Chakradhar Reddy Nallu

nchakradhar44@gmail.com | 656-204-2489 |GitHub

Summary

Data Engineer with around 3 years of extensive experience in designing and optimizing **data solutions**, scalable **data models**, and ETL workflows using AWS and Azure platforms. Proficient in Python, SQL, and Shell Scripting, with a strong track record of integrating diverse data sources to deliver actionable insights. Successfully developed interactive dashboards for real-time business intelligence and implemented secure, robust solutions aligned with **data governance** policies and best practices. Adept at managing cloud infrastructure and big data technologies to drive data-driven decision-making in dynamic environments.

Professional Experience

Maslow AI | Data Engineer | New Jersey, US **Sep 2024 – Dec 2024**

- Developed and implemented data ingestion processes from **structured** (SQL, JSON) and **unstructured** (logs, streaming data) sources using **Apache Spark** and **Scala**, ensuring high efficiency and reliability in processing large-scale data sets.
- Designed, built, and maintained big data infrastructure using Hadoop and Spark, optimizing performance and scalability.
- Ensured data quality and integrity by performing data validation, automated testing, and implementing error-handling mechanisms.
- Developed **scalable ETL workflows** using **AWS Glue** and **EC2**, automating data transformation processes for diverse datasets.
- Utilized **AWS services** (**S3**, **DynamoDB**, **CloudWatch**, **Lambda**, **SQS**) to build scalable and cost-effective data engineering workflows.
- Leveraged **AWS IAM** for secure **role-based access control**, ensuring compliance with **data governance policies**.
- Engineered solutions for **real-time data processing**, leveraging **cloud computing frameworks** to meet dynamic application demands.
- Collaborated with cross-functional teams, including data analysts and data scientists, to develop scalable data workflows, improving pipeline consistency and optimizing performance.

Environment: AWS (S3, Glue, Lambda, DynamoDB, CloudWatch, EC2, SQS, IAM), Apache Spark, ETL, Real-Time Data Processing.

Aarmec technology | Data Engineer | India **Jan 2021- Agu 2023**

- Optimized scalable data pipelines with Apache Spark and AWS, reducing processing time by 40% and infrastructure costs by 25%
- Automated workflows with Shell Scripting and Python, decreasing manual effort by **50%** and improving operational efficiency.
- Designed a data pipeline that supported real-time machine learning predictions, enhancing business decision-making speed by 20%.ata
- Improved query performance by **30%** through database optimization (Teradata, Oracle) and implemented Redis for **45%** faster data retrieval.
- Documented and maintained technical specifications, architecture diagrams, and data flow processes to enhance team collaboration and knowledge sharing.
- Developed and maintained **CI/CD workflows** with **Jenkins** and **Docker**, enabling continuous integration and deployment of big data.
- Utilized **Python** (**Pandas, NumPy**) for automation, data manipulation, and streamlining complex data processing tasks.
- Collaborated with **cross-functional** teams to gather business requirements and translate them into scalable data solutions, effectively communicating technical concepts to non-technical stakeholders to ensure alignment and project success.

Environment: Apache Spark, Shell Scripting (Linux), Oracle, AWS (S3, Lambda, RDS, EMR), SQL, Apache Hive, Data Lakes, Pandas, NumPy.

Technical Skills

Programming and Scripting Languages: Python, Java, Scala, SQL, Shell Scripting, HTML5, CSS, Bash, PowerShell.
Database Technologies: MySQL, PostgreSQL, PL/SQL, NoSQL, SQL Server, Oracle, MongoDB, Amazon DynamoDB.
Cloud Technologies: AWS, Microsoft Azure (Databricks, Data Factory, Synapse, Data Lake), Google Cloud Platform(GCP).
Big Data Tools: Apache Spark, Hadoop (HDFS, MapReduce), Kafka, Hive, OLAP(Snowflake), Databricks, Delta Lake.
Data Modeling and Data Warehouse: Star and Snowflake Schema, Dimensional Modeling, Amazon Redshift, Azure Synapse, BigQuery.
Development Tools and Workflow Orchestration: Git, Docker, Kubernetes, Visual Studio Code, Apache Airflow.
Data Visualization Tools: Tableau, Power BI, QuickSight. **Operating Systems:** Linux (Ubuntu), Windows Server.
Data Workflow and Real-Time Processing: Apache Kafka, AWS Kinesis, Spark Streaming, Flink, PySpark, Automation with Python.
Methodologies and Soft Skills: Agile, Scrum, Waterfall, Software Development Life Cycle (SDLC), Problem-solving, Collaboration, Communication, Critical Thinking, Project Management.

Academic Projects

Reddit Data Pipeline Engineering **Sep 2024 – Nov 2024**

- Built a scalable ETL pipeline with Airflow, AWS, and Docker, improving query performance by 35% and reducing manual effort by 85%.
- Optimized data storage with Amazon S3 and Redshift, enabling seamless processing of large-scale Reddit datasets.

Technologies Used: Apache Airflow, Celery, PostgreSQL, Amazon S3, AWS Glue, Amazon Athena, Amazon Redshift, Docker, Python

YouTube Data Analysis Project **Mar 2024 – May 2023**

- Designed an ETL pipeline with AWS and PySpark, cutting query time by 40%. Automated data workflows improved efficiency by 90%.
- Developed interactive dashboards in QuickSight to visualize engagement trends across 20+ key performance indicators.

Technologies Used: AWS Lambda, AWS Glue, Amazon S3, AWS Athena, Amazon QuickSight, PySpark, Python (NumPy, Pandas), Parquet.

Education

MS in computer science- University of South Florida, FL (GPA: 3.88) **Aug 2023 - Dec 2024**

Achievements

- AWS Solution Architect, Developer Associate, and Cloud Practitioner certification.
- Awarded 5 star in HackerRank Problem Solving for Python and SQL.