SAI CHANDAN REGONDA

Location: WI | **Phone:** 414-210-0477 | **Email:** saichandanregonda21@gmail.com | **LinkedIn:** Sai Chandan

SUMMARY

Results-driven Data Engineer with 4+ years of experience designing and delivering scalable data solutions in cloud-native environments (AWS, Azure, Snowflake). Proven expertise in building robust ETL/ELT pipelines, automating workflows, and enabling real-time analytics for high-volume, business-critical operations. Strong background in SQL, Python, and big data technologies such as Spark, EMR, and Airflow. Adept at translating ambiguous business requirements into structured data models and actionable insights to support decision-making in fast-paced, metrics-driven environments. Known for embracing Leadership Principles, primarily *Deliver Results*, *Earn Trust*, and *Dive Deep,* while implementing data governance, privacy controls, and self-service analytics solutions that scale. Certified in AWS Solutions Architect, Azure Data Engineer, and SnowPro Core.

TECHNICAL SKILLS

Programming Languages: Python, SQL, R, C, Java, JavaScript, Shell Scripting, Scala, HTML, CSS
Frameworks & Tools: Apache Airflow, Apache Spark, Hadoop, SnapLogic, Terraform, Dagster.
Databases: MySQL, MongoDB, PostgreSQL, MS SQL Server, Snowflake
Cloud Platforms: AWS (Lambda, Glue, S3, Redshift), Azure, GCP
CI/CD Tools: Git, Jenkins
Methodologies: Agile (Scrum)
Data Visualization Tools: Tableau, Power BI, MS Excel, SSRS

EDUCATION

Master of Science in Computer Science   Sep 2022 - Dec 2023
University of Wisconsin, Milwaukee
Bachelor of Technology in Computer Science   June 2018 -May 2022
Gokaraju Rangaraju Institute of Engineering and Technology, India

CERTIFICATIONS

- **Snowpro Core Certification -** Snowflake
- **Azure Data Engineer Associate:** Microsoft
- **AWS Solutions Architect:** AWS
- **AWS Academy Cloud Foundation: AWS**
- **Data Analytics Essentials: Cisco**
- **Programming Essentials in Python: Cisco**

EXPERIENCE

Technical Analyst
Northwestern Mutual, WI   March 2025 – Current
Northwestern Mutual is a leading financial services company that provides insurance, investment, and wealth management solutions. Its Data Engineers drive innovation by building robust data pipelines and platforms that enable data-driven decision-making in financial planning, risk management, and client services.

- Designed and deployed 15+ scalable ETL/ELT pipelines using **Databricks, Apache Airflow, Python**, and **Azure Blobs**, processing over 8 TB of structured and semi-structured data from legacy systems and APIs with **99.8**% accuracy.
- Modeled and partitioned datasets in **Snowflake** and **Azure Synapse Analytics**, applying dimensional modeling (star/snowflake schema), materialized views, and clustering to optimize read performance for analytical workloads.
- Implemented advanced transformation logic using **SQL, Spark**, and **Python**, including data cleansing, deduplication, and enrichment to ensure high-quality input for reporting and ML pipelines.
- Automated data pipeline deployment with **GitHub Actions** and **Jenkins**, integrating with **unit tests** and data validation scripts to support robust **CI/CD** for production-grade data systems.
- Enabled data observability using **Azure Monitor** and custom Python scripts, tracking pipeline health, schema drift, and record anomalies, reducing **incident resolution** time by **40%.**
- Enforced **data governance** policies using Snowflake row-level security (RLS), IAM roles, and column-level masking to ensure compliance and secure access to PII and financial datasets.
- Delivered **20+** interactive dashboards in **Power BI** and **Tableau**, enabling self-service access to **KPI**s and real-time insights for recruiting, finance, and operations teams.
- Collaborated with **data scientists**, **product owners**, and **business analysts** to design reusable data assets and analytical datasets supporting machine learning models and business reporting.
- Part of end-to-end delivery of data engineering projects, from **requirement gathering** and architecture design to **production deployment** and performance optimization, aligning with **Agile sprints**.

Data Engineer   April 2024 – March 2025
ENSIGN Group, TX
The Ensign Group is a leading provider of skilled nursing and rehabilitative care services across the U.S. They focus on improving patient outcomes through innovative technologies. Data Engineers are key in optimizing data infrastructure and enabling real-time insights for operational efficiency.

