Vishwajeet Vijay Pawar

Data Engineer | AI & ML learning enthusiast

LinkedIn | GitHub | Portfolio | +353 894631972 | vpawar6254@gmail.com | Cork, Ireland(Open to relocate)

Summary:

Data Engineer & **MSc AI** candidate with professional expertise in **Python, Scala,** and **Azure Cloud**. Proven track record at UST Global optimizing **scalable ETL** processes using **Azure Data Factory** and **Databricks**. Leveraging **advanced Big Data**, **NLP**, and **LLM** training to build **high-performance infrastructure** for actionable business insights.

Skills:

Python, Scala, SQL, Java, JavaScript, TypeScript**,** Spark (PySpark, Scala Spark), Hadoop, Airflow, SSIS, ETL Workflows, Microsoft Azure, GCP, Databricks, GitHub**,** ML, Deep Learning, NLP, LLM, Machine Vision, Metaheuristic Optimization**,** MySQL, MSSQL, Cassandra, CosmosDB**,** React.js, HTML, CSS, SCSS, Bootstrap, Microsoft 365(Excel, Word, PowerPoint), Jira

Certificates:

A-Z AI 2025 (Jun 25'), Airflow management (Mar 25'), Cassandra (Nov 24'), Spark with Scala (Aug 24'), GCP for Data Engineering (Oct 24'), Web Development (Dec 22')

Professional Experience:

Developer 1 – Software Engineering, US Technology (UST Global), Thiruvananthapuram, India   Jun 2024 – Jun 2025
- **Pipeline Architecture**: Orchestrated scalable data pipelines for **two major client projects** using Azure Data Factory (**ADF**), automating complex **data ingestion** into **GCP** buckets using **Airflow, reducing manual intervention** by 100%
- **ETL Optimization**: Engineered high-performance transformation workflows on Azure **Databricks** using **Python** (**PySpark**), tuning **Spark** jobs to significantly **reduce processing latency** by 20% and **compute costs**
- **Data Modelling**: Designed efficient **data structures** and executed **advanced SQL** queries against Azure **Cosmos DB** and **MSSQL**, transforming raw data into actionable assets for **business intelligence**

Full Stack Web Developer Intern, KWYK PVT LTD, Pune, India   Mar 2022 – Jun 2022
- **Full-Stack Application Development**: Designed and deployed a **responsive** delivery-focused **web application**, utilizing **HTML5**, **CSS3**, and **Bootstrap** to ensure a **seamless** cross-platform **user experience**
- **Backend & Database Architecture**: Engineered efficient server-side processing logic using **PHP** and managed relational data structures within **MySQL** to **optimize data retrieval** and **backend performance**
- **Workflow Automation**: Implemented **JavaScript** functions to **automate** critical operational **tasks**, including **real-time email** notifications and **dynamic UI** animations, enhancing user engagement

Crew Member, La Tana Pizzeria, Cork, Ireland   Oct 2025 – Dec 2025
- **Operational Management & Inventory Control: Optimized** operational **workflows** by meticulously **managing stock** records and inventory logistics to ensure resource availability, demonstrating **strong organizational skills** and **attention to detail**.
- **Client Relations & Team Collaboration:** Facilitated seamless service operations by utilizing **effective communication** to handle customer inquiries and **coordinate with team members**, ensuring prioritized task management and **high** levels of **customer satisfaction**.

Projects:

Airbnb Insight Vault | Github **|** PySpark, Azure tools, ETL, Databricks, SQL**,** Data Analysis and Visualization
- **End-to-End Cloud Data Architecture**: Architected a scalable data solution on Azure, integrating Blob Storage, Data Lakes, and Cosmos DB to **efficiently process** and store **large-scale** Airbnb **datasets** for **high-performance retrieval**.
- **Automated ETL Pipeline Development**: Engineered robust **ETL pipelines** and **dataflows** using ADF, automating **data ingestion** and **transformation** processes to ensure data **integrity** and **reduce manual latency**.
- **Advanced Analytics & Visualization**: Leveraged PySpark on Azure Databricks to **execute complex data cleaning** and aggregation, developing interactive dashboards that **visualized** key market **trends** and **actionable insights**.

MediSpark Insights | Github **|** ScalaSpark, PySpark, SQL, Databricks, Data Analysis and Visualization
- **Big Data Engineering**: Engineered **high-performance** data **pipelines** using Apache Spark (Scala/PySpark) on Databricks, **optimizing** distributed **computing tasks** for efficient **large-scale dataset processing**.
- **ETL & Insight Generation**: Orchestrated end-to-end ETL workflows to **transform raw data** into structured intelligence, deriving actionable business insights and enabling data-driven **decision-making**.

Education:

M.Sc. in AI, Munster Technological University, Cork, Ireland   **Sep 2025 – Sep 2026**
Modules: Big Data, Metaheuristic Optimization, Machine Learning, Deep Learning, Machine Vision, NLP, LLMs

B.Tech. in CSE, MIT World Peace University, Pune, India (9.01 CGPA)   **Jul 2019 – Jun 2023**
Modules: Big Data, OOPs, Data Structures, Data Warehousing and Management, DBMS, Web Technology, Distributed Computing, AI, UIUX