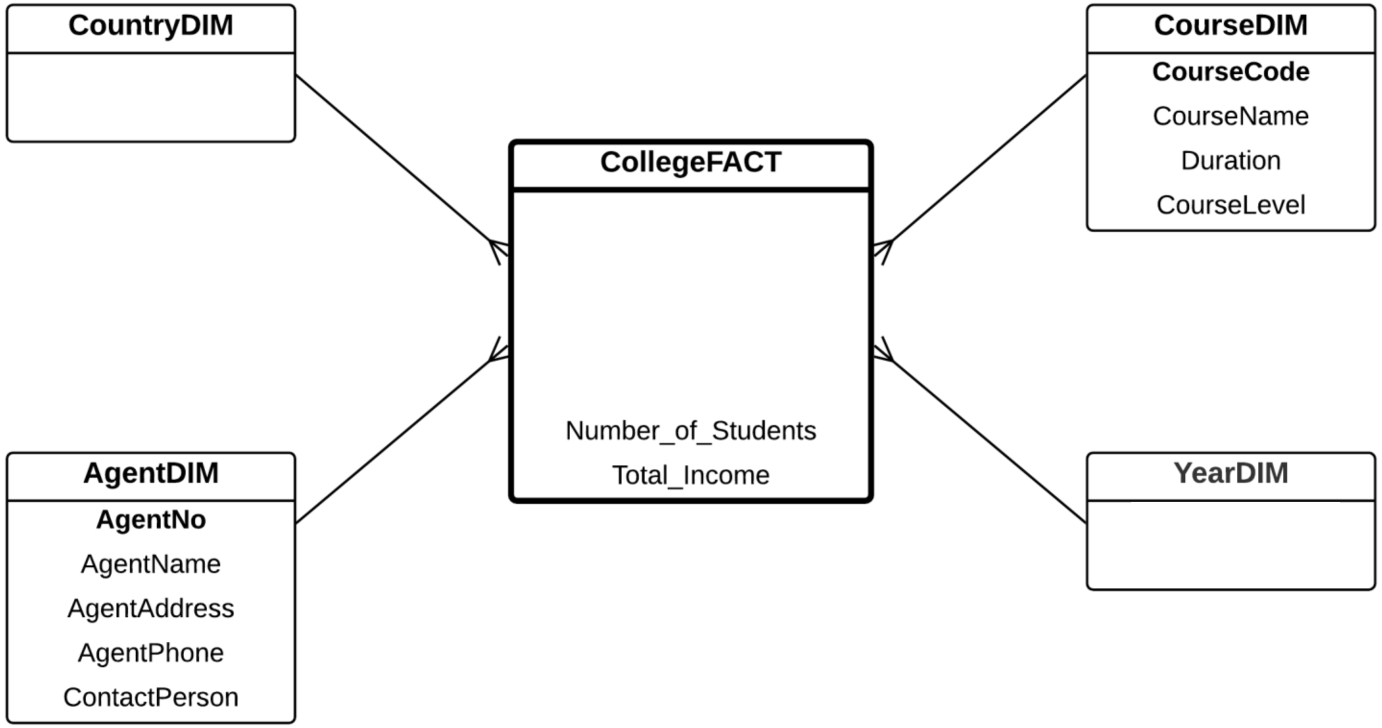
ENROLMENT
PK EnrolmentNo
EnrolmentYear
EnrolmentStatus
FK StudentID
FK AgentNo
FK CourseCode

COURSE
PK CourseCode
CourseName
Duration
CourseLevel
FK CampusID

AGENT
PK AgentNo
AgentName
AgentAddress
AgentPhone
ContactPerson

CAMPUS
PK CampusID
CampusName
CampusAddress

Case Study #1: International College



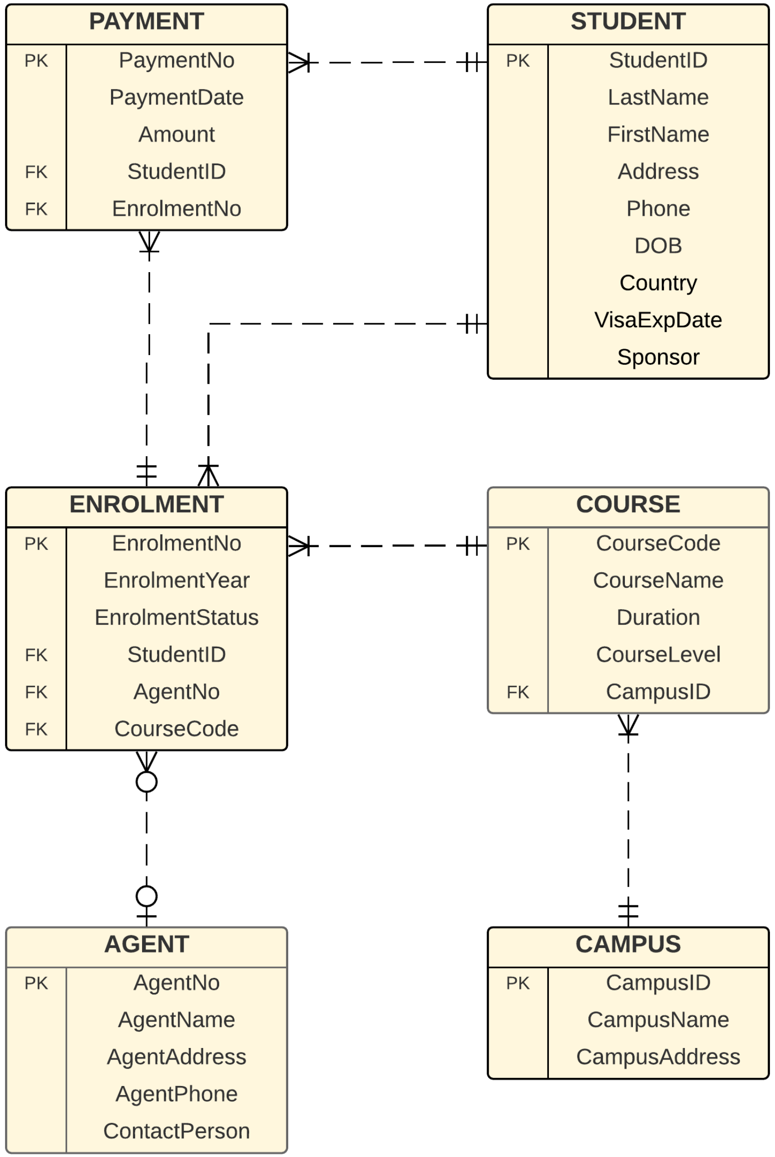
CountryDIM

CourseDIM
- **CourseCode**
- CourseName
- Duration
- CourseLevel

CollegeFACT
- Number_of_Students
- Total_Income

AgentDIM
- **AgentNo**
- AgentName
- AgentAddress
- AgentPhone
- ContactPerson

YearDIM



PAYMENT
- PK PaymentNo
- PaymentDate
- Amount
- FK StudentID
- FK EnrolmentNo

STUDENT
- PK StudentID
- LastName
- FirstName
- Address
- Phone
- DOB
- Country
- VisaExpDate
- Sponsor

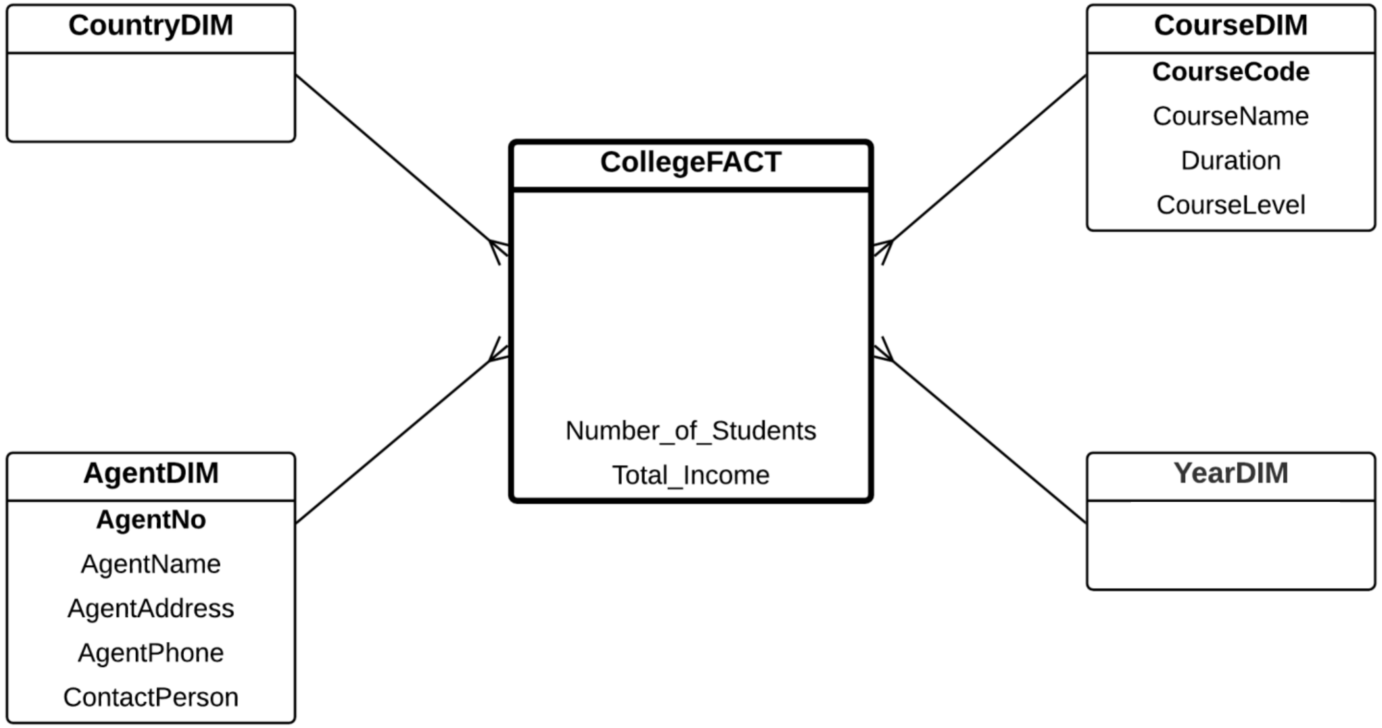
ENROLMENT
- PK EnrolmentNo
- EnrolmentYear
- EnrolmentStatus
- FK StudentID
- FK AgentNo
- FK CourseCode

COURSE
- PK CourseCode
- CourseName
- Duration
- CourseLevel
- FK CampusID

AGENT
- PK AgentNo
- AgentName
- AgentAddress
- AgentPhone
- ContactPerson

CAMPUS
- PK CampusID
- CampusName
- CampusAddress

Case Study #1: International College



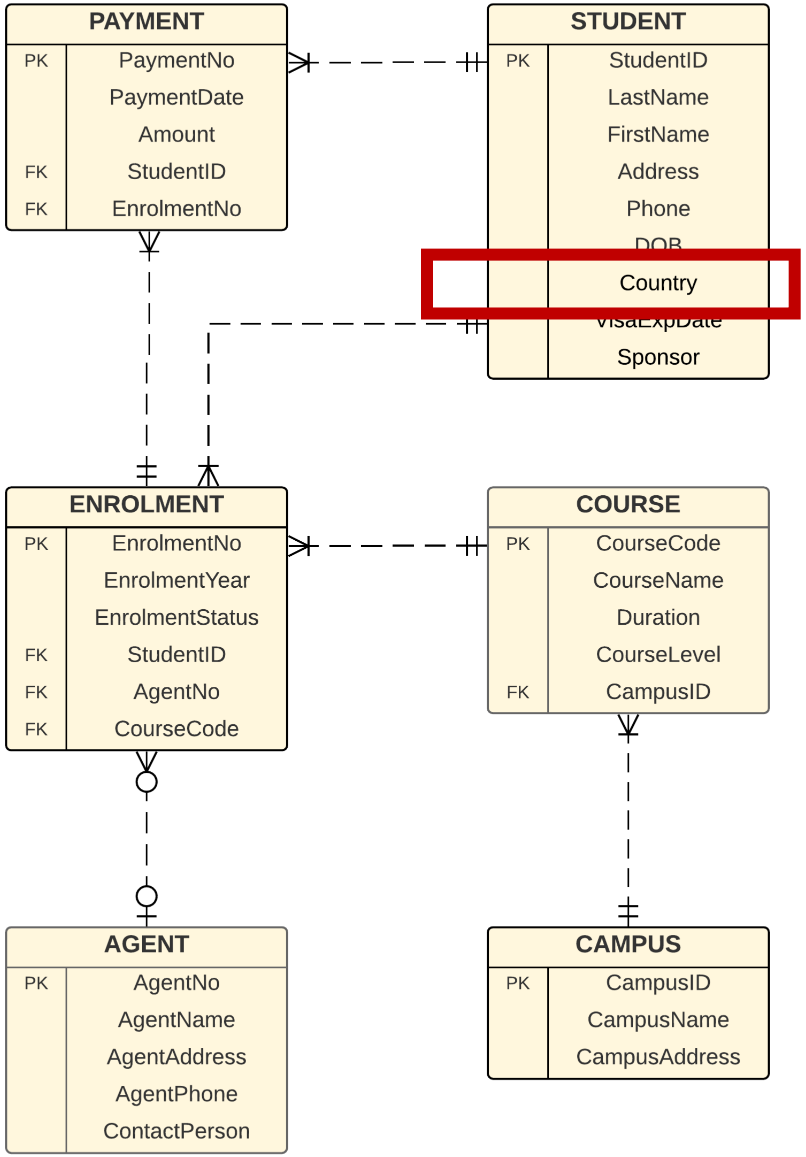
CountryDIM

CourseDIM
CourseCode
CourseName
Duration
CourseLevel

CollegeFACT
Number_of_Students
Total_Income

AgentDIM
AgentNo
AgentName
AgentAddress
AgentPhone
ContactPerson

YearDIM



PAYMENT
PK PaymentNo
PaymentDate
Amount
FK StudentID
FK EnrolmentNo

STUDENT
PK StudentID
LastName
FirstName
Address
Phone
DOB
Country
VisaExpDate
Sponsor

