

Business Scenario

You are working as a Data Engineer at a telecom company called **Beejan Technologies**. Every day, thousands of customers complain about issues like poor network, incorrect billing, or bad customer service. These complaints come through different channels: social media, call center log files, SMS, and website forms.

The management is frustrated. Data is stored in different formats. The reporting team manually compiles spreadsheets. No single pipeline exists for this data flow. Reports are delayed. Teams work in silos.

They ask you to design a solution to bring all this data together, clean it, enrich it and make it ready to provide actionable insights.

Your first job is to design a **conceptual** end-to-end data pipeline to solve this problem.

This conceptual data pipeline will lay the foundation for building the actual pipeline. So everything at this stage should be about concepts not tools.

To help you achieve success on this project, your manager **Najeeb** has provided you with the following guiding questions to help shape your design thinking.

Steps	Questions
1. Source Identification	<ul style="list-style-type: none"> - What are the data sources? - What formats and frequency (batch or streaming or both)?
2. Ingestion Strategy	<ul style="list-style-type: none"> - Will you use API ingestion, file uploads, or streaming? - How will real-time data (e.g. Twitter) be handled?
3. Processing/Transformation	<ul style="list-style-type: none"> - How will you clean and standardize the data? - How will you classify complaints into categories?
4. Storage Options	<ul style="list-style-type: none"> - Do you need a data lake? A warehouse? Both? - What format will the cleaned data be stored in (Parquet, JSON, etc.)?
5. Serving	<ul style="list-style-type: none"> - How will the data be queried? - How will the downstreams users use this data?
6. Orchestration & Monitoring	<ul style="list-style-type: none"> - How often will this pipeline run? - How will failures be detected or notified?
7. Dataops	<ul style="list-style-type: none"> - Where will the pipeline run? - How will you make it available in production?

By the end of this task, you should have a solid blueprint required for building the actual pipeline.

Note: Because this is an imaginary company, it is fine for you to make assumptions, just be clear and communicate any assumptions you make.

Expected Output

- A diagram of the conceptual pipeline (you can use tools like draw.io, lucid.app or anything you are comfortable with).
- A not more than 2-page written explanation covering:
 - Design choices
 - Assumptions/thought process
 - Challenges or unknowns
 - Any other information
- Don't mention any tools.

Duration

1 week (28/08/2025)

Mode of Submission

Share a link to your solution. This can be a GitHub readme, PDF, presentation slide etc.

The submission form will be provided.