SRIRAM REDDY BOLLAM

Fremont, CA, 94536

sriramreddy045@gmail.com linkedin.com/in/sriram-reddy1 +1-(571)-977-0683

Summary

Accomplished Big Data Analytics Engineer with 3 years of experience in designing and implementing efficient ETL processes, leveraging **Python**, **SQL**, **distributed processing**, **Hadoop**, **Spark**, and **cloud-based data platforms**. Excelled in collaborating with cross-functional teams to deliver high-quality solutions, resulting in a **40% reduction in data processing time**. Proficient in both big data engineering and **data analytics**, with strong expertise in **statistical analysis** and **machine learning**, driving actionable insights and optimizing data workflows to support business objectives.

Education

George Mason University **January 2021 – May 2022**

M.S., Data Analytics Engineering (Spec: Big Data Analytics, Machine Learning) *VA, USA*

BML Munjal University **August 2016 – June 2020**

B.S., Computer Science and Engineering *Gurgaon, India*

Technical Skills

Programming Languages: Python, Java, SQL
Cloud Platforms: AWS , Azure , Databricks
Big Data Technologies: Hadoop, PySpark, Spark Optimizations, Hive, HBase, Cassandra DB, Kafka, Spark Streaming, MapReduce
Data Processing & Analysis: Pandas, NumPy, Data Pre-processing, Statistical Analysis, Hypothesis Testing, A/B Testing, Machine Learning, SciPy, Matplotlib
Databases: PostgreSQL, MySQL, SQL Server, Snowflake
Data Visualization: Power BI, QuickSight, Tableau
DevOps & Version Control: DevOps, Change Management, Jenkins, Docker, Git
Miscellaneous: Data Structures and Algorithms, Distributed Systems, Unix, Data Modeling, Airflow, VS Code, Jupyter Notebook, Anaconda, ETL, ELT, Data Warehousing, SSIS, SSRS, SSMS.

Professional Experience

Data Engineer, Datics Inc **November 2023 – Present**

San Jose, CA, USA

- Defined and optimized **ETL pipelines** using **Azure Data Factory** for 10TB daily data processing, reducing errors by 5,000 records per batch and improving pipeline efficiency by 30% using **Azure Monitor**.
- Developed **PySpark** data pipelines in **Azure Databricks**, transforming various file formats and implementing **Delta Lake**, improving data consistency and query performance by 40%.
- Created **Power BI** dashboards for inventory management with **Azure Databricks**, and executed ETL processes to **Snowflake** and **S3**, reducing processing time by 6 hours.
- Optimized **Spark jobs** with join strategies like **broadcast hash join**, enhancing performance and efficiency. Utilized Databricks utility functions and **AQE** for data skewness management.

Software Engineer, Amazon **September 2022 – October 2023**

East Palo Alto, CA, USA

- Spearheaded **ETL pipeline optimization** with **AWS Glue**, **Lambda**, **PySpark**, and **S3**, automating 95% of 3+ million daily sales records processing, saving 12 hours of manual work weekly, and minimizing data transmission errors.
- Engineered reusable PySpark ETL modules to standardize transformation logic, accelerating development by 50% and enhancing code maintainability.
- Optimized SQL queries by rewriting joins, adding indexes, and programming stored procedures, reducing CPU utilization by 30% and improving response times by 40%. Analyzed data to provide actionable insights to stakeholders.
- Enhanced **ETL data pipeline** performance using **Spark** optimization techniques such as broadcast joins, partitioning, bucketing, and compression techniques , reducing job runtime by 50% and improving data processing efficiency.
- Structured SQL stored procedures to automate data preparation for AWS QuickSight dashboards, reducing dashboard build time by 30% and ensuring timely data availability for reporting.
- Led the database system upgrade to the latest version with best DevOps practices, ensuring minimal disruption and reducing customer downtime by 40%, while ensuring compliance with SLA requirements.
- Refined and automated **AWS RDS PostgreSQL extension tests** using a **CI/CD pipeline**, reducing regression testing time by 50% during major version releases, ensuring efficient and reliable updates.
- Managed the software development lifecycle (SDLC) in an agile environment, delivering projects on time and responding to evolving project requirements.

Data Analyst Intern, Accure Analytics **January 2022 – May 2022**

Fairfax, VA, USA

- Collaborated with a team of 5 to scrape and collect image data from medical sources for building an AI-powered **skin disease detection model**.
- Engineered a comprehensive **data pre-processing pipeline** to normalize image sizes, convert color spaces, and augment data, improving model accuracy by 18% over baseline by addressing issues with images exposing other body parts.

- Implemented quality assurance checks and data validation protocols, reducing data discrepancies by 30% and improving training program performance.
- Created interactive **Power BI dashboards** to track model performance by disease type, body region, age groups, and gender, enabling stakeholders to identify gaps and make informed decisions.

Data Analyst Intern, Innomatics **January 2020 – December 2020**

Hyderabad, Telangana, India

- Revamped data system by integrating **Hive** with **HBase** for rapid data retrieval and efficient aggregation, reducing data retrieval time by 60%.
- Conducted in-depth analysis of learner engagement and course completion data using SQL and Python, enabling targeted interventions that increased course completion rates by 20%.
- Developed and maintained interactive dashboards with Tableau and QuickSight, providing real-time insights on training program performance and learner progress, which improved decision-making and reduced data discrepancies by 15%.

Academic Projects

F1 Racing Data Analysis

- Orchestrated **data engineering** for Formula 1 data using **Azure Databricks**, **Delta Lake**, **Unity Catalog**, and **Azure Data Factory**. Designed an **ETL pipeline** for incremental loads, ensuring **data integrity** and **scalability**.
- Developed **Lakehouse architecture**, reducing latency by 45%, and established interactive **dashboards** with **Power BI**, **PySpark**, **Spark SQL**, and **Azure KeyVault** to enhance security and reporting.

Crime Data Analysis in Chicago

- Analyzed crime rates using **Python**, **Pandas**, **SQL**, **R**, and **Excel**, improving predictive accuracy by 25%. Conducted **hypothesis testing** (Chi-Square Test) on arrests and domestic crimes.
- Generated a prediction model with **Machine Learning** algorithms (**Random Forest**, **Logistic Regression**, **XGB Classifier**), achieving 87.8% accuracy and creating visualizations to support law enforcement.

Certifications

AWS Certified Cloud Practitioner - *AWS*
DP900 – Microsoft Azure Data Fundamentals - *Microsoft*
DP203 – Microsoft Azure Data Engineer Associate - *Microsoft*
Generative AI Fundamentals - *Databricks*