Ramu Siripurapu **Mobile**: **+91 8500037019**
Email: ram9766@gmail.com

Professional Experience

- **8+ years of experience** in software development, with **5+ years focused on Big Data technologies** including **Hadoop, Hive, Apache Spark, Scala, PySpark, and Databricks**.
- Strong ability to **analyze and understand business applications, data flows, and relationships between datasets**, ensuring alignment with functional requirements.
- Proficient in **ingesting and transforming data from various sources into data lakes** (HDFS/ADLS) across **UAT and production environments**.
- Experienced in **scheduling and orchestrating data workflows** using **Control-M**, including **migrating job schedules from Zena to Control-M**.
- Hands-on experience in **transferring files from edge nodes to Azure Data Lake Storage (ADLS)** and **building Delta tables** on top of ingested data.
- Developed and scheduled **Databricks notebooks and jobs** for end-to-end data processing pipelines.
- Proficient in building **Delta Live Table (DLT) pipelines** on **Databricks Unity Catalog clusters** for structured and scalable data processing.
- Experience in **managing DLTs using Volumes** to streamline configuration and reusable components.
- Worked with **Confluent Kafka** in **Databricks** for batch ingestion and integration with Delta Lake.
- Skilled in **task estimation, documentation**, preparing **status reports**, handling **change requests**, and following **code review protocols**.
- Technically proficient in **Pyspark, Shell Scripting**, and scheduling tools including **Control-M** and **Zena**.
- Strong **team player** with the capability to work independently under minimal supervision when needed.
- Known for being **responsible, adaptable**, and **committed to quality delivery** in dynamic environments.
- Experience in **CI/CD practices** using **Jenkins** and **Urban Code** for multi-environment deployments, and **Git/GitHub** for version control and code collaboration.

Education

- **B.Tech (Electrical and Electronics)** from GITAM Deemed University,2016.

Experience:

- Cognizant, India Senior Associate Nov'21-Till Now
- Tata Consultancy Services, India System Engineer Dec'16-Oct'21

Technical skills

- **Programming Languages:** Scala, PySpark
- **Scripting Languages:** Shell Scripting
- **Databases:** SQL, Hive
- **Big Data & Cloud Technologies:** Apache Spark, Hive, Azure Data Lake Storage (ADLS), Azure Databricks, Unity Catalog, Delta Live Tables (DLT)
- **Scheduling & Workflow Tools:** Control-M, Zena
- **DevOps & Deployment Tools:** Jenkins, UrbanCode Deploy
- **Version Control:** Git (GitHub)
- **Utilities & Other Tools:** WinSCP, PuTTY, JIRA, SQL Server Management Studio (SSMS)
- **Integrated Development Environments (IDEs):** IntelliJ IDEA, Visual Studio Code

Certifications

- Microsoft Azure Data Engineering Associate **(DP-203)**
- Microsoft Azure Fundamentals Certification **(AZ-900)**
- Databricks certified Data Engineer Associate

Professional Experience

Cognizant Technology Solutions – Senior Associate
Nov 2021 – Present

Title:	**DAS Membership Data-GPD Ingestion**
Client	**Health Care Service Corporation(HCSC).**
Technologies used	**Spark,Scala,Hive, Shell scripts,Databricks,ADLS, DLT,Volumes**

Description:

- The project involves ingesting over 1,000 types of files from various lines of business (e.g., Medicare, Medicaid, MedSupp) into the enterprise data lake via an edge server and eGateway. Files are audited, validated, and then loaded into raw and curated layers in HDFS/ADLS to support downstream analytics and business processes.

Roles & Responsibilities:

- Developed scalable ingestion frameworks to move files from edge nodes to **HDFS/ADLS**, including **audit validation** based on file patterns and metadata.
- Built transformation pipelines using **Scala** to load structured data into **Hive raw and curated tables**.
- Implemented **incremental data loads** by extracting selective columns from source systems and exporting to CSV.
- Developed **PySpark scripts** to parse JSON configurations and dynamically generate **Control-M variables**.

- Automated ingestion pipelines with **dynamic HDFS path creation** in PySpark.
- Migrated multiple batch jobs from **Zena to Control-M**, ensuring validation and scheduling consistency.
- Designed and implemented logic to **flatten complex JSON structures** and load data into Hive layers based on mapping documents.
- Created and managed **Delta tables** on ADLS to enable efficient querying and historical data tracking.
- Built **DLT pipelines** using Unity Catalog to ingest and transform data into **bronze and silver layers**.
- Developed and scheduled **Databricks notebooks and jobs** to automate batch processing and validations.
- Ingested **Kafka messages in batch mode** into Delta Live Tables for near-real-time data capture.
- Utilized **Volumes** in Databricks to manage configuration files and enable modular pipeline architecture.
- Authored **Shell scripts** for ingestion validations and pre-processing across various file types.
- Contributed to full-cycle development including **HQL, Shell, and Spark-Scala coding**.
- Managed deployment using **Jenkins and Urban Code** across multiple environments.
- Used **Git/GitHub** for version control, code review, and release branch management.

Tata Consultancy Services – System Engineer
Dec 2016 – Oct 2021

- Contributed to multiple big data projects involving data lake ingestion, data transformation, and job orchestration.
- Developed Spark jobs and shell scripts for business rule application, reconciliation, and data delivery.
- Applied JSON-based transformation and validation rules dynamically using Spark-Scala.
- Scheduled and monitored workflows using Control-M and managed deployment across environments.

Declaration:

I hereby declare that all the information mentioned above is true to the best of my knowledge.

Date :

Place :Hyderabad Ramu Siripurapu.