

# SAHITHYA ARVETI NAGARAJU

📞 716 247 7348 ✉️ [sahithyaarvetinagaraju@gmail.com](mailto:sahithyaarvetinagaraju@gmail.com) [in linkedin](#) [🌐 sahithya-portfolio](#) [🐙 github.com/SAHITHYA21](#)

## Education

### University at Buffalo (GPA 3.97/4.00)

*Master of Science in Data Science*

**Jan. 2024 – May 2025**

*Buffalo, New York*

### Siddaganga Institute of Technology (CGPA 9.44/10.00)

*Bachelor of Science in Information Science and Engineering*

**Aug. 2017 – Jun. 2021**

*India*

## Technical Skills

**Programming Languages:** Java, Python, Golang, R, MATLAB, Shell Scripting, C++, PHP, Data Structures, Algorithms, Object Oriented Programming

**Database Management:** MongoDB, PostgreSQL, MySQL, Hadoop, Spark, Pyspark, Sonwflake, Databricks

**Machine Learning & Analytics:** numPy, Pandas, scikit-learn, SQLite3, TensorFlow, PyTorch, Matplotlib

**Cloud & DevOps:** GCP, AWS, Kubernetes, Docker, Jenkins, Terraform

**Data Visualization & Monitoring:** Tableau, AppDynamics, Splunk, Grafana

**Tools & Platforms:** Excel, Linux, Unix, Kafka, GitHub, Bitbucket, Jira, ServiceNow, ElasticSearch

## Professional Experience

### University at Buffalo

*Graduate Teaching Assistant*

**Jan 2025 - May 2025**

*Buffalo, NY, United States*

- Collaborated with Professor Khinkis on refining lecture materials for advanced statistical methods and data mining (using R), incorporating real-world case studies that improved student comprehension scores by 15% according to course evaluations.
- Supported 78 students in understanding Time Series analysis, grading assignments and projects, and providing timely feedback.

### Sabre

*Software Engineer II*

**Aug 2021 – Jan 2024**

*Bengaluru, India*

- Migrated GetThere services from legacy web servers to Google Cloud Platform using Terraform, Docker, and Kubernetes, improving scalability, optimizing traffic, and cutting infrastructure costs by 22%.
- Led endpoint migrations for Oracle and Nexus in a live, distributed production environment by designing and executing a zero-downtime migration strategy, resulting in 100% service continuity.
- Managed production and certification deployments on Linux systems, resolving 95% of incidents within SLA through effective root cause analysis and release management.
- Automated infrastructure management tasks using Python and Shell scripting, significantly reducing manual intervention.
- Monitored server health across 150+ nodes using Grafana, AppDynamics, and Splunk, enabling proactive incident resolution and 24/7 reliability.
- Built and maintained a disaster recovery stack aligned with DevOps and CI/CD principles, reducing recovery time.
- Collaborated with cross-functional teams including database, product, and support teams to ensure seamless deployments and post-deployment validation.
- Designed and implemented data synchronization workflows to securely transfer data from on-premise systems to GCP Cloud Storage, supporting data archival and downstream processing.

### Sabre

*Software Engineer (Intern)*

**Jan 2021 – July 2021**

*Bengaluru, India*

- Created a ServiceNow API based application using Python (Flask) and Jenkins to automate change record submissions, resulting in a 60% reduction in manual processing time.
- Designed AppDynamics dashboards to track server sessions and booking counts for validating successful deployments across servers.
- Participated in daily stand-ups and sprint planning, collaborating with senior engineers to deliver reliable release automation.
- Provided support for deployment testing, change verification, and performance validation under production-like conditions.

### Samsung PRISM

*Research Intern*

**Aug 2020 - Aug 2021**

*India*

- Designed and trained an LSTM-based neural machine translation model to convert Hindi to English using TensorFlow and Keras.

