

(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

Sub-Section: Use Cases of Operational DB vs DWH

Use case (Operational DB)

- A telecommunication company named **XTec**.
- They have lots of systems. One of this systems is a CRM system as example of operational DB.
 - The CRM system handles the customer activities with the company including (sales, change in customer plans, and other activities).
 - This system has a backend database (MySQL).
 - CRM team can report their sales and customer activities from their database.
 - Product owner can take a decision based on their system backend reports.

Use case (DWH)

- What is the need for DWH?
 - This company has other systems
 - ✍️ ➕ billing, charging, signaling.
 - They need to report information related to the CRM, billing, and signaling source systems in one report.
 - So, they need to ingest (transfer) the data from the source systems to one single database.
 - The decision from the DWH is a **global and strategical decision.**
 - If the company needs to build a machine learning model which needs data from different sources. They need to load the data from a centralized database rather than read each source alone.

Use case (DWH)

The Full picture required a DWH. However, we still need the other operational databases for product development perspective.

Use case (ODS)

- Why do we need the ODS?
- How does it fit in our system?

Use case (ODS)

XTec has a call center system which handles the customer inquiries.

This system requires the some data related to usage, customer information, billing details to be calculated and accumulated in **real-time** to be able to give the customer the right answer for his inquires.

Use case (ODS)

- So, What is the challenge for this system?
 - It needs specific information from different source systems.
 - It requires to track the source system database changes or update in real-time.
 - It's functionality is based on the aggregate data not the transactions
✍️ ➡️ (It needs the total outgoing calls till time or it needs the total charging amounts from prepaid or the available limits from billing if it is postpaid).

Use case (ODS)

- ODS is based on change data capture (CDC). This approach used to determine the data change and apply action based on this change.
- ODS uses the real-time aggregations to support the online systems from different source systems.