Rama Krishna P

poluru.ramakrishna99@gmail.com | +1 669-200-9286 | **Open to relocation, USA** | Portfolio | LinkedIn

WORK EXPERIENCE

Data Engineer, ECC, Burlingame, California   May 2023 – Feb 2024
- Implemented a BI solution on **Azure**, utilizing ADLS, Lakehouse, Azure Active Directory, ADB, and ADF, achieving **99%** uptime and ensuring **96%** of data integrity for reports and dashboards.
- Engineered a **data warehouse** 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**, **AWS**, and **GCP**, guiding cost-effective and compatible solutions.

Data Engineer II, Microsoft, Hyderabad, India.   Jul 2021 - Jul 2022
- 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 deliver **business intelligence** insights.
- 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 by gathering requirements from **DataOps, Stakeholders** and demoing incremental builds (CI/ CD) on Azure DevOps.

Data Engineer I, Microsoft, Hyderabad, India.   Jun 2020 - Jun 2021
- Migrated on-premises **business intelligence** system to Azure, developing 40+ databricks notebooks **(PySpark, SQL)** with **zero** pipeline failures and data discrepancies on the **production environment** on pilot deployment.
- Initiated a **Python, Scala, Spark SQL** based validation framework to automate data validation across the data engineering lifecycle, reducing error rates by **85%** 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 teams and systems.

Associate Data Engineer, MAQ Software, Hyderabad, India.   Dec 2019 - May 2020
- Elevated user experience and reduced report load time to **9 seconds**, streamlined analytics in **Power BI** along with RLS thereby enhancing data accessibility and security, developed **DAX** measures for 5 **reports** and **dashboards**.
- Designed and executed SSIS ETL packages, cutting data processing time by **11%** over multiple terabytes, and utilized SSAS for advanced data analytics through tabular models, 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
- Integrated workflows utilizing flat files (JSON, CSV), and **REST APIs**. Worked on **SQL** tasks including DDL, DML, TCL, **window functions**, subqueries, **CTEs**, and focused on development of **stored procedures** and **views.**
- 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 formats and technologies:** Parquet, ORC, Hive, Apache Spark, HDFS, Hadoop, Map Reduce, Kafka.
- **Cloud:** AWS Glue, EMR, Lamda, Azure Databricks, Azure Synapse, Microsoft Fabric, GCP, Snowflake.
- **Databases and Data warehouses:** Google Big Query, MySQL, MS SQL Server, MongoDB, Postgres.
- **Programming:** C, Python, Java, Pandas, NumPy.
- **ETL and Visualization:** Airflow, DBT, SSIS, SSAS, Azure Data Factory, Power BI, Tableau, QuickSight.
- **Tools and Operating Systems:** Git, Visual Studio Code, Linux.

CERTIFICATIONS AND ACHEIVEMENTS
- Microsoft Certified Azure **Data Engineer Associate**, Databricks Certified **Data Engineer Associate**.
- 70-761: Querying Data with Transact-SQL, 70- 778: Analyzing and Visualizing Data with Microsoft Power BI.
- Best student award (SJSU). Star of the Quarter (MAQ Software), DataLemur, Microsoft Learning Badges.

EDUCATION
Master of Science, Data Analytics, San Jose State University, San Jose, California, USA.   **2022 - 2024**
Bachelor of Technology, Computer Science. JNTU (MREC), Hyderabad, India.