(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

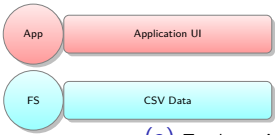
Videos classification

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

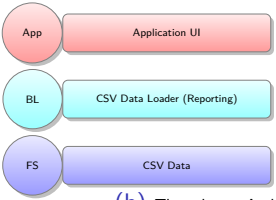
Table: Video classification
The green circle ● means short video.
The blue circle ● means medium video.
The red circle ● means long video

Section: Data Abstraction

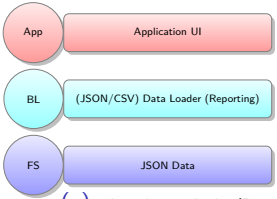
Motivation to Data Layers (Use Case)



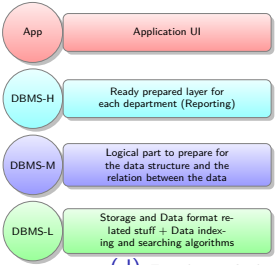
(a) Two layers Arch. (Data & UI)



(b) Three layers Arch. (Data & BL & UI)



(c) Three layers Arch. (Data (multi-sources) & BL & UI)



(d) Four layers Arch. (DB (L, M, H) & UI)

Figure: Data Abstraction Journey

Motivation to Data Layers (Solution Thinking)

- How can we think about a data solution or challenges in the data products?

Motivation to Data Layers (Solution Thinking)

- How can we think about a data solution or challenges in the data products?
  - Requirements analysis.

Motivation to Data Layers (Solution Thinking)

- How can we think about a data solution or challenges in the data products?
  - Requirements analysis.
  - Identify the problem (challenges).

Motivation to Data Layers (Solution Thinking)

- How can we think about a data solution or challenges in the data products?
  - Requirements analysis.
  - Identify the problem (challenges).
  - Think about how to overcome the challenges.

Motivation to Data Layers (Solution Thinking)

- How can we think about a data solution or challenges in the data products?
  - Requirements analysis.
  - Identify the problem (challenges).
  - Think about how to overcome the challenges.
  - Ask your self the following questions:

Motivation to Data Layers (Solution Thinking)

- How can we think about a data solution or challenges in the data products?
  - Requirements analysis.
  - Identify the problem (challenges).
  - Think about how to overcome the challenges.
  - Ask your self the following questions:
    - Can we solve the problem using the current data structure by adding new features?

Motivation to Data Layers (Solution Thinking)

- How can we think about a data solution or challenges in the data products?
  - Requirements analysis.
  - Identify the problem (challenges).
  - Think about how to overcome the challenges.
  - Ask your self the following questions:
    - Can we solve the problem using the current data structure by adding new features?
    - What if we enhance/change the data structure or modeling?

Motivation to Data Layers (Solution Thinking)

- How can we think about a data solution or challenges in the data products?
  - Requirements analysis.
  - Identify the problem (challenges).
  - Think about how to overcome the challenges.
  - Ask your self the following questions:
    - Can we solve the problem using the current data structure by adding new features?
    - What if we enhance/change the data structure or modeling?
    - Could it help if we change the backend engine (DBMS system)?

Motivation to Data Layers (Solution Thinking)

- How can we think about a data solution or challenges in the data products?
  - Requirements analysis.
  - Identify the problem (challenges).
  - Think about how to overcome the challenges.
  - Ask your self the following questions:
    - Can we solve the problem using the current data structure by adding new features?
    - What if we enhance/change the data structure or modeling?
    - Could it help if we change the backend engine (DBMS system)?
- To answer these questions you need to understand the **data layers**.

Data Layers (Abstraction)

- Any data product (database) contains multi-layers.

Data Layers (Abstraction)

- Any data product (database) contains multi-layers.
- Each layer responsible for different tasks and operations.

Data Layers (Abstraction)

- Any data product (database) contains multi-layers.
- Each layer responsible for different tasks and operations.
- Each layer interacts with (hardware or software or mixed).

Data Layers (Abstraction)

- Any data product (database) contains multi-layers.
- Each layer responsible for different tasks and operations.
- Each layer interacts with (hardware or software or mixed).
- Eliminate the complexity of data interactions; not all internal processes are shared or available for the user.

Data Layers (Abstraction)

- Any data product (database) contains multi-layers.
- Each layer responsible for different tasks and operations.
- Each layer interacts with (hardware or software or mixed).
- Eliminate the complexity of data interactions; not all internal processes are shared or available for the user.
- The developer for each layer hides irrelevant internal details from the developer (users).

Data Layers (Abstraction)

- Any data product (database) contains multi-layers.
- Each layer responsible for different tasks and operations.
- Each layer interacts with (hardware or software or mixed).
- Eliminate the complexity of data interactions; not all internal processes are shared or available for the user.
- The developer for each layer hides irrelevant internal details from the developer (users).
- The process of **hiding** irrelevant details from the developer (user) is called data **abstraction**.

Data Layers (Abstraction)

Definition

Data Abstraction and Data Independence: DBMS comprises complex data-structures. To make the system efficient in terms of retrieval of data and reduce complexity in terms of usability of users, developers use abstraction i.e., hide irrelevant details from the users. This approach simplifies database design.

- There are 3 levels of data abstraction.

Data Layers (Abstraction)

Definition

Data Abstraction and Data Independence: DBMS comprises complex data-structures. To make the system efficient in terms of retrieval of data and reduce complexity in terms of usability of users, developers use abstraction i.e., hide irrelevant details from the users. This approach simplifies database design.

- There are 3 levels of data abstraction.
  - Physical Level

Data Layers (Abstraction)

Definition

Data Abstraction and Data Independence: DBMS comprises complex data-structures. To make the system efficient in terms of retrieval of data and reduce complexity in terms of usability of users, developers use abstraction i.e., hide irrelevant details from the users. This approach simplifies database design.

- There are 3 levels of data abstraction.
  - Physical Level
  - Logical/Conceptual Level.

Data Layers (Abstraction)

Definition

Data Abstraction and Data Independence: DBMS comprises complex data-structures. To make the system efficient in terms of retrieval of data and reduce complexity in terms of usability of users, developers use abstraction i.e., hide irrelevant details from the users. This approach simplifies database design.

- There are 3 levels of data abstraction.
  - Physical Level
  - Logical/Conceptual Level.
  - View Level.

Data Layers (Abstraction)