

(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: Source System Integration Process

Source System Integration Process

- In some companies, they hire or dedicate a team for this part (business analyst, system analyst, data analyst, or demand team).

Source System Integration Process

- In some companies, they hire or dedicate a team for this part (business analyst, system analyst, data analyst, or demand team).
- Before we start, we need to document all the communications into any format.



Source System Integration Process

- In some companies, they hire or dedicate a team for this part (business analyst, system analyst, data analyst, or demand team).
- Before we start, we need to document all the communications into any format.
 - Confluence pages, Word, or Excel sheet.

Source System Integration Process

- In some companies, they hire or dedicate a team for this part (business analyst, system analyst, data analyst, or demand team).
- Before we start, we need to document all the communications into any format.
 - Confluence pages, Word, or Excel sheet.
 - Make the discussion online and put comments to make the history available always.

Source System Integration Process

- In some companies, they hire or dedicate a team for this part (business analyst, system analyst, data analyst, or demand team).
- Before we start, we need to document all the communications into any format.
 - Confluence pages, Word, or Excel sheet.
 - Make the discussion online and put comments to make the history available always.
 - We need to clarify all the tasks and what is the expected output,
  (analysis means to document data structure, format, column names, etc.).

Source System Integration Process

- Requirements gathering.

Source System Integration Process

- Requirements gathering.
- Identify the stakeholders (Data owner(s)).

Source System Integration Process

- Requirements gathering.
- Identify the stakeholders (Data owner(s)).
- Data Analysis includes but not only (format, latency, and column definitions).

Source System Integration Process

- Requirements gathering.
- Identify the stakeholders (Data owner(s)).
- Data Analysis includes but not only (format, latency, and column definitions).
- Check the source system access and perform connectivity assessment.

Source System Integration Process

- Requirements gathering.
- Identify the stakeholders (Data owner(s)).
- Data Analysis includes but not only (format, latency, and column definitions).
- Check the source system access and perform connectivity assessment.
- Initiate the technical discussion about the best way to ingest the data.

Source System Integration Process

- Requirements gathering.
- Identify the stakeholders (Data owner(s)).
- Data Analysis includes but not only (format, latency, and column definitions).
- Check the source system access and perform connectivity assessment.
- Initiate the technical discussion about the best way to ingest the data.
- Data Ingestion method and format.

Source System Integration Process

- Requirements gathering.
- Identify the stakeholders (Data owner(s)).
- Data Analysis includes but not only (format, latency, and column definitions).
- Check the source system access and perform connectivity assessment.
- Initiate the technical discussion about the best way to ingest the data.
- Data Ingestion method and format.
- Sign or confirmation for every point between the stakeholders.

Source System Integration Process

- Requirements gathering.
- Identify the stakeholders (Data owner(s)).
- Data Analysis includes but not only (format, latency, and column definitions).
- Check the source system access and perform connectivity assessment.
- Initiate the technical discussion about the best way to ingest the data.
- Data Ingestion method and format.
- Sign or confirmation for every point between the stakeholders.
- This layer deliver a data analysis (Source system interface) document.