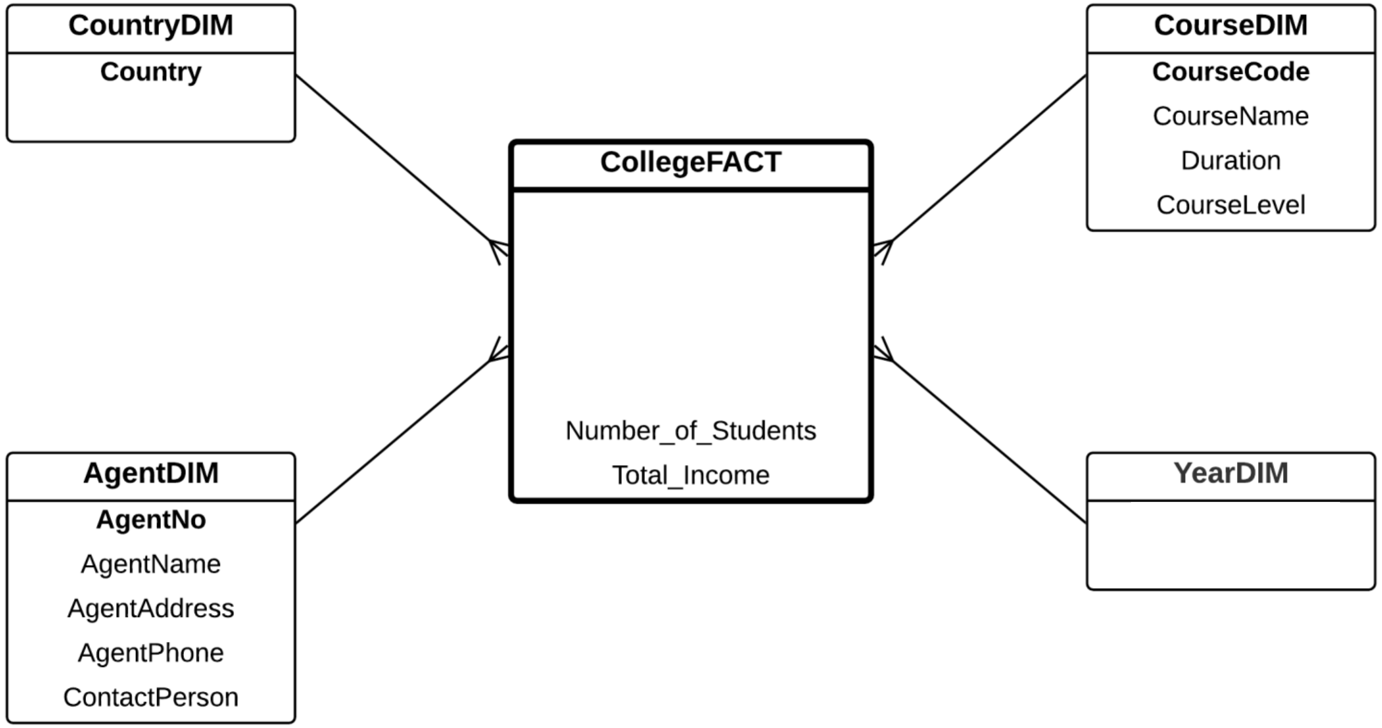
ENROLMENT
PK EnrolmentNo
EnrolmentYear
EnrolmentStatus
FK StudentID
FK AgentNo
FK CourseCode

COURSE
PK CourseCode
CourseName
Duration
CourseLevel
FK CampusID

AGENT
PK AgentNo
AgentName
AgentAddress
AgentPhone
ContactPerson

CAMPUS
PK CampusID
CampusName
CampusAddress

Case Study #1: International College



CountryDIM

- **Country**

CourseDIM

- **CourseCode**
- CourseName
- Duration
- CourseLevel

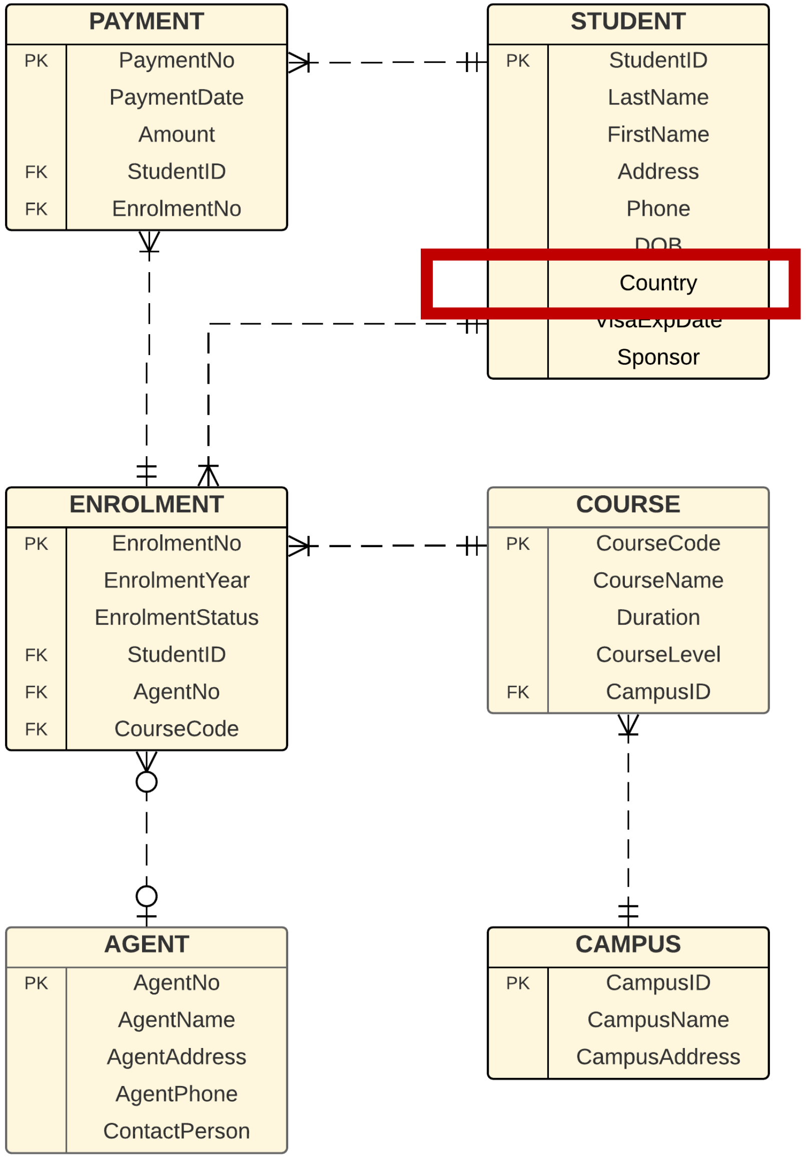
CollegeFACT

- Number_of_Students
- Total_Income

AgentDIM

- **AgentNo**
- AgentName
- AgentAddress
- AgentPhone
- ContactPerson

YearDIM



PAYMENT	
PK	PaymentNo
	PaymentDate
	Amount
FK	StudentID
FK	EnrolmentNo

STUDENT	
PK	StudentID
	LastName
	FirstName
	Address
	Phone
	DOB
	Country
	VisaExpDate
	Sponsor

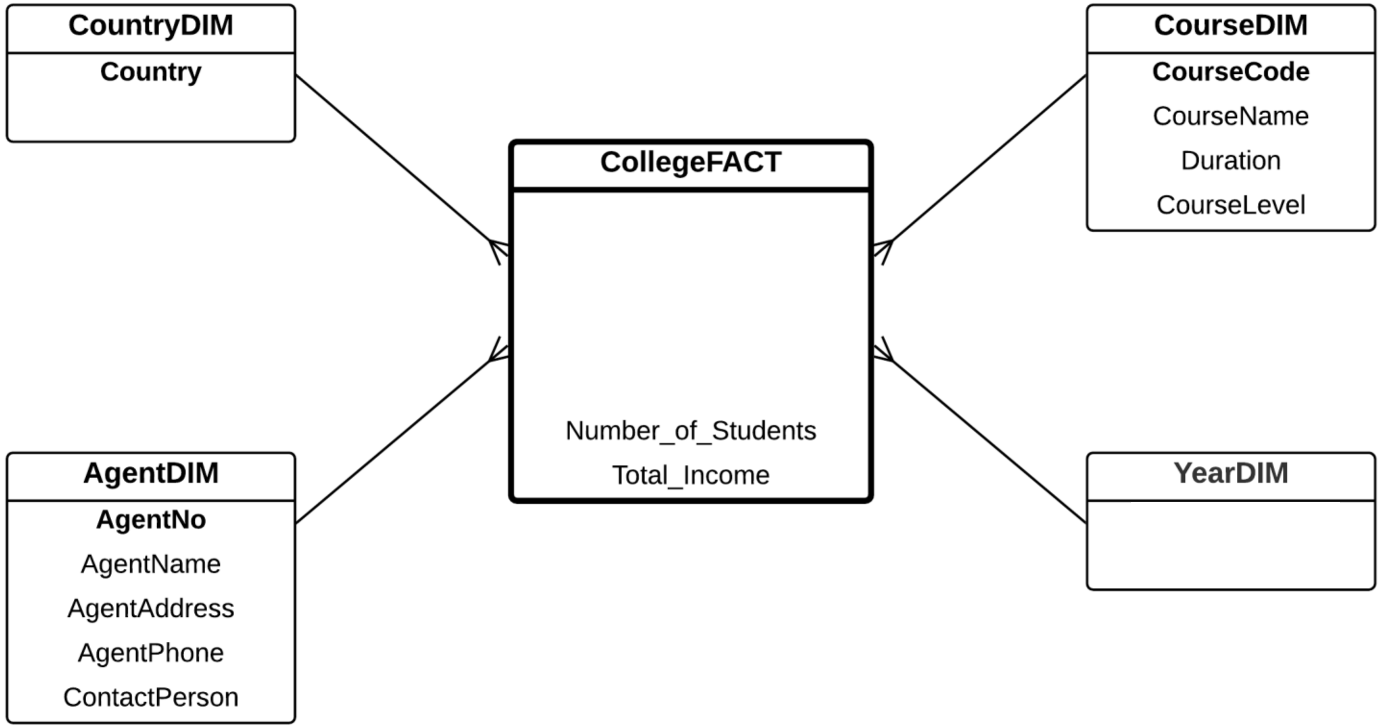
ENROLMENT	
PK	EnrolmentNo
	EnrolmentYear
	EnrolmentStatus
FK	StudentID
FK	AgentNo
FK	CourseCode

COURSE	
PK	CourseCode
	CourseName
	Duration
	CourseLevel
FK	CampusID

AGENT	
PK	AgentNo
	AgentName
	AgentAddress
	AgentPhone
	ContactPerson

CAMPUS	
PK	CampusID
	CampusName
	CampusAddress

Case Study #1: International College



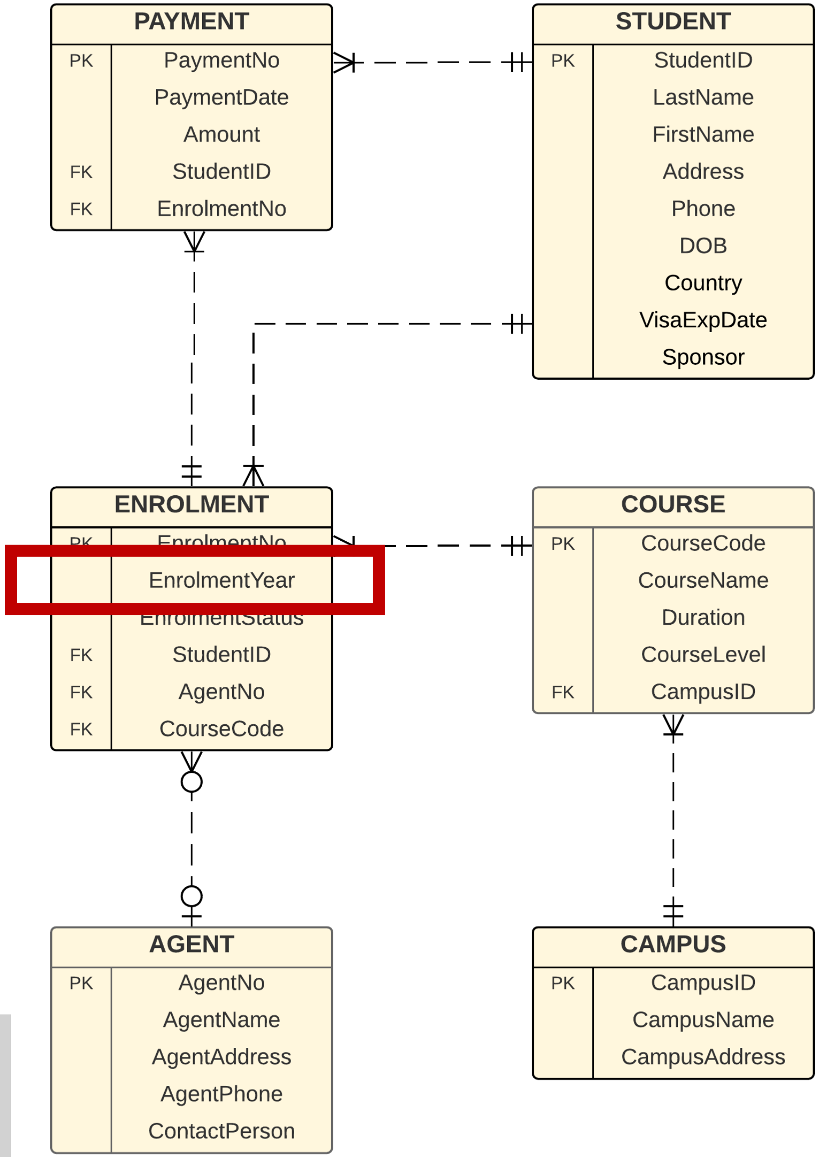
CountryDIM
Country

CourseDIM
CourseCode
CourseName
Duration
CourseLevel

CollegeFACT
Number_of_Students
Total_Income

AgentDIM
AgentNo
AgentName
AgentAddress
AgentPhone
ContactPerson

YearDIM



PAYMENT
PK PaymentNo
PaymentDate
Amount
FK StudentID
FK EnrolmentNo

STUDENT
PK StudentID
LastName
FirstName
Address
Phone
DOB
Country
VisaExpDate
Sponsor

