BHARGAV REDDY MUNNANGI

945-274-1994| Dallas, Texas, 75201 | bhargavreddy.m18@gmail.com
https://www.linkedin.com/in/mbhargavreddy | mygreatlearning.com/eportfolio/bhargav-reddy

SUMMARY

Experienced Data Engineer with 6+ years of expertise in building scalable data pipelines, automating ETL workflows, and delivering analytics solutions for large-scale industrial and infrastructure projects. Proficient in PySpark, SQL, and Power BI, with hands-on experience across AWS, Azure, and on-premises systems. Skilled at translating complex operational data into actionable insights to drive efficiency, cost savings, and informed decision-making.

PROFESSIONAL EXPERIENCE

Larsen & Toubro, Chennai, India Jul 2018 – Aug 2022

Data Engineer

- Designed and implemented scalable ETL pipelines across AWS and Azure ecosystems using tools like AWS Lambda, AWS Batch, PySpark, Azure Databricks, and Azure Data Factory (ADF), enabling efficient ingestion and transformation of multi-terabyte datasets from diverse sources.
- Developed a fault-tolerant AWS data pipeline to process Machine-Readable Files (MRFs) from insurers, optimizing data ingestion, and reducing storage costs by 60% using partitioned Parquet formats on S3.
- Built an AI agent using RAG pipelines with FAISS to generate SQL queries from natural language inputs, integrated with D3.js dashboards for real-time visual insights.
- Automated data orchestration using Apache Airflow, streamlining scheduling, monitoring, and execution of complex ETL workflows across cloud environments.
- Engineered serverless ingestion pipelines with Azure Logic Apps and Azure Functions, enabling cost-efficient integration with external data providers.
- Enhanced data quality by implementing PySpark-based transformation pipelines to clean, enrich, and aggregate large datasets before loading into AWS Redshift and Azure Synapse, improving analytics performance by 30%.
- Designed a star schema and implemented SCDs to support historical tracking and improve performance in sales and inventory analytics use cases.
- Conducted in-depth consumer analytics and segmentation using Python, SQL, and ML models (e.g., Decision Trees, Classification), contributing to a 20% increase in customer retention and optimized marketing spend.
- Led version control and collaborative development via GitHub and GitLab, introducing peer reviews and CI/CD automation to maintain code stability.
- Partnered with cross-functional stakeholders to translate business requirements into robust data solutions, improving the accuracy and impact of data-driven decisions.

Larsen & Toubro, Delhi, India Jul 2016 – Jun 2018

Data Analyst

- Analyzed large-scale operational and project performance data using Python and SQL to uncover trends, inefficiencies, and improvement opportunities across infrastructure and engineering domains.
- Conducted in-depth time series and statistical analysis on equipment usage, project timelines, and cost fluctuations, enabling proactive decision-making and enhancing operational efficiency by 20%.
- Automated data pipelines to collect, clean, and transform real-time data from multiple sources, significantly reducing manual reporting effort and ensuring data consistency across teams.
- Designed and maintained interactive Power BI dashboards to visualize key metrics such as project progress, resource utilization, and contractor performance—improving executive visibility and reporting accuracy.
- Built predictive models using Scikit-learn and Python to forecast project delays, resource bottlenecks, and budget overruns, supporting data-driven planning and risk mitigation.
- Collaborated with cross-functional teams, including engineering, procurement, and finance, to gather requirements and translate them into analytical solutions that supported strategic initiatives.
- Implemented anomaly detection frameworks to flag irregularities in inventory usage, machinery performance, and site productivity, contributing to preventive maintenance and cost savings.
- Led user acceptance testing (UAT) and validation of analytics solutions to ensure alignment with on-ground data and business objectives across infrastructure projects.
- Managed project timelines and stakeholder communications using Agile practices and tools such as Jira, ensuring timely delivery of analytics insights in alignment with business goals.

PROJECT EXPERIENCE

Smart Construction Analytics Platform

- Designed a real-time IoT analytics system with Spark, Kafka, and AWS to monitor construction equipment efficiency, reducing equipment failures by 25%.
- Developed interactive Power BI dashboards to track fuel consumption, machine usage, and maintenance schedules, improving project timelines.

Automated ETL Pipeline with Apache Airflow and Google BigQuery

- Developed and deployed an ETL pipeline using Apache Airflow to orchestrate data extraction from cloud sources and transformation in Google BigQuery.
- Implemented dynamic scheduling to handle time-sensitive data flows and optimized data loading performance, resulting in a 25% reduction in processing time.
- Ensured data quality by incorporating monitoring and alerting mechanisms, which proactively notified stakeholders of pipeline failures or data anomalies.

Serverless Data Processing and Analytics on AWS

- Designed and implemented a serverless ETL pipeline using AWS Lambda, AWS Glue, and AWS Redshift to ingest, transform, and load large datasets from multiple sources into a centralized data warehouse.
- Applied schema-on-read techniques to process semi-structured data (JSON, Parquet), reducing storage costs by 40% and enhancing query performance by leveraging partitioned tables.
- Integrated AWS Athena for ad-hoc querying and AWS QuickSight for real-time business intelligence, delivering actionable insights within minutes.

Automated Invoice Processing using Optical Character Recognition (OCR)

- Developed an OCR-based invoice processing system using Python, Tesseract, and AWS Textract to extract key data points (invoice number, date, amount) from scanned documents.
- Implemented an automated ETL pipeline to clean, validate, and load extracted data into AWS RDS for further processing, reducing manual data entry time by 70%.
- Integrated the solution with a Power BI dashboard for financial reporting and audit trail monitoring, ensuring improved data accuracy and compliance.

AI-Based Sentiment Analysis for Customer Feedback

- Built a sentiment analysis model using NLP techniques with the Hugging Face library and Python to analyze customer feedback from surveys and social media.
- Employed pre-trained transformer models (BERT) to perform text classification, achieving an accuracy of 92% in categorizing sentiments as positive, negative, or neutral.
- Integrated the solution into a real-time reporting dashboard in Power BI, providing executives with instant insights into customer satisfaction trends.

TECHNICAL SKILLS

Languages: Python (NumPy, Pandas, Matplotlib, Seaborn, Scikit-learn), MySQL, NoSQL
Tools: Apache Spark, Kafka, AWS (S3, Glue, Redshift, EMR), Google BigQuery, Azure Data-Lake, Azure Data Factory, Azure Synapse, Snowflake, MySQL, MongoDB, HDFS, Airflow, DBT, Apache Beam, Impala, Hadoop, Hive, Flume
Visualization & Tools: Power BI, Tableau, Looker, KNIME, Appian, UI Path, SAS, Alteryx, Excel, Solver
Core Competencies: Data Warehousing, ETL Pipelines, Predictive Analytics, NLP, Data Governance

EDUCATION

The University of Texas at Dallas Aug 2023 – May 2025
Master of Science, Business Analytics & Artificial Intelligence
The University of Texas at Austin Jan 2022 – Jan 2023
Post Graduate Program, Data Science & Business Analytics

AWARDS

Engineering Innovation Award, *Larsen & Toubro* Dec 2021
Outstanding Performer, *Larsen & Toubro* Oct 2021