Bhavana Rajashekar
Data Engineer

Boston, MA |(667) 345-4217 | Email: bhavanarajashekar4@gmail.com

SUMMARY

● Data engineer with 3+ years of experience designing, developing, and optimizing data solutions to support data-driven decision-making and business objectives.

● Expertise in data integration, ETL processes, and database management, ensuring efficient and reliable data pipelines for data analysis and reporting.

● Proficient in big data technologies such as Hadoop, Spark, and Kafka, handling large volumes of data and implementing scalable data processing solutions.

● Skilled in working with various database systems like SQL Server, Oracle, and MySQL, ensuring data integrity and efficient data retrieval for analysis.

● Experienced in cloud platforms such as AWS, Azure, and Google Cloud, leveraging cloud services to build scalable and cost-effective data infrastructures.

● Strong knowledge of data modeling and data warehousing concepts, designing and optimizing data schemas and structures for efficient data storage and retrieval.

● Proficient in data visualization tools like Tableau and Power BI, transforming complex data into insightful visualizations and reports for business stakeholders.

● Experienced in implementing data governance frameworks, ensuring compliance with regulatory standards and maintaining data privacy and security.

● Skilled in performance tuning and optimization of ETL processes and SQL queries, improving data processing speed and overall system efficiency.

● Collaborative team player with excellent communication and problem-solving skills, working closely with cross-functional teams to understand data requirements and deliver effective data solutions.

SKILLS

Methodologies:	SDLC, Agile, Waterfall
Programming Language:	Python, R, SQL, SAS
Packages:	NumPy, Pandas, Matplotlib, SciPy, Scikit-learn, TensorFlow
Visualization Tools:	Tableau, Power BI, Excel
IDEs:	Visual Studio Code, PyCharm, Eclipse
Database:	MySQL, Microsoft SQL Server, MongoDB, Oracle Database, Postgres, Elastic Search
Cloud Technologies:	Amazon Web Services (AWS), Azure, Google Cloud
ML Algorithms:	Linear Regression, Logistic Regression, Decision Trees, Supervised Learning, Unsupervised Learning, Classification, SVM, Random Forests, Naive Bayes, KNN, K Means
Other Technical Skills:	Google Analytics, DAX, Power Query, Alteryx, SAS, JIRA, SAP, SSIS, SSRS, Machine Learning Algorithms, Probability distributions, Confidence Intervals, Hadoop, Spark, Kafka, Snowflake, Redshift, Big Query, Airflow, DBT, Informatica, Talend, Git , Advance Analytics, Data Mining, Data Visualization, Data warehousing, Data transformation, Data, Storytelling, Association rules, Clustering, Classification, Regression, A/B Testing, Forecasting & Modelling, Data Cleaning, Data Wrangling.
Operating Systems:	Windows, Linux, MacOS

EDUCATION

Master of Information Science - University of Maryland, Baltimore County
Bachelor of Engineering in IT – RNS Institute of Technology, India

CERTIFICATION

Cloud Certification in AWS

EXPERIENCE

Data Engineer | Commonwealth Corporation, Boston | June 2023 – Present

- Identified and catalog various data sources within the healthcare organization, including **Electronic Health Records (EHR)**, laboratory results, billing data, and external data sources.
- Developed **ETL processes** to extract data from disparate sources, ensuring data integrity and consistency and implemented incremental data extraction mechanisms to capture updates in real-time.
- Standardized data formats, resolve inconsistencies, and handle missing or erroneous data during the transformation phase and applied data cleansing and validation procedures to ensure high data quality.
- Used tools like **Apache NiFi or Apache Camel** to facilitate smooth data flow between systems.
- Deployed machine learning models using **Amazon SageMaker** to analyze and predict client's fiscal data to forecast future spending, profits, and generated reports of recommendations to decrease hospital drug expenditure
- Monitored and managed cloud resources using **AWS CloudWatch**, setting up alerts and dashboards for real-time visibility into system performance.
- Designed and implemented a scalable and optimized data warehouse schema to store integrated healthcare data on **AWS Redshift.**
- Implemented robust security measures to ensure compliance with healthcare data privacy regulations (e.g., HIPAA) also applied encryption, access controls, and audit trails to protect sensitive patient information.
- Used **AWS Glue** for data ingestion, enabling automatic discovery, cataloging, and transformation of healthcare data and leveraged **AWS DataSync** for securely transferring healthcare data to and from AWS.
- Extracted large patient's files from **NoSQL database (MongoDB)** and processed them with Spark using mongo Spark connector.
- Designed visually appealing and interactive dashboards in **Tableau** that present key healthcare metrics and utilized color coding, charts, and graphs to enhance data visualization.
- Designed high-availability and disaster recovery solutions using **AWS Backup** and **AWS Elastic Disaster Recovery**, ensuring data resilience.
- Worked very closely with healthcare interoperability and messaging standards, like **HL7 2.x, FHIR, HIPAA, Radiology, Patient Access**, etc.

Data Engineer | AIESEC, India | July 2020 – Jan 2022

- Involved in designing and implementation of a distributed data processing framework using **Apache Spark**, achieving a **50% reduction in data processing time** for real-time analytics.
- Support data governance policies and security measures with **Azure Security Center**, implementing best practices for data protection and risk management to protect sensitive data.
- Conducted comprehensive data profiling and analysis, identifying, and resolving data inconsistencies to **improve data reliability by 25%.**
- Developed **Spark/Scala**, Python for **regular expression (regex)** project in the **Hadoop/Hive** environment with Linux/Windows for big data resources.
- Managed security groups on Azure, focusing on high-availability, fault-tolerance, and auto scaling using **Terraform** templates. Along with Continuous Integration and Continuous Deployment with **Azure Functions** and **Azure DevOps**.
- Developed data monitoring and alerting mechanisms using ELK stack, ensuring real-time data accuracy and anomaly detection.
- Worked with Data NoSQL Warehouse in the development and execution of data conversion, data cleaning and standardization strategies and plans as several small tables are combined into one single data repository system **MDM (Master Data Management).**
- Used **Azure Data Catalog** with crawler to get the data from **Azure Blob Storage** and perform SQL query operations using **Azure Synapse Analytics**.
- Implemented data integration and analytics solutions using **Snowflake**, enhancing data processing capabilities and performance.