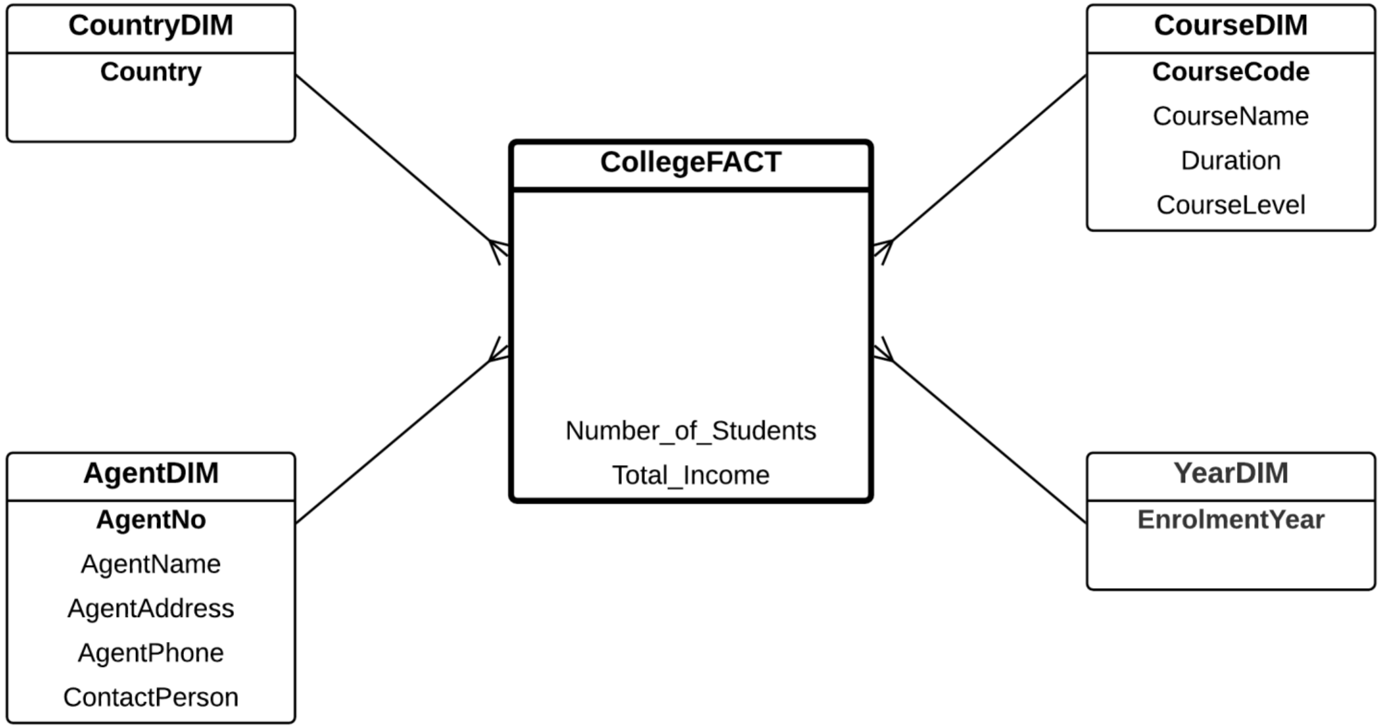
ENROLMENT
PK EnrolmentNo
EnrolmentYear
EnrolmentStatus
FK StudentID
FK AgentNo
FK CourseCode

