

# (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

<sup>1</sup>Big Data & Analytics Department, Epam Systems

The Definitive Guide to Big Data Engineering Tasks

# Videos classification


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

**Table:** Video classification

- The green circle ● means short video.
- The blue circle ● means medium video.
- The red circle ● means long video

## Dimensions Types: Fast (Rapidly) Changing Dimension (Mini Dimension)

# Fast Changing Dimension (Mini Dimension)

- When we have a dimension with one or more of its attributes changing very fast.
-  *It causes a performance issue if we tried to handle this case similar SCD Type 2 because of the rapidly changing in this dimension and the table will includes a lot of rows for this dimension.*
- We solve this case by separation the attributes into one or more dimensions. This technique also called ***mini-dimensions***.

# Fast Changing Dimension (Mini Dimension)

- How to implement FCD (Mini Dimension)?

*Hint: Search for the mini-dimension relation table.*

Patient_id	Name	Gender	BirthDate	Weight	B_Presure	UpdateDt
123	Anna	F	1968-01-12	50	110.0	2019-01-01
123	Anna	F	1968-01-12	55	130.0	2019-01-07
123	Anna	F	1968-01-12	59	115.0	2019-01-14
123	Anna	F	1968-01-12	65	120.0	2019-01-21

Table: Patient Profile Dimension

Patient_id	Name	Gender	BirthDate
123	Anna	F	1968-01-12

Patient_Key	Weight	B_Presure
1	50	110.0
2	55	130.0
3	59	115.0
4	65	120.0

Table: Patient Profile Dimension After Removing FCD and Split it into Junk-Dimension table

# 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

# Fast Changing Dimension (Mini Dimension)

Patient_id	Patient_Key	Start_Date	End_Date
123	1	2019-01-01	2019-01-07
123	2	2019-01-07	2019-01-14
123	3	2019-01-14	2019-01-14
123	4	2019-01-21	null

Table: Patient Mini Dimension

