Phanindra Kumar Mulamreddy

mulamreddyp7@gmail.com | (667) 464-0071 | LinkedIn

Objective

Experienced **AWS Data Engineer (3+ Years)** with a strong proficiency in **Python**, **ETL** processes, data modeling, Big Data Analytics and Data Transformation. Skilled in managing real-time data pipelines, ensuring data integrity and reliability. Proven track record of implementing efficient data distribution and storage solutions in cloud platforms, such as **Google Cloud** and **AWS**.

- Extensive experience in Amazon Web Services **(AWS)** Cloud services such as **EC2**, **VPC**, **S3, IAM**, **EBS**, **RDS**, **ELB, VPC, Route53**, Ops Works, **DynamoDB,** Autoscaling, **CloudFront**, **CloudTrail**, **CloudWatch**, **CloudFormation, Elastic Beanstalk, AWS SNS, AWS SQS, AWS SES, AWS SWF** & **AWS** Direct Connect, Trend Analysis.
- Experienced in Data Analysis, Design, Development, Implementation and Testing using Data Conversions, Extraction, Transformation and Loading **(ETL)** and **SQL** Server, **ORACLE**, and other relational and non-relational databases.
- Hands-on experience with **GCP** services like Cloud Function, Data Proc, **Kubernetes**, Cloud Storage, **Big Query**, Cloud Run, Cloud Registry, Cloud Composer, Cloud Monitor, etc.
- Strong proficiency in **SQL** concepts, Presto **SQL**, **Hive SQL**, **Python (Pandas, NumPy, SciPy, Matplotlib)**, **Scala**, **Java**, and **PySpark** to cope with the increasing volumes of data.
- Experience in testing **Big Data Hadoop (HDFS, Hive, Sqoop and Flume), Master Data Management (MDM) and Tableau Reports.**
- Hands On experience on **Spark Core, Spark SQL, Spark Streaming** and creating the **Data Frames** handle in **SPARK** with **snowflake**.
- Strong experience in writing scripts using **Python API, PySpark** and **Spark API** for analyzing data.
- Experience in using **Docker** and **Ansible** to fully automate the deployment and execution of the benchmark suite on a cluster of machines.

SKILLS

- **Programming:** Python, SQL, PySpark, Shell Scripting, BASH Scripting, Rust
- **AWS:** EMR, EC2, EBS, RDS, S3, Athena, Glue, Elasticsearch, Lambda, SQS, DynamoDB, Redshift, PostgreSQL
- **Cloud Platforms:** Google Cloud Platform (GCP), Amazon Web Services (AWS)
- **Databases:** MySQL, Microsoft SQL Server (MSSQL), PostgreSQL, MongoDB (NoSQL), Snowflake
- **Libraries:** Pandas, NumPy, Scikit-learn, PyTorch, SciPy, Matplotlib, TensorFlow, Network
- **Reporting & Visualization:** Tableau, Looker Studio (Data Studio), Quicksight, BI
- **Big Data Technologies:** Hadoop, Spark (pyspark), dbt, Apache Kafka, Apache Spark, Pig, Oozie, YARN
- **ETL Tools:** Apache Nifi, Apache Airflow, Protegrity, Informatica.
- **Methodologies:** Agile, Waterfall, Scrum, Machine learning
- **Operating Systems:** UNIX, LINUX, Ubuntu, windows
- **Other Skills:** Git, Data structures, Data management, Real-time data processing, Debugging, Docker, CI/CD, Kubernetes, Security fundamentals

EDUCATIONS

- **Master's in data science** from University of Maryland Baltimore County, Baltimore, MD**.**
- **B. tech in Mechanical Engineering** from VR Siddhartha Engineering College, INDIA

CERTIFICATION

- AWS Data Engineer Associate
- Runner Up in **National Go-Kart Competition 2019 (SKDC, KKC)**

WORK EXPERIENCE

AWS Data Engineer | Broadcom - Plano TX | March 2023 – Present

Responsibilities

- **Developed** and **deployed 20+ AWS** Lambda functions using **Python** to handle data transformations on **large datasets** in **EMR clusters**, improving real-time data processing by **30%**.
- **Created data marts** to support scalable data storage and facilitate advanced reporting, improving access to critical datasets.
- **Designed and implemented a Security Framework** for fine-grained access control to objects in **AWS S3**, utilizing **AWS Lambda** and **DynamoDB**.