- Achieved a BLEU score of 0.97 through advanced preprocessing, embedding techniques, and hyperparameter tuning.
- Co-authored and published research in the International Journal of Intelligent Systems and Applications in Engineering (IJISAE).
- Conducted literature reviews and model benchmarking to ensure accuracy, efficiency, and model generalization.

## Center for System Design

May 2019 - Jul 2019

Intern

India

- Built an interactive simulation tool for asymmetric roof trusses using React, HTML, CSS, and JavaScript.
- Enabled real-time structural parameter manipulation and visualization, enhancing design comprehension for civil engineering students and researchers.
- Collaborated with faculty and peers to refine simulation accuracy and improve frontend responsiveness.
- Documented development process and presented project outcomes to academic stakeholders.

## Projects

---

### Cloud-Native ETL Pipeline | *Data Warehouse, Databricks, Snowflake, Airflow, Python, Docker, SQL* Mar 2025

- Built an automated ETL pipeline using Snowflake's TPC-H dataset, modular dbt models, and Apache Airflow DAGs to support scalable transformation workflows.
- Added rigorous data quality validation through dbt's generic and singular tests, ensuring pipeline integrity and consistency.

### Real Estate Price Prediction | *Python, SQLite3, numpy, scikit, MLFlow, Git, Digital Ocean, Docker, Streamlit* Mar 2024

- Built an end-to-end regression pipeline for predicting housing prices using scikit-learn models (like Regression, KNN, Random Forest), exploratory data analysis, and feature engineering.
- Containerized the app with Docker, deployed on Digital Ocean, and integrated a user interface via Streamlit to enable real-time predictions.

### Online Shopping System | *Python (numpy, pandas, Faker), MySQL* Nov 2024

- Designed and deployed a scalable e-commerce relational database supporting 3,500+ users with <4ms query performance using optimized indexes, stored procedures, and CTEs.
- Integrated real-time inventory triggers and synthetic test data using Faker for full lifecycle transaction simulation and analysis.

### A Predictive Assessment System for Diabetes Risk Factors | *Python, PySpark, PySpark's MLlib* Nov 2024

- Created a distributed machine learning workflow with PySpark MLlib to handle large medical datasets and train Logistic Regression and Gradient Boosted models.
- Achieved 8% performance improvement and 40% reduction in execution time for risk prediction of diabetes based on demographic and clinical indicators.

### Ollama Local LLM RAG | *Python, Generative AI, LLM, Llama, RAG* Aug 2024

- Developed a local Retrieval-Augmented Generation (RAG) pipeline using Llama3 and mx-bai-embed-large embeddings (1024-d) to enable intelligent PDF querying with 90% accuracy.
- Reduced hallucinations by 25% by applying cosine similarity (top-k=3) with threshold filtering (score>0.5), ensuring high-precision knowledge retrieval.

### Web Portal for Activity Points | *HTML, CSS, JavaScript, PHP, MySQL* Aug 2020

- Developed a responsive web portal to manage AICTE-mandated student activity points, integrating HTML/CSS, JavaScript, PHP, and MySQL to streamline data entry and score tracking for students and faculty.
- Designed and implemented system architecture with data flow, use case, and sequence diagrams, improving maintainability and reducing manual tracking errors by over 80%.

### Bank Marketing Campaigns | *Tableau, Data Visualization* Jul 2024

- Designed a Tableau dashboard to evaluate term deposit campaign performance, highlighting conversion patterns and segment-specific engagement.
- Enabled decision-makers to forecast marketing outcomes and optimize targeting using interactive filters and historical trend visuals.

## Awards & Certifications

---

- Received the "Say Thanks" Award at Sabre for demonstrating passion and accountability on high-impact projects.
- Won Sabre's internal coding competition, CodeIt, for outstanding problem-solving and coding skills.
- Completed Coursera certifications in Jenkins, Google Cloud Platform, Google Compute Engine, and Kubernetes fundamentals.