

# SAI PAVAN BANALA

+18135731604 | bsaipavan98@gmail.com | Tampa, USA

## CERTIFICATIONS

---

AWS Certified Data Engineer-Associate | AWS Cloud Quest: Cloud Practitioner | Hacker Rank Certified SQL Advanced

## EDUCATION

---

**University of South Florida**  
*Master's, Information Systems*

**August 2022 - May 2024**  
*GPA: 3.9*

## PROFESSIONAL EXPERIENCE

---

### Infosys

*Senior Analytics Engineer*

**Hyderabad, TG, India**

*September 2021 - June 2022*

- Seamlessly executed data ingestion and processing operations using Apache Spark, interfacing with Snowflake and data lake S3, thereby streamlining data workflows and accelerating processing time for extensive datasets.
- Automated large data modeling tasks using EMR Cluster with Spark, resulting in a 50% reduction in data processing time, enhancing data quality.
- Used AWS Lambda functions for automating data processing, leading to a 34% improvement in efficiency.
- Designed and deployed a Kafka-based system to handle high-throughput data ingestion from the in-house IOT systems and real-time processing, enhancing the efficiency and scalability of our data pipeline.
- Developed and maintained complex ETL workflows in Airflow, scheduling and monitoring data processing tasks to ensure timely and accurate data delivery across various systems.
- Achieved almost 99% uptime for the database system by implementing database replication for high availability.
- Skilled in developing comprehensive data models to optimize data storage and retrieval processes.
- Used Spark Data Frame persistency and caching mechanisms to reduce data processing overhead and improve query performance.
- Skilled in data profiling and management using SQL queries, with a focus on creating, testing, and refining queries to support data profiling and optimize data quality and integrity.
- Installed and configured EC2 instances, setting up Linux environments to facilitate various operations. I also created and managed EMR clusters for big data processing.
- Skilled in data visualization to effectively communicate insights and drive strategic initiatives.
- Worked extensively using Agile methodologies to enhance project flexibility and team collaboration.

### Infosys

*Analytics Engineer*

**Hyderabad, TG, India**

*September 2019 - August 2021*

- Created integration objects to ingest data from external storage sources like S3, Salesforce, MySQL Server for efficient loading into Snowflake. Performed necessary transformations using dbt, and integrated with GitHub for version control.
- Automated data ingestion processes between S3 and Snowflake using Event triggers and Snowpipe, improving data availability and freshness for critical business dashboards and decision-making.
- Proficient in utilizing AWS Glue Catalog for effective metadata management and data cataloging, and performed quick data analytics using Amazon Athena, enabling fast and efficient querying of large datasets stored in S3.
- Implemented and optimized RESTful API to integrate diverse data sources seamlessly, enhancing efficiency and reliability.
- Managed and analyzed large volumes of log data daily using Splunk, resulting in proactive issue resolution and a notable improvement in incident response times.
- Experienced in utilizing a data warehouse to store and retrieve structured and unstructured data for analysis and reporting.
- Reduced manual workflows by 20% by converting business requirements into technical specifications, demonstrating strong communication skills and project management skills in a cross-functional team.
- Developed interactive Tableau dashboards showcasing key operational metrics and trends, enhancing data analytics reporting and compliance, and facilitating prompt data-driven decisions.

### Infosys

*Systems Engineer Intern*

**Mysore, KA, India**

*December 2018 - May 2019*

- Completed extensive training in Python, SQL, and Data Structures, alongside in-depth immersion in Big Data technologies, focusing on Apache Spark, Hadoop, and cloud-based data platforms.
- Participating in data engineering projects such as building data warehouses or designing databases.
- Contributing to implementing data transformations to support business objectives and analytical needs.

**BSNL**

*Analytics Intern*

**Vijayawada, AP, India**

*May 2017 - July 2017*

- Conducted data collection and preprocessing for network performance metrics within BSNL infrastructure.
- Collaborated closely with cross-functional teams to design and deploy innovative network monitoring tools, leveraging advanced analytics to enhance reliability and performance.
- Assisted in the documentation of procedures and protocols for network management, ensuring streamlined operations and effective communication across departments.

## PROJECTS & OUTSIDE EXPERIENCE

---

### **Credit Default Analysis (Spark, DBFS, Databricks, SQL, ML Models)**

**Tampa, FL, USA**

*Graduate Student*

*April 2024 - May 2024*

- Leveraged Spark, Databricks, and advanced ML models to predict credit default likelihood using key demographic factors, enabling financial institutions to proactively manage risk and optimize lending strategies.
- Conducted thorough data preprocessing and feature engineering to enhance model accuracy, providing actionable insights that significantly improve decision-making in lending practices.

### **Data Storytelling with Tableau: State Dynamics (NY, CL, FL, TX) (Python, Tableau, Prep Builder)**

**Tampa, FL, USA**

*Graduate Student*

*October 2023 - November 2023*

- Utilized Python, Tableau, and Prep Builder to explore trending states and uncover correlations between crime rates and degree completions, identifying correlation rates between 0.016 and 0.024.
- Developed interactive visualizations to effectively communicate insights, aiding in data-driven decision-making for policy and educational stakeholders.

### **Visualization of industry Mix in St. Petersburg Corridors (Power BI, Python, Excel, DAX)**

**Tampa, FL, USA**

*Graduate Student*

*March 2023 - May 2023*

- Analyzed 2023 industry data for St. Petersburg city corridors using Python and Excel, creating a comprehensive Power BI report to visualize the industry mix for stakeholders.
- Developed an interactive dashboard, which was featured in a university blog, showcasing the distribution and insights for strategic planning and community development.

### **FoodGasam Website (HTML, CSS, Python, SQLite)**

**Tampa, FL, USA**

*Graduate Student*

*March 2023 - May 2024*

- Developed a full stack Flask web application for an Indian food restaurant, enabling online orders and admin functionality using HTML, CSS, Python, and SQLite.
- Implemented robust backend features with the Flask framework to streamline restaurant operations and enhance user experience.

### **A Machine Learning Model: I94 Traffic Analysis (Python, Azure ML)**

**Tampa, FL, USA**

*Graduate Student*

*September 2022 - November 2022*

- Developed a machine learning model for I94 traffic analysis using Python for data mining and Azure ML, achieving 89% accuracy in predictions.
- Leveraged Azure ML to preprocess data, build, and deploy the model, enhancing traffic management insights and operational efficiency.

## SKILLS

---

**Database:** MySQL, Oracle, MSSQL, PostgreSQL, MongoDB

**Big Data:** Hadoop, Apache Spark, Apache Hive, MapReduce, Sqoop, Databricks, AWS (S3, EC2, EMR, Lambda, Glue, RDS, Redshift), Apache Kafka

**Programming:** SQL, Scala, NoSQL databases (MongoDB, Cassandra), Python (Pandas, NumPy), Linux/Unix

**Cloud:** AWS(S3,EC2, EMR, Cloudwatch, Redshift), Databricks, Snowflake

**Data Processing & Analysis:** ETL, Informatica, Airflow, OLAP, OLTP, Tableau, Power BI