(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

Chapter: Introduction To Data Management and Data Warehouse

Chapter Objectives

- Be familiar with data management life-cycle.

Chapter Objectives

- Be familiar with data management life-cycle.
- Understand the data abstraction and the data layer.

Chapter Objectives

- Be familiar with data management life-cycle.
- Understand the data abstraction and the data layer.
- Motivation to DWH.

Chapter Objectives

- Be familiar with data management life-cycle.
- Understand the data abstraction and the data layer.
- Motivation to DWH.
- What are the different types of DWH?

Chapter Objectives

- Be familiar with data management life-cycle.
- Understand the data abstraction and the data layer.
- Motivation to DWH.
- What are the different types of DWH?
- Usecases for DWH. How is it different from the operational DB?

Chapter Objectives

- Be familiar with data management life-cycle.
- Understand the data abstraction and the data layer.
- Motivation to DWH.
- What are the different types of DWH?
- Usecases for DWH. How is it different from the operational DB?
- Explain the data Encoding and Formats.

Chapter Objectives

- Be familiar with data management life-cycle.
- Understand the data abstraction and the data layer.
- Motivation to DWH.
- What are the different types of DWH?
- Usecases for DWH. How is it different from the operational DB?
- Explain the data Encoding and Formats.
- Show what the challenges of building a DWH are?

Chapter Objectives

- Be familiar with data management life-cycle.
- Understand the data abstraction and the data layer.
- Motivation to DWH.
- What are the different types of DWH?
- Usecases for DWH. How is it different from the operational DB?
- Explain the data Encoding and Formats.
- Show what the challenges of building a DWH are?
- What are the data modeling and its design?

Section: Data Management

Data Management

- Data are a product.

Data Management

- Data are a product.
- Data product has a life-cycle as following (simplified):

Data Management

- Data are a product.
- Data product has a life-cycle as following (simplified):
  - **Question**, Idea, or service.

Data Management

- Data are a product.
- Data product has a life-cycle as following (simplified):
  - **Question**, Idea, or service.
  - **Identify** the source of information and the data type.

Data Management

- Data are a product.
- Data product has a life-cycle as following (simplified):
  - **Question**, Idea, or service.
  - **Identify** the source of information and the data type.
  - **Document** all details regarding the data including quality, security, efficiency, and access (consideration during the cycle).

Data Management

- Data are a product.
- Data product has a life-cycle as following (simplified):
  - **Question**, Idea, or service.
  - **Identify** the source of information and the data type.
  - **Document** all details regarding the data including quality, security, efficiency, and access (consideration during the cycle).
  - Delivery automation (Tools and Process). AKA **DevOps** cycle.

Data Management

- Data are a product.
- Data product has a life-cycle as following (simplified):
  - **Question**, Idea, or service.
  - **Identify** the source of information and the data type.
  - **Document** all details regarding the data including quality, security, efficiency, and access (consideration during the cycle).
  - Delivery automation (Tools and Process). AKA **DevOps** cycle.
  - Data Architecture (model design and rules).

Data Management

- Data are a product.
- Data product has a life-cycle as following (simplified):
  - **Question**, Idea, or service.
  - **Identify** the source of information and the data type.
  - **Document** all details regarding the data including quality, security, efficiency, and access (consideration during the cycle).
  - Delivery automation (Tools and Process). AKA **DevOps** cycle.
  - Data Architecture (model design and rules).
  - **Extraction**, **Transformation**, and **Loading** Process.

Data Management

- Data are a product.
- Data product has a life-cycle as following (simplified):
  - **Question**, Idea, or service.
  - **Identify** the source of information and the data type.
  - **Document** all details regarding the data including quality, security, efficiency, and access (consideration during the cycle).
  - Delivery automation (Tools and Process). AKA **DevOps** cycle.
  - Data Architecture (model design and rules).
  - **Extraction**, **Transformation**, and **Loading** Process.
  - Business Intelligence (**BI**) or data discovery (continues process).

Data Management

- Data are a product.
- Data product has a life-cycle as following (simplified):
  - **Question**, Idea, or service.
  - **Identify** the source of information and the data type.
  - **Document** all details regarding the data including quality, security, efficiency, and access (consideration during the cycle).
  - Delivery automation (Tools and Process). AKA **DevOps** cycle.
  - Data Architecture (model design and rules).
  - **Extraction**, **Transformation**, and **Loading** Process.
  - Business Intelligence (**BI**) or data discovery (continues process).
  - **Integration** and publishing.

Data Management

- Data are a product.
- Data product has a life-cycle as following (simplified):
  - **Question**, Idea, or service.
  - **Identify** the source of information and the data type.
  - **Document** all details regarding the data including quality, security, efficiency, and access (consideration during the cycle).
  - Delivery automation (Tools and Process). AKA **DevOps** cycle.
  - Data Architecture (model design and rules).
  - **Extraction**, **Transformation**, and **Loading** Process.
  - Business Intelligence (**BI**) or data discovery (continues process).
  - **Integration** and publishing.
  - Data retention or **archiving** process
    (Hot or Cold storage).

Data Management Life-Cycle