VENKATA KRISHNA POKALA

(813)-647-8587 — subbupokala143@gmail.com — LinkedIn: vk167 — GitHub — Portfolio — Tampa, FL

PROFESSIONAL SUMMARY

Results-driven Software Engineer with 3+ years of experience specializing in data engineering, ETL processes, and full-stack development. Proven expertise in database management, big data technologies, and cloud platforms with a strong foundation in software reliability engineering.

TECHNICAL SKILLS

- **Data Engineering:** ETL Pipelines, Data Modeling, Data Integration, Data Warehousing
- **Databases:** Apache Cassandra, MongoDB, PostgreSQL, MySQL, NoSQL
- **Big Data Technologies:** Hadoop, Apache Kafka, Apache Airflow, Spark
- **Cloud Platforms:** GCP (BigQuery, Cloud Storage), AWS (Redshift, S3)
- **Programming Languages:** Python, Java, SQL, PL/SQL
- **Tools & Technologies:** Docker, Kubernetes, Jenkins, Git
- **Data Analysis:** Tableau, Power BI, Statistical Analysis
- **Development Methodologies:** Agile, DevOps, SDLC

WORK EXPERIENCE

Software Engineer-DSE, Infosys, Hyderabad, India — Jun' 2022 – Jul' 2023

- Architected and implemented ETL pipelines using Apache Kafka and AWS services, processing 500GB+ daily data with 40% improved efficiency
- Developed scalable data architecture using Apache Cassandra and MongoDB, handling 1TB+ of data while maintaining sub-second query response times
- Implemented real-time data monitoring and alerting system using Splunk and custom dashboards, reducing system downtime by 30%
- Optimized SQL and NoSQL database queries through indexing and query restructuring, resulting in 40% performance improvement
- Created automated data validation frameworks using Python and Java, ensuring 99.9% data accuracy across 100M+ daily records
- Designed and implemented microservices-based data integration platform using Spring Boot and Docker, improving system scalability by 35%
- Participated in weekly code reviews and assisted in onboarding 2 junior developers through knowledge sharing and pair programming
- Collaborated with cross-functional teams to define data models and implement data governance policies

Software Engineer - Data Systems, OpenText, Hyderabad, India — Apr' 2020 – May' 2022

- Developed and maintained data pipelines using Python and Apache Airflow, processing 500GB+ daily data with 99.9% reliability
- Implemented data modeling and warehouse solutions in PostgreSQL and AWS Redshift, reducing data retrieval time by 30%
- Built real-time data streaming solutions using Apache Kafka, handling 50K+ events per second
- Created automated data quality validation frameworks using Python and SQL, reducing data errors by 40%
- Designed and implemented REST APIs for data access layer, serving 1000+ requests per minute with 99.9% uptime
- Contributed to migration of legacy data systems to cloud-based solutions (AWS), helping achieve 35% cost reduction and 45% improved performance
- Implemented data security measures including encryption and access controls, ensuring compliance with data protection regulations
- Developed documentation for data pipelines and processes, reducing onboarding time for new team members by 50%

Data Engineering Intern, Verzeo, India — Jan' 2019 – Jun' 2019

- Designed and implemented ETL processes using Python and SQL, improving data accuracy by 25%
- Built data preprocessing pipelines for machine learning models using pandas and NumPy
- Created interactive data visualization dashboards using Tableau, reducing reporting time by 30%

- Developed automated data quality checks using Python scripts, ensuring 99% data accuracy
- Assisted in implementing data warehousing solutions and optimizing database queries
- Collaborated with data scientists to prepare and clean datasets for machine learning models

EDUCATION

- **MASTER OF SCIENCE** Computer Science, University of South Florida — Tampa, FL GPA-3.85
- **BACHELOR OF TECHNOLOGY** Computer Science, Vignan University — India GPA-9.4

RELEVANT COURSEWORK: Data Mining, Database Management Systems, Big Data Analytics, Machine Learning, Advanced Data Structures and Algorithms

PROJECTS

Data Pipeline Optimization Project Aug' 2024 – Dec'2024

- Designed and implemented scalable ETL pipelines using Apache Kafka and GCP BigQuery
- Developed data quality frameworks reducing error rates by 40%
- Implemented real-time data processing solutions handling 1M+ records per hour

Enterprise Data Integration Platform Aug' 2023 – Dec' 2023

- Built end-to-end data integration solution using Python and Apache Airflow
- Implemented data validation and transformation logic for multiple data sources
- Developed automated testing framework for data quality assurance

CERTIFICATIONS

- Microsoft Azure Data Engineer Associate (DP-203)
- Google Cloud Platform Data Engineer

RESEARCH

- Published paper on "Prediction of Breast Cancer: Comparative Review of Machine Learning Techniques" (2022)