

Pushkar Patil

Boston, MA, USA (Open to relocate) | pushkar.patil1269@gmail.com | (857)3130289 | [linkedin.com/in/pushkarpatil1/](https://www.linkedin.com/in/pushkarpatil1/)

Summary

Data Engineer with 3+ years experience building scalable data pipelines, automating ETL processes. Proficient in Python, NoSQL, SQL, PySpark, and cloud platforms (GCP, AWS, Azure). Expertise in data warehousing (Big Query, Snowflake), data orchestration (Airflow, Kinesis), data cataloging, microservices, and multi-source data integration for customer-first solutions

Experience

Data Engineer | StudentLandings

Feb 2025 – Present | Boston, MA

- Amplified international student housing platform's search speed 60%, integrated Elasticsearch to resolve NoSQL inefficiencies, optimized Redshift cost estimation and caching
- Elevated UI performance 5%, engineered ETL pipelines with AWS Glue for transformations, Amazon S3 for staging, Amazon Athena for queries, OpenStreetMap for geospatial context, enabling faster feature deployment across the platform
- Drove customer engagement, seasonal planning accuracy by 20% via methodologies and cross-team strategy sessions

Data Analyst/ Engineer | Institute for Experiential AI at NEU

Jan 2024 – Aug 2024 | Boston, MA

- Enhanced analytics for 9M users by optimizing 1M records/min of sensor data (AQI) using ArcGIS and Big Query, improving urban planning, emissions analysis, public health insights
- Cut data latency by 15%, constructing scalable ETL/ELT pipelines using Pub/Sub (real-time ingestion), GCP Cloud Storage (data staging), DBT, PySpark, Spark SQL (transformations), streamlining 16 workflows for cross-functional stakeholders
- Achieved a 35% inefficiency reduction by aligning data with stakeholders in Looker, ArcGIS dashboards, resulting in \$2.5K/month cost savings, ensuring regulatory compliance

Data Engineer I | Asfaleia

Dec 2020 – Aug 2022 | Mumbai, IN

- Minimized operational costs 20% for multiple clients, deployed integrated IT solutions via Asfaleia's Retail & Ecommerce Consultancy, honed retail operations, inventory management, customer engagement
- Tripled throughput, slashed latency 25%, saved \$10K annually, orchestrated automated batch workflows using Airflow, PySpark on AWS Glue, AWS Lambda, Snowflake, leveraged hybrid schema modeling to streamline high-volume data processing
- Elevated order processing, improved forecast accuracy by 20% using Tableau, led team meetings, facilitated executive presentations

SQL Developer | Spanish BOLO

Jan 2020 – Sept 2020 | Hyderabad, IN

- Resolved feature deployment delays, architected AWS workflows using Athena, Step Functions, S3, DynamoDB, boosted revenue forecasting 30% for a subscription-based Spanish learning platform
- Increased subscriptions 7%, simplified real-time pricing with SQL queries, refined joins using composite and hash indexes, triggers, procedures, slashed execution time 10%, powered real-time pricing via AWS Kinesis
- Boosted inventory updates, revenue projections 20%, developed dashboards in PowerBI, coordinated cross-team reviews

Skills

Data Engineering Tools: Talend, Fast API, Databricks, Unix, AWS (Glue, Lambda, S3, Kinesis, Cognito, Elasticsearch, DynamoDB, RDS, Athena), Apache (Airflow, Kafka, Spark, Hadoop, NiFi, Flink), GCP (BigQuery, Pub/Sub, CloudComposer, Dataflow), Alteryx

Cloud Platforms: Snowflake, Oracle Cloud, Docker, Kubernetes, Terraform, Git

Databases: PostgreSQL, MySQL, Oracle DB, MongoDB, Firebase, Cassandra

Programming Languages: Python (NumPy, Pandas, Scikit-learn, TensorFlow, Matplotlib, Seaborn), Java, Scala

Visualization: Looker Studio, Tableau, Power BI

Education

Master of Science in Information Systems, Northeastern University

Sep 2022 – Dec 2024 | Boston, MA

Bachelor of Science in Computer Science, Mumbai University

Jun 2017 – Oct 2020 | Mumbai, IN

Projects

LogiFlow: Logistics Data Pipeline (GCP, Pulumi, FastAPI, Apache Beam, Docker, Airflow)

- Engineered logistics analytics pipeline on GCP with Python, Dataflow, Cloud Scheduler, BigQuery, increased throughput 40%
- Automated ETL workflows for supply chain data using Mage.ai, Cloud Scheduler, Dataflow, improved analytics accuracy 35%

Fraud Detection System (Databricks, Delta Lake, Apache Kafka, Flink, Presto, Pulumi, LLM)

- Architected real-time fraud detection pipelines with Delta Lake, Flink, Kafka, slashing query latency by 25% using Presto
- Deployed PyTorch-based LLM for anomaly detection, leveraging GPT 2 for pattern recognition and enhancing fraud detection accuracy

PodcastGPT: RAG-Based AI for Intelligent Podcast Search (FAISS, OpenAI API, LangChain, FastAPI)

- Innovated PodcastGPT with FAISS, OpenAI GPT, LangChain, accelerating query retrieval speed by 15%
- Orchestrated scalable API using FastAPI, Docker, AWS Lambda, Step Functions, optimizing podcast search efficiency