COURSE
PK CourseCode
CourseName
Duration
CourseLevel
FK CampusID

AGENT
PK AgentNo
AgentName
AgentAddress
AgentPhone
ContactPerson

CAMPUS
PK CampusID
CampusName
CampusAddress

Case Study #1: International College



CountryDIM

Country

CollegeFACT

Number_of_Students

Total_Income

CourseDIM

CourseCode

CourseName

Duration

CourseLevel

AgentDIM

AgentNo

AgentName

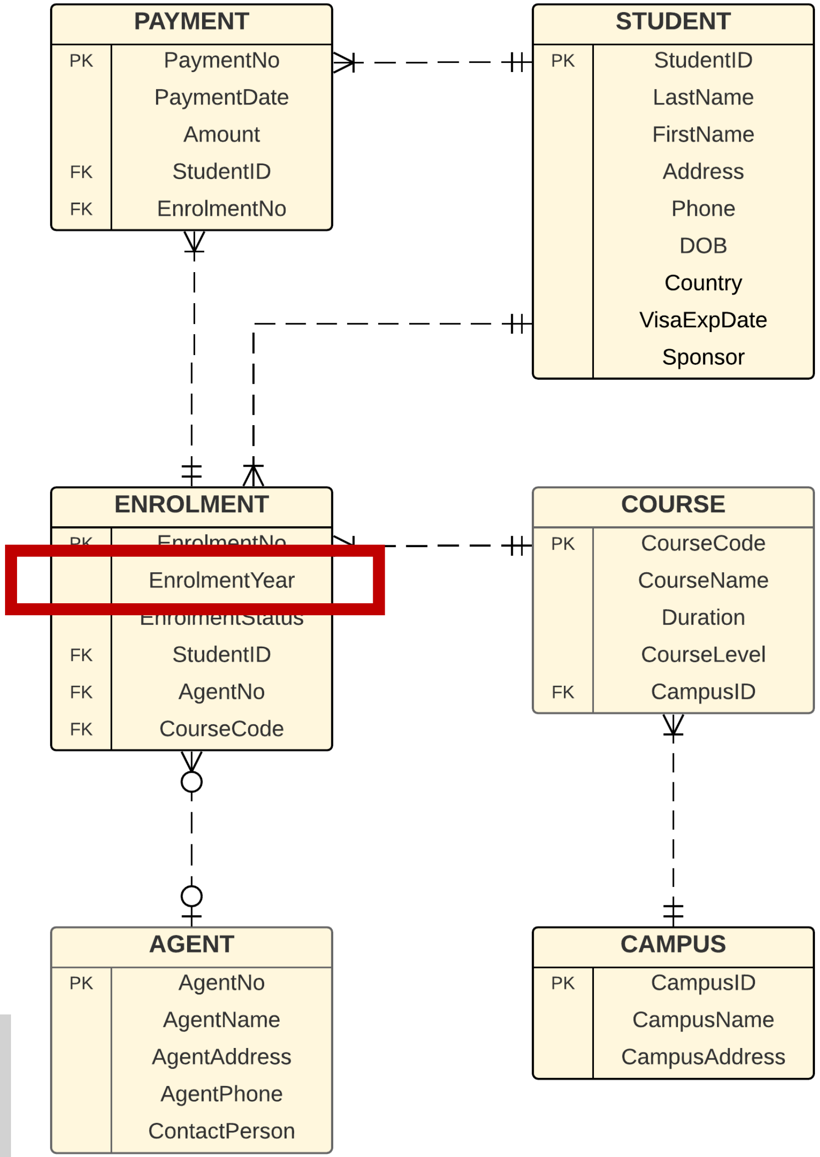
AgentAddress

AgentPhone

ContactPerson

YearDIM

EnrolmentYear



PAYMENT

PK PaymentNo

PaymentDate

Amount

FK StudentID

FK EnrolmentNo

STUDENT

PK StudentID

LastName

FirstName

Address

Phone

DOB

Country

VisaExpDate

Sponsor

ENROLMENT

PK EnrolmentNo

EnrolmentYear

EnrolmentStatus

FK StudentID

FK AgentNo

FK CourseCode

COURSE

PK CourseCode

CourseName

Duration

CourseLevel

FK CampusID

AGENT

PK AgentNo

AgentName

AgentAddress

AgentPhone

ContactPerson

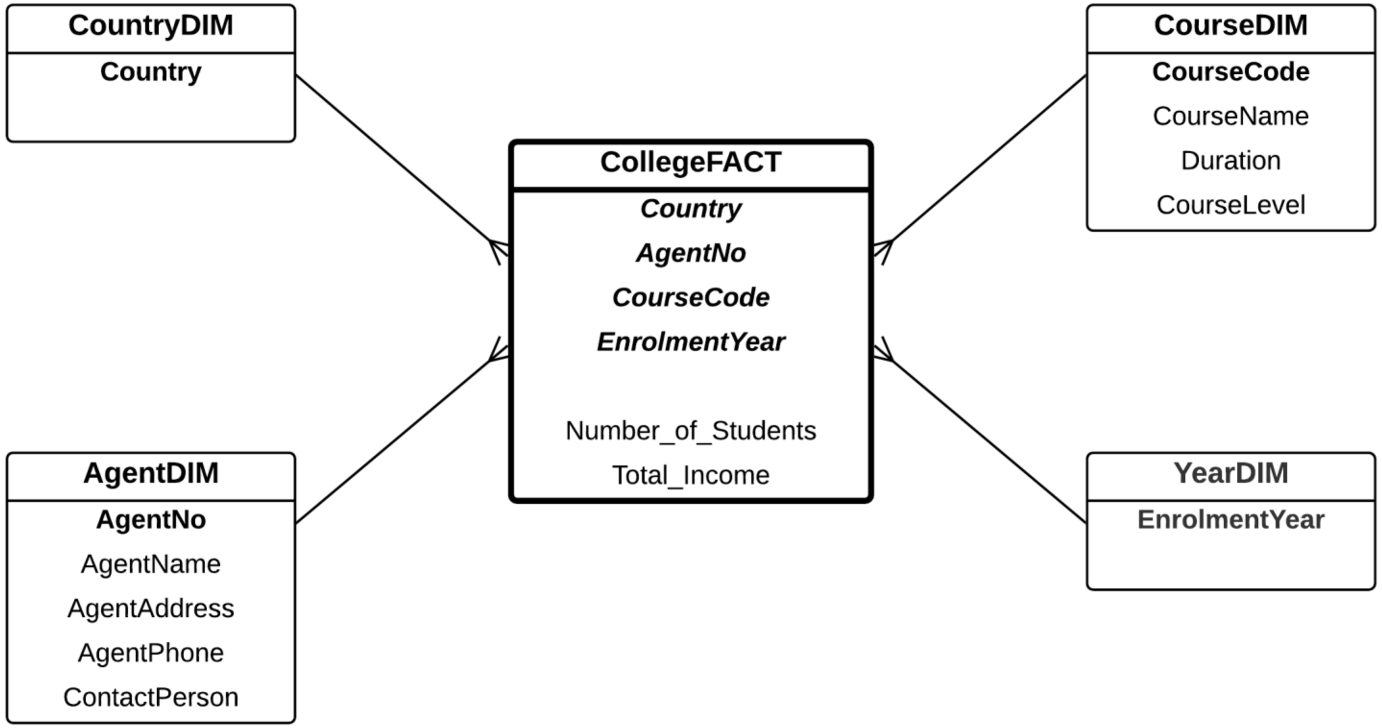
CAMPUS

PK CampusID

CampusName

CampusAddress

Case Study #1: International College



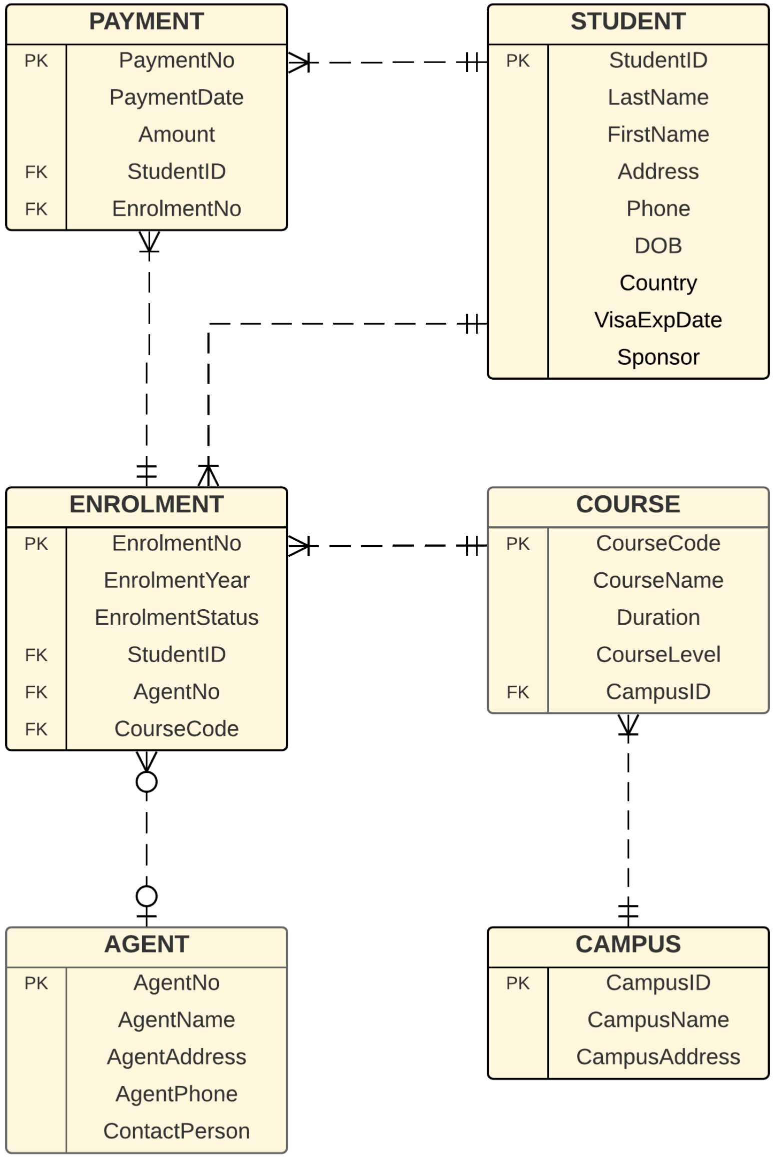
CountryDIM
- **Country**

CollegeFACT
- ***Country***
- ***AgentNo***
- ***CourseCode***
- ***EnrolmentYear***
- Number_of_Students
- Total_Income

CourseDIM
- **CourseCode**
- CourseName
- Duration
- CourseLevel

AgentDIM
- **AgentNo**
- AgentName
- AgentAddress
- AgentPhone
- ContactPerson

YearDIM
- **EnrolmentYear**



PAYMENT

PK	PaymentNo
	PaymentDate
	Amount
FK	StudentID
FK	EnrolmentNo

STUDENT

PK	StudentID
	LastName
	FirstName
	Address
	Phone
	DOB
	Country
	VisaExpDate
	Sponsor

ENROLMENT

PK	EnrolmentNo
	EnrolmentYear
	EnrolmentStatus
FK	StudentID
FK	AgentNo
FK	CourseCode

COURSE

PK	CourseCode
	CourseName
	Duration
	CourseLevel
FK	CampusID

AGENT

PK	AgentNo
	AgentName
	AgentAddress
	AgentPhone
	ContactPerson

CAMPUS

PK	CampusID
	CampusName
	CampusAddress

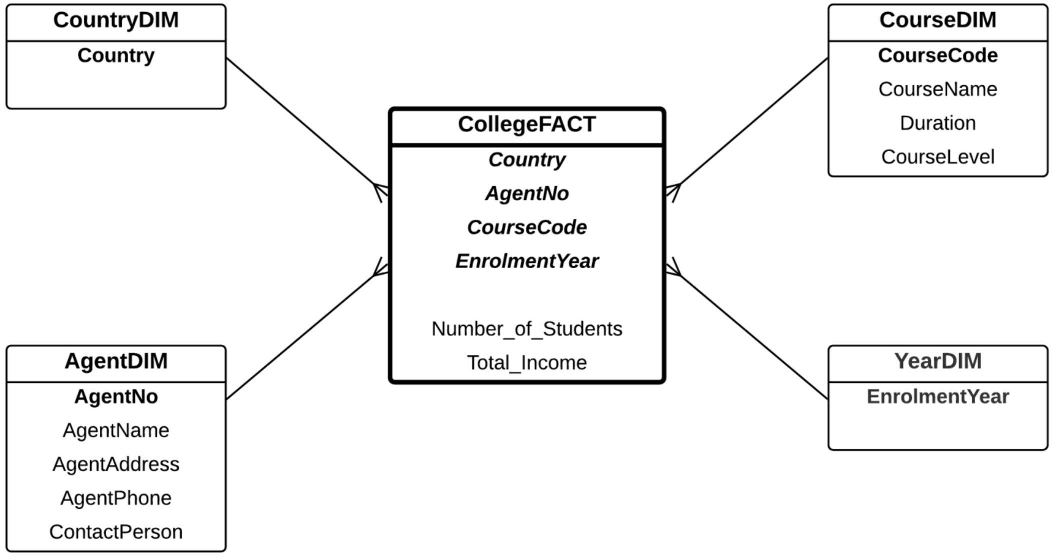
Case Study #1: International College

- To create AgentDIM:

```
– create table AgentDim as
  select * from Agent;
```

- To create CountryDIM:

```
– create table CountryDim as
  select distinct Country
  from Student;
```



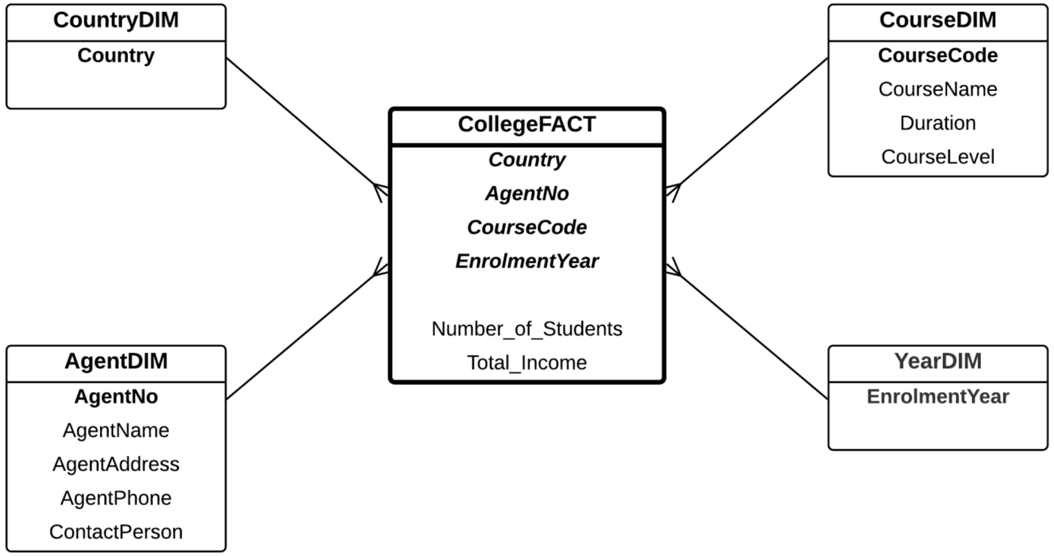
Case Study #1: International College

- To create CourseDIM:

```
- create table CourseDim as
  select CourseCode, CourseName, Duration, CourseLevel
  from Course;
```

- To create YearDIM:

```
- create table YearDim as
  select distinct EnrolmentYear
  from Enrolment;
```



Case Study #1: International College

- To create CollegeFACT:

```
- create table CollegeFact as
  Select S.Country, E.AgentNo, E.CourseCode, E.EnrolmentYear,
  count(S.StudentID) as Number_of_Students,
  sum(P.Amount) as Total_Income
  from Student S, Enrolment E, Payment P
  where E.EnrolmentNo = P.EnrolmentNo
  and E.StudentID = S.StudentID
  group by S.Country, E.AgentNo, E.CourseCode, E.EnrolmentYear;
```

Case Study #1: International College

- To create CollegeFACT:

  -
    ```
    create table CollegeFact as
    Select S.Country, E.AgentNo, E.CourseCode, E.EnrolmentYear,
    count(S.StudentID) as Number_of_Students,
    sum(P.Amount) as Total_Income
    from Student S, Enrolment E, Payment P
    where E.EnrolmentNo = P.EnrolmentNo
    and E.StudentID = S.StudentID
    group by S.Country, E.AgentNo, E.CourseCode, E.EnrolmentYear;
    ```

Case Study #1: International College

- To create CollegeFACT:

```
- create table CollegeFact as
  Select S.Country, E.AgentNo, E.CourseCode, E.EnrolmentYear,
  count(S.StudentID) as Number_of_Students,
  sum(P.Amount) as Total_Income
  from Student S, Enrolment E, Payment P
  where E.EnrolmentNo = P.EnrolmentNo
  and E.StudentID = S.StudentID
  group by S.Country, E.AgentNo, E.CourseCode, E.EnrolmentYear;
```

Case Study #1: International College

- To create CollegeFACT:

  -
    ```
    create table CollegeFact as
     Select S.Country, E.AgentNo, E.CourseCode, E.EnrolmentYear,
     count(S.StudentID) as Number_of_Students,
     sum(P.Amount) as Total_Income
     from Student S, Enrolment E, Payment P
     where E.EnrolmentNo = P.EnrolmentNo
     and E.StudentID = S.StudentID
     group by S.Country, E.AgentNo, E.CourseCode, E.EnrolmentYear;
    ```

