

OJASSHUKLA

Wollongong, NSW | 0424494110 | ojasshukla01@gmail.com | [LinkedIn](#) | [GitHub](#)

PROFESSIONAL PROFILE

Experienced Senior Data Engineer with over five years of expertise in designing, optimizing, and managing cloud-based data architectures. Specializing in ETL/ELT processes, data modeling, data migration, and automation, I have led large-scale data solutions across startups and mid-sized organizations. I excel in developing cost-efficient, scalable cloud data platforms and implementing infrastructure-as-code on GCP, AWS, and Azure. My expertise extends to data wrangling, data governance, streaming data architectures, and working with Databricks and Snowflake. Adept at stakeholder communication and delivering client-focused solutions, I have successfully driven business intelligence strategies, financial data processing, and enterprise data warehousing.

CORE COMPETENCIES

- Data Pipeline Optimization
- GIS Data Processing
- Data Warehousing and Governance
- Cloud Computing
- Infrastructure-as-code
- Database Design
- Containerization
- ETL Processes
- Automation
- BI and Visualization
- DevOps and CI/CD
- Streaming Data Architectures

WORK EXPERIENCE

INDEPENDENT DATA ENGINEER, WOLLONGONG NSW

FEBRUARY 2025 TO PRESENT

DATA ENGINEER

- Built a Unified Immigration Data Platform, an open-source end-to-end data solution simulating visa lifecycle, coaching analytics, financial support, and post-arrival services. Technologies: GCP (BigQuery, Pub/Sub), dbt, DuckDB, Terraform, FastAPI
- Designed modular data transformation frameworks to support real-time and batch data ingestion, integrating with REST APIs, flat files, and cloud storage
- Deployed event-driven microservices on GCP using FastAPI, Docker, and Cloud Functions, reducing pipeline response time and improving observability
- Created metadata-driven pipeline templates with automated schema validation and audit logging to enable scalable onboarding of new data sources
- Implemented infrastructure-as-code (IaC) using Terraform for cloud-agnostic deployment across AWS and GCP with CI/CD integrations via GitHub Actions
- Trying hands-on with ML workflows, including visa approval prediction and IELTS score analytics using Scikit-learn, Pandas, and BigQuery ML.
- Experimenting with automated model retraining pipelines using dbt snapshots and Cloud Scheduler to simulate real-world deployment flows.
- Exploring LLM-powered data workflows and prompt engineering using LangChain, Vertex AI, and Snowflake Cortex for intelligent data extraction.
- Consulted for a fintech prototype involving loan disbursement analytics and creditworthiness scoring, building pipelines from raw financial applications to actionable dashboards
- Developed reusable dbt models and Jinja macros for transformation layers, boosting code reuse and testing coverage across datasets
- Actively pursuing Google Cloud Professional Data Engineer certification to enhance production-readiness in cloud-native data architectures
- Mentored junior engineers via open GitHub contributions and peer feedback loops to improve code quality, deployment hygiene, and data modeling

ALUMINIUM STEWARDSHIP INITIATIVE (ASI), WOLLONGONG NSW
FEBRUARY 2023 TO JANUARY 2025

DATA ENGINEER

- Built an end-to-end Data Warehouse on GCP, integrating multiple heterogeneous data sources into a centralized database, enabling real-time insights for stakeholders.
- Reduced ETL processing time by 40% by optimizing Python scripts using Pandas, DuckDB, BigQuery and dbt, improving data transformation efficiency.
- Built both batch and stream data pipelines using Python, Apache Kafka, and GCP Pub/Sub, enabling real-time and scheduled data integration from multiple sources (APIs, databases, and flat files).
- Built portable data infrastructure compatible with both AWS and GCP environments using Terraform, Docker, and Kubernetes for vendor-agnostic deployment.
- Led mentorship programs, training junior engineers on data pipeline development, coding best practices, and cloud automation.
- Designed and deployed RESTful APIs using FastAPI for seamless integration between data pipelines and frontend applications, ensuring low-latency access to processed datasets.
- Automated infrastructure deployment using Terraform and Infrastructure-as-Code, reducing manual interventions by 70%.
- Implemented automated data quality validation using Great Expectations, ensuring 99.9% data accuracy across ingestion pipelines.
- Collaborated with finance & product teams to provide actionable insights, reducing operational costs by 15% through optimized query strategies.

