TOPIC NAME : Facts Elaboration

DAY :

TIME : DATE : / /

Facts

- Additive
- Semi Additive
- Non Additive

Additive facts can be aggregated accross all dimensions.

Example, If you have sales data, you can easily aggregate it to get monthly or yearly sales by summing up the daily values.

Semi-Additive Facts:

They cannot be aggregated accross all dimensions. For example, while you can sum up the inventory quantity for a product over time, you cannot sum it accross all products, as it doesn't make any sense

Non-additive Facts:

They cannot be aggregated at any level. For example, average temperature is a non additive fact because you cannot accurately average the averages. If you have daily temperature data, you can find the avg for each day, but averaging those averages wouldn't provide a meaningful result.

GOOD LUCK

TOPIC NAME : _____ DAY : _____

TIME : DATE : / /

NULL in Facts:

- SQL, BI tools can properly handle null values. So, we don't need to take special care for it.
- Sometimes we might need to replace nulls with 0's.
- But if we have a suppose foreign key column ~~fas~~ and that column contains null values, we have to replace that null values with some dummy values. Also we should add that dummy value in the primary key column also of that table whose foreign key are we talking about. We should not keep any null values in the foreign key column.

Year to Date Facts:

- We should not store Year-to-date, month-to-date, etc in our fact tables. We should keep the fact values based on daily grain.
- Later we should use the BI tools to convert them into year to date or month to date values.

Types of Fact Tables

- Transactional
- Periodic Snapshot
- Accumulating Snapshot

GOOD LUCK

TOPIC NAME : _____ DAY : _____

TIME : DATE : / /

Transactional Fact Table:

→ Here 1 row = measurement of one event or transaction

→ A transaction that take place at a specific time.

Example:

	FK	FK	Measure
Sales_id	Product_id	Date_id	Units
1	3	20220101	1
2	5	20220102	1
3	2	20220102	2
4	3	20220103	5

Characteristics:

- Most common and very flexible
- Typically additive
- Tend to have a lot of dimensions (FK) associated
- They can be enormous in size and can have a rapid growth.

Periodic Snapshot Fact Table:

→ 1 row = summarizes measure of many events / transactions.

→ Summarized measure over a period (1 day, 1 week etc.)

GOOD LUCK

TOPIC NAME : _____ DAY : _____

TIME : DATE : / /

Example: Sales transaction per week

	Measure	Measure	Measure
Week-id	Revenue	Sales	Cost
1	323	123	12
2	541	322	31
3	242	108	12
4	352	212	51
5	312	198	25

∴ Here period = 1 week (Per week)

Characteristics:

- Lots of measures, very few dimensions (FK)
- Tend to be not as enormous in size.
- Table does not grow so rapidly.
- Table grows continously per period. It maintains a control.
- Typically additive
- No events = 0 or Null

Accumulating Snapshot fact Table:

→ 1 row = summarizes measure of many events ~~per~~/ transaction

→ Summarized measures of lifespan of 1 product/Process (exp: Order fullfillment)

→ Definite beginning and definite ending (steps in between)

→ Date/time keys associated with role playing dimension.

GOOD LUCK

TOPIC NAME : _____ DAY : _____

TIME : DATE : / /

PK	FK	Measure	FK	FK	FK	FK	Measure
Orderid	Order Date	No. of prod	Product_id	Production start	Production end	Shipping Date	Damaged Produc
1	20220102	100	32	20220103	20220110	20220113	3
2	20220103	100	32	20220104	20220112	20220113	4
3	20220103	100	32	20220103	20220112	2022014	1
4	20220104	100	32	20220108	20220110	2022017	6

Characteristics:

- Number of dimension is very high
- Good performance
- Here we found the measures of process in lifespan.

Factless Fact Table:

Sometimes in Fact tables, only the dimensional aspects of an event are recorded.

Example

Reg_id	Entry Date	Dept_id	Region_id	Manager_id	Position_id
1	20220103	1	2	3	10
2	20220103	3	3	4	112
3	20220304	4	6	3	202
4	20220104	4	8	6	117
5	20220405	3	4	8	18

GOOD LUCK

In the above table, there is no such measures / facts available We are only storing dimensional aspects there. So, it can be taken as a factless fact table.

In the factless fact tables we can do query like →

→ How many employees has been registered last month?

→ How many employees has been registered in a certain region?

Steps to create a fact table:

There are 4 key decisions.

First of all we have to keep in mind the business needs.

Then we have to take the 4 key decisions.

1) Identify business process for analysis.
   Example → Sales, Order processing

2) Declare the grain.
   Example → Transaction, Order, Order lines, Daily, Daily + location

3) Identify dimensions that are relevant.
   Example → Try to come up with the questions. What, when, how and why
   Time, Locations, Products, Customers. etc.

4) Identify facts for measurement.

TOPIC NAME : _____ DAY : _____

TIME : DATE : / /

Natural VS Surrogate Key:

Natural keys:

→ They come out of source system

Surrogate Keys:

→ They also called Artificial Keys

→ They consist of integer numbers.

→ They are actually PK or FK ~~also~~

→ Their column name can have _PK or _FK as suffix

→ They can be created by database or ETL Tools.

Benefits of Surrogate Keys:

- Improve Performance (less storage / better joins)
- Can handle dummy values by adding (-1, 999, etc) for nulls or missing values.
- Integrate multiple source systems.
- Easier update

GOOD LUCK