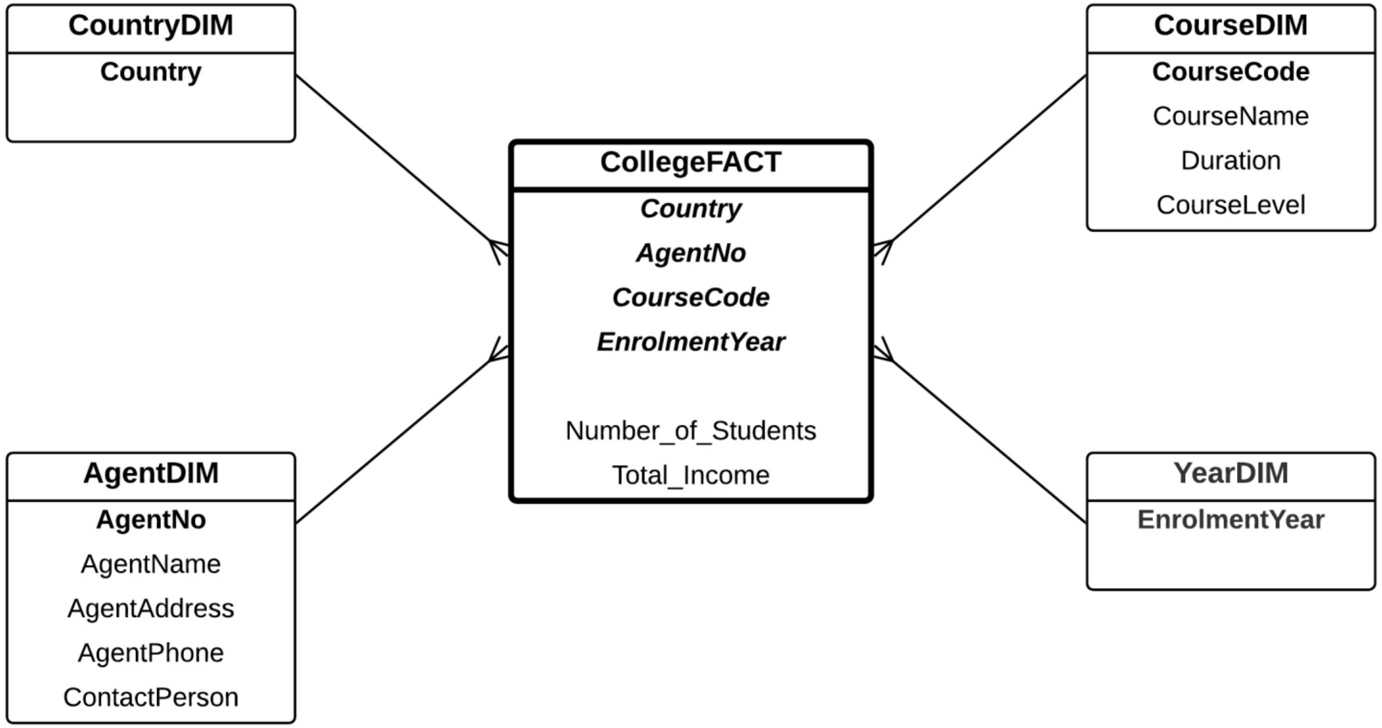
Case Study #1: International College

- To create CollegeFACT:

  -
    ```
    create table CollegeFact as
     Select S.Country, E.AgentNo, E.CourseCode, E.EnrolmentYear,
     count(S.StudentID) as Number_of_Students,
     sum(P.Amount) as Total_Income
     from Student S, Enrolment E, Payment P
     where E.EnrolmentNo = P.EnrolmentNo
     and E.StudentID = S.StudentID
     group by S.Country, E.AgentNo, E.CourseCode, E.EnrolmentYear;
    ```

Case Study #1: International College

- College Star Schema
  - **Fact**:
    - Number of Students
    - Total Income
  - **Dimensions**:
    - Country
    - Agent
    - Course
    - Year



CountryDIM
Country

CourseDIM
CourseCode
CourseName
Duration
CourseLevel

CollegeFACT
Country
AgentNo
CourseCode
EnrolmentYear
Number_of_Students
Total_Income

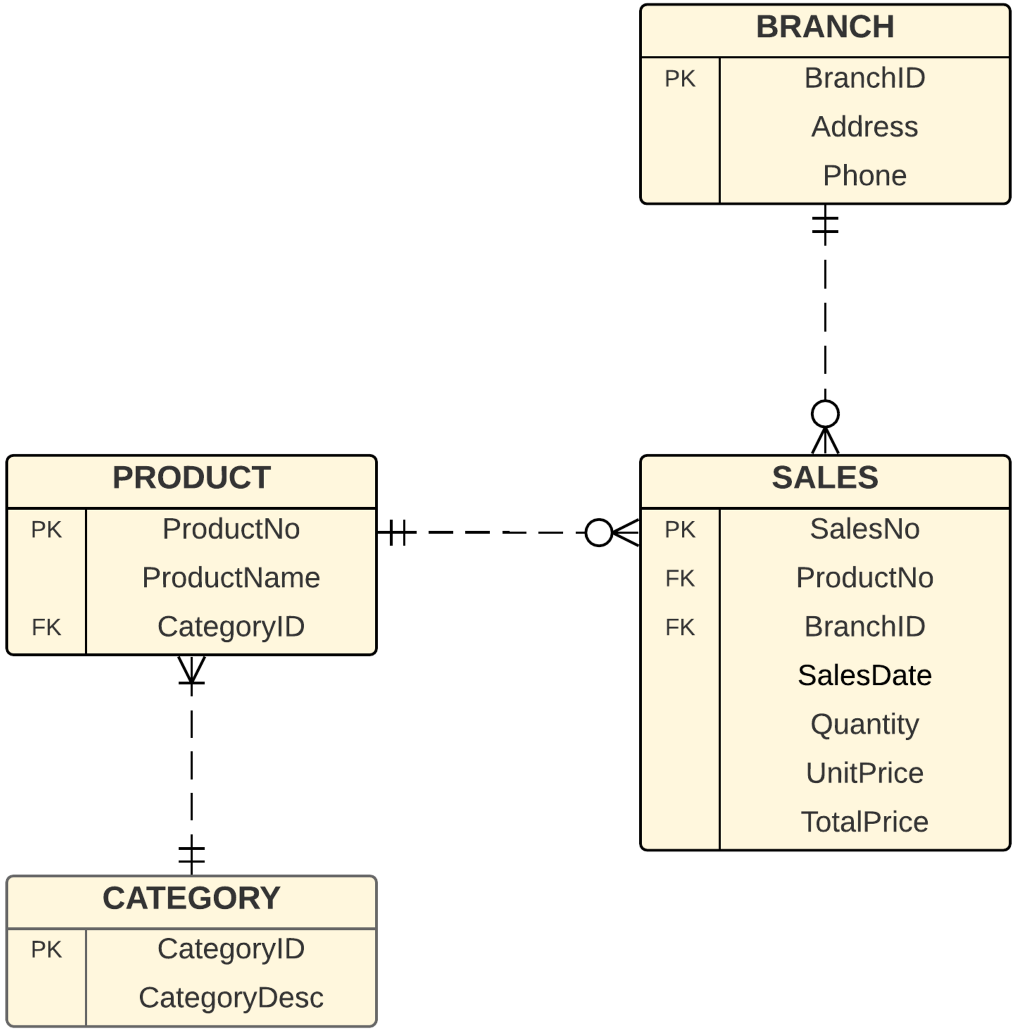
AgentDIM
AgentNo
AgentName
AgentAddress
AgentPhone
ContactPerson

YearDIM
EnrolmentYear

Transformation Process
Case Study #2

Case Study #2: Sales

- Suppose that we would like to analyze Total Sales from various point of views, such as Quarter, Branch, and Product Category.

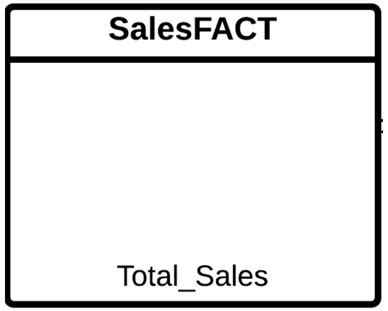


Case Study #2: Sales

- Sales Star Schema
  - **Fact**:
    - Total Sales
  - **Dimensions**:
    - Time
    - Branch
    - Product Category

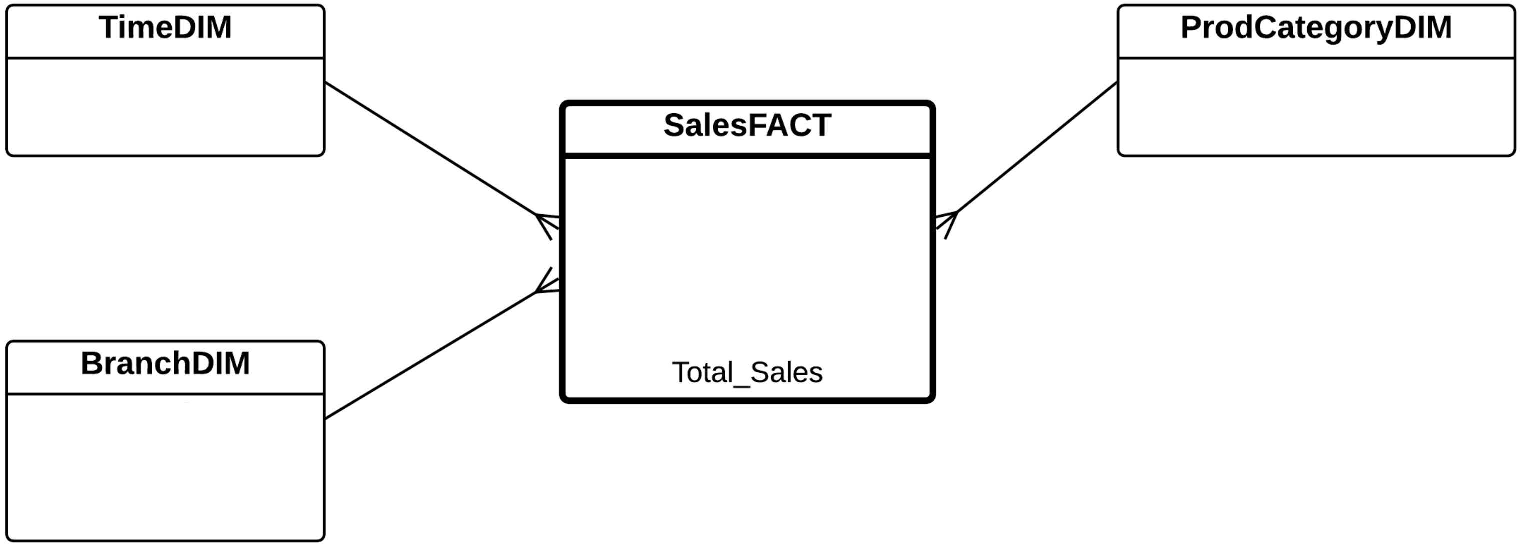
Case Study #2: Sales

- Sales Star Schema
  - **Fact**:
    - Total Sales
  - **Dimensions**:
    - Time
    - Branch
    - Product Category

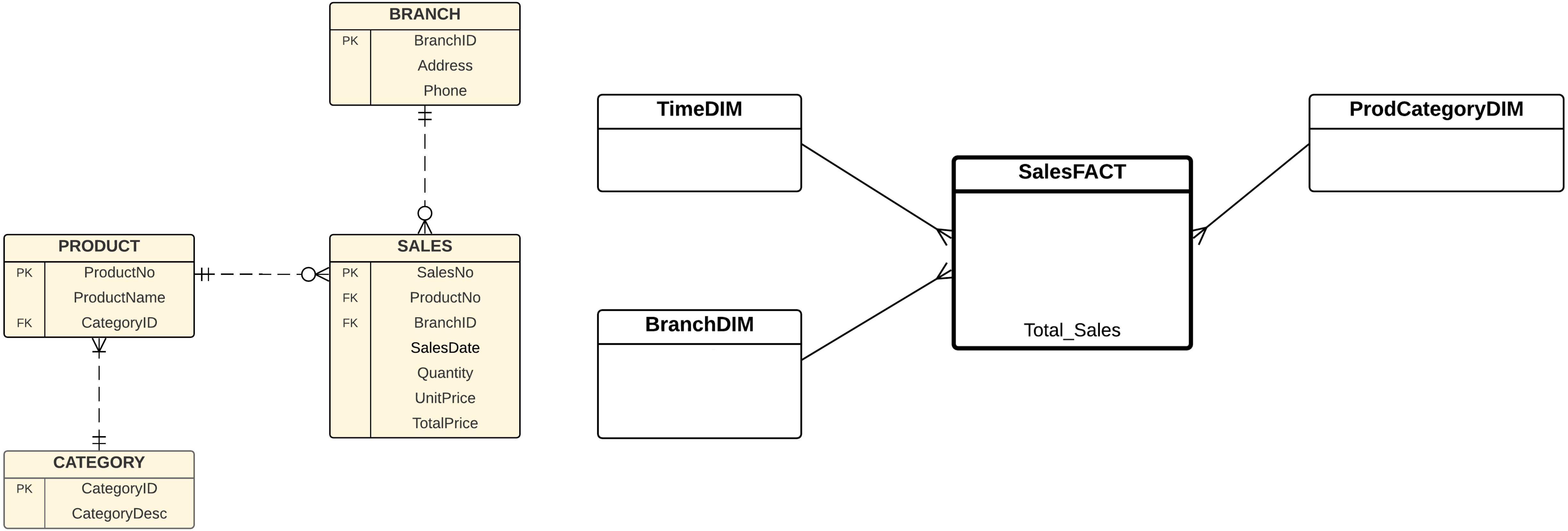


Case Study #2: Sales

