

Raju Bachu

(901)-546-3848 | bnraju834@gmail.com | [LinkedIn](#)

Professional Summary

Data Engineer with 4+ years of experience in big data technologies and machine learning to extract insights from large, complex datasets. Expertise lies in building scalable data pipelines, optimizing workflows, and communicating technical findings clearly to stakeholders. Passionate about driving data-based solutions to inform business strategies and process optimizations. Thrives in fast-paced environments.

Education

Master of Science, Computer Science,
Southern University A&M college| GPA: 3.9/4.0

Aug 2021 -Dec 2022

Work Experience

Data Engineer, Best Buy

Jan 2023 – present

- Extracted diverse datasets from various sources including relational databases, APIs, JSON, and CSV files, ensuring comprehensive data collection for business analysis.
- Engineered robust ETL (Extract, Transform, Load) pipelines to efficiently ingest raw data into Azure Data Lake Storage (ADLS), employing Azure services like Azure Data Factory for seamless integration.
- Applied complex transformations using SQL and Python within Azure Databricks, aggregating and cleansing data to meet specific business requirements and enhance data quality.
- Created Pipelines in ADF using Linked Services/Datasets/Pipeline/ to Extract, Transform and load data from different sources like Azure SQL, Blob storage, Azure SQL Data warehouse, write-back tool, and backward.
- Developed Kafka producers and consumers to ingest and integrate data from various sources and systems, including real-time event streams and log data.
- Implemented delta loading strategies within Azure Delta Lake architecture, ensuring efficient data storage, versioning, and reliable updates for real-time analytics.

Environment: python, pyspark, SQL, spark SQL, Adobe Marketing Cloud, AWS Kinesis, Azure Databricks, Azure Azure synapse Analytics, Data Factory, Azure Blob storage, snowflake, Amazon CloudWatch, Hive, AWS Glue, Git, Apache Kafka, Amazon Redshift, Rally.

Data Engineer, JB Hunt

Sep 2021 - Dec 2022

- Managed and Upheld SQL databases in the trucking and logistics industry, ensuring safe data storage and availability for a large customer base. Employed Python data tools like Pandas, Scikit-Learn, Matplotlib, and TensorFlow for efficient transportation data management and analytics.
- Worked on data archival by transferring data across platforms, validating data while transferring, and archiving data files for various DBMS by creating dynamic SSIS packages.
- Extensive experience in using SQL Management Studio, SQL Server Business Intelligence Solutions like SSRS and SSIS.
- Used SSMS to manage SQL Server databases, including creating, modifying, and deleting database objects such as tables, views, stored procedures, and indexes.

Environment: python, SQL, Containers, Spark, Mongo DB, SSIS, SSRS, Hadoop, Pyspark, snowflake, Teradata, Hive, HDFS, Python, Kafka, RESTFUL APIs.

Data Engineer, Tech Mahindra

Nov 2019-May2021

- Designed and implemented ETL processes to extract data from NoSQL databases such as DynamoDB. Transformed the data and then loaded it into the Redshift data warehouse in order to consolidate large volumes of data for more efficient analysis.
- Performed A/B testing using Programming languages like Python to evaluate new app features, increasing usage by 20%.
- Exhibited strong proficiency in DevOps practices, including coding, scripting, database management, and infrastructure-as-code (IaC) with tools like Terraform. Skilled in building and optimizing CI/CD pipelines to ensure efficient and reliable software delivery.

Environment: python, spark, SQL, Amazon S3, Amazon Elastic MapReduce (EMR) , Amazon VPC ,Microsoft Excel, Pandas, VPC, AWS IAM, Amazon EC2,Amazon CloudWatch, AWS Glue, Git, Apache Spark, Agile.

Technical Skills and Software proficiencies

C, Python, SQL, MYSQL, Oracle Database, anaconda, GIT, Windows, iOS, UNIX, S3, Azure services (Data Factory, Datalake, Blobstorage), AWS(Redshift ,EMR)Apache spark, Hadoop, SSMS, SSIS, Data Visualization: Power BI, Matplotlib, Seaborn VPC, Agile, CloudWatch, EC2, IAM.