(Big) Data Engineering In Depth

From Beginner to Professional

Moustafa Alaa
Senior Big Data Engineer
MoustafaAlaa Moustafa Alaa @Moustafa_alaa22
Garage Education
mustafa.alaa.mohamed@gmail.com

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

Data solution thinking (Summary)

Let's answer our previous question. How can we solve data challenges?

Data solution thinking (Summary)

- Let's split the problem based on the data layers.

Data solution thinking (Summary)

- Let's split the problem based on the data layers.
  - View layer

Data solution thinking (Summary)

- Let's split the problem based on the data layers.
  - View layer
    - When we need to add/remove/create new reports, it is usually a view layer.

Data solution thinking (Summary)

- Let's split the problem based on the data layers.
  - View layer
    - When we need to add/remove/create new reports, it is usually a view layer.
    - We don't need to change the logical or physical layer to support the view layer.

Data solution thinking (Summary)

- Let's split the problem based on the data layers.

Data solution thinking (Summary)

- Let's split the problem based on the data layers.
  - Logical Layer

Data solution thinking (Summary)

- Let's split the problem based on the data layers.
  - Logical Layer
    - When you have missing sources into your logical layer, and you need to add this source and its structure.

Data solution thinking (Summary)

- Let's split the problem based on the data layers.
  - Logical Layer
    - When you have missing sources into your logical layer, and you need to add this source and its structure.
    - There is a performance issue in the existing reports, and you need to change the model.
      ✍ ➔ reduce the join by creating a new join table (*materialized view*).

Data solution thinking (Summary)

- Let's split the problem based on the data layers.
  - Logical Layer
    - When you have missing sources into your logical layer, and you need to add this source and its structure.
    - There is a performance issue in the existing reports, and you need to change the model. ✍ ➔ reduce the join by creating a new join table (*materialized view*).
    - Update the data type or the existing relation, which could help to fix some data or performance issues.

Data solution thinking (Summary)

- Let's split the problem based on the data layers.

Data solution thinking (Summary)

- Let's split the problem based on the data layers.
  - Physical Layer

Data solution thinking (Summary)

- Let's split the problem based on the data layers.
  - Physical Layer
    - When our problem is hard or impossible to fix by optimizing the query (view)/ logical layer, it is time for physical change.

Data solution thinking (Summary)

- Let's split the problem based on the data layers.
  - Physical Layer
    - When our problem is hard or impossible to fix by optimizing the query (view)/ logical layer, it is time for physical change.
    - If we need to change your storage/compression/structure/access technique.

Data solution thinking (Summary)

- Let's split the problem based on the data layers.
  - Physical Layer
    - When our problem is hard or impossible to fix by optimizing the query (view)/ logical layer, it is time for physical change.
    - If we need to change your storage/compression/structure/access technique.
    - If we need to change the data orientation structure from row to column or key-value storage, It is time to change the physical layer.

Data solution thinking (Summary)

- Let's split the problem based on the data layers.

Data solution thinking (Summary)

- Let's split the problem based on the data layers.
  - https://beginnersbook.com/2015/04/levels-of-abstraction-in-dbms/

Data solution thinking (Summary)

- Let's split the problem based on the data layers.
  - https://beginnersbook.com/2015/04/levels-of-abstraction-in-dbms/
  - https://www.guru99.com/dbms-data-independence.html

Data solution thinking (Summary)

- Let's split the problem based on the data layers.
  - https://beginnersbook.com/2015/04/levels-of-abstraction-in-dbms/
  - https://www.guru99.com/dbms-data-independence.html
  - https://www.geeksforgeeks.org/data-abstraction-and-data-independence/