

(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

Sub-Section: Extraction Layer

Extraction Layer

- In some companies, they hire or dedicate a team for this part (extraction or ingestion team), but in other companies, it is part of the data engineering team.
- This layer takes the output analysis and decisions from the previous layer (source system analysis) and implement the extraction (quality from the previous team output highly affect this team).
- There is a lot of consideration this team needs to take care of or deal with, but we can summarize it in the following:
 - Data latency analysis as it affects the tool and the methodology (stream or batch).
 - Data extraction method (push or pull).
 - Data size and format compared with the available resources for this project.
- This layer output is a minimal data cleansing (no transformation) into the staging/landing layer.