UNIVERSITY OF WOLLONGONG/DEMENTIA TRAINING AUSTRALIA (DTA), WOLLONGONG NSW, AUSTRALIA
JUNE 2022 TO JANUARY 2023

DATA ENGINEER

- Developed automated ELT workflows, reducing manual data handling by 80% and ensuring real-time analytics.
- Migrated critical financial and operational data to BigQuery, improving accessibility and performance for stakeholders.
- Integrated Cloud Functions & Apache Airflow to orchestrate scheduled and event-driven transformations.
- Enhanced data modeling techniques to support complex analytics and forecasting.
- Improved dashboard performance by 35% by optimizing SQL queries for Tableau, Power BI, and Looker.

HERE TECHNOLOGIES, MUMBAI, MAHARASHTRA
FEBRUARY 2019 TO DECEMBER 2021

DATA SPECIALIST

- Developed and optimized Hadoop-based ETL pipelines, utilizing AWS services such as S3, EC2, Glue, and Lambda, along with Apache Kafka for real-time data streaming.
- Automated infrastructure management using Terraform and CI/CD pipelines to improve deployment efficiency.
- Implemented Docker and Kubernetes for scalable and resilient data processing, enhancing cloud-based workflows.
- Ensured compliance with industry regulations by strengthening data security and governance.
- Designed and optimized large-scale distributed data processing workflows using Apache Spark, improving data transformation speeds and efficiency.
- Developed Hive-based data warehousing solutions to enhance query performance and structured data accessibility across the organization.
- Managed HDFS-based storage, optimizing data retention and retrieval strategies to improve system efficiency.
- Integrated GIS datasets into Hadoop workflows to enable geospatial analytics and location-based decision-making.
- Leveraged Tableau and Power BI to create interactive dashboards, supporting business intelligence and reporting needs.

EDUCATION

MASTER OF COMPUTER SCIENCE, 2022

University of Wollongong, Wollongong NSW, Australia

BACHELOR OF COMPUTER SCIENCE, 2017

Mumbai University, Mumbai, India

DIPLOMA IN COMPUTER TECHNOLOGY, 2013

Maharashtra State Board, Mumbai, India

PROJECTS AND OPEN SOURCE CONTRIBUTIONS

1. Open Compliance ESG

[GitHub Repository](#)

Designed and implemented an open-source platform to automate ESG (Environmental, Social, and Governance) metric processing and compliance validation.

Technologies: Python, Streamlit, FastAPI, DuckDB, dbt, GitHub Actions

2. Data Engineering Case Studies

[GitHub Repository](#)

Compiled modular projects showcasing best practices in ETL pipeline design, transformation logic, and orchestration across varied domains.

Technologies: Python, dbt, Apache Airflow, SQL, BigQuery

3. SOP Generator Web App

[Frontend](#) | [Backend](#)

Developed a full-stack tool to generate Statements of Purpose (SOPs) dynamically based on user inputs, primarily for academic applications.

Technologies: FastAPI, React.js, Docker, GitHub Actions

4. Health Analytics BI Dashboard

[GitHub Repository](#)

Built an interactive dashboard to visualize and analyze healthcare datasets for operational and strategic insights.

Technologies: Power BI, SQL

5. Web-Based Intrusion Detection System (Master's Project – University of Wollongong)

[GitHub Repository](#)

Created a lightweight IDS platform capable of flagging suspicious web traffic based on ML classification.

Technologies: Python, Flask, Scikit-learn, HTML/CSS

6. SQL Injection Simulation (Master's Project – University of Wollongong)

[GitHub Repository](#)

Simulated and documented common SQL injection scenarios and mitigation strategies within a sandboxed learning environment.

Technologies: PHP, MySQL, JavaScript

AFFILIATIONS AND CERTIFICATIONS

- Google Cloud Certified – Professional Data Engineer (In Progress)
- Organized State-Level Robowars Competition (LTCoE, 2016-2017)
- Core Committee Member at Tech-Zephyr (2015-2017)

VOLUNTEERISM

- Community Development Volunteer, Rashtrajyot (NGO): Contributed to various community development initiatives focused on education, healthcare, and sustainable development projects.

ADDITIONAL INFORMATION

Languages: English, Hindi, and Marathi

Technical Proficiencies:

- **Languages:** Python, SQL, JavaScript
- **Cloud & Databases:** GCP (BigQuery, Dataflow, Pub/Sub), AWS (Redshift, S3, Lambda, Glue), PostgreSQL, DuckDB
- **ETL & Automation:** Terraform, dbt, Apache Airflow, Cloud Functions, Pandas
- **Analytics & Visualization:** Tableau, Power BI
- **DevOps & Infrastructure:** Docker, Kubernetes, CI/CD, Kafka
- **Geospatial Analytics:** QGIS, ArcGIS, Spatial Databases, Route Optimization

Interests: Literature (fiction, poetry), competitive gaming matches (DOTA2, strategy games), badminton (doubles, recreational), swimming (freestyle, endurance).