- Designed and deployed scalable **ETL/ELT** pipelines in **Azure Data Factory** and Snowflake, processing 2M+ records daily with 99.9% accuracy, reducing pipeline latency by 40%.
- Led a data migration project, transferring 5+ TB of critical business data from on-premises systems to **Azure Data Lake** and Snowflake, ensuring zero data loss and enabling 20% faster query execution, leading to improved analytics decision-making for business teams.
- Optimized Snowflake performance by implementing clustering, partitioning, and query tuning, reducing compute costs by 20% and improving query execution speed by 30%, enabling real-time insights for enterprise reporting.
- Integrated Databricks with Snowflake for scalable big data processing, reducing batch processing time by 25% and improving the efficiency of machine learning model training for a retail analytics project.
- Developed real-time data streaming architecture using Azure Event Hub, Snowflake Streams, and Apache Kafka, enabling scalable and efficient data processing for live business insights.
- Automated CI/CD pipeline deployments for Snowflake and Azure Data Engineering workflows using Jenkins and Git while implementing data governance policies, ensuring 100% compliance with row-level security (RLS), data masking, and Python-based validation scripts.

Data & Analytical Engineer   **Feb 2021 - Jul 2022**
Dream Ambassadors, India
Dream Ambassadors is a purpose-driven organization that empowers individuals through data-driven insights and innovative solutions. Data Engineers at Dream Ambassadors are vital in building scalable data systems that support impactful decision-making and drive meaningful change.

- **Collaborated with stakeholders** to gather business requirements, translating them into technical data solutions that improved data-driven decision-making and reporting **efficiency by 30%.**
- **Analyzed and transformed large datasets** using **SQL and Python in Snowflake**, uncovering **actionable insights** that contributed to a **20% increase in operational efficiency** by optimizing key business processes.
- **Designed and maintained 5+ real-time dashboards and reports** in **Tableau and Power BI**, integrating **AWS and Snowflake**, enabling **instant visibility into key metrics** and resulting in a **15% increase in operational efficiency**.
- Conducted advanced data analysis using **SQL** and **Python**, deriving actionable insights that improved forecasting **accuracy by 10% and reduced data inconsistencies by 25%.**
- **Performed statistical analysis and predictive modeling** (regression, clustering, and time-series forecasting) using **SQL and Python in Snowflake on AWS**, improving **demand forecasting accuracy by 12%** and enhancing **resource allocation.**

PROJECTS
Smart Bot for Handwritten Digit String Recognition
- Developed a **Python-based CNN** model for real-time handwritten digit string recognition, increasing cheque processing accuracy by 25% and security.
- Initiated audio confirmation, boosting transaction verification efficiency by 30%. Featured in IEEE publications.

Brain Tumor Detection Using CNN
- Designed and optimized a robust CNN architecture using **TensorFlow/Keras,** achieving 97.5% accuracy.
- Executed comprehensive data augmentation techniques, increasing the quality of training datasets by 50%; this enhancement led to a significant 30% improvement in model accuracy for predictive analytics projects.

Resume Screening Bot Using LLMs (GenAI)
- Developed an AI-powered resume screening bot using GPT-4 to evaluate candidate resumes against job descriptions, delivering a relevance score and explanation.
- Implemented semantic search using FAISS and OpenAI embeddings to match resume content with job role keywords and responsibilities.
- Integrated LangChain with Streamlit to create an interactive frontend for users to upload resumes and job descriptions for instant analysis.
- Engineered prompts for GPT to generate natural-language feedback on resume-job fit, highlighting gaps and strengths for user improvement.
- Deployed the application on a local server, demonstrating capabilities in GenAI, vector databases, and LLM orchestration.

AI-Powered Data Quality Anomaly Detection
Personal Project | Python, Scikit-learn, Airflow, Snowflake, Azure Monitor
- Designed and implemented an ML-based anomaly detection system to monitor ETL pipeline health by identifying schema drift, null spikes, and outliers in time-series data.
- Utilized Isolation Forest and Z-score models in Python/Scikit-learn to detect data anomalies, improving data reliability and reducing downstream data issues by 40%.
- Automated pipeline monitoring using Apache Airflow, triggering alerts in Azure Monitor on anomaly detection, enabling real-time data observability.
- Integrated Snowflake as the central data warehouse and used SQL + Python for feature engineering and anomaly tagging.
- Enabled proactive issue resolution by delivering weekly anomaly trend reports to stakeholders, improving ETL trust and governance.