Sriharshini Addagudi

7736836327 ◇ harshini.addagudi@gmail.com ◇ United States ◇ LinkedIn

SUMMARY

IT professional with 5+ years as a Data Engineer with expertise in SQL, PL/SQL, and NoSQL programming and a solid background in developing and maintaining data pipelines, creating impactful analytics, and implementing machine learning models. My skill set spans diverse industries and technologies, including Azure, AWS, GCP, and various data visualization tools like Tableau and Power BI. With extensive ETL experience, I excel in designing, developing, and optimizing end-to-end data integration processes. Additionally, I am well-versed in the project life cycle and Agile methodology, ensuring efficient project delivery and continuous improvement through iterative development and collaboration.

SKILLS

Programming Languages SQL, Java, R, Python (Matplotlib, Seaborn, NumPy, SciPy, Pandas, Keras, Scikit-Learn)
Databases SQL Server, MS-Access, Oracle, Teradata, PostgreSQL, Mongo DB, MySQL, SAP HANA
Cloud Platforms AWS, Azure, GCP
ETL SSIS, Informatica, Talend, SAP Data Services, Databricks
Big Data Hadoop, Hive, HDFS, HBase
Others Git, Tableau, Informatica, SSRS, Power BI, JIRA, Kafka, Kubernetes, Jenkins (CI/CD), Cloudera

EXPERIENCE

Graduate Assistant Dec '22 — Dec '23
Oklahoma City University United States

- Managed and analysed large datasets, ensuring data quality and integrity.
- Created SQL, PL/SQL, and NoSQL programming for data manipulation and extraction.
- Developed and maintained database objects such as stored procedures, materialized views, and functions.
- Implemented impactful analytics using Tableau, enhancing decision-making processes.
- Conducted data process analysis using DFDs and UML to document and improve system-level information flows.
- Collaborated with cross-functional teams to translate business requirements into effective data management solutions.
- Utilized GIT for version control, managing code changes effectively and ensuring project consistency.

Senior Data Engineer Dec '21 — Sep '22
Citi Bank Hyderabad, India

- Developed and maintained advanced data pipelines using PySpark and Spark SQL within a medallion architecture, facilitating efficient data flow from claims data sources including SAP Hana, Salesforce, and Blob storage to Azure Synapse Warehouse.
- Spearheaded the migration from Hive metastore to Unity Catalog's metastore, overseeing testing and validation of pipeline transitions specific to claims data. Managed the transition from Azure Data Factory to Databricks workflows and from interactive to job clusters, improving performance by 30% and achieving an additional 15% cost savings through optimal resource utilization.
- Developed and automated processes to clean, transform, and prepare claims data for analytics, ensuring data accuracy and consistency and reducing data inconsistencies by 40%.
- Developed and implemented machine learning models to detect fraudulent activities within claims data, resulting in a 30% increase in fraud detection accuracy.
- Monitored and ensured the health of claims data pipelines using Grafana, reducing performance bottlenecks by 40%. Implemented CosmosDB solutions and scheduled triggers according to claims processing requirements, enhancing pipeline reliability by 35% and improving data-driven decision-making efficiency.
- Designed and implemented interactive dashboards and reports using Power BI and Tableau, providing actionable insights and improving data visualization for claims processing stakeholders.
- Utilized DAX (Data Analysis Expressions) commands such as CALCULATE, SUMX, and FILTER to create complex measures and calculations specific to claims data, enhancing report capabilities and increasing report generation efficiency by 50%. This resulted in a 25% boost in stakeholder satisfaction and decision-making speed.
- Developed and maintained Terraform configurations to define, deploy, and manage cloud infrastructure as code, ensuring consistent and repeatable deployments for data engineering environments.
- Automated Terraform workflows using CI/CD tools such as Jenkins, GitLab CI, GitHub Actions, and Azure DevOps to streamline data pipeline deployment processes.
- Environment: PySpark, Spark SQL, SAP Hana, Salesforce, Azure Synapse Warehouse, Azure Data Factory, Databricks, Unity Catalog, Grafana, CosmosDB, Terraform.

Data Engineer Jan '20 — Dec '21
Tata Consultancy Services Hyderabad, India

- Worked collaboratively with Application teams to develop and deploy dashboard solutions on the data platforms.
- Produced documentation, metrics, graphs, and tables for past reports, presentations, and publications.

- Assisted in the development of ETL processes using SSIS and SQL for data warehousing and reporting, improving data processing efficiency by 20%.
- Supported data modeling and database design activities to ensure optimal data structure and performance, reducing query times by 25%.
- Utilized JIRA for task management and progress tracking in Agile projects, increasing project completion rate by 15%.
- Contributed to data governance initiatives, focusing on data quality and security, achieving a 98% data quality score.
- Enviroment: SISS, SQL, JIRA

Data Engineer May '18 — Dec '19
Telstra Hyderabad, India

- Developed and tested Store procedures, Functions and packages in PL/SQL for Data ETL.
- Tested Complex ETL Mappings and Sessions based on business user requirements and business rules to load data from source flat files and RDBMS tables to target tables.
- Developed ETL jobs as per the requirements to update the data into the staging database (Postgres) from various data sources
- Documented program development, logic, coding, testing, changes and corrections.
- Created complex Cognos reports using calculated data items, multiple lists in a single report.
- Designed and developed scalable ETL processes using AWS Glue and Redshift, automating data integration from various sources into the data warehouse. Reduced manual intervention by 40%.
- Enhanced data query performance by integrating AWS Athena for direct querying of structured and unstructured datasets in S3, while implementing AWS IAM policies for robust data governance.
- Automated data processing tasks with AWS Lambda, increasing processing speed and reducing manual intervention.
- Environment: ETL, Cognos, PL/SQL, SQL, MS Office, MS Excel,AWS Glue, Redshift, Athena, IAM, Lambda, CloudWatch, S3

EDUCATION

Masters in Computer Science, Oklahoma City University Aug '22 — Dec '23
Oklahoma, United States

- Developed supervised machine learning models using Scikit-learn, TensorFlow for tasks such as classification and regression, employing algorithms such as decision trees, random forests, support vector machines (SVM), and neural networks.
- Built Speed Tracking System. Utilized machine learning algorithms for efficient vehicle detection, tracking, and speed calculation.
- Utilized regression analysis to predict future outcomes and trends based on historical data, providing valuable insights for decision-making.

Bachelors in Information Technology, Kakatiya Institute of Technology and Science Aug '15 — May '19
Warangal, India

AWARDS

On The Spot Nov '21
Tata Consultancy Services

Best Support Team Member Jan '22
Tata Consultancy Services