

Rama Krishna Poluru

San Jose, CA | +1 5103998039 | ramakrishna.poluru1999@gmail.com
| [LinkedIn](#) | Website | [GIT](#) | [Medium](#) | [DataLemur](#) | Slack | Reddit |

WORK EXPERIENCE

Data Engineer Intern, ECC, Burlingame, California

May 2023 – Present

- Developed and automated **SQL jobs**, advanced SQL queries to draw insights in **Power BI**, resulting in a **45%** reduction in time spent on manual data processing.
- Collaborated with **Safety, Business Proposal, and Project teams**; optimized data flow to enhance analytics, resulting in a **20% faster delivery** of key reports and a **15% reduction** in data discrepancies.
- Spearheading the design and development process of data warehouse, NLQ chatbot from legacy on premise database for business needs on **Azure** which also involved **data integration, data ingestion, data modeling**, security, management.

Data Engineer II, Microsoft, Hyderabad, India.

Jan 2022 - Jul 2022

- Implemented and maintained large scale BI solution using Azure services, **data lake, Azure Databricks**, Integrating **ADB with ADF** pipelines and BI system resulting in **1.63%** surge in revenue and **2%** upsurge in customer retention
- Coordinated a team of **6** and addressed **cross functional teams** in an **Agile** environment owning product backlogs including requirements gathering from **DataOps** and demoing incremental builds (**CI/ CD**) on **Azure DevOps**.
- Enhanced data architecture, independently designing efficient ETL processes. Actively improved and maintained existing **production environments**, ensuring seamless data flow to downstream users and systems.

Data Engineer I, Microsoft, Hyderabad, India.

Jun 2020 - Dec 2022

- Migrated on-premises data warehouse to azure cloud-based solution, developing 40+ notebooks (**Python, Spark SQL**) with **zero pipelines failures and data discrepancies** on the production environment on pilot deployment.
- Streamlined **ELT** pipelines on **Azure Data Factory** and 150+ **databricks** notebooks for 5 dependent snapshot systems to process upwards of **20M + records** every 6 hours to capture businesses past and current trends.
- Initiated a validation framework utilizing **Python and Scala** to validate data across stages of data engineering lifecycle by automating manual process and reducing manual efforts of **6 teams** by **9hrs** daily.
- Optimized the execution time of **5 big data** batch pipelines by **10%** by implementing spark techniques, efficient coding standards and aligning data transformation logic aligning to true business needs.
- Maintained data platform **robust against data quality issues** by conducting daily, weekly tests and continuously optimizing performance by reducing time complexity and cost azure resources utilized.

Associate Data Engineer, MAQ Software, Hyderabad, India.

Dec 2019 - May 2020

- Elevated user experience by reducing report load time to **9 seconds**, streamlined analytics by architecting **Power BI** reports with **DAX** measures for **5 KPI dashboards** with **RLS**, and enhanced data accessibility and security.
- Engineered a data warehousing solution with **star and snowflake** schemas, reducing data retrieval times by **30%** and providing actionable insights, substantially aiding services and marketing departments.
- Designed and executed **SSIS ETL** packages, cutting data processing time by **11%** over multiple terabytes, and utilized **SSAS** for advanced **data analytics** through multidimensional cubes, bolstering data-driven decision-making.

ACADEMIC AND PERSONAL PROJECTS

- **United States H-1B Visa Data System for Operational Usage and Analytics** (*AWS Lambda, AWS Glue, Snowflake, MySQL, Tableau Python, Pandas, OLAP, Kimball dimensional modeling, ERD, Star Schema*) ([Medium](#)) ([Git](#))
- **Uncovering Coordinated Campaigns: Artificial Inflation of Trend Popularity on Twitter.** (*AWS EMR, PySpark, Python, AWS Redshift, QuickSight*) ([Medium](#)) ([Git](#))
- **Forecasting Drought Area Percentage in California using Machine Learning Algorithms** (*Decision Tree, Random Forest, LSTM, ANN, Jira, CRISP-DM, WBS, Gantt & PERT chart*) ([Git](#))
- **Real-Time Stock Market examination with Kafka** (*EC2, Athena, Apache Kafka*) ([Git](#))

SKILLS

- **Big Data Technologies:** Hive, Spark, HDFS.
- **Cloud:** Glue, EMR, Lambda, Azure Databricks, Azure Synapse, Microsoft fabric.
- **Databases and Data warehouses:** Snowflake, Google Big Query, MySQL, MS SQL Server, MongoDB (NoSQL).
- **Programming:** C, Python, Pandas, NumPy.
- **ETL and Visualization:** Airflow, DBT, SSIS, SSAS, Azure Data Factory, Power BI, Tableau, QuickSight.
- **Tools and Operating Systems:** Git, Visual Studio Code, Ubuntu.

EDUCATION

San Jose State University, San Jose, California, USA

Aug 2022 - May 2024

Master of Science, Data Science & Analytics

Jawaharlal Nehru Technological University, Hyderabad, India.

Aug 2016 - Sep 2020

Bachelor of Technology, Computer Science & Engineering.