

Nakshatra Rane
TY AIEC 2
24-08-2025

ADT23SOCB0630
Roll no: 14
Data Engineering

Case Study

Mini Project: Data Lifecycle for Healthcare – *Patient Monitoring & Hospital Resource Utilization*

Problem Statement:

Hospitals face challenges in monitoring patients' health parameters, tracking hospital resource usage (beds, ventilators, medicines), and making real-time decisions. A data lifecycle approach helps in capturing, processing, storing, analyzing, and visualizing this information for better healthcare management.

Data Lifecycle Stages:

Data Capture (Collection):

- Sources: IoT-enabled medical devices (BP monitors, heart rate sensors, ventilators), hospital ERP system, patient EHRs.
- Method: APIs from hospital systems, CSV exports, IoT streams (MQTT/Kafka).

Data Storage:

- Database: MySQL/PostgreSQL for structured patient and hospital data.
- Data Lake: Azure Data Lake / AWS S3 for unstructured logs (sensor readings, medical reports).

Data Cleaning & Transformation (ETL):

- Tools: Python (Pandas, NumPy), Azure Data Factory / Talend.
- Tasks: Handle missing values (e.g., empty BP readings), normalize units (Celsius → Fahrenheit), remove duplicates.

Data Processing & Integration:

- Integration of patient vitals + hospital resource data into a central warehouse (Snowflake/BigQuery).
- Real-time streams handled using Apache Kafka or Azure Stream Analytics.

Data Analysis:

- Predictive analytics for patient risk scores (ML models using Python/Scikit-learn).
- Trend analysis of medicine consumption and bed occupancy.

Data Visualization:

- **Power BI Dashboard** showing:
 - Patient vitals (live monitoring of heart rate, oxygen, BP).
 - Hospital occupancy (beds, ICU, ventilators availability).
 - Medicine stock levels & predicted shortages.
 - KPI alerts (patients at high risk, occupancy > 90%).

Data Governance & Security:

- HIPAA-compliant encryption of patient data.
- Role-based access: Doctors → patient vitals, Admins → hospital resource data.

Data lifecycle stages from raw collection to secure governance

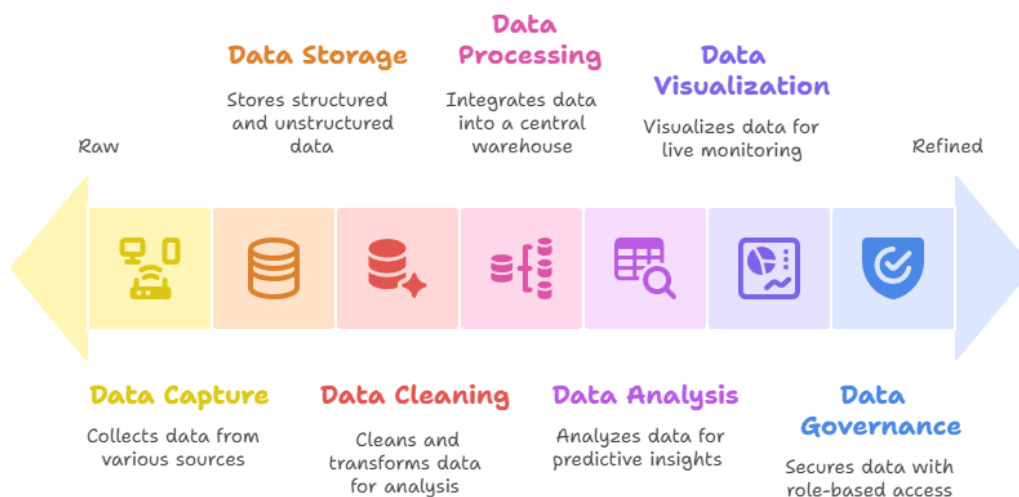


Image 1: Data lifecycle stages

Proposed Mini Project Workflow:

- Step 1 – Capture:** Collect live patient vitals (IoT + hospital records).
- Step 2 – Store:** Load into SQL database + Data Lake.
- Step 3 – Clean/Transform:** Use Python ETL pipeline.
- Step 4 – Process:** Merge patient + hospital resource data.
- Step 5 – Analyze:** Apply predictive analytics for risks & shortages.
- Step 6 – Visualize:** Build Power BI dashboards for doctors & admins.

Mini Project Workflow for Healthcare Data

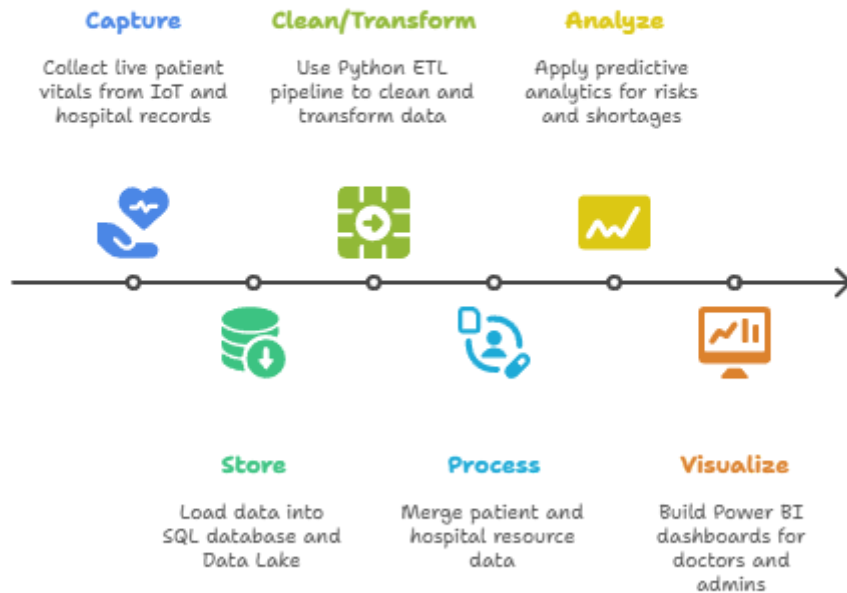


Image 2: Mini Project Workflow

Expected Outcomes:

- Real-time monitoring of patient health.
- Efficient hospital resource management.
- Early alerts for critical patients and medicine shortages.
- Data-driven decision-making improving patient care & hospital efficiency.