

EDUCATION

UNIVERSITY OF SOUTH FLORIDA | M.S. Business Analytics and Information Systems

Aug 2022 - May 2024

TECHNICAL SKILLS

- **BIG DATA:** Hadoop, Spark, Hive, Kafka, Data Warehousing, Avro, Parquet, Unix, Airflow.
- **CLOUD:** Azure Databricks, Azure Synapse Analytics, Azure Stream Analytics, Azure Mapping Data Flows, Azure Data Factory (ADF), Azure DevOps, CI/CD, Snowflake, AWS S3, AWS Lambda, AWS EC2.
- **PROGRAMMING:** SQL, Python, Spark SQL, PySpark, T-SQL, Azure SQL.
- **MACHINE LEARNING:** Supervised, Unsupervised Machine Learning techniques, NLP, Deep Learning.
- **DATA ANALYTICS:** Statistics, Excel (Pivots, PowerPivot, VBA).
- **DATA VISUALIZATION:** Power BI, Tableau, JIRA reporting.
- **DATABASES:** SQL Server, MySQL DB, PostgreSQL DB, Pinecone Vector Database, MongoDB.

WORK EXPERIENCE**STUDENT ASSISTANT | USF IT, TAMPA, USA**

Aug 2022 – May 2024

- Designed a robust **data model in Snowflake**, tailored for handling **Type 2 slowly changing dimensions (SCD 2)**, resulting in a 20% improvement in historical data accuracy and analysis efficiency.
- Spearheaded **batch data loading processes in Snowflake** from multiple sources, reducing loading time by 30% and enhancing data availability and reliability for analytical insights.
- Integrated **Snowflake with Power BI** to create interactive reports, boosting user engagement by 25%. Established **row-level security in Power BI**, ensuring compliance with access control policies and reducing data security incidents by 15%.

AZURE DATA ENGINEER | ACCENTURE SOLUTIONS PRIVATE LTD | BENGALURU, INDIA

Oct 2015 - May 2020

- Spearheaded the utilization of **Agile principles** to convert business requirements into scalable programming logic, ensuring that semi-structured data seamlessly loaded into Impala target tables, reducing data processing time by 40%.
- Leveraged **Azure Databricks and Azure Data Factory** to implement diverse data loading schedules, including hourly, monthly, and quarterly intervals, enhancing operational flexibility and efficiency.
- Designed and implemented **data pipelines for Near Real-Time Streaming data ingestion** and processing using **Apache Kafka and Apache Spark**, enabling timely insights and decision-making.
- Directed the effective utilization of **Machine Learning strategies on Geospatial Data**, resulting in a \$5 million profit margin enhancement and optimized resource allocation, reducing operational costs by 15% while maintaining service quality standards.
- Developed interactive **Power BI** visualizations on telecom datasets using, empowering clients with insights into telecom signal strength for informed, data-driven decision-making.
- Performed a comprehensive data analysis and **predictive modeling to forecast customer Churn Rates**, resulting in a 15% reduction in customer attrition and increased revenue retention.
- Collaborated cross-functionally with stakeholders to identify **Key Performance Indicators (KPIs)** and develop dashboards in **Power BI** for real-time monitoring, optimizing decision-making processes and operational efficiency.

PROJECTS**ON PREMISE STRUCTURED DATA MIGRATION TO AZURE CLOUD PLATFORM: ([GitHub](#))**

- Created a comprehensive data migration solution leveraging **Azure Data Factory** to migrate on-premise structured data into **Azure cloud**, followed by data transformations using **Azure Databricks** and loading into the gold layer of **Azure Storage Gen 2**.
- Engineered custom views and reports in **Power BI** by leveraging **Azure Synapse Analytics**; stored access credentials securely in **Azure Key Vault** and incorporated a robust security protocol through **Azure Active Directory**.

AZURE DATABRICKS CI/CD PIPELINE IMPLEMENTATION AND ENVIRONMENT SETUP: ([GitHub](#))

- Achieved seamless **DEV-to-PROD** deployment for **Azure Databricks** notebooks via end-to-end **CI/CD pipelines**, integrating **Azure DevOps** and utilizing **YAML** for pipeline creation, reducing deployment time by 30% and ensuring 100% successful deployments.
- Demonstrated environment setup, branch protection, and permissions configuration for seamless dev-to-prod deployment.

TWITTER SENTIMENT ANALYSIS ON THE RUSSIA-UKRAINE CRISIS: ([GitHub](#))

- Optimized processing time by 40% with **Azure Databricks**, handling 3 million tweets. Utilized **Logistic Classifier Model** and achieved 82% Accuracy, ensuring precise classification of positive and negative sentiments.

AZURE CLOUD BASED MOVIE RECOMMENDATION SYSTEM: ([GitHub](#))

- Orchestrated a comprehensive data pipeline in **Azure Data Factory (ADF)** to handle movie data, incorporating tasks such as metadata extraction, schema validation, file copying, and notebook execution.
- Developed a scalable **Azure Databricks Notebook** utilizing **Collaborative Filtering with the ALS (Alternating Least Squares) algorithm**, enhancing movie recommendation accuracy by 25%.
- Integrated **Azure Data Factory (ADF)** with **Azure Logic Apps**, automating email notifications for pipeline events, ensuring timely responses to successful recommendations or errors, resulting in a 50% reduction in response time to issues and improved operational efficiency.

CRICKET DATA ANALYSIS USING AWS AND SNOWFLAKE: ([GitHub](#))

- Created an **AWS Lambda service** to extract and store over 50000 API records into **AWS S3 bucket**, automating data loading into **Snowflake using Apache Airflow** on **AWS EC2 instance**, achieving a 100% success rate.
- Flattened and processed **deeply nested JSON data** for efficient storage into **Snowflake**, enabling integration with **Power BI** to create over 3 dynamic reports, with automated email notifications on successful data loads from **S3 buckets into Snowflake**.

AMAZON SALES DATA ANALYSIS FOR MOBILE PHONES USING AWS AND SNOWFLAKE: ([GitHub](#))

- Engineered a data pipeline to process Amazon sales data for mobile phones over a year, ingesting **CSV, JSON, and Parquet formats via AWS Lambda into S3**.
- Utilized **Snowflake** to transform and analyze the data and developed visualizations using **Power BI**.
- Automated the entire workflow with **Apache Airflow** on an **EC2 instance**, ensuring seamless and efficient data processing.