Prathyusha Thummalapalli

+18132918953 | tprathyusha1705@gmail.com | GitHub | LinkedIn

EDUCATION

University of South Florida, GPA 3.93/4.00   **Aug 2022 - May 2024**
Master of Science, Business Analytics and Information Systems
G Narayanamma Institute of Technology and Sciences, GPA 3.85/4.00   **Aug 2015 - May 2019**
Bachelor of Technology, Computer Science and Engineering

PROFESSIONAL EXPERIENCE

Graduate Assistant, USF, Tampa   **Dec 2022- May 2024**
- Developed an application using Python for data filtering and targeted communication by retrieving data from course-level API and integrating Canvas messaging API leveraging Postman and SQL, reducing manual efforts by 70%.
- Instructed and guided 800+ students on MS Office and Tableau, fostering a data-driven culture.
- Collaborated with cross-functional teams to establish scalable data solutions, fostering a data-driven culture while contributing innovative ideas for enhancing student learning.

Consultant, Syniti, Hyderabad   **Apr 2021 - Jul 2022**
- Migrated on-premises file parsing scripts to PySpark on Databricks, achieving a significant reduction in a single day's data parsing time from 2 hours to 15 minutes.
- Engineered ETL pipeline in Azure Data Factory, slashing processing time by 30% for parsed data files. Loaded the data into Azure Synapse for swift analytics and KPI tracking, expediting fault detection.
- Implemented Spark Structured Streaming with autoloader for 40% faster file ingestion into Azure Data Lake V2. Integrated Azure Stream Analytics for Precise Change Data Capture (CDC). Automated workflows through Azure Logic Apps.
- Optimized database structures by introducing star schema, leading to a 25% reduction in query response times and enhancing overall system efficiency.
- Created and delivered over 50 customized ad hoc reports, enabling data-informed decision-making, while educating sales teams on insights resulting in a 20% increase in repeat business from clients.
- Integrated Azure DevOps for streamlined CI/CD pipelines, deploying data artifacts like notebooks and ETL processes. This reduced error rates by 25% through effective branch utilization for development and testing.
- Built a PySpark machine learning model on Databricks for predicting battery health using regression and time series analysis. Visualized trends, triggered customer alerts, and saved $10 million through timely battery replacements.

EIM Analyst, DMR-A Syniti Company, Hyderabad   **Jul 2019 - Mar 2021**
- Automated Cron jobs with Glue Data crawlers, using scheduled and conditional triggers, improving job efficiency by 30% and enabling timely data loading to S3. Incorporated Athena to analyze data for retrieving EPS statistics.
- Enriched the raw unstructured data into structured normalized data using Spark jobs as a consumer for Kafka brokers. This minimized data duplication and manual checks on the data from approximately 1Hr to 5min.
- Orchestrated event-driven step functions in VPC to create serverless architecture for loading the client's EPS pipeline metadata using lambda functions into Dynamo DB with DAX clusters which enhanced the computing costs by 30%.
- Facilitated data engineers in transforming the data loading pipeline from Redshift to Snowflake by designing UDFs on file formats (JSON, Parquet) from S3, resulting in a 40% cost reduction.
- Improved performance by migrating Python code from on-premises to AWS EMR, reducing the run time from 2 hours to 35 minutes with PySpark Data Frames on Databricks.
- Leveraged PySpark on Databricks integrated with AWS to perform data validation checks for the files which are more than 1TB in size present in S3 after successful job completions.
- Implemented a Hybrid architecture for on-premises pipelines, enhancing data loading to MySQL on-premises. Leveraged Self-Hosted Integration Runtimes and AWS FSx, resulting in a 40% speed improvement.

Intern, DMR-A Syniti Company, Hyderabad   **Jan 2019 - Jun 2019**
- Streamlined ETL process by utilizing APIs to extract JSON format from the client's internal website and loading the data into Data Frames reducing the processing time from 1 hour to just 5 minutes.
- Transformed the Databricks notebooks from pandas to PySpark and Spark SQL, optimizing data processing for >10TB datasets. Slashed workflow time from 90 hours to just 10 hours, achieving a remarkable efficiency boost.
- Integrated SQS and SNS in AWS, leveraging Lambda functions for automated data loading from client servers to S3. Reduced manual file collection time from 6 hours to 2 hours.

ACADEMIC PROJECTS

Visualization of Plastic Waste Production and Plastic Waste Mismanagement
A comprehensive data visualization project utilizing **Tableau** to illustrate the production and mismanagement of plastic waste. The project included Data Collection and Aggregation from multiple sources, Geospatial Analysis, and Comparative Analysis.
Visualization of a Vehicle dataset
Created a Power BI report to present the focus areas in terms of model/zone/defect type as there are quality issues in manufacturing to reduce the rework costs and ensure a better product for customer.
House-A-Bull
Built a web application to help students of USF find accommodation using Python Flask, HTML, CSS, Bootstrap, and JavaScript. Added dynamic visualizations using chart.js to show the analytics as per the search results.
Text Analysis on Annual Reports
Created a dataset from unstructured data of annual reports of 25 companies which is extracted by **Python Scripts** and performed **data wrangling**. Analyzed the resulting dataset and ran a topic model to determine the topics in shareholder letters.
Movielens
Executed a comprehensive data warehousing project involving **dimensional modeling**, Talend ETL for loading 25 million Movielens data into **Snowflake**, SQL analytics, Tableau visualization, and **predictive modeling** with Lenskit's Item-Item and ALS algorithms on ratings data. Demonstrated expertise in database design, analytics, and predictive modeling for valuable insights from large data.

TECHNICAL SUMMARY

Certifications- MCSA Certified - Querying Microsoft SQL Server 2012/2014, Microsoft Certified- Azure Data Fundamentals
Languages & Tools - SQL, Python, R, PySpark, Microsoft Azure, AWS, Azure Databricks, Airflow, MS Office, Tableau, Power BI
Core Competencies- ETL, Data Migration, Big Data Analytics, Data Conversion, Project Management, Data Warehousing