

ZEESHAN YALAKPALLI

Cloud Data Engineer | AI-Powered Pipelines | GCP Specialist

+91-8088730318 • zeeshanyalakpalli@gmail.com • [GitHub](#) • [LinkedIn](#) • Bengaluru, India

PROFESSIONAL SUMMARY

Final-year Computer Science student and aspiring Cloud Data Engineer with hands-on experience building production-grade data pipelines on GCP. Proficient in real-time streaming (Apache Beam, Pub/Sub), batch processing (BigQuery, Dataflow), workflow orchestration (Apache Airflow), and AI-augmented pipelines using Gemini 2.5. Demonstrated ability to design, containerise, and deploy automated cloud-native data systems from scratch. Author of 2 peer-reviewed research papers presented at international conferences.

EDUCATION

Bachelor of Engineering in Computer Science & Engineering

HKBK College of Engineering, Bengaluru (Affiliated to VTU)

2022 – 2026 | CGPA: 8.75/10

EXPERIENCE

Software Engineer Intern — Hostup Cloud Technologies Pvt. Ltd

On-site · Bengaluru

Aug 2025 – Oct 2025

- Developed and deployed cloud-native web applications using Next.js — implemented authentication, ticketing systems, and a bare-metal grid management tool, reducing page load times by 35%.
- Collaborated with senior engineers to architect backend features, integrate REST APIs, and deliver production deployments using standard CI/CD pipelines.
- Gained hands-on exposure to modern cloud infrastructure management, deployment automation, and scalable application design patterns.

PROJECTS

AI-Powered E-commerce Data Quality Monitor *Python · Gemini 2.5 Flash · BigQuery · Docker · Cloud Run · Cloud Scheduler · GCS · Streamlit · Plotly*

- Built an end-to-end anomaly detection pipeline scanning 200+ e-commerce orders across 5 anomaly categories; used Gemini 2.5 Flash API to auto-generate plain-English root cause explanations per anomaly.
- Deployed as a containerised Cloud Run Job scheduled daily via Cloud Scheduler (6 AM IST) — fully serverless; stored results in BigQuery with run_id pipeline lineage tracking enabling cross-run trend analysis.
- Built a live Streamlit + Plotly dashboard (dark theme) visualising anomaly breakdowns and Gemini explanations directly from BigQuery in real time.

Healthcare Data Pipeline — Batch & Streaming *Apache Beam · GCP Dataflow · Cloud Pub/Sub · BigQuery · Docker · Python*

- Built end-to-end batch and streaming pipelines using Apache Beam Python SDK, processing 4,500+ US hospital records with DirectRunner locally and deploying to GCP Dataflow for cloud execution.
- Implemented HIPAA-style PII de-identification using salted SHA-256, dead-letter routing for invalid records, and anomaly detection flagging 265 high-risk hospitals into a dedicated BigQuery sink.
- Built real-time admission event processing via Cloud Pub/Sub with 5-minute fixed and 3-minute sliding windows, handling late-arriving data using Beam watermarks with 2-minute allowed lateness.

Real-Time Crypto Streaming Pipeline *GCP · Cloud Functions · Pub/Sub · BigQuery · Cloud Scheduler · Python*

- Engineered an automated pipeline ingesting BTC/ETH/SOL prices every 60 seconds, achieving under 2-second end-to-end latency with a rigorous data quality validation layer maintaining 100% data integrity.
- Configured Pub/Sub dead-letter topics with retry policies for fault-tolerant delivery; orchestrated execution via Cloud Scheduler with zero manual intervention across continuous ingestion.

BigQuery Data Warehouse — NYC Taxi Dataset *BigQuery · SQL · Data Modeling · Partitioning · CTEs · Window Functions*

- Designed a two-layer warehouse (RAW → CURATED) processing millions of NYC Taxi records; applied partitioning, clustering, CTEs, and window functions to optimise query performance and reduce scan costs.

- Authored modular SQL transformations including joins, aggregations, JSON processing, and fact table design following star schema principles.

AI-Powered Bug Bounty Platform (Final Year Project) *Python · Flask · React.js · TF-IDF · Scikit-learn · JWT · REST APIs*

- Designed a full-stack vulnerability scanner integrating a Python ML microservice using TF-IDF vectorisation and classifiers to detect SQL Injection, XSS, and insecure form patterns across web targets.
- Engineered a parallel URL scanning data pipeline achieving ~5s per scan with structured JSON output; exposed results via JWT-authenticated REST APIs consumed by a React.js real-time security dashboard.
- **Research:** Peer-reviewed preprint (DOI: 10.20944/preprints202506.1363.v1); accepted for oral presentation at 5th SGCNSP-2025 (Singapore) and published in Springer's Lecture Notes in Electrical Engineering.

TECHNICAL SKILLS

Data Engineering	Apache Beam, Apache Airflow (DAGs, XCom, operators, scheduling, Docker), ETL/ELT pipeline design, Stream & batch processing, Windowing & watermarks, Dead-letter queuing, PII masking, Data quality frameworks, Pipeline lineage tracking
Cloud (GCP)	BigQuery, Cloud Dataflow, Cloud Pub/Sub, Cloud Run (Jobs & Services), Cloud Functions, Cloud Scheduler, GCS, Artifact Registry, Vertex AI, Cloud Monitoring, IAM
AI & LLM	Gemini 2.5 Flash API, Vertex AI Studio, Prompt engineering, LLM pipeline integration, google-genai SDK, Anomaly explanation automation
Data Warehousing	BigQuery RAW → CURATED architecture, Star schema, Fact & dimension tables, Partitioning, Clustering, Materialized views, Query cost optimisation
SQL	Window functions, CTEs, Joins (inner/left/semi/anti), Aggregations, JSON processing, Query optimisation, BigQuery dialect
DevOps & IaC	Docker, CI/CD, GitHub Actions, GitLab, Cloud Run Jobs, Artifact Registry, Environment variable management, Kubernetes (conceptual)
Languages	Python, SQL, Java, JavaScript, TypeScript, HTML5 & CSS
Frameworks & Tools	Streamlit, Plotly, Flask, React.js, Next.js, Node.js, Angular, Tailwind CSS
Databases	BigQuery, PostgreSQL, MySQL, MongoDB, Firebase, Redis
Tools	VS Code, Cursor, Git, GitHub, Postman, Linux, Jupyter Notebook, Docker Desktop

ACHIEVEMENTS & RESEARCH

- **Research Published:** AI-Powered Automated Bug Bounty Platform — peer-reviewed preprint (DOI: 10.20944/preprints202506.1363.v1), June 2025.
- **Research Published:** Accepted for oral presentation at 5th SGCNSP-2025 (Singapore) and published in Springer's Lecture Notes in Electrical Engineering, December 2025.
- Achieved Grade A+ in Competitive Programming assessments.
- Recipient of the Maymar Charitable Trust Scholarship and Zenith Foundation Academic Award.
- Secured 1st Prize in college Science Fest for design and presentation of a Wind Turbine Model.