

Nauman Shahid

Principal Data Engineer

contact@nauman.cc · +971 50 730 6079 · Abu Dhabi, UAE · Singaporean · linkedin.com/in/nshah1d · github.com/nshah1d

PROFESSIONAL SUMMARY

Principal Data Engineer with over a decade of experience engineering production-grade ELT pipelines, normalising complex multi-source schemas, and delivering scalable business intelligence systems across enterprise and SME engagements. Technical practice is built on consolidating fragmented operational data into automated pipeline architectures that eliminate manual reporting cycles. System design is consistently underpinned by three core principles: first-principles design, privacy-first data governance, and complete data sovereignty. Professional development is building toward AI systems architecture, with a longer-term focus on explainable infrastructure, transparent RAG pipelines, and auditable intelligent systems.

CORE COMPETENCIES

Data Pipeline Engineering	Python, SQL, Apache Kafka, Apache Spark, Apache Airflow, ELT/ETL Architecture, Batch & Streaming Pipeline Design, CI/CD (GitHub Actions), pandas, SQLAlchemy, Idempotent Loading
Data Architecture & Modelling	PostgreSQL, DuckDB, BigQuery, MySQL, NoSQL, Schema Normalisation, Query Optimisation, Partitioned Tables, Offline-First Architecture, Physical Data Modelling
Business Intelligence	Power BI, Tableau, KPI Dashboard Design, Statistical Modelling, ydata-profiling, Data Profiling, scikit-learn, NumPy, scipy
Data Quality & Governance	Privacy-First Data Governance, Data Health Auditing, Deterministic Deduplication, Heuristic Field Mapping, Master Data Management (MDM), Regular Expressions
AI Systems & Development	Prompt Engineering, RAG Pipeline Concepts, Generative AI Workflows, Chain-of-Thought Prompting, Flask, Django, Testing Frameworks

PROFESSIONAL EXPERIENCE

Principal Data Engineer · MindFlex Jan 2026 – Present

- Architect and deliver Python ELT pipelines consolidating 4–6 fragmented SME data sources into centralised PostgreSQL and DuckDB analytical layers, automating reporting cycles and eliminating an average of 12 hours of weekly manual compilation per engagement.
- Execute Data Health Audits using ydata-profiling to systematically identify and remediate structural data quality issues, resolving schema anomalies and duplication errors affecting 30–40% of ingested historical records per dataset.
- Design and deliver sector-specific Power BI dashboards processing 150,000+ monthly transaction records across the active client portfolio, synthesising fragmented POS, inventory, and accounting data into automated executive intelligence.

Data Systems & Operations Manager · Private Family Office Apr 2022 – Oct 2025

- Built a self-hosted reader for exported chat archives that never guesses a date or drops a line, joins overlapping exports with line-level provenance, and searches 500,000 messages in under half a second.
- Built a loss-preserving contact data platform for VCF and CSV libraries that patches only edited properties, searches 100,000 contacts in the browser and turns concurrent writes into detected conflicts.
- Designed and enforced offline-first data governance standards across sensitive executive operations, maintaining complete data sovereignty with zero reliance on third-party cloud infrastructure across 3.5 years of continuous multi-territory operation.

Academic Tutor · The University of Sydney Jul 2019 – Dec 2021

- Instructed undergraduate cohorts in Object-Oriented Programming and Calculus, translating abstract computational theory into applied algorithmic and data structure skills.
- Designed rigorous analytical problem sets to evaluate data structure implementation and algorithmic efficiency, building systematic frameworks for assessing technical competency.
- Conducted structured code reviews providing actionable feedback on logic, efficiency, and correctness, training cohorts to write reliable, maintainable software against defined specifications.

Business Analyst · Birdi Oct 2018 – Mar 2021

- Contributed technical architecture definitions and data migration feasibility assessments to a successful \$1.2M NSW Spatial Services contract bid, translating complex spatial data requirements into a structured commercial framework.
- Formulated data-driven financial models evaluating hardware integration costs and projected system loads, supporting critical pricing and scoping decisions across multiple contract submissions.
- Executed market data analysis and technical feasibility case studies to align core platform capabilities with enterprise spatial applications, directly informing product strategy across the fiscal year.

