Sujal Dhungana

Perth, Australia ✉ sujal@dataprofessionacademy.com 📱 0410677503 in/dhunganasujal ☁ sujal.datasparta.com

SUMMARY

Senior Data Engineer with deep expertise in designing, building, and automating robust, petabyte-scale distributed data systems. Specialising in writing production-quality software to manage the lifecycle of core data services like Kafka, OpenSearch, and Kubernetes. Deeply committed to "Keep it simple", "Everything as Code", and "Own what you build" principles, empowering teams to deliver secure, scalable, and maintainable pipelines that have a real-world impact.

SKILLS

Key Skills: Data Engineering, DataOps, Data Infrastructure Management, Data Warehousing, Data Modelling, Event-Driven Architecture, Distributed Systems

Languages: Python, SQL, PySpark, R, C, C++, Bash, JavaScript

Platforms & Data Orchestration: Amazon Web Services (EKS, ECS, Lambda, Step Functions), Databricks, Kubernetes, ArgoCD, Helm, Ansible

Data Processing & Warehousing: Python, SQL, AWS Glue, DBT, ClickHouse, Snowflake, Apache Iceberg, AWS MSK(Managed Kafka), AWS Opensearch, DynamoDB, AWS RDS

GitOps & Infrastructure as Code: Cloudformation, CDK, Terraform, Databricks Asset Bundle, CI/CD, Azure DevOps, Gitlab, Artifactory

Monitoring & Visualisation Tools: Prometheus, Grafana, PowerBI, Quicksight, Qlik

EXPERIENCE

Senior Data Engineer
Western Power   September 2025 - Present, Perth, WA

- **Engineered a large-scale data architecture** to process massive volumes of network telemetry, demographic, and meteorological datasets. Designed and maintained over 1,000 interdependent dbt models to clean, transform, and feed data into time-domain machine learning models for 10-year network demand forecasting.
- **Championed an "everything-as-code" and CI/CD culture** across the team. Replaced error-prone, manual UI development with Databricks Asset Bundles and dbt, enabling automated, seamless deployment of pipelines and scheduled jobs across multiple environments.
- **Built internal developer tooling and helper scripts** to automate enterprise data governance. Authored a comprehensive Data Engineering Starter Guide that went beyond documentation, providing the codebase required to make the right engineering practices the easiest path for developers.
- **Led technical onboarding and team mentorship**, successfully scaling the data engineering function from 1 to 3 engineers. Instilled a culture of pragmatic excellence, code quality, and end-to-end ownership.

Resident Platform Engineer (Contractor)
Telstra   October 2024 - August 2025, Perth, WA

- **Architected and operated an event-driven data ingestion platform (AWS MSK / Kafka)** processing over 1TB/hour of critical security telemetry. Engineered near-real-time pipelines using Vector to denoise, transform, and extract essential information from high-volume logs before routing to Clickhouse and OpenSearch sinks, with minimal impact on source-to-sink latency.
- **Reduced compute costs by over 50% (saving >$600,000 annually)** without impacting system reliability. Achieved this by decoupling I/O-bound and CPU-bound operations, optimising Kafka components, and implementing event-based scaling for Kubernetes resources.
- **Secured the end-to-end data architecture** by enforcing strict VPC network isolation, mandatory TLS encryption, SASL/SCRAM client authentication, and securing all management APIs with OAuth2.
- **Built internal Python tooling and RESTful APIs** to automate the provisioning and management of Kubernetes applications, reducing secure data pipeline creation time from days to hours.
- **Maintained platform uptime above SLA targets** for mission-critical pipelines through rigorous change management and automated rolling updates. Led on-call incident response, conducted deep-dive root cause analysis, and authored Standard Operating Procedures (SOPs) to prevent data loss and recovery from incidents.

Principal Data Engineer
Data Vanguards Pty Ltd.   April 2024 - September 2025, Perth, WA

- **Architected a warehouse-agnostic Data Platform (Redshift, Athena, Snowflake)** as Technical Lead. Built OAuth2-secured APIs and integrated pipeline metadata to enable context-aware AI agents to securely query and generate analytical reports.
- **Developed a high-frequency, serverless data ingestion pipeline** for IoT telemetry using AWS Lambda, SQS, DynamoDB, and Redshift. Enabled near real-time analytics while enforcing least-privilege IAM security and implementing structured logging for end-to-end traceability.

