Team Insight Hub

Ghaidaa Hisham Ibrahim (Team Leader)

Abdelrahman Khaled Abdelrahman

Esraa Gebreal Abd Al-Gawad

Fatma Sief El-naser

Seif Ismail Ahmed

Bassant Hesham

Abstract

Insight Hub is a data engineering project designed to build a fully automated and scalable ETL pipeline for collecting, processing, and classifying data about sensors used in automated vehicles. The project provides an accessible and organized platform that supports students, researchers, and automotive companies in identifying and comparing sensor technologies such as LiDAR, radar, ultrasonic, and camera modules.

Using Microsoft Azure's data ecosystem, the project automates every stage of the ETL process — from ingestion to visualization — following the Bronze–Silver–Gold architecture. The final deliverable will be a centralized, continuously updated Power BI dashboard that provides real-time analytics and sensor insights, supporting both educational and industrial applications.

Introduction

Autonomous vehicles rely on an array of advanced sensors for environment perception, localization, and decision-making. However, information about these sensors — their specifications, use cases, and compatibility — is scattered across different manufacturers and databases.

This creates a major barrier for students, researchers, and developers who need reliable, structured data to support innovation.

Insight Hub aims to bridge this gap by developing an automated data pipeline that collects, cleans, classifies, and visualizes information about automotive sensors. The solution leverages **Azure cloud tools** to ensure scalability, automation, and accuracy in every stage of data handling.

Problem Statement

Currently, there is no unified platform that provides organized and up-to-date information about sensors used in autonomous vehicles. Most available data is fragmented, presented in inconsistent formats, or requires manual collection.
This fragmentation slows down research, development, and academic work, especially for students who lack access to centralized technical resources.
Therefore, this project focuses on creating a cloud-based ETL system that automates data ingestion, transformation, and visualization to simplify sensor analysis and decision-making.

Project Description

Insight Hub will create a fully automated, end-to-end data engineering pipeline using Microsoft Azure. The project follows a three-layer ETL architecture (Bronze–Silver–Gold):

- Bronze Layer: Ingest raw sensor data from open datasets, manufacturer APIs, and CSV files.
- Silver Layer: Clean, validate, and standardize data using Azure Databricks and store it in Azure SQL Database.
- Gold Layer: Perform advanced analytics using Azure Synapse and visualize results with Power BI.

This structure ensures reliability, scalability, and easy maintenance. The final dashboard will allow users to search, filter, and compare sensors by type, accuracy, range, cost, and function — providing valuable insights for both academic and professional users.

Group Members & Roles

Name	Role	Responsibilities
Ghaidaa	Team Leader / Data Architect	Oversee all project phases, design the ETL pipeline structure, manage team coordination, and ensure Azure services are properly integrated.
Abdelrahman	Data Ingestion Engineer (Bronze Stage)	Handle data sourcing and ingestion from APIs, CSVs, and web datasets using Azure Data Factory and Blob Storage.
Fatma	Data Transformation Engineer (Silver Stage)	Clean, transform, and validate data using Azure Databricks; ensure consistency and format standardization.
Esraa	Database Engineer	Design and manage Azure SQL Database schemas, optimize storage performance, and ensure data reliability.
Sief	Data Analytics & Visualization Engineer (Gold Stage)	Develop dashboards and analytics using Azure Synapse Analytics and Power BI; interpret insights for end users.
Bassant	Quality Assurance & Documentation Specialist	Conduct testing, track project milestones, maintain documentation, and ensure project deliverables meet quality standards.

Objectives

General Objective:
To design and implement an automated, scalable data engineering pipeline for collecting and classifying sensor data used in autonomous vehicles.

Specific Objectives:

1. Implement a full ETL pipeline based on the Bronze–Silver–Gold architecture.
2. Collect and organize data from multiple automotive sensor sources.
3. Clean, transform, and store structured data using Azure Databricks and SQL Database.
4. Develop interactive Power BI dashboards for visualization and comparison.
5. Automate the entire workflow for scalability and real-time updates.

Tools & Technologies

Category	Technology / Tool	Purpose
Cloud Platform	Azure Resource Manager	Manage cloud resources and deployment
Data Storage	Azure Blob Storage	Store raw (Bronze) sensor data
Data Processing	Azure Databricks	Transform and clean data (Silver stage)
Database	Azure SQL Database	Store standardized data

Category	Technology / Tool	Purpose
Data Orchestration	Azure Data Factory	Automate and schedule ETL workflows
Data Warehouse	Azure Synapse Analytics	Analytical processing and reporting
Visualization	Power BI	Create dashboards and visual insights
Development & Collaboration	Python, SQL, GitHub	Scripting, queries, and version control

Procedures

The project will progress through **three major ETL stages**, each with specific goals and deliverables:

Stage 1 – Bronze (Data Ingestion)

Description: Collect raw sensor data from multiple open sources and APIs.
Tools: Azure Resource Manager, Azure Blob Storage, Azure Data Factory
Milestones:

- Week 1–2 → Azure setup and Blob Storage configuration
- Week 3 → Automated ingestion pipelines using Data Factory

Stage 2 – Silver (Data Cleaning and Transformation)

Description: Standardize, clean, and integrate the data for consistency.
Tools: Azure Databricks, Azure SQL Database
Milestones:

- Week 4–5 → Develop cleaning scripts and transformation logic
- Week 6 → Store structured data in SQL Database

Stage 3 – Gold (Data Analytics and Visualization)

Description: Perform analysis and visualization of final data.
Tools: Azure Synapse Analytics, Power BI
Milestones:

- Week 7–8 → Create analytical models in Synapse Analytics
- Week 9 → Develop and deploy Power BI dashboards

Milestones & Deadlines

Milestone	Description	Deadline
1. Project Initialization	Environment setup and resource configuration	Week 1–2
2. Data Ingestion (Bronze)	Connect data sources and automate ingestion	Week 3
3. Data Cleaning (Silver)	Clean and transform data using Databricks	Week 4–6
4. Database Integration	Load structured data into Azure SQL Database	Week 6
5. Analytics & Visualization (Gold)	Develop analytical models and Power BI dashboards	Week 7–9
6. Final Review	Testing, documentation, and project submission	Week 10

Conclusion

The **Insight Hub** project demonstrates the practical application of modern data engineering concepts by building a **fully automated and scalable cloud-based pipeline**. By integrating various Azure services through the **Bronze–Silver–Gold ETL architecture**, the project ensures reliable data collection, transformation, and visualization.
The final product — a Power BI dashboard for automotive sensor data — will support academic research, industrial innovation, and future development in intelligent transportation systems.