Yukta Vajpayee

+1(408)581-4452 yukta.vajpayee@sjsu.edu linkedin.com/yuktavajpayee30 github.com/Yukta30

EDUCATIONAL QUALIFICATIONS

San Jose State University August 2025 - Present
Masters of Science, Computer Science San Jose, California

- **Relevant Coursework:** Data Structures and Algorithms, Operating Systems, Object-Oriented Design, Machine Learning, Computer Networks, Database Systems, Web Programming, Software Engineering, Cloud Computing, Distributed Systems

SRM Institute of Science and Technology May 2022
B-Tech in Computer Science (CGPA: 9.58 / 10.00) Chennai, Tamil Nadu

- **Relevant Coursework:** Data Structures and Algorithms (C++), Software engineering and Project Management, Intro to CS II (C++), Object Oriented Programming Language, Operating Systems, Advanced Programming Practices (Python)

TECHNICAL SKILLS AND COMPETENCIES

Languages: Java, Python, C/C++, Scala, JavaScript, HTML/CSS, SQL, Go, TypeScript, OpenCV, Ruby, Kotlin
Technologies: Kafka, Delta Lake, Trino, Athena, PinotDB, DynamoDB, MySQL, PostgreSQL, MongoDB, SQLite, REST APIs, JSON, Docker, Kubernetes, E2E Integration, Django, RAG Model, Hive
DevOps/Cloud: AWS (Lambda, Redshift, Athena, EMR), Apache Spark, Apache Flink, Airflow, Prometheus, Kibana, Git, VS Code, IntelliJ, PyCharm, NetBeans, Jira, Confluence, Miro

WORK EXPERIENCE AND INTERNSHIPS

Hewlett Packard Enterprise – Aruba Networking August 2022 – July 2025
Cloud Developer - I [Data Engineer] Bangalore, Karnataka
Network-as-a-Service (NaaS)

- Automated manual workflows by building end-to-end **orchestration pipelines with Python and REST APIs,** reducing operational **effort by 85%** and improving SLA compliance.
- Designed and implemented a **geo-aware service leveraging GPS data and rules-based engines to auto-generate incident tickets** for temperature-driven device outages, enabling faster Problem Management response through **ITSM tools like ServiceNow**.
- Engineered a real-time monitoring system to reflect **live status (online/offline) of 100K+ devices** using scalable cloud-native infrastructure.

Data Engineering Enhancements

- Migrated multiple **Django APIs** from web servers to standalone **Python script-based execution** environments to prevent timeout issues and improve reliability.
- Transitioned **Apache Flink** jobs from **AWS EMR clusters to Kubernetes (K8s)** for improved containerization, reduced infrastructure cost, and better horizontal scalability.
- Optimized scheduling of **Apache Airflow and Apache Spark** jobs by managing resource pools, and migrated reporting tables from Amazon **Athena to Delta Lake on S3** to revamp query latency and data freshness.

Gen-AI Related

- Engineered 100+ automated evaluation tests using **RegEx and embedding similarity metrics to reduce hallucinations** and ensure reliable RAG model outputs.
- Applied **Retrieval-Augmented Generation (RAG)** pipelines to dynamically filter and retrieve structured data from **100K+ records**, cutting manual query **effort by 40%** and improving data engineering workflows.

Hewlett Packard Enterprise – Aruba Networking Jan 2022 – July 2022
Software Development Intern Bangalore, Karnataka

- Successfully executed multiple Proof of Concepts (POCs) and developed multiple microservices, demonstrating proficiency in back-end development and integration.
- Designed an automation service from scratch, enhancing internal management processes and saving **20+ hours of manual work** weekly across teams.

TATA Steel – TISCO June 2020 – August 2020
Data Engineering Intern Jamshedpur, Jharkhand

- Optimized valve monitoring system by developing Python scripts to process raw CSVs into interactive visualizations, improving performance insights and decision-making.
- Designed anomaly detection logic using statistical methods and threshold-based rules, enabling early identification of valve irregularities and reducing downtime risk.

TECHNICAL PROJECTS

DataMorph – Adaptive ETL Framework for Sensor Streams | Python, Apache Airflow, PostgreSQL, Kafka, Pandas, Docker

- Built an Airflow ETL pipeline ingesting data from 10K+ devices, **cutting failures by 40% via dynamic validation.**
- Simulated real-time ingestion using Kafka, achieved fault tolerance with retry logic and alert mechanisms.
- Project Link

LivePulse – Real-Time Data Lake and Device Status Analytics | Scala, Apache Spark, Delta Lake, Databricks SQL, Kubernetes

- Constructed a **Spark pipeline processing 5K+ events/sec**, ensuring instant status detection and minimal downtime.
- Leveraged Delta Lake on **100GB+ datasets**, enabling BI dashboards that slashed reporting latency by 95%.
- Project Link

AlertCast – Real-Time Device Health Ticketing System | Go, Kafka, Kafka Streams, PostgreSQL, Redis, Docker

- **Engineered a Kafka pipeline processing 200+ device** events daily, auto-classifying alerts and reducing resolution time by 35%.
- Developed severity-based rules **generating 50+ high-priority tickets monthly**, enabling faster triage and analytics.
- Project Link