Pavan Sabnaveesu

📞 **832 274-8181**| ✉️ sabnaveesuprasanth@gmail.com | 🔗 **LinkedIn**

Professional Summary

Strategic data engineer with 5+ years of experience in optimizing data processes and driving business value through actionable insights. Proven ability to reduce operational costs, improve decision- making speed, and minimize risks, while fostering collaboration within diverse teams to support business growth.

Technical Skills

🧠 **Programming & Scripting:** Python, SQL, Java NoSQL (MongoDB, Cosmos DB), BigQuery, Pandas, NumPy
📇 **Data Management:** Snowflake, BigQuery, Azure Synapse, Power BI, Tableau, Excel
☁️ **Cloud Services:** AWS (S3, Glue, Redshift), Azure (Data Factory, Data Lake), GCP(Vertex AI)
🗄️ **Data Engineering:** Apache Kafka, Spark, Flink, PySpark, Databricks, dbt, Airflow
🧰 **Tools & DevOps:** GitHub, Jenkins, Docker, Kubernetes, DevOps, CI/CD, MLflow

Work Experience

Lead Data Engineer Cyber Nirvana(Contract)   **February 2025 – Present**

- Architected a Kafka-based event processing pipeline integrated with a quantized LLaMA and Cosmos DB, enabling real-time fraud detection by processing 10M transactions daily with sub-50ms latency, reducing fraud by 17%
- Engineered real-time order tracking using Apache Flink, dbt, and fine-tuned BERT, processing 1M retail transactions daily with sub-second latency, improving demand forecasting accuracy by 32% for supply chain optimization
- Designed data ingestion pipelines with Kafka, Azure Event Hubs, and quantized GPT-4O, slashing processing time by 40%, accelerating ML inference for real-time customer analytics in a retail loyalty program

Data Engineer, NextRow Digital   *July 2021 – Dec 2022*

- Optimized and automated 20+ ETL pipelines using ADF, Databricks, and Airflow, reducing data time by 40% through advanced SQL tuning, indexing, and Spark-based transformations across Data Lake and Synapse
- Led a team of 5 and mentored members on Power BI tools, engineering principles, best practices, overseeing deployment, resolving issues, data modeling, ensuring robust pipelines and trained clients on new features
- Led the data engineering team, working with stakeholders to identify customer needs, and collaborated with data scientists to deploy ML models into production using MLflow and Feature Store
- Automated app deployment using Docker, Kubernetes, and CI/CD pipelines via Azure DevOps and GitHub, provisioning scalable infrastructure with Terraform to improve deployment efficiency and system uptime by 25

Data Engineer, Meslova Systems   *Sept 2018 - June 2021*

- Created data governance and security models, mapping 30+ documents and 20 data dictionaries. Streamlined pipelines and metrics in Confluence, boosting collaboration and reducing onboarding by 30%
- Optimized complex SQL queries, improving data retrieval speed by 5x (from 5s to 1s) by minimizing joins, indexing, and using subqueries enhancing overall application performance
- Redesigned database schemas into BCNF to streamline customer order management, reducing redundancy and ensuring data integrity, which improved query performance by 30%
- Accelerated manufacturing analytics by 35% using Spark to process 10TB of production data, leveraging distributed computing for real-time inventory insights
- Transformed customer insights into Snowflake, reducing query times by 50% with optimized schemas and partitioning for scalable Power BI reporting

Education

Texas A&M University   **CGPA: 3.9/4.0**
Master of Science   January 2023 – Dec 2024
Graduate Research Assistant   February 2023 – Dec 2024

Projects

- Designed and implemented scalable data lake and warehouse solutions using AWS S3, Glue, Redshift, and Snowflake, improving data retrieval and ETL time by 50% through efficient partitioning and schema optimization
- Designed a real-time ICU telemetry pipeline with Docker, Kafka and Kubernetes, managed Tableau analytics via REST API to hospital systems, processing 500 000 vitals daily in under 100ms and boosting clinical alerts by 25%