- **Built and managed ETL pipelines** in **AWS Glue**, successfully migrating **50+ terabytes** of data from SQL Server to **AWS Redshift**, reducing data loading time.
- **Implemented CI/CD pipelines** using **Docker**, and **Apache Airflow**, automating deployments and reducing manual data processing efforts.
- Monitored **real-time streaming applications** using **Apache Spark**, **Kafka**, **Scala**, and **Hive** to perform streaming ETL, enabling machine learning model integration.
- **Leveraged Python libraries** like **Pandas, NumPy**, and **Matplotlib** for large-scale data analysis and manipulation, increasing data processing efficiency.
- **Automated ETL workflows** using **AWS Glue**, reducing manual intervention by **20%**, and transforming datasets across various AWS environments.
- Deployed and orchestrated **machine learning pipelines** using **Apache Spark** on **AWS EMR**, enabling parallel processing of large datasets for model training and predictions.
- Leveraged **AWS SageMaker** to automate hyperparameter tuning, model training, and deployment, streamlining the **end-to-end machine learning lifecycle**.
- **Built cloud-based Data Lakes** using Data Warehousing tools i.e., **AWS Redshift** and **Snowflake**, enabling scalable storage for data pipelines and reducing data retrieval latency by **30%**.
- **Implemented rigorous data integrity checks and quality control measures**, ensuring reliability and accuracy across large data pipelines..
- **Collaborated on defining data requirements and optimizing internal processes**, driving process improvements and operational efficiencies in the data management strategy.

Data Engineer | Cognizant Technology Solutions - India | February 2021 – July 2022

Responsibilities

- Collaborated with **Database Administrators** and **Cloud teams** (AWS, SQL Server, Oracle) to ensure seamless integration of database tables, columns, and metadata across **DEV, QUAL, and PROD environments** in **AWS Cloud** and **Snowflake** reducing data inconsistencies.
- **Built and optimized ETL pipelines** using **AWS Glue**, transferring **millions of records daily** from multiple sources into **AWS S3**, ensuring **99.9% data accuracy** and **100% reliability** across production environments.
- Developed **Apache Spark** scripts, including **UDFs**, for **data aggregation**, complex transformations, and querying large datasets, enhancing data processing capabilities.
- Utilized **SQL** and **Python** to load data from **APIs** into containers, transferring it to **Snowflake**, optimizing data storage costs and retrieval times.
- Migrated and optimized **Hadoop datasets** into **AWS S3**, improving storage performance by **50%** and cutting infrastructure costs by **20%**.
- Implemented **Protegrity** for **tokenization of PII data elements**, ensuring secure migration from on-premise Hadoop to **AWS** environments.
- Performed **data transformations** and **data cleansing** using **Apache Hive**, improving overall data quality, leading to better data readiness for advanced analytics.
- Collaborated with **BI teams**, utilizing **Tableau** for creating dynamic reports and dashboards, providing actionable insights through **Exploratory Data Analysis (EDA)**.
- Integrated real-time data processing and enrichment using **Apache Spark**, enabling advanced analytics and timely decision-making.

Operations Research Analyst | TATA Capital - India | APR 2020 – JUN 2021

Responsibilities

- Conducted in-depth **data analysis** on **large datasets (100K+ rows)** using **R, Python,** and **MATLAB**, providing actionable insights that improved business strategy and process optimization, resulting in significant **increase in operational efficiency**.
- Applied **mathematical modeling**, **linear programming**, and **quantitative analysis** to solve complex business problems, improving decision-making speed by **30%** and reducing operational costs.
- Designed and maintained **10+ dashboards** and **operational reports** to monitor **key performance indicators (KPIs)**, tracking efficiency, profitability, and resource utilization, improving real-time reporting accuracy.
- Utilized **SQL** to gather, clean, and analyze large datasets, enabling detailed operational analysis and improving decision-making.
- Conducted sensitivity analysis and risk assessment to evaluate how changes in input variables affect business outcomes, improving **decision-making** under uncertainty.
- Applied **optimization algorithms** to develop models for **logistics**, **scheduling**, and **resource planning**, increasing **operational efficiency** and reducing project delivery times by **15%**.
- Generated statistical reports and visualizations using tools like **BI** and **Tableau** to communicate findings and recommendations effectively to stakeholders.