- Sales Star Schema
  - **Fact**:
    - Total Sales
  - **Dimensions**:
    - Time
    - Branch
    - Product Category



Case Study #2: Sales



BRANCH

PK	BranchID
	Address
	Phone

PRODUCT

PK	ProductNo
	ProductName
FK	CategoryID

SALES

PK	SalesNo
FK	ProductNo
FK	BranchID
	SalesDate
	Quantity
	UnitPrice
	TotalPrice

CATEGORY

PK	CategoryID
	CategoryDesc

TimeDIM

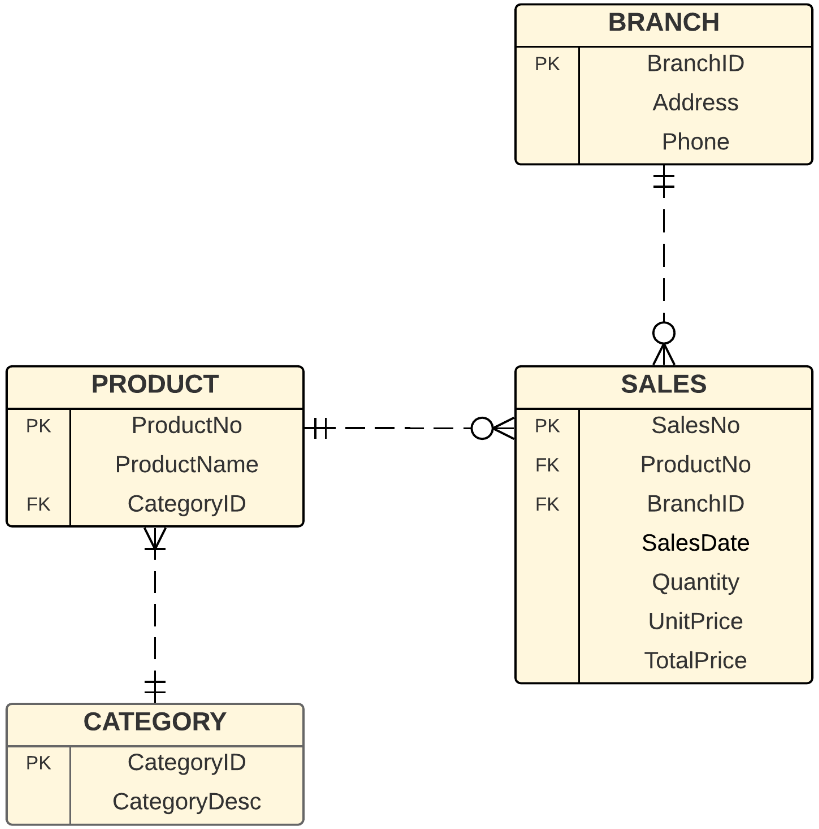
ProdCategoryDIM

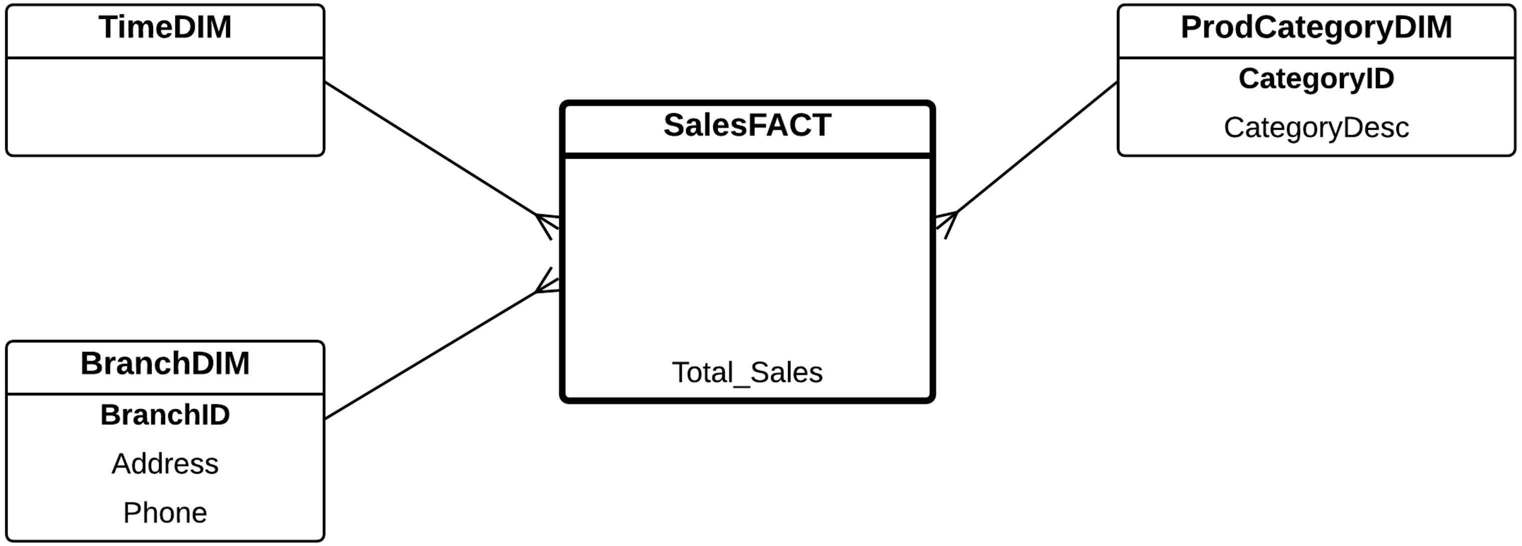
SalesFACT

Total_Sales

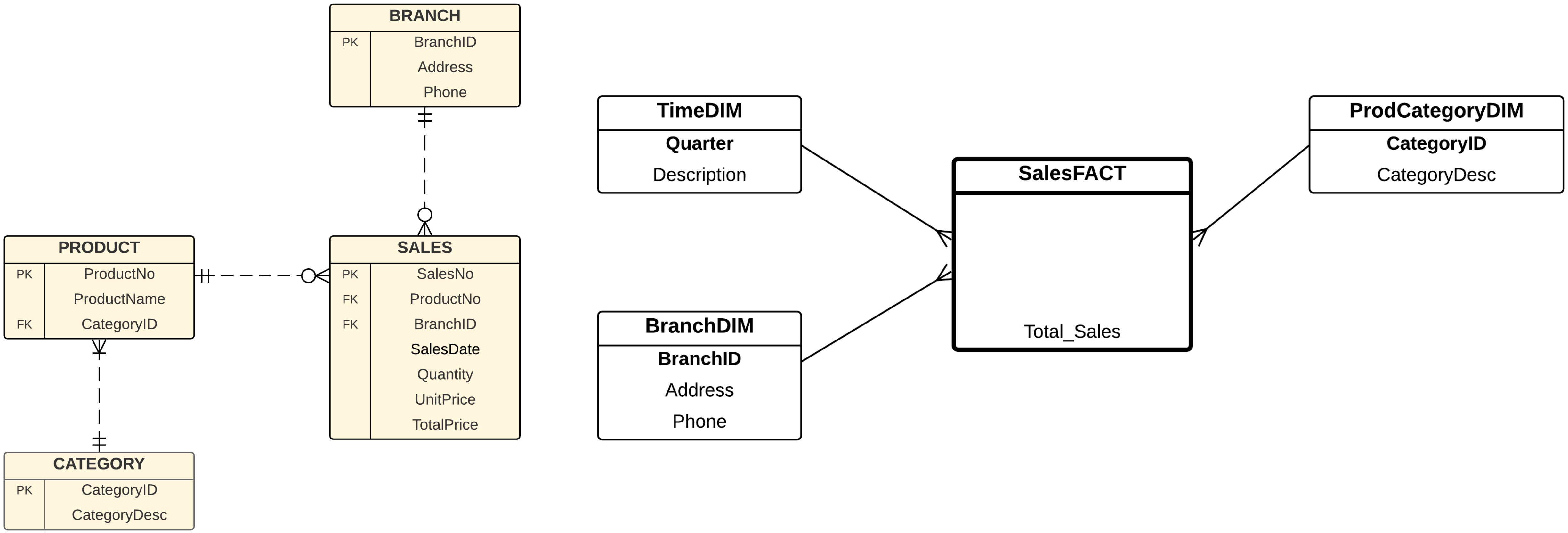
BranchDIM

Case Study #2: Sales





Case Study #2: Sales



BRANCH
PK BranchID
Address
Phone

PRODUCT
PK ProductNo
ProductName
FK CategoryID

SALES
PK SalesNo
FK ProductNo
FK BranchID
SalesDate
Quantity
UnitPrice
TotalPrice

CATEGORY
PK CategoryID
CategoryDesc

TimeDIM
Quarter
Description

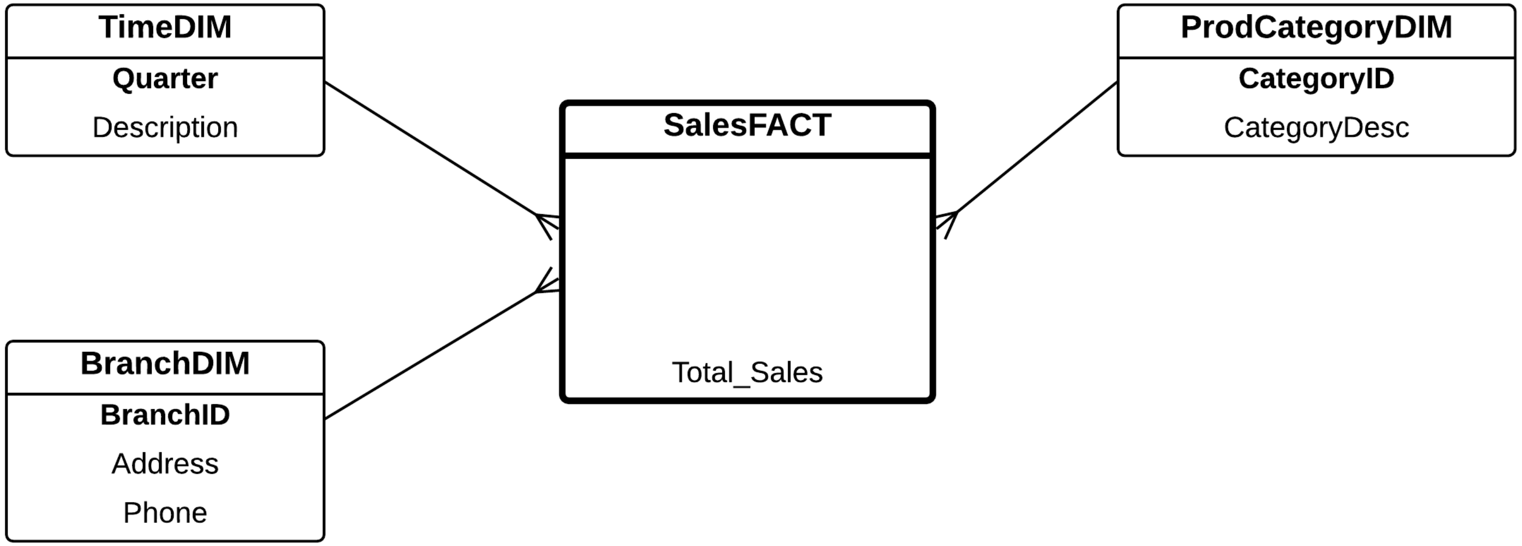
BranchDIM
BranchID
Address
Phone

SalesFACT
Total_Sales

ProdCategoryDIM
CategoryID
CategoryDesc

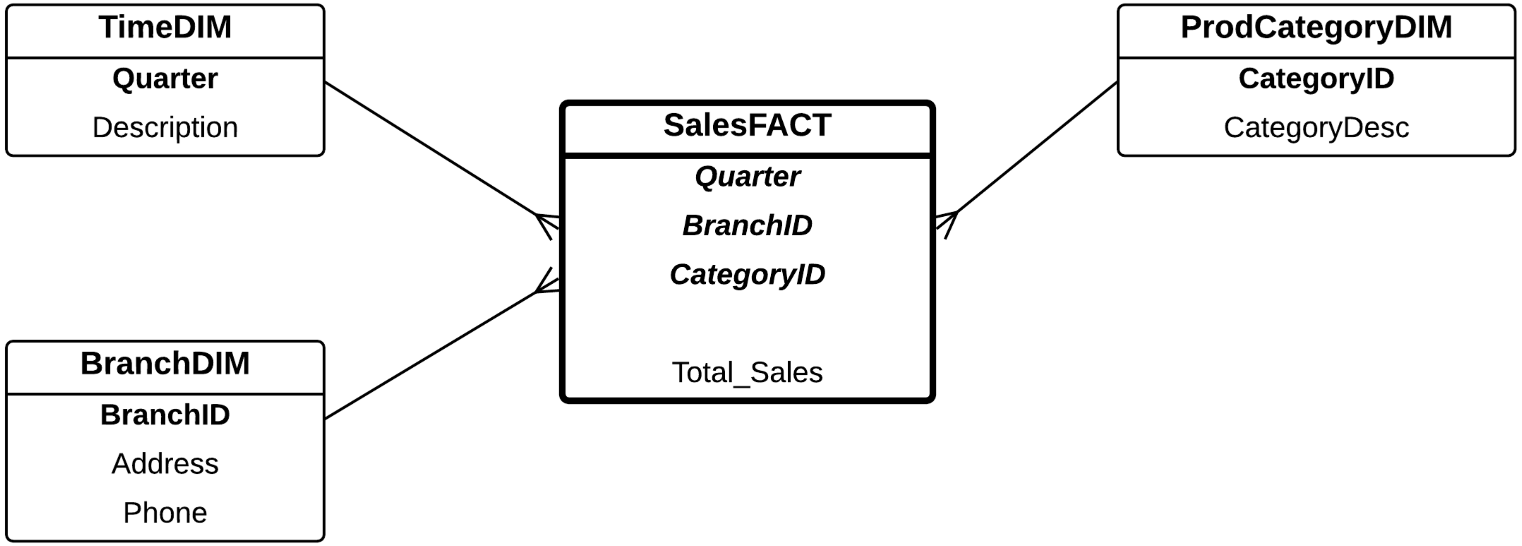
Case Study #2: Sales

Quarter	Description
1	Jan-Mar
2	Apr-Jun
3	Jul-Sep
4	Oct-Dec



Case Study #2: Sales

Quarter	Description
1	Jan-Mar
2	Apr-Jun
3	Jul-Sep
4	Oct-Dec



TimeDIM
Quarter
Description

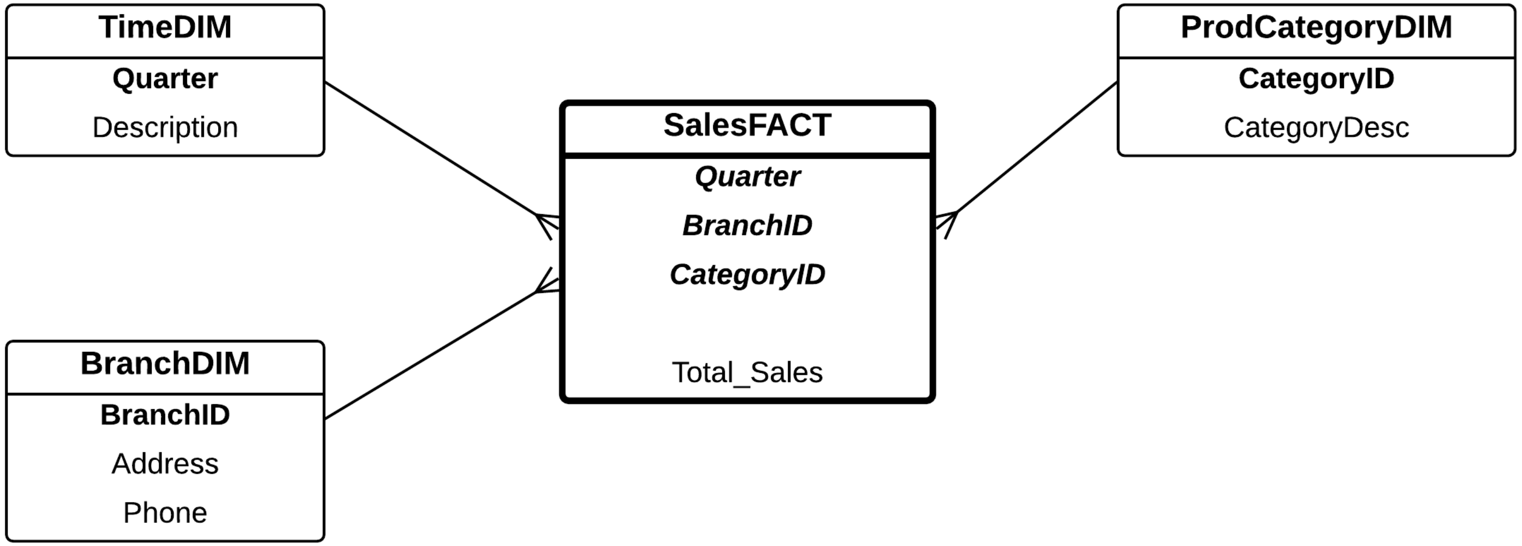
BranchDIM
BranchID
Address
Phone

SalesFACT
Quarter
BranchID
CategoryID
Total_Sales

ProdCategoryDIM
CategoryID
CategoryDesc

Case Study #2: Sales

- Sales Star Schema
  - **Fact**:
    - Total Sales
  - **Dimensions**:
    - Time
    - Branch
    - Product Category



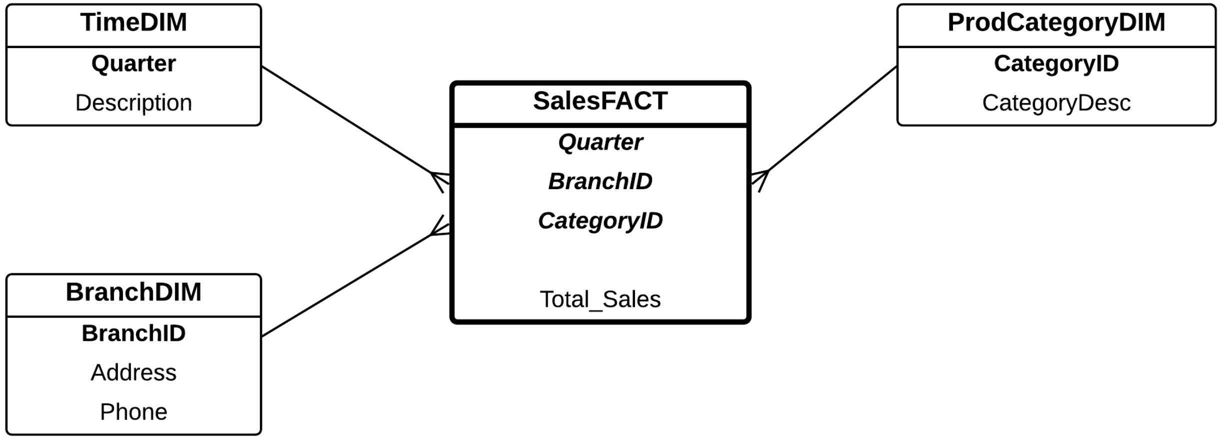
Case Study #2: Sales

- To create ProdCategoryDIM:

