Cloud-Native Air Pollution & Respiratory Death Analysis

IEEE-Style Project Documentation and Case Study

Muhammad Haider Abbas | MSc Computer Science & Technology | Ulster University Birmingham Campus



Figure 1. Project visual identity used in the portfolio website.

Abstract

Cloud-Native Air Pollution & Respiratory Death Analysis is documented as a professional portfolio case study by Muhammad Haider Abbas. The project demonstrates practical capability in Data / Cloud through a structured problem-solution design, implementation direction, testing plan and future development roadmap. The report explains the business or academic motivation, system requirements, architecture, workflow, quality considerations and expected professional value of the project. It is prepared for use as downloadable portfolio evidence for recruiters, clients and academic reviewers.

Keywords: Data / Cloud, Academic Project, Data Engineer, Azure, PySpark, Big Data, Analytics

I. Project Profile

Field	Details
Project Title	Cloud-Native Air Pollution & Respiratory Death Analysis
Category	Data / Cloud
Year	2026
Role	Data Engineer
Status	Academic Project
Tools	Azure Data Lake Gen2, Synapse Spark, PySpark, Plotly
Portfolio Slug	cloud-air-pollution-respiratory-analysis

II. Introduction and Problem Context

This document presents the technical and professional documentation for Cloud-Native Air Pollution & Respiratory Death Analysis. The project is positioned as a academic project in the Data / Cloud category. The work highlights the author's ability to combine technical implementation, documentation discipline and outcome-focused project thinking. The project summary is: A cloud data project analysing average PM2.5 pollution and respiratory deaths using Azure and PySpark.

A. Problem Statement

Environmental and health datasets need scalable processing before meaningful insights can be extracted.

B. Proposed Solution

Raw CSV data is ingested, transformed with Spark, stored as Parquet and analysed through country-level aggregation and visualisation.

III. Aim and Objectives

To design a cloud-native analytics pipeline for exploring relationships between PM2.5 air pollution and respiratory mortality indicators.

- Ingest environmental and health data into cloud storage.
- Clean and transform data using scalable Spark processing.
- Store analytical outputs in efficient formats such as Parquet.
- Visualise trends and correlations for public-health interpretation.

IV. Requirements Analysis

ID	Requirement	Type	Priority
FR-01	Azure Data Lake storage	Functional / project requirement	Required
FR-02	Synapse Spark processing	Functional / project requirement	Required
FR-03	PySpark cleaning and joins	Functional / project requirement	Required
FR-04	Country/year aggregation	Functional / project requirement	Required
FR-05	Plotly dashboard direction	Functional / project requirement	Required
FR-06	Correlation analysis	Functional / project requirement	Required
NFR-01	Usability and clear portfolio presentation	Non-functional	Required
NFR-02	Maintainability and future extensibility	Non-functional	Required
NFR-03	Professional documentation and evidence quality	Non-functional	Required

V. System Architecture

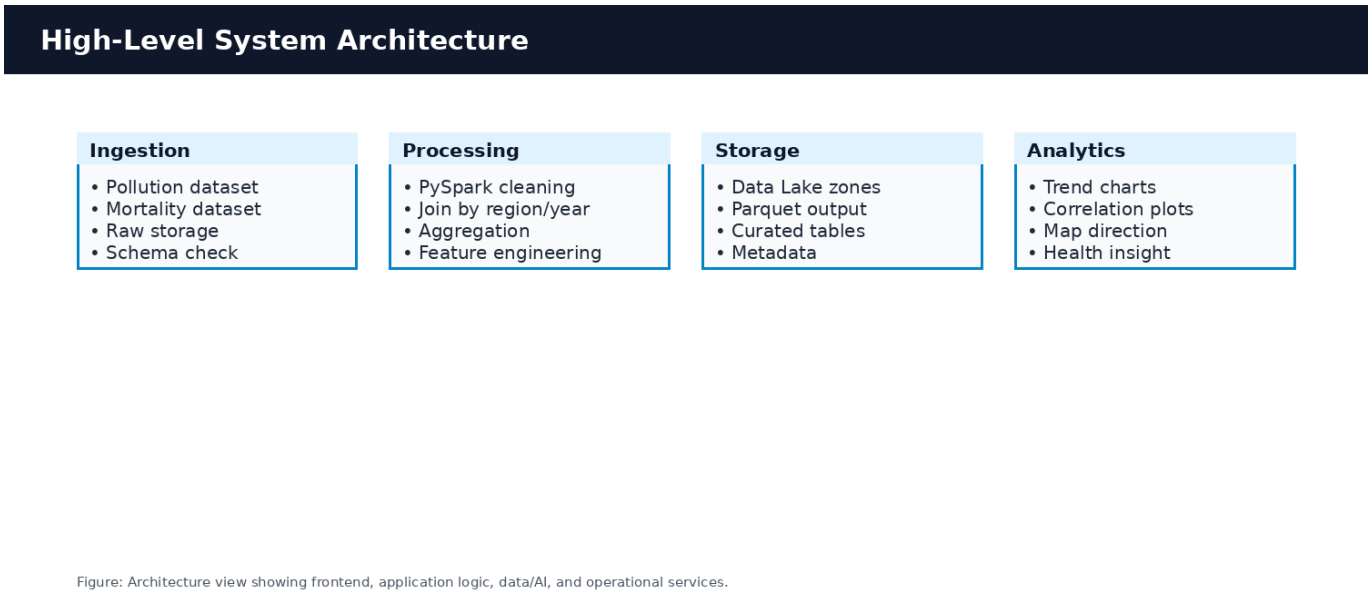


Figure: Architecture view showing frontend, application logic, data/AI, and operational services.

