**Pavan Sabnaveesu**

📞 **832 274-8181|** ✉️ pavansabnaveesu@gmail.com **|**  [**LinkedIn**](https://www.linkedin.com/in/prasanth-sabnaveesu/)

**Professional Summary**

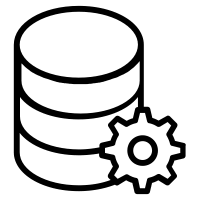
Strategic data engineer with 5+ years of experience in optimizing data processes and driving business value through actionable insights. Proven ability to reduce operational costs, improve decision- making speed, and minimize risks, while fostering collaboration within diverse teams to support business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

* + Enabled real-time fraud detection by developing a Kafka-based event processing pipeline with highly scalable and low-latency Cosmos DB, identifying transaction anomalies in MS to minimize financial risk and prevent fraud
  + Enhanced real-time order tracking using Flink and dbt, processing 1M transactions daily with sub-second latency for accurate demand forecasting
  + Designed real-time data ingestion pipelines using Kafka and Azure Event Hubs, reducing data processing time by 40%, enabling faster machine learning model inference and analytics

***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**

* + Created enterprise Power BI solutions, data visualization tools, and key metrics for manufacturing departments. Established a unified data warehouse and documented metric definitions for consistent usage
  + 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