  -
    ```
    create table ProdCategoryDim as
      select * from Category;
    ```

- To create BranchDIM:

  -
    ```
    create table BranchDim as
      select * from Branch;
    ```



Case Study #2: Sales

- To create TimeDIM:

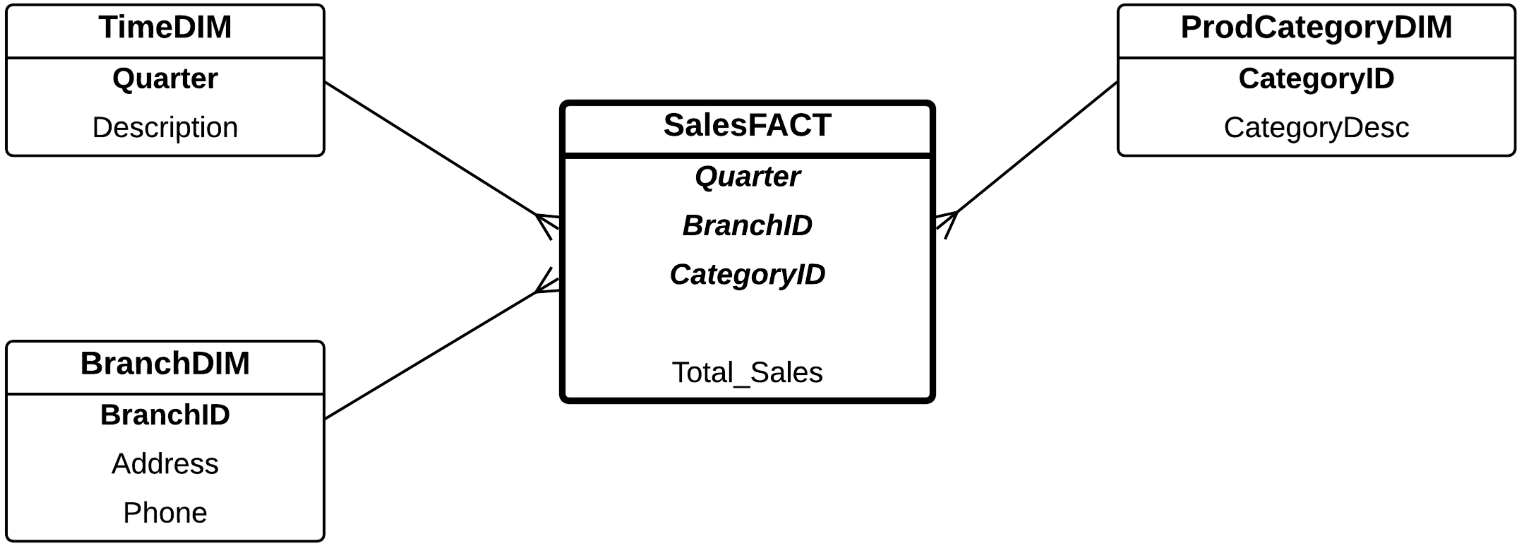
  ```
  - create table TimeDim
     (Quarter number(1),
     Description varchar2(20));
  ```

- To insert the values in TimeDIM:

  ```
  - Insert into TimeDim values (1, ’Jan-Mar’);
    insert into TimeDim values (2, ’Apr-Jun’);
    insert into TimeDim values (3, ’Jul-Sep’);
    insert into TimeDim values (4, ’Oct-Dec’);
  ```

Case Study #2: Sales

- Sales Star Schema
  - **Fact**:
    - Total Sales
  - **Dimensions**:
    - Time
    - Branch
    - Product Category



Case Study #2: Sales

- To create Temporary Fact for SalesFact:

```
- create table TempFact as
  select
    S.SalesDate,
    B.BranchID,
    C.CategoryID,
    S.TotalPrice
  from Branch B, Sales S, Product P, Category C
  where B.BranchID = S.BranchID
  and S.ProductNo = P.ProductNo
  and P.CategoryID = C.CategoryID
  and to_char(S.SalesDate, ’YYYY’) = ’2020’;
```

Case Study #2: Sales

- To create Temporary Fact for SalesFact:

```
– create table TempFact as
  select
    S.SalesDate,
    B.BranchID,
    C.CategoryID,
    S.TotalPrice
  from Branch B, Sales S, Product P, Category C
  where B.BranchID = S.BranchID
  and S.ProductNo = P.ProductNo
  and P.CategoryID = C.CategoryID
  and to_char(S.SalesDate, ’YYYY’) = ’2020’;
```

Case Study #2: Sales

- To create Temporary Fact for SalesFact:

```
- create table TempFact as
  select
    S.SalesDate,
    B.BranchID,
    C.CategoryID,
    S.TotalPrice
  from Branch B, Sales S, Product P, Category C
  where B.BranchID = S.BranchID
  and S.ProductNo = P.ProductNo
  and P.CategoryID = C.CategoryID
  and to_char(S.SalesDate, ’YYYY’) = ’2020’;
```

Case Study #2: Sales

- To create Temporary Fact for SalesFact:

```
– create table TempFact as
  select
     S.SalesDate,
     B.BranchID,
     C.CategoryID,
     S.TotalPrice
  from Branch B, Sales S, Product P, Category C
  where B.BranchID = S.BranchID
  and S.ProductNo = P.ProductNo
  and P.CategoryID = C.CategoryID
  and to_char(S.SalesDate, ’YYYY’) = ’2020’;
```

Case Study #2: Sales

- To create Temporary Fact for SalesFact:

```
– create table TempFact as
  select
     S.SalesDate,
     B.BranchID,
     C.CategoryID,
     S.TotalPrice
  from Branch B, Sales S, Product P, Category C
  where B.BranchID = S.BranchID
  and S.ProductNo = P.ProductNo
  and P.CategoryID = C.CategoryID
  and to_char(S.SalesDate, ’YYYY’) = ’2020’;
```

no aggregation function, and no group by

Case Study #2: Sales

- To alter the TempFact:

```
– alter table TempFact;
  add (Quarter number(1));
```

- To update the TempFact to turn SalesDate into Quarter:

```
– update TempFact
  set Quarter = 1
  where to_char(SalesDate, ’MM’) >= ’01’
  and to_char(SalesDate, ’MM’) <= ’03’;
```

Case Study #2: Sales

- To update the TempFact to turn SalesDate into Quarter:

  -
    ```
    update TempFact
    set Quarter = 2
    where to_char(SalesDate, ’MM’) >= ’04’
    and to_char(SalesDate, ’MM’) <= ’06’;

    update TempFact
    set Quarter = 3
    where to_char(SalesDate, ’MM’) >= ’07’
    and to_char(SalesDate, ’MM’) <= ’09’;

