MADHURI PALURI

Newark, California

📞 551-344-7396 ✉ palurimadhuri22@gmail.com 💼 linkedin.com/in/madhuri-paluri-822716198/ ⬛ github.com/MadhuriPa98

Professional Summary

Data Engineer with 5+ years of experience in designing and implementing scalable data pipelines, optimizing SQL queries, and managing cloud-based data solutions. Proven expertise in using AWS services, Databricks, and various big data tools to streamline data processing and enhance ETL efficiency, resulting in significant cost savings and operational improvements. Demonstrated ability to develop real-time data solutions, create interactive dashboards, and integrate machine learning models to drive business insights and improve decision-making. Recognized for problem-solving skills and awarded for impactful contributions to team success and client satisfaction

Education

Stevens Institute of Technology — **Aug. 2022 – May 2024**
MS Thesis in Computer Science — *Hoboken, New Jersey*

Thesis Title : A Efficient Sub-Sentence classification using student teacher networks

Technical Skills

Programming Languages: Python, Java, C, HTML/CSS, JavaScript, SQL
Developer Tools: VS Code, Eclipse, Google Cloud Platform, Android Studio
Technologies/Frameworks: Linux, Jenkins, GitHub, PyTorch, Tensorflow
Cloud Technologies: DataBricks, Ansible, Terraform, Kinesis, EMR, DynamoDB, Jenkins, sagemaker, CloudWatch, AWS Glue, AWS Lambda
BigData Tools: Spark, Kafka, Airflow, Snowflake, Hive, Sqoop, Terraform, Presto, Reltio, Redshift
DataBases: Oracle, MySQL, PostgreSQL, SQL Server, MongoDB

Experience

Stevens Institute of Technology — **August 2023 – Current**
Data Operations Analyst — *Hoboken, NewJersey*

- Created Pipelines in ADF using Datasets/Pipeline/ to extract, transform, and load data from different sources API, Blob storage
- Implemented several scripts in PySpark on databricks platform like filters, merge, aggregations by ensuring data passes 90% quality checks
- Utilized Azure synapse analytics along with stream anlaytics to capture learning management systems(LMS) activity of students and enhanced engagement by 20% in year with LMS by building effective ML solutions using captured data.
- Built custom assignment tracker pipeline in canvas using airflow by creating DAGs and setting dependencies of tasks and monitoring them using logs achieveing task success rate 85% by implementing data quality checks and retry logics
- Enhanced PySpark scripts written in databricks by employing optimizations and saved 150k by changing autoscaling options, governing cluster configuration operations

Tata Consultancy Services — **July 2019 – July 2022**
Data Engineer — *Hyderabad, India*

- Enhanced SQL Query optimization by 45% using suitable indexing, Window functions and Common table expressions (CTEs)
- Orchestrated data syncing as parquet from source into AWS S3 using python and loaded into AWS Glue Data Catalog crawlers to create raw tables for files
- Developed scalable data pipelines to asynchronously ingest and process 1k+ events per minute, using Python, AWS lambda which is serverless (framework)
- Enhanced AWS Glue scripts with PySpark to create dataframes and implement business requirement transformations on top it with SparkSQL
- Optimized the Glue workloads for different type of data loads by choosing right compression, cluster type, instance type storage type to analyze data with low cost and high scalability.
- Solved complex data integration issues using Data build tool(DBT) by implementing data lineage tracking features, resulting in a 50% reduction in data pipeline errors and a 30% increase in overall ETL efficiency, significantly enhancing data accuracy and reliability for business decision-making
- Developed real-time streaming pipelines to ingest clickstream data into Snowflake using Kafka and S3 and collaborated with ML team for building personalised recommendation engine based on clickstream data

- Developed RESTful APIs in flask and integrated with JSON serialization to eliminate data silos which enable real-time data synchronization, resulting in a 60% reduction in manual data entry and a 40% improvement in operational efficiency

Accenture   **May 2018 – April 2019**
Data Engineer   *Banglore, India*
- Designed mappings for moving data from STAGING layer to the Data warehouse and loading data to fact and dimension tables using STAR schema architecture resulting in less than 10% data leakage
- Engineered end to end ETL pipelines while ingesting 10gb data from external APIs to snowflake data warehouse by using snow pipe or external pipelines and loaded the data to amazon S3 buckets
- Worked with Snowflake cloud data warehouse for integrating data which include loading nested JSON formatted data into snowflake table.
- Collaborated with SRE/Infrastructure team to orchestrate CI/CD pipelines using gitHub actions and deployed them using kubernetes
- Automated weekly report generation and visualization dashboards for communicating the performance indicators with business leaders using Tableau

Projects

JOE CRM – A replacement for Traditional CRM systems | *Groq, Llama3, LECL, BraveAPI, VERCEL*   **May 2024**
- Revolutionized traditional data driven CRM systems to action driven using LLAMA3 70B which creates a campaign plan for prospective customers by sending highly customized emails.
- Used Langchain Expression language (LECL) for multi-step conversations thereby enabling sophisticated prompts and deployed it using FASTAPI.
- Used Brave API to get the relevant history of customers and curated prompts for GROQ powered LLAMA3 which helps in analyzing users for customer success managers.
- Leveraged VERCEL for building action centric front end user interface and deployed JOE CRM on AWS.

Real time fleet monitoring app for ride sharing company | *AWS Lambda, Kinesis, Quicksight*   **December 2021**
- Triggered data events with AWS lambda and performed several transformations, filtering and moved the data to kinesis firehose for real time delivery.
- Developed Dashboards using Amazon QuickSight by dividing the fleet into different zones and providing the data responsible for fleet drop to app company which resulted in 20% surge in business

Certifications

- **AWS Certified Data Engineer**, AWS.
- **Microsoft Certified Data Engineer DP-203**, Microsoft.
- **ML Devops Engineer**, Udacity

Achievements

- Received **Best Employee award** while working with TCS for making impact in the team through solving production issues which fostered client business by saving $400K to client
- Achieved **4th** place in Hackathon out of 55 teams conducted by **META** in year 2024 where we developed end to end automated CRM application.