

EDUCATION

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

Aug 2022 - May 2024

TECHNICAL SKILLS

- **BIG DATA:** Hadoop, Spark, Data Warehousing, Avro, Parquet, Unix, Airflow.
- **CLOUD:** Snowflake, AWS S3, AWS Lambda, AWS EC2, AWS Redshift, AWS Kinesis, Azure Databricks, Azure Synapse Analytics, Azure Stream Analytics, Azure Mapping Data Flows, CI/CD.
- **PROGRAMMING:** SQL, Python, PySpark, T-SQL, Azure SQL.
- **MACHINE LEARNING:** Supervised, Unsupervised Machine Learning techniques, NLP, Deep Learning, Statistics.
- **DATA VISUALIZATION:** Power BI, Tableau, JIRA reporting.
- **DATABASES:** MySQL 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.
- Initiated and optimized **batch data loading** processes in **Snowflake** from diverse sources, cutting loading time by 30% and improving data availability and reliability for critical 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.

AWS 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 **Amazon Redshift, AWS Glue, and AWS Lambda** to implement diverse data loading schedules, including hourly, monthly, and quarterly intervals, enhancing operational flexibility and efficiency by 100%.
- Architected and managed **near real-time** data ingestion and processing pipelines using **AWS Kinesis and Databricks**, boosting data availability by 35% and empowering timely and informed decision-making processes.
- 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.
- Crafted detailed **Power BI** reports on telecom signal strength; empowered clients with critical insights, resulting in a 30% increase in data-driven decision-making and strategic planning capabilities.
- Performed a comprehensive data analysis and **predictive modeling** to forecast **customer Churn Rates** using **AWS SageMaker**, resulting in a 15% reduction in customer attrition and increased revenue retention.
- Pioneered the creation of real-time monitoring dashboards in **Power BI**, collaborating with stakeholders to identify critical **KPIs**, optimizing operational efficiency, and boosting decision-making speed by 20% within the first quarter.

PROJECTS

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**.
- Incorporated **Snowflake** to transform and analyze the data and developed 3 visualizations using **Power BI**.
- Automated the entire workflow with **Apache Airflow** on an **EC2 instance**, ensuring seamless and efficient data processing with 100% accuracy.

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

- Architected an end-to-end data migration framework using **Azure Data Factory** and **Azure Databricks**, transitioning on-premise data to **Azure cloud**; enhanced data processing speed by 50%.
- 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.

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%.
- Streamlined email notification process for pipeline events by integrating **Azure Data Factory** and **Azure Logic Apps**; accelerated response times by 50% and enhanced overall operational workflow.