    update TempFact
    set Quarter = 4
    where Quarter is null;
    ```

Case Study #2: Sales

- To create SalesFACT:

  -
    ```
    create table SalesFact as
    select
      Quarter,
      BranchID,
      CategoryID,
      sum(TotalPrice) as Total_Sales
    from TempFact
    group by Quarter, BranchID, CategoryID;
    ```

Two Column Table Methodology

MONASH
University

Two Column Table Methodology

One Fact Measurement:

When creating a star schema, you need to imagine that the data you want to analyse consists of **two columns**.

The first column is the **category** (e.g. A, B, C, D), and the second column is the statistical **numerical figure** (e.g. F).

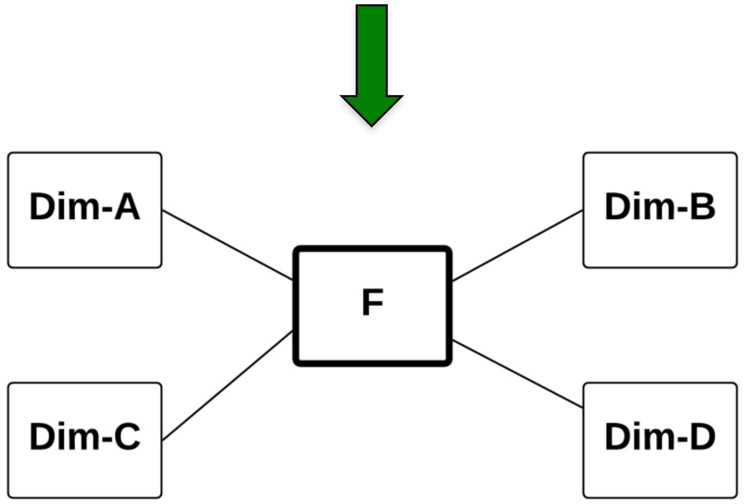
The second column (e.g. F) has to be consistent throughout all the two-column tables.

A	F
x	4
y	3

B	F
r	5
s	3

C	F
k	1
m	1

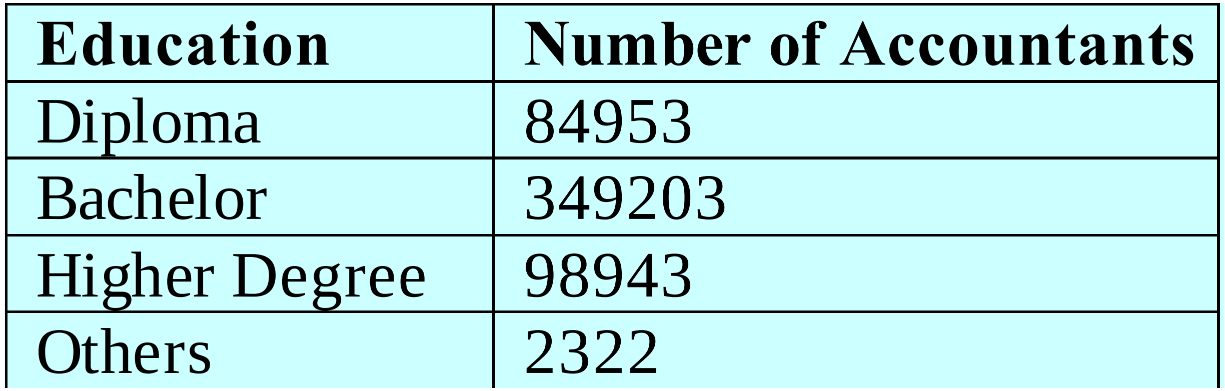
D	F
p	2
q	5



Two Column Table Methodology

- **Case Study 1: Analysis of Accountants**

  Suppose the CPA organization would like to analyze its members (i.e. accountants) in a particular city. Assume that the organization has the full details of its members.



Education	**Number of Accountants**
Diploma	84953
Bachelor	349203
Higher Degree	98943
Others	2322

Two Column Table Methodology

- We can also look at the figures from the gender point of view, like:

Gender	Number of Accountants
Male	434322
Female	89932

- Another way to analyze number of accountants is form the type of the accountant job itself; something like:

Type	Number of Accountants
Government	3843
Private Business	45303
Personal	45930
etc	
etc	

- Note that the figures are fictitious, and the “Types” of Accountants (indicating different roles of accountants) are also fictitious.

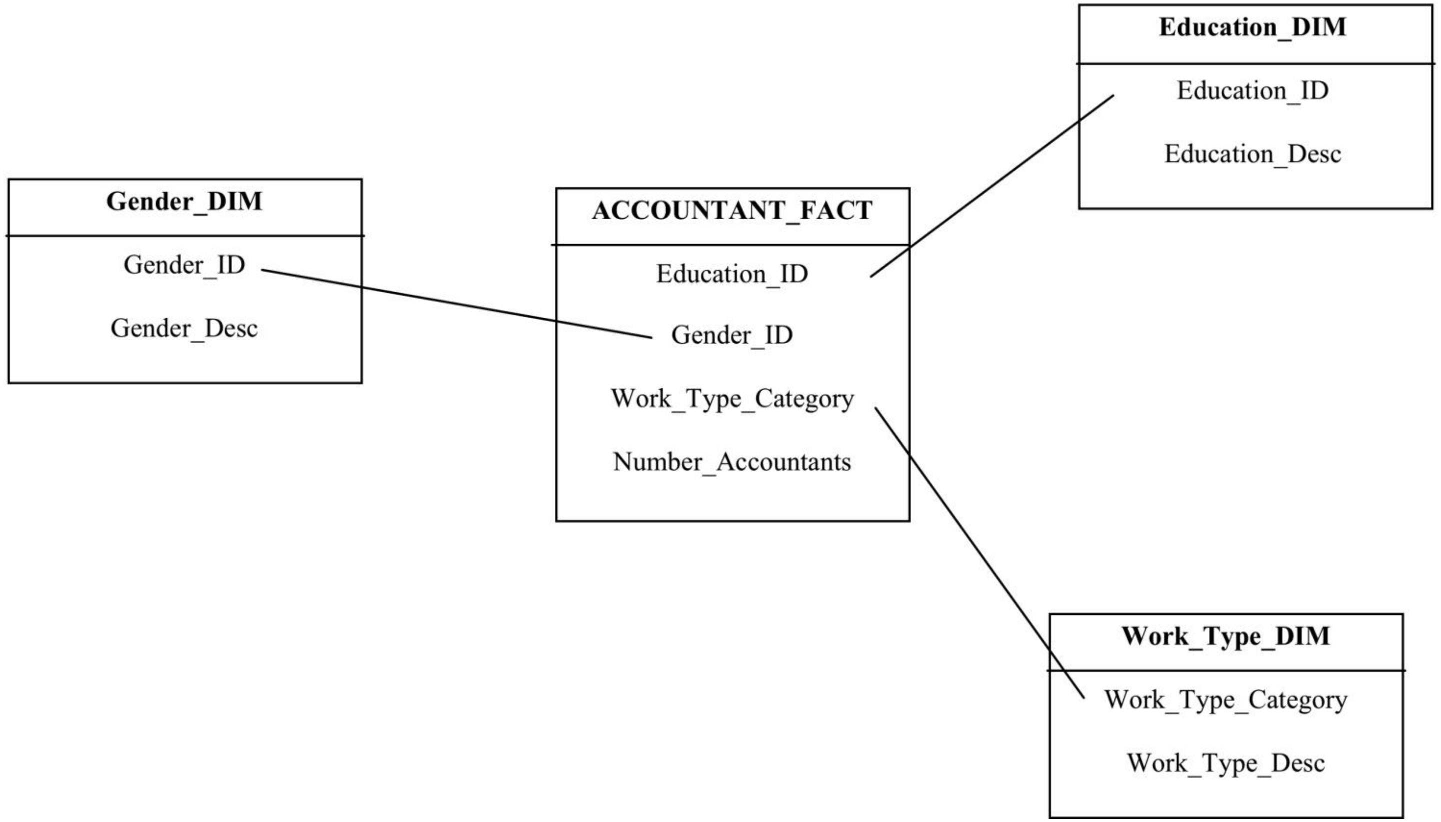
Two Column Table Methodology

- You can further identify other example to analyze number of accountants. In the above three tables, the first angle to look at the number of accountants is from the educational background, the last one is from the type of the accountant itself, whether it is a private business accountant, etc.

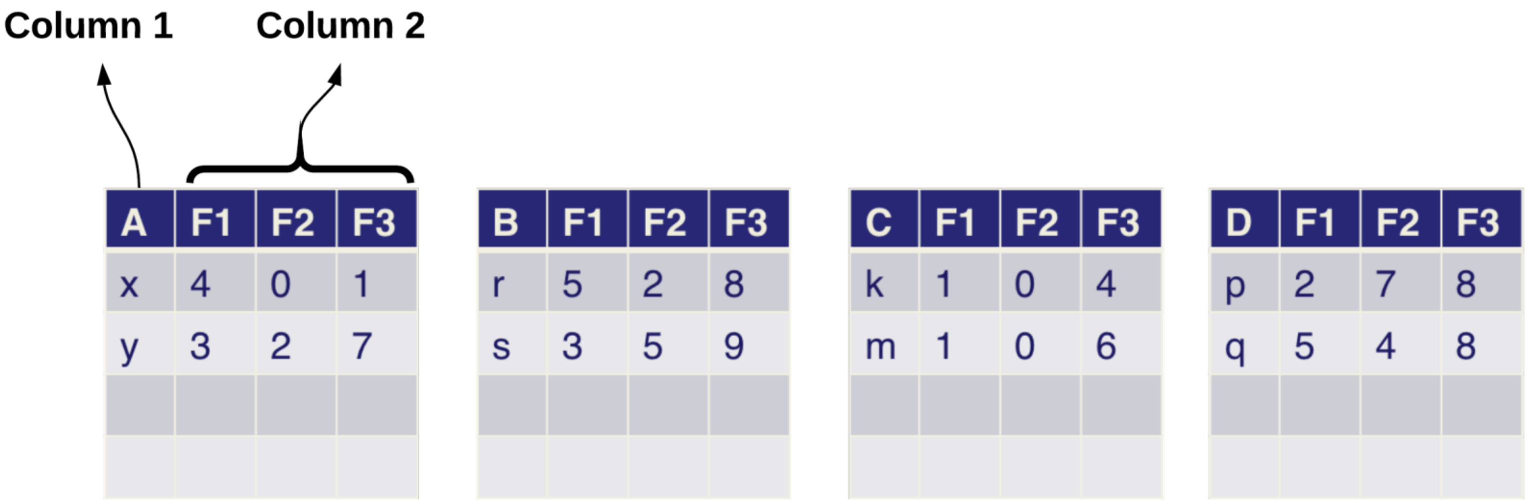
- As you can see, the second column is CONSISTENTLY UNIFORM. In the above example, it is number of accountants. The first column changes depending on from which angle that you want to see.

Two Column Table Methodology

- Therefore, in this case study, the star schema could look like the following:



Two Column Table Methodology

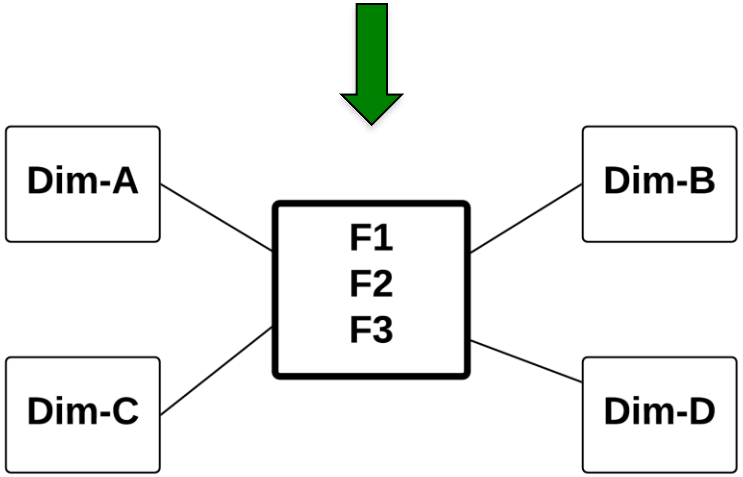

Column 1
Column 2

A	F1	F2	F3
x	4	0	1
y	3	2	7

B	F1	F2	F3
r	5	2	8
s	3	5	9

C	F1	F2	F3
k	1	0	4
m	1	0	6

D	F1	F2	F3
p	2	7	8
q	5	4	8



Multiple Fact Measurements:

The second column in the two-column tables, which is the numerical fact measurement (e.g. column F) can actually be multiple columns (call them: F1, F2, F3), as long as all of these columns (e.g. F1, F2, F3) relate to all of the categories (e.g. A, B, C, D).

Two Column Table Methodology

- **Case Study 2: Student Enrollment**

  The University Administrator(s) needs to keep track of the number of enrollment for particular unit or campus and the students' performance each year in order to maintain the University performance. The head of admin has assigned you the task of developing a small Data Warehouse in which to keep track the enrollment and performance statistics.

Two Column Table Methodology

- For example:

	F1	F2
Subject	**Number of Students**	**Total Score**
Database	8	539
Java	5	327
SAP	1	63
Network	2	105

- Another example could be something like this:

	F1	F2
Semester	**Number of Students**	**Total Score**
One	9	618
Two	7	416

Two Column Table Methodology

- In analyzing number of students (apart from the subject and semester as shown above), you could also see the number of student from another angle, for example from the campus and grade:

	F1	F2
Campus	**Number of Students**	**Total Score**
Main	9	658
City	5	271
DE	2	105

Two Column Table Methodology

- For example:

	F1	F2
Grade	**Number of Students**	**Total Score**
HD	3	253
D	4	300
C	4	256
P	2	105
N	3	120

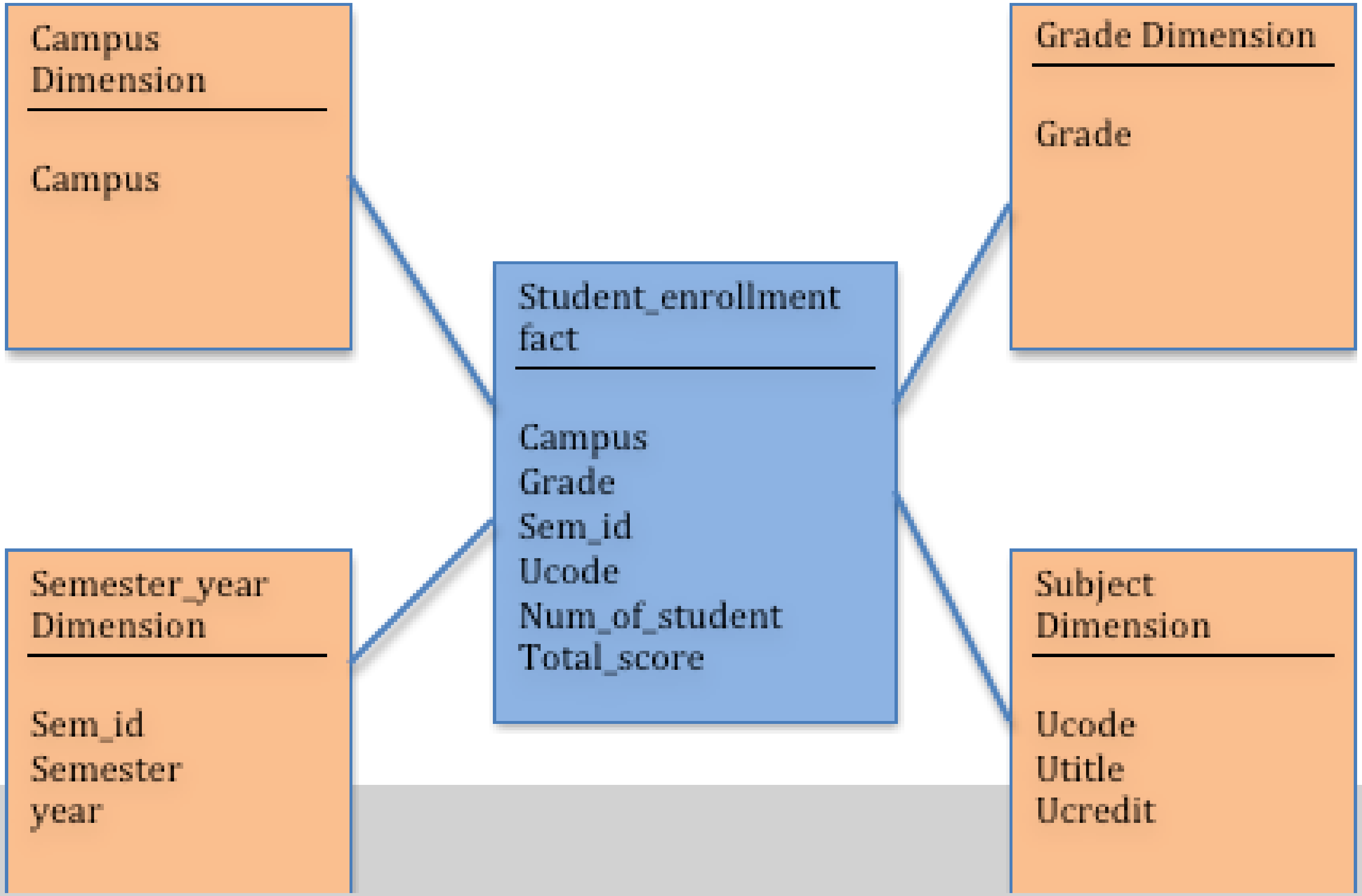
Two Column Table Methodology

- The first columns of the above examples are the **dimensions**, whereas the other columns that contain the statistical/summarized/aggregated values is the **fact**.

- In the above example, the fact is then STUDENT_ENROLLMENT_FACT, and the dimensions are SUBJECT, SEMESTER, GRADE and CAMPUS.

Two Column Table Methodology

- The star schema for the STUDENT ENROLLMENT is shown as follows:



Campus Dimension

Campus

Grade Dimension

Grade

Student_enrollment fact

Campus
Grade
Sem_id
Ucode
Num_of_student
Total_score

Semester_year Dimension

Sem_id
Semester
year

Subject Dimension

Ucode
Utitle
Ucredit