



Shivakumar Pasem

Data Engineer & Analyst

✉ pasemshivakumar706@gmail.com ☎ 7743208269  LinkedIn  Portfolio

Professional Summary

Data Engineer & Analyst with 3+ years of experience designing cloud-based data pipelines, AI models, and real-time dashboards. Skilled in Python, Power BI, Azure, AWS, and NLP tools like Hugging Face and DistilBERT. Proven track record in automating workflows, reducing manual effort, and driving insights from billions of clinical, financial, and customer records across healthcare, B2B, and B2C domains.

Skills

Languages & Scripting:

- Python (Pandas, NumPy, PySpark, Scikit-learn), SQL, PL/SQL, Shell

Data Platforms & Warehousing:

- Databricks, Snowflake, MySQL, PostgreSQL, Azure SQL Database, Oracle

Cloud & Big Data:

Microsoft Azure, AWS (S3, EC2, RDS)

Compliance:

- HL7 (Clinical Data Integration), HIPAA (Regulatory Compliance), Epic EMR (Reporting Automation)

Data Engineering & Integration:

- Azure Data Factory, Apache Airflow, Oracle Integration Cloud, Jenkins, dbt

Visualization & Reporting:

- Power BI (Executive KPIs, Sales Dashboards), Tableau (Customer Journey, Market Insights), Excel (Advanced Excel – VLOOKUP, PivotTables, Macros, Power Query)

AI & Machine Learning:

- Churn Prediction, Forecasting, Sentiment Analysis, Hugging Face Transformers, DistilBERT, Generative AI (LLMs), Text Classification, EDA, A/B Testing

Professional Experience

Data Engineer, Cardinal Health 09/2024 – Present | USA

- Analyzed and deployed 22+ ETL pipelines with PySpark and Azure Data Factory, processing over 3.2 TB of HL7 and Epic EMR data for care operations, finance, and compliance reporting.
- Configured AWS S3 archival workflows and lifecycle rules to support long-term storage for over 1.5 billion medical records, reducing query latency for historical audits by 62 seconds.
- Spearheaded the creation of 12 real-time Power BI dashboards for clinical, finance, and executive departments, reducing reporting request volume by over 45 weekly tickets and saving 20 weekly hours.
- Reduced ETL refresh times from 3 hours to 1.3 hours through Databricks pipeline optimization, improving reporting accuracy for 6 departments.
- Deployed NLP pipelines using Hugging Face Transformers to summarize 350,000+ physician notes annually, saving medical analysts over 180 hours in manual review.
- Maintained Apache Airflow DAGs monitoring 40+ daily batch processes, achieving 99.97% SLA uptime over 12 months.
- Boosted 15 CI/CD pipeline deployments using Jenkins with Zero rollback incidents across all releases.
- Developed an end-to-end HIPAA-compliant pipeline to ingest, clean, and visualize 800,000+ patient records across 10 hospitals.

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

- Developed and delivered 10+ Oracle ERP integrations to automate procurement and finance processes across 5 internal departments and 3 client systems.
- Eliminated over 60 hours/month of manual data entry by automating reconciliation workflows using Python and PL/SQL scripts.
- Created Power BI dashboards monitoring daily operational KPIs, used by 4 regional managers to guide resource planning for both B2B and B2C clients.
- Conducted daily monitoring of production pipelines, flagging anomalies in under 2 minutes using JSON schema validation and alerting logic.
- Ran sentiment analysis models on internal support tickets (~200,000 records), helping managers reprioritize features that impacted 3,000+ users.

Data Analyst, Wipro

04/2020 – 04/2021 | India

- Built 18 interactive dashboards in Power BI and Tableau that drove \$1.2M in improved inventory forecasting and customer conversion across 2 B2B logistics clients.
- Boosted ETL transformation for over 1.6 million daily records using Python and Advanced Excel (PivotTables, VLOOKUP, Power Query), cutting manual data prep by 2.5 hours per day.
- Analyzed 25 GB/month of telecom user behavior to identify churn risk thresholds; insights directly informed pricing changes across 4 product tiers.
- Designed a classification model using DistilBERT and NLTK to automatically tag sentiment across 200,000+ support reviews; improved triage time by 42 minutes per support cycle.

Projects

Cournot oligopoly model, 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.

ReviewSense – Sentiment Analysis for E-commerce Reviews, 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.

Data Pipeline for Healthcare Operations Analytics, 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.

Certificates

- Databricks Certified Data Engineer Associate (Issued by Databricks)
- Microsoft Azure Data Fundamentals (DP-900)
- AWS Certified Data Analytics – Specialty (DAS-C01)

Education

Master of science in Data science, University of Massachusetts

01/2023 – 12/2024 | Dartmouth, USA

Bachelor of Technology in Information Technology,

Jawaharlal Nehru Technological University

08/2017 – 09/2021 | India