

Rama Krishna Poluru

ramakrishna.poluru1999@gmail.com | +1 5103998039 | San Jose, CA. | [Portfolio](#) | [LinkedIn](#) | [GIT](#) |

WORK EXPERIENCE

Data Engineer Intern, ECC, Burlingame, California

May 2023 – Present

- Implemented a BI solution with **Azure**, utilizing data lakes, Azure Active Directory, ADB, and ADF, achieving **99%** uptime and ensuring **96%** of data integrity for reports and dashboards.
- 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.
- Drove the strategic **data architecture** evaluation for the development of a NLQ chatbot across Azure **Synapse**, Microsoft **Fabric**, and other services, guiding cost-effective and compatible solutions.

Data Engineer II, Microsoft, Hyderabad, India.

Jul 2021 - Jul 2022

- Migrated on-premises data warehouse to azure cloud-based solution, developing 40+ notebooks (**PySpark**, **Spark SQL**) with **zero** pipeline **failures** and data discrepancies on the production environment on pilot deployment.
- Optimized execution time of **3** big data **batch pipelines** by **60%** by employing Spark techniques, efficient coding standards, and refining data transformation logic to meet business needs and minimizing data **redundancy**.
- Maintained **data platform** integrity and performance through daily and weekly testing, peer code reviews, and continuous optimization of notebook execution times, significantly reducing Azure costs by nearly **\$0.2M** annually.
- 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.

Data Engineer I, Microsoft, Hyderabad, India.

Jun 2020 - Jun 2021

- Streamlined **ELT** pipelines using **Azure Data Factory** and 150+ **Azure Databricks** notebooks for 5 dependent systems to process upward **20M +** transactional records every **6** hours to enhance **business intelligence** processes.
- Initiated a **Python** and **Scala** based validation framework to automate data validation across the data engineering lifecycle, reducing error rates by **15%** and saving **9 hours** of daily manual effort for **6** teams.
- Independently enhanced **data pipelines** through efficient **ETL** process design, actively optimizing and maintaining production environments for uninterrupted data flow to downstream users and systems.

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.
- 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.
- Optimized data analysis time by developing **advanced SQL** queries and stored procedures, enhancing Power BI report efficiency with a **45%** reduction in processing and refresh times.

Junior Data Engineer, Fratello Innotech, Hyderabad, India.

May 2019 - Nov 2019

- Enhanced billing automation for major retailers (Amazon, Flipkart, Snapdeal) using APIs, Selenium, and **Airflow**, boosting efficiency by **40%** and cutting labor by **60%**.
- Streamlined custom **ETL** process for a finance client using AWS **Glue** and ingested in AWS **RedShift**.
- Performed auditing, inspecting and visualizing **S3** server access logs using AWS **Lambda**, and AWS **Athena**.

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](#))
- **Real-Time Stock Market examination with Kafka** (*EC2, Athena, Apache Kafka*) ([Git](#))

SKILLS

- **Big Data Technologies:** Hive, Spark, HDFS.
- **Cloud:** AWS Glue, AWS EMR, AWS Lambda, Azure Data Bricks (ADB), 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 (ADF), Power BI, Tableau, QuickSight.
- **Tools and Operating Systems:** Git, Visual Studio Code, Linux.

CERTIFICATIONS AND ACHIEVEMENTS

- [Best student award](#) in Database Systems and Analytics class (SJSU). Star of the Quarter (MAQ Software).
- [70-761](#): Querying Transact-SQL. [70-778](#): Visualizing Data - Power BI, [DataLemur](#), [Microsoft Learning Badges](#)

EDUCATION

Master of Science, Data Science & Analytics, San Jose State University, San Jose, California, USA.

Bachelor of Technology, Computer Science. JNTU, Hyderabad, India.