PROFESSIONAL EXPERIENCE (continued)

- Business Analyst Intern · Fleet Space Technologies

Jun 2018 – Sep 2018

 - Authored the technical architecture sections of a Victorian Government IoT trial submission, synthesising hardware constraints, connectivity data, and deployment specifications into a coherent commercial proposal.
 - Conducted technical analysis of IoT connectivity constraints across four agricultural trial regions, quantifying Portal device deployment requirements and pricing models for the government submission.
- Python Developer Intern · NCSOFT West

Jan 2017 – Jun 2017

 - Engineered Python parsing algorithms to extract and filter raw game server telemetry, transforming high-volume JSON event streams into structured, analytically consumable datasets.
 - Designed backend data schemas to standardise incoming telemetry structures, enabling reliable extraction and filtering of sensitive metadata from high-volume server event logs.
- Digital Solutions Engineer · Creative Graphics

Mar 2013 – Dec 2016

 - Managed the full technical lifecycle of business intelligence tools across a client portfolio spanning nearly four years, transforming multi-year market research datasets into interactive dashboards and executive reporting frameworks.
 - Translated complex client briefs into functional data architectures, building digital products capable of ingesting, processing, and presenting dynamic operational datasets to stakeholder specification.

TECHNICAL PROJECTS

- F1 Analytics: Post-Race Engineering Analysis Platform

Nov 2024 – Present

Python 3.12, PostgreSQL, FastF1, pandas, NumPy, SQLAlchemy

 - Developing a Formula 1 analysis platform that turns race-weekend timing, telemetry and race context into an evidence-led post-race engineering debrief, through a Race Debrief dashboard and a cross-event Season Index.
 - Engineered an evidence-preserving race-data warehouse: a season-partitioned PostgreSQL star schema with idempotent, single-transaction session loads and resumable backfills, preserving lap validity, purpose, traffic and lineage.
- WhatsApp Archive Viewer: Provenance-Preserving Chat Archive Platform

Feb 2025 – Present

JavaScript (ES modules, Web Workers), PHP 8.2

 - Engineered a strict parser for iOS and Android WhatsApp exports that never guesses an ambiguous date or drops a line, and reconciled overlapping exports with exact source-line provenance.
 - Scaled in-browser reading with a Web Worker engine, loading 500,000 messages in under 5 seconds and searching full history in under 0.5 seconds, measured by an automated Chromium gate.
- Contacts Hub: Loss-Preserving Contact Data Platform

Aug 2025 – Present

JavaScript (ES modules, Web Workers), PHP 8.2

 - Engineered a loss-preserving vCard 3.0/4.0 and CSV editing engine that patches only edited properties into the original bytes, with in-browser search across 100,000 contacts in under 0.6 seconds.
 - Built revision-checked PHP persistence (exclusive lock, SHA-256 base-revision check, exact backup, atomic rename and read-back) that turns lost or concurrent writes into detected conflicts.
- Covert Messenger: Concealed Private Messaging Platform

Mar 2026 – Present

JavaScript (ES modules), PHP 8, MySQL 8

 - Engineered an ordered change feed in which every send, edit, deletion and reaction commits under one database lock, so both clients converge without a missed or reordered change, verified by a 104-check PHP gate.
 - Built the chat behind a working Formula 1 dashboard, with a stateless HMAC-SHA256 device-bound token, a PBKDF2-hashed entry lockout and no chat data left in the browser once the tab closes.

EDUCATION

- Bachelor of Science in Computer Science

In Progress

University of the People, USA
- General Certificate of Education (Advanced Level)

2007

Tanglin Trust School, Singapore

CERTIFICATIONS

Database Engineer Professional Certificate	Meta	Jun 2026
Python Development Professional Certificate	Microsoft	May 2026
Business Intelligence Professional Certificate	Google	Mar 2026
Prompting Essentials	Google	Feb 2026
Data Science Professional Certificate	IBM	Feb 2026