JAHNAVI VANKAYALAPATI

Katy, TX, 77494 | 6154913515 | jahnavivch@gmail.com | LinkedIn

Dynamic Generative AI and Data Engineer with 4+ years of experience in developing scalable data pipelines, AI-driven automation, & cloud-based analytics solutions across healthcare, fintech, and enterprise domains. Skilled in ETL orchestration (Airflow, AWS Glue), integrating OpenAI and Cognitive Services to optimize workflows, and proficient in Python, SQL, Azure, AWS, Snowflake, Power BI, and Tableau. Demonstrated capacity to use creativity, automation, and cross-functional cooperation to improve data quality, turn complex data into actionable insights, and produce quantifiable outcomes.

Core Competencies

Generative AI | Data Engineering | ETL Pipelines & Automation | Data Analytics | KPI Reporting | Data Visualization (Power BI, Tableau) | Cloud & Data Warehousing (Azure, AWS, Snowflake, BigQuery) | Large Language Models (LLMs) | Natural Language Processing (NLP) | RAG Systems | Machine Learning | Big Data Processing (Spark, Airflow, Glue, dbt) | API & Microservices Development | Containerization (Docker, Kubernetes) | Cross-Functional Collaboration | Data Governance & Quality | Data Security & Compliance (HIPAA, IAM) | Strategic Problem Solving | Mentoring & Coaching | Agile/Scrum Environments

TECHNICAL KNOWLEDGE

Programming Languages:	Python (Pandas, NumPy, Scikit-learn), SQL, Java (Basics), Shell Scripting.
Database:	Snowflake, Google BigQuery, MySQL, Oracle PL/SQL, MongoDB (Star & Snowflake Schema Design), DynamoDB.
Cloud Platforms:	Azure (OpenAI, Cognitive Services, Functions, Speech Services), AWS (S3, Glue), Google Cloud Platform (BigQuery).
Frameworks:	Apache Spark, Apache Airflow, AWS Glue, dbt, Flask, React.js.
CI/CD Tools:	Git, GitHub, GitLab, Jenkins, JIRA, Confluence.
Web Development:	HTML, CSS, JavaScript, React.js, Streamlit.
Security:	Identity and Access Management (IAM), HIPAA Compliance, Cloud Security.
Tools:	Power BI, Tableau, Excel (Power Query, PivotTables), Docker, Kubernetes, Agile/Scrum.

PROFESSIONAL EXPERIENCE

SBS Corp, *Texas (USA)*   **Oct 2023 – Present**

Generative AI / OpenAI Developer

- Architected and deployed a scalable data processing system on Azure, streamlining a critical workflow that improved data availability and reduced manual processing time by 40%.
- Enhanced and supported production ETL pipelines for processing complex, unstructured data, ensuring high data quality and achieving 95% accuracy for downstream AI models and analytics.
- Systematized the generation of synopses, taglines, and SEO keywords for over 100 podcast episodes, reducing manual effort by 60% and boosting content accessibility in 10 languages.
- Produced clean, efficient, and thoroughly tested code while actively identifying inefficiencies in existing systems, improving project efficiency by 30% and minimizing deployment time by 20%.
- Promoted robust development practices through code reviews for new data pipelines and microservices within a geographically distributed multi-functional team.
- Optimized data processing frameworks and systems by integrating Azure Speech Services and GPT-3.5, improving project efficiency by 30% and minimizing deployment time by 20%.
- Built a React-based web app for easy audio file uploads, enabling intelligent, end-to-end processing from transcription to content creation via Azure Functions.

Environment & Tools: Azure Cognitive Services, Azure OpenAI, Azure Functions, Azure Speech Services.

Accenture, *Hyderabad (India)*   **Jan 2022 – Aug 2023**

Data Engineer

- Built a scalable Power BI dashboard integrating EHR, claims, and patient feedback across 50+ healthcare facilities, improving visibility into readmissions and provider performance.
- Used SQL (Snowflake) and Python (Pandas, NumPy) to clean and transform 10M+ records, reducing ETL load time by 30%

while ensuring HIPAA compliance.
- Refined predictive models with Scikit-learn and deployed using Apache Spark, improving the accuracy of patient admission forecasting by 25%.
- Streamlined ETL pipelines using Apache Airflow, AWS Glue, and dbt, enabling real-time ingestion of healthcare data and reducing manual reporting by 40%.
- Integrated Apache Kafka for streaming patient feedback and real-time data ingestion into analytics workflows.
- Used MongoDB to store and retrieve semi-structured survey data for performance tracking dashboards.
- Containerized ML workflows using Docker and deployed with Kubernetes on a test cluster for scalable model training and inference.
- Led data quality initiatives, performed schema validation, and presented insights to executive leadership, reducing patient wait times by 15%.

Environment & Tools: SQL (Snowflake), Python (Pandas, NumPy, Scikit-learn), Apache Airflow, AWS Glue, Apache Spark, dbt, Apache Kafka, MongoDB, Docker, Kubernetes, Power BI, Excel, EHR Systems, Git, HIPAA-Compliant Cloud Infrastructure.

Wipro, *Hyderabad (India)*   **July 2021 – Dec 2021**
Data Analyst
- Created and managed Jira tickets by identifying issues, assigning priorities, and tracking progress through completion.
- Performed Excel data analysis using PivotTables, charts, and conditional formatting to generate key business insights.
- Expanded and designed interactive dashboards in Tableau, connecting multiple data sources for real-time analytics.
- Wrote and optimized SQL queries to extract, join, and aggregate data for reporting and performance tracking.
- Combined insights from Excel, Tableau, and SQL to support data-driven decisions and business process improvements.

Environment & Tools: SQL, Oracle PL/SQL, Tableau, Python (Basics), Excel, JIRA, Power BI, Data Pipelines, Client Reporting Tools.

ACADEMIC PROJECTS

- Improved a React.js web app using Google's T5-Small for automated summarization of U.S. Congressional bills; fine-tuned the model with custom preprocessing to improve ROUGE scores and output accuracy.
- Built a Streamlit multi-PDF Q&A system using LangChain, OpenAI embeddings, and FAISS, enabling natural language queries with context-based responses.
- Enhanced Random Forest models using MPI & Multiprocessing, improving training efficiency by 60% and analyzing 1M+ healthcare records for predictions.
- Cloud-native React.js app deployed on AWS (S3, CloudFront, Cognito, API Gateway) with DynamoDB and AWS Bedrock integration, enhancing scalability and user experience by 30%.

EDUCATION

Middle Tennessee State University, Tennessee, USA   **Aug 2023 - May 2025**
Master of Science in Computer Science   GPA: 3.75/4

Graduate Assistant for Accounting Services department (Systematized financial reporting workflows using SQL, Argos, Banner, and Python to enable real-time insights and developed dynamic departmental forms to enhance data collection efficiency)

Relevant Coursework: Database Management Systems, Data Mining, Intro to Artificial Intelligence, Selected Topics in Database Systems, Data Understanding, Data Exploration, Parallel Processing Concepts, Software Design & Development, Web Technologies

Jawaharlal Nehru Technological University, Kakinada, India
Bachelor of Technology in Computer Science and Engineering

Relevant Coursework: Artificial Intelligence, Machine Learning, Database Management Systems, Data Warehousing and Mining, Computer Networks, Compiler Design, Operating Systems, Object-Oriented Programming (Java), Design and Analysis of Algorithms, Data Structures, Programming in C/C++, Cloud Computing, Information Security, Data Communication, Python Programming