Figure 2. High-level architecture or production pipeline for the project.

The architecture separates user-facing presentation, project logic, data or model processing, and operational output. This structure makes the project easier to explain, test and extend inside a professional portfolio environment.

VI. Workflow Design

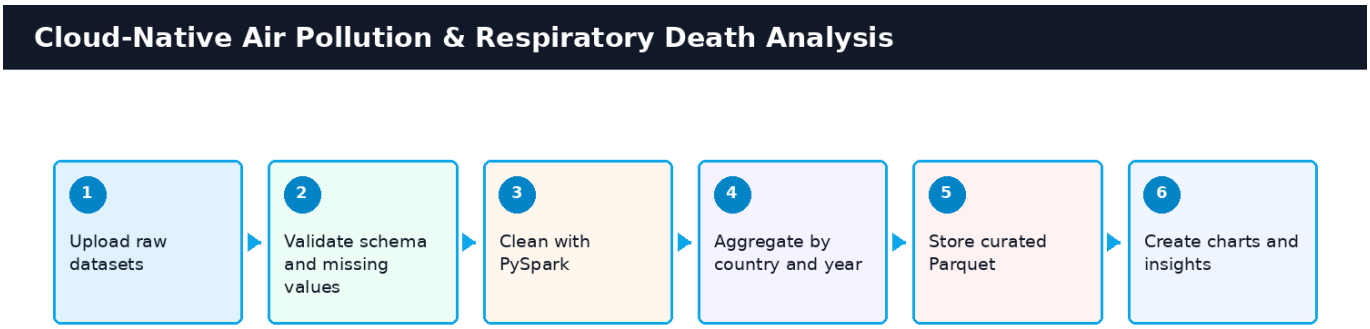


Figure: Generated project workflow / architecture artefact for portfolio documentation.

Figure 3. Project workflow from input or user action to final output.

VII. Data, Components and Implementation View

Component	Layer	Responsibilities	Notes
C-01	Ingestion	Pollution dataset, Mortality dataset, Raw storage, Schema check	Core component group
C-02	Processing	PySpark cleaning, Join by region/year, Aggregation, Feature engineering	Core component group
C-03	Storage	Data Lake zones, Parquet output, Curated tables, Metadata	Core component group
C-04	Analytics	Trend charts, Correlation plots, Map direction, Health insight	Core component group

Implementation should be documented through screenshots, code snippets, database schema extracts, test outputs or demonstration media when available. For deployed or live systems, screenshots and logs should be attached as appendices to increase evidence strength.

VIII. Testing and Validation Strategy

Test Area	Test Method	Expected Result
Schema validation	Check required columns	Expected fields present
Transformation	Run PySpark notebook	Output tables created
Aggregation	Group by region/year	Numeric values consistent
Visualisation	Render Plotly charts	Charts communicate trends

The testing strategy focuses on validating the main user or system journey, confirming that core requirements behave as expected, and recording limitations clearly. This supports honest portfolio evidence and avoids overclaiming features that are still prototypes.

IX. Security, Ethics and Professional Considerations

- Use transparent data sources.
- Clearly separate evidence from assumptions.
- Maintain professional documentation quality.

X. Results, Outcome and Future Work

A. Outcome

A strong big data portfolio project connecting cloud infrastructure with real-world health analytics.

B. Future Enhancements

Add automated ETL pipeline, Power BI dashboard and predictive modelling.

Roadmap Stage	Enhancement Direction
Short term	Add missing screenshots, demo links and downloadable supporting files.
Medium term	Connect live/source URLs and replace placeholder evidence with real artefacts.
Long term	Use analytics, user feedback or formal evaluation results to strengthen the project case study.

XI. References

[1] World Health Organization, "Ambient (outdoor) air pollution," https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health

[2] Microsoft, "Introduction to Azure Data Lake Storage Gen2," https://learn.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-introduction

[3] Microsoft, "Apache Spark in Azure Synapse Analytics," https://learn.microsoft.com/en-us/azure/synapse-analytics/spark/apache-spark-overview

Appendix A. Portfolio Download Mapping

Item	Value
Recommended PDF path	docs/cloud-air-pollution-respiratory-analysis-project-documentation.pdf
Recommended Word path	docs/cloud-air-pollution-respiratory-analysis-project-documentation.docx
Admin field: Documentation	docs/cloud-air-pollution-respiratory-analysis-project-documentation.pdf
Admin field: Extra Project File	docs/cloud-air-pollution-respiratory-analysis-project-documentation.docx