- **Standardised data ingestion by implementing a Medallion architecture.** Built reusable, modular pipelines that reduced the time required to onboard new data sources by 60% while significantly lowering ongoing platform maintenance.
- **Redesigned a legacy AWS ingestion pipeline** to resolve performance bottlenecks, reducing cloud compute costs by 90% (saving >$30,000 annually). Established an end-to-end observability stack using Prometheus, Grafana, and custom CloudWatch metrics to monitor system health and validate savings.
- **Engineered a fully automated, end-to-end MLOps pipeline** to train, deploy, and operate machine learning models, delivering real-time item recommendations for a large-scale national auction platform.

Founder | Instructor

Data Profession Academy   **August 2023 - Present, Perth, WA**
- **Founded and instructed a 12-week data analytics bootcamp**, teaching Python, advanced SQL, and data modeling to 20+ students. Mentored the cohort through complex technical projects, resulting in the majority securing first industry data roles within 3 months.
- **Presented at the Perth Data Engineering Meetup** on "Building Reusable Pipelines," sharing technical frameworks for designing simple, maintainable data architectures that reduce engineering overhead and improve scalability.
- **Spoke at the Google Developer Group (GDG) Perth** on practical AI integration, demonstrating how software engineers can embed AI tools into daily development workflows to solve complex technical challenges.
- **Organised and hosted online webinars and 1-day technical workshops**, independently creating curriculum to teach foundational data engineering and analytics skills to the broader tech community.

Senior Data Engineer

Water Corporation   **May 2022 - April 2024, Perth, WA**
- **Engineered an automated MLOps pipeline (AWS SageMaker, Step Functions, CodeBuild)** to deploy and operate machine learning models for real-time data streams. Productionized a predictive model (awarded top paper at OzWater 2024), reducing deployment time from days to minutes while enforcing strict cross-account deployment governance.
- **Designed a config-driven data ingestion platform** to load diverse data sources into Apache Iceberg on S3. Built automated dimensional modelling and Slowly Changing Dimensions (SCD), cutting pipeline creation time from 2 days to 2 hours and reducing cloud compute costs by 50%.
- **Automated data validation and quality assurance testing.** Replaced manual validation workflows with automated test scripts, reducing QA execution time from days to hours and ensuring data accuracy and freshness in AWS Redshift.
- **Built large-scale PySpark ETL pipelines in AWS Glue** to process high-volume telemetry and relational data (PI, GIS, SQL Server). Delivered a highly reliable data product used to execute objective risk assessments for $2.3 billion in physical assets.
- **Developed an automated, self-serve SQL reporting platform** that eliminated production bottlenecks and reduced data model build times by 50% for downstream analytics teams.

Technology Consultant

Visagio   **August 2020 - May 2022, Perth, WA**
- **Engineered a data-driven time and attendance forecasting tool** for a tier-one mining company. Processed historical consumption and real-time workforce data to generate operational forecasts, driving daily plan optimisations that provided savings of $40 million.
- **Built a strategic scenario-modelling engine** to process complex demographic datasets and generate 5-year and 10-year workforce projections for long-term enterprise resource planning.
- **Refactored mission-critical regulatory reporting pipelines** and visualisation layers (Power BI, Spotfire). Applied strict data modelling and transformation best practices to simplify the architecture, improving query responsiveness and long-term maintainability.

EDUCATION

Master of Predictive Analytics

Minor in Finance and Investment · Curtin University · Bentley, Western Australia · 2020 · 8.0/10.0

Bachelor of Electronics and Communications Engineering

Tribhuvan University · Kathmandu, Nepal · 2017 · 8.0/10.0

INVOLVEMENT

Python Instructor

Nepalese Engineers in WA · July 2021 - August 2021
- Conducted a 5-week-long Python programming masterclass with more than 20 participants. The masterclass covered the basics of Python programming and specific use cases for different engineering fields, including electrical, mechanical and other engineering.

Student Volunteer

John Curtin Leadership Academy · January 2020 - December 2020
- Led a volunteer team to help the Curtin C3 project conduct various activities, including a "Design-A-Thon" conducted in May 2020.
- Led a team to organise, advertise and coordinate an online event "WA's Sustainable Future" with speakers from Deloitte, Curtin University and CleanState and 50+ participants.