SHIVAKUMAR PASEM

pasemshivakumar706@gmail.com (774) 320-8269 LinkedIn

PROFESSIONAL SUMMARY

Data Analyst & Engineer with 3 years of experience building scalable ETL pipelines, automating data integration workflows, and developing interactive dashboards across platforms like Azure, Databricks, and Power BI. Skilled in SQL-based reporting, Python scripting, and cloud data tools to deliver actionable insights and real-time analytics. Proven track record in collaborating with cross-functional teams, maintaining data quality, and streamlining deployment using CI/CD pipelines. Focused on transforming raw datasets into high-impact business intelligence solutions

KEY SKILLS

Programming and Scripting
Python (Pandas, NumPy, Scikit-learn, NLTK, PySpark), SQL, PL/SQL, JSON

Data Engineering and Integration
Oracle Integration Cloud, Jenkins, Apache Airflow, Azure Data Factory

Cloud and Big Data Tools
Microsoft Azure, Databricks, PySpark, Snowflake

Data Analysis and Visualization
Tableau, Power BI, SQL-based Reporting, Exploratory Data Analysis (EDA)

Machine Learning and NLP
Scikit-learn, DistilBERT, Hugging Face Transformers, Text Preprocessing

Databases and Warehousing
Oracle, MySQL, Azure SQL Database

Version Control and Automation
Git, Jenkins (CI/CD Pipelines)

CERTIFICATIONS

- Databricks Certified Data Engineer Associate (Issued by Databricks)
- Microsoft Azure Data Fundamentals (DP-900)

PROFESSIONAL EXPERIENCE

Data Engineer, Cardinal Health 05/2024 – Present

- Engineered high-performance ETL pipelines using PySpark, Azure Data Factory, and SQL, accelerating data availability across clinical analytics systems.
- Optimized Databricks workflows to improve batch processing efficiency by 35%, reducing delays in daily data refresh jobs.
- Delivered real-time KPI dashboards in Power BI for clinical and executive teams, automating executive-level reporting workflows by 40%.
- Implemented robust data quality checks and pipeline monitoring with Apache Airflow DAGs, ensuring consistent batch job execution and reliable data refreshes.
- Orchestrated Jenkins-based CI/CD workflows, improving deployment reliability and enabling efficient rollback mechanisms.

Junior Data Engineer, Infosys 05/2021 – 11/2022

- Designed and implemented 10+ Oracle ERP integrations, streamlining procurement and finance workflows and reducing manual reconciliation efforts across departments.
- Automated data workflows using Python and PL/SQL, saving over 15 hours/week and improving processing consistency.
- Built data-rich reporting interfaces in Power BI and Tableau from Oracle back-end systems, cutting reporting cycle time by 30% and improving executive visibility into operations.
- Led cross-functional delivery across QA, analysts, and client teams for 3 enterprise rollouts, ensuring on-time go-lives with zero postdeployment defects.
- Constructed CI/CD pipelines in Jenkins with integrated JSON schema validation, reducing integration errors by 15% and speeding up release cycles.

Data Analyst, Wipro 04/2020 – 04/2021

- Created interactive executive dashboards in Power BI and Tableau to monitor KPIs for logistics and telecom clients, supporting strategic planning and accelerating time-to-insight.
- Developed Python-based data transformation scripts for cleaning and aggregation tasks, boosting ETL pipeline efficiency by 25% and reducing processing anomalies.
- Conducted exploratory data analysis (EDA) on high-volume transactional and usage datasets, uncovering behavioral patterns and pricing insights that informed marketing and retention strategies.
- Collaborated with cross-functional teams and product owners to translate evolving business needs into scalable analytics solutions using Azure SQL Database and Python.

EDUCATION

Master of Science in Data Science, 01/2023 – 12/2024
University of Massachusetts Dartmouth, USA

Bachelor of Technology in Information Technology, 08/2017 – 09/2021
Jawaharlal Nehru Technological University, India

PROJECTS

Cournot Oligopoly Market Simulation Technologies, Python, Pandas, NetworkX, Excel

- Simulated competitive market behavior using Cournot Oligopoly theory, modeling firm interactions through a Barabási–Albert network in NetworkX.
- Conducted exploratory analysis and visualized strategic outcomes to study pricing decisions, market equilibrium, and dynamic competition.
- Combined data wrangling (Pandas) with Excel-based visualization to bridge economic theory with business-ready insights.

Data Pipeline for Healthcare Operations Analytics Technologies, PySpark, Databricks, Azure Data Factory, Power BI

- Created an end-to-end pipeline to process patient admission, discharge, and treatment data for analytics use.
- Crafted operational dashboards to support hospital staffing and resource allocation strategies, enabling data-informed scheduling.
- Automated task dependencies and data SLA monitoring using Airflow, improving pipeline observability.

ReviewSense – Sentiment Analysis for E-commerce Reviews Technologies,
Python, DistilBERT, Hugging Face Transformers, NLTK

- Built a sentiment analysis pipeline using DistilBERT, improving classification accuracy for customer reviews in an e-commerce context.
- Leveraged Hugging Face Transformers and NLTK for text preprocessing, model fine-tuning, and performance optimization, enabling actionable insights into user satisfaction and product feedback.
- Contributed to more data-informed engagement strategies by uncovering trends in customer sentiment across product categories