(Big) Data Engineering In Depth

From Beginner to Professional

Mostafa Alaa Mohamed
Senior Big Data Engineer
 MoustafaAlaa Moustafa Alaa @Moustafa_alaa22
 mustafa.alaa.mohamed@gmail.com

[1]Big Data & Analytics Department, Epam Systems

The Definitive Guide to Big Data Engineering Tasks

Videos classification

Watching Method / Audience	Computer	Mobile/Tablet	Just listening
Developer	●		
DevOps	●		
Business	●		

Table: Video classification
The green circle ● means short video.
The blue circle ● means medium video.
The red circle ● means long video

Dimensions Types: Heterogeneous Dimensions

Heterogeneous Dimensions

- This type works when we have a case that a company selling different product to the same base of customer. Every product has it different attributes.

Heterogeneous Dimensions

- This type works when we have a case that a company selling different product to the same base of customer. Every product has it different attributes.

- One famous example of this type assume an insurance company has two types of product like health and car. In this case Car insurance has different attributes than the health insurance.

Heterogeneous Dimensions

- This type works when we have a case that a company selling different product to the same base of customer. Every product has it different attributes.

- One famous example of this type assume an insurance company has two types of product like health and car. In this case Car insurance has different attributes than the health insurance.

- If we tried to model this two different products this type name Heterogeneous dimensions.

Heterogeneous Dimensions

- There are different scenario to implement this type

Heterogeneous Dimensions

- There are different scenario to implement this type

  Separate Dimensions Split each one in separate dimensions and facts. It will be less data and business will do this analysis from two separate facts.

Heterogeneous Dimensions

- There are different scenario to implement this type

Separate Dimensions Split each one in separate dimensions and facts. It will be less data and business will do this analysis from two separate facts.

Merge Attributes We will merge all the attributes in one single table and we will add the common attributes and null for un related attributes. Implementing this scenarios when we have less different of attributes. However, this implementation is not recommended because of the table size, performance, and maintenance.

Heterogeneous Dimensions

- There are different scenario to implement this type

Separate Dimensions Split each one in separate dimensions and facts. It will be less data and business will do this analysis from two separate facts.

Merge Attributes We will merge all the attributes in one single table and we will add the common attributes and null for un related attributes. Implementing this scenarios when we have less different of attributes. However, this implementation is not recommended because of the table size, performance, and maintenance.

Generic Design In this approach we will create a single fact table and single dimension with the common attributes. The problem of this design we will report or care about the common attributes only.