

# DEPI – IoT Data Pipeline Project Documentation

## 1. Project Planning

### 1.1 Project Title

Real-Time IoT Data Pipeline for Sensor Monitoring

### 1.2 Project Overview

This project presents the design and implementation of a complete end-to-end IoT data pipeline capable of generating, processing, storing, and visualizing real-time sensor data.

The system simulates temperature and humidity readings, performs ETL cleaning, streams the data through Kafka, stores it in Azure SQL Database, and provides real-time insights through a Power BI dashboard.

The solution demonstrates modern data engineering practices, including data streaming, batch/real-time processing, database modelling, and dashboard analytics.

### 1.3 Problem Statement

Organizations handling IoT devices need a reliable pipeline to ingest sensor data, detect abnormal readings, and visualize system health in real time.

This project solves:

- Lack of centralized real-time monitoring.
- Difficulty detecting anomalies automatically.
- Absence of unified storage for historical data.

### 1.4 Project Objectives

- Develop a reliable IoT data generator for continuous sensor simulation.
- Build an ETL pipeline to clean and validate incoming readings.
- Implement Kafka for scalable real-time data streaming.
- Build a consumer that processes messages and generates alerts.
- Store processed sensor data in an Azure SQL cloud database.
- Design a Power BI dashboard for real-time visualization and analysis.

### 1.5 Technologies Used

- **Python** (Data generation, ETL, processing)
- **Docker** (Containerized producer, consumer, Kafka)
- **Kafka** (Messaging and real-time streaming)
- **Azure SQL Database** (Cloud storage)
- **Power BI** (Dashboard visualization)
- **Figma** (UI/UX wireframes & ERD)

---

## 2. Stakeholder Analysis:

### 2.1 Internal Stakeholders:

Stakeholder	Role	Interest	Impact
Team Members	Development & testing	High	High
Team Leader	Coordination, delivery quality	Very High	High
Instructor / Mentor	Guidance & evaluation	Medium	High

### 2.2 External Stakeholders:

Stakeholder	Role	Interest	Impact
DEPI Evaluators	Assess the project	High	High
DEPI Management	Ensure learning outcomes	High	Very High

---

## 3. Database Design

### 3.1 Database Description

The database stores both raw and processed sensor readings in a structured format, enabling analytics, reporting, and anomaly detection.

### 3.2 Schema Structure

#### SensorData Table

Column	Type	Description
id	INT (PK)	Unique record ID
timestamp	DATETIME	Reading timestamp
temperature	FLOAT	Measured temperature
humidity	FLOAT	Measured humidity
status	NVARCHAR(50)	Normal / Alert

### AlertLogs Table

Column	Type	Description
alert_id	INT (PK)	Unique alert ID
sensor_id	INT (FK)	Linked reading from SensorData
alert_message	NVARCHAR(100)	Alert description
alert_time	DATETIME	Time of alert generated

### 3.3 ERD Diagram (Figma Link)

<https://www.figma.com/design/1tYSHGD2XNwaJZTkzNzLg2/IoT-Database-Design---ERD>

---

## 4. UI/UX Design

### 4.1 Dashboard Vision

The Power BI dashboard provides a real-time monitoring interface for IoT data, giving quick insights about environmental conditions and anomaly detections.

### 4.2 Dashboard Features

- Time-series temperature tracking
- Time-series humidity analysis
- KPI cards (Average Temperature, Average Humidity, Alerts Count)
- Alert distribution visualization
- Filters for date/time and device selection
- Minimalistic layout suitable for live monitoring

### 4.3 Design Requirements

- Clear, readable typography
- Modern dark-mode interface
- Logical grouping of charts and KPIs
- User-centered layout with intuitive navigation

### 4.4 Figma UI Link

<https://www.figma.com/design/UVpyJMRvIXmzIBgz9s4fVQ/IoT-Dashboard---UI-UX-Design>

---

## 5. System Architecture (Optional Section)

*This section enhances your PDF but is not mandatory.*

Pipeline flow:

Generator → ETL → Producer → Kafka → Consumer → Azure SQL → Power BI Dashboard

---

## 6. Conclusion

The IoT Data Pipeline project demonstrates practical data engineering skills through real-time data ingestion, processing, cloud storage, and interactive visualization. This system can be used as a foundation for larger IoT platforms requiring scalable and reliable monitoring solutions.