

SAI NIKHIL ARATIPAMULA

Tallahassee, FL

📞 8503393663 ✉️ snaratipamula@gmail.com www.linkedin.com/in/sainikhilaratipamula <https://sainikhilaratipamula.github.io/>

Education

Florida State University | GPA : 3.7

Master of Science in Data Science

Aug 2023 - May 2025

Tallahassee, Florida

Work Experience

Evergreen Solutions

Data Engineering Intern

Oct 2024 - Present

Tallahassee, Florida

- Engineered an end-to-end data pipeline using PySpark for compensation analysis, enhancing processing efficiency by 40%.
- Refined data management workflows by implementing Python and SQL techniques, ensuring accuracy and consistency.
- Formulated an Advanced Excel template to automate data ingestion from multiple sources, improving operational efficiency.
- Created interactive Power BI dashboards to deliver actionable insights for executives and clients, enabling data-driven decision-making.

Florida State University

Research Intern - Data Engineering

Jul 2024 - Oct 2024

Tallahassee, Florida

- Built and managed ETL pipelines using AWS and GCP to efficiently process and analyze large-scale graph datasets.
- Orchestrated automated workflows using Apache Airflow, Cloud Composer, and AWS Step Functions, reducing processing time by 30% while ensuring scalability and compliance with data governance standards.
- Managed cloud data storage and retrieval with AWS Redshift, DynamoDB, S3, and GCP BigQuery, optimizing query performance and enabling real-time data insights across terabyte-scale datasets.

LTIMindtree (Client : Citibank)

Senior Data Engineer

Aug 2022 - Jul 2023

Jacksonville, Florida - Remote

- Developed and implemented ETL pipelines using PySpark in Databricks and AWS Glue, resulting in a 50% increase in data processing efficiency and a 40% improvement in data quality across various structured and unstructured datasets by reducing cloud costs by 30%.
- Constructed and fine-tuned complex SQL queries to extract and transform data from multiple databases, leveraging Python for automation to minimize manual work and improve operational efficiency by 30%.
- Deployed data workflows and scheduling using Apache Airflow and Shell scripting, reducing manual intervention by 50% and optimizing pipeline execution, leading to a 30% reduction in processing time.
- Improved data storage and processing capabilities by implementing Hadoop, Hive, and HDFS, enhancing data scalability and throughput by 40%, ensuring efficient reporting and analytics for stakeholders.
- Collaborated with key stakeholders to gather data requirements, ensuring integrity and accuracy through validation, and quality checks.
- Resolved data performance and pipeline issues while providing ongoing support and optimizing data systems for improved efficiency.

Larsen and Toubro Infotech

Data Engineer

Jun 2022 - Aug 2022

Mumbai, India

- Established and engineered ETL processes using PySpark to ingest and transform large datasets across multiple platforms (local storage, HDFS, Hive, and MySQL), improving data processing efficiency by 50% and supporting real-time analytics.
- Enhanced and optimized Apache Kafka and Spark Streaming for real-time data ingestion, enabling seamless integration and streaming of large datasets, which enhanced system performance by 30% and ensured timely data availability.
- Executed data ingestion workflows across S3 and Hadoop environments, using PySpark to manage large datasets. This resulted in a 30% reduction in manual tasks and increased storage efficiency by 40% through compression and partitioning techniques.
- Created Tableau dashboards for business insights and presented clear visualizations to stakeholders, driving data-driven decisions.

Skills

- Technical Skills:** Python, SQL, AWS (S3, Redshift, Glue, Lambda, CloudWatch), Azure (Data Lake, Synapse, Data Factory), GCP (BigQuery, Cloud Storage, Dataflow), Apache Spark, Hadoop, Databricks, Airflow, Kubernetes, Docker, Git, GitHub, Hive, Kafka, MySQL, MongoDB, Power BI, Tableau, Advanced Excel (Pivot Tables), Ab Initio, Autosys, Confluence, Linux, Unix, Bash, Shell Scripting, C, C++, Java, R, HTML, CSS, ServiceNow, Agile, Jira, Kanban.

Projects

Azure End-to-End Data Pipeline for Customer and Sales Insights

Oct 2024 - Nov 2024

- Built a scalable data pipeline using Azure Data Factory and Databricks, automating data ingestion, transformation, and storage across Bronze, Silver, and Gold layers in Azure Data Lake.
- Streamlined data processing in Azure Synapse Analytics, implementing partitioning and incremental loading to enhance query performance and enable real-time reporting in Power BI.

Uber Data Analytics Pipeline Using Google Cloud Platform

Feb 2024 - Mar 2024

- Designed an end-to-end ETL pipeline on Google Cloud Platform (GCP) utilizing Google Storage for data ingestion, Mage for data orchestration, and BigQuery for efficient data processing, enabling scalable analytics on Uber trip records.
- Restructured data transformation workflows using Python and SQL within Mage pipelines, ensuring data integrity and creating structured datasets for visualization in Looker Studio, enhancing business insights and reporting efficiency.