

Tarun Punhani

(469) 427-5968 | tpunhani@gmail.com | Bentonville, AR, USA | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

PROFESSIONAL SUMMARY

Data Engineer with 10+ years of overall software engineering experience, including about 4 years focused on data engineering and data science — building streaming pipelines, cloud data platforms, and ML forecasting systems — at Walmart. Led the company's Azure-to-GCP streaming migration and delivered \$140K+ in annual cost savings through infrastructure optimization. Currently builds and maintains a Databricks/PySpark demand forecasting platform and develops internal AI developer tools to improve team productivity. Combines hands-on pipeline engineering with cross-functional initiative ownership (backend, UI, and data workstreams) and Agile/Scrum leadership.

WORK EXPERIENCE

Walmart Inc., Bentonville, AR, US

Jan 2023 – Present

Software Engineer III

GCP Migration & Cloud Architecture

- Led end-to-end migration of streaming data pipelines from Azure to GCP, architecting a GCP streaming pipeline integrating Kafka Connect, Google Cloud Storage, and Databricks Autoloader for Delta format processing.
- Designed and implemented automated Autoloader setup with CI/CD workflows, enabling consistent and repeatable deployments across environments.
- Diagnosed and resolved a major streaming bug causing data processing delays between the data lake and Databricks, tracing the root cause to a Spark job configuration issue.
- Migrated legacy BigQuery cron jobs into the company's central repository as part of a product consolidation strategy.
- Added data latency tracking columns to BigQuery ingestion pipelines, improving observability and real-time monitoring.
- Created a reusable GCP onboarding template, reducing future pipeline integration effort and standardizing deployment practices.
- Built an ingestion pipeline to process data from Microsoft SharePoint to Google Cloud Storage using the Microsoft Graph API.

Forecasting Pipeline Engineering & AI Tooling

- Maintain and support a demand forecasting platform — a suite of 10+ interdependent Databricks/PySpark notebooks (ensemble forecasting, linear-optimization-based inbound/outbound reconciliation, CatBoost sub-category disaggregation) — with 100+ automated test cases covering the pipeline.
- Built a CatBoost-based machine learning model to forecast demand split across merchandise categories, analyzing historical sales data and event patterns to achieve 80% forecast accuracy within a single cycle.
- Designed and published notebook-diff-doctor, an internal AI developer tool that automatically fixes formatting issues in Databricks notebooks so code changes are easy to review — adopted company-wide through an internal developer tools catalog.
- Onboarded automated email notifications across the forecasting pipeline suite using an internal company email service, removing a third-party vendor dependency and improving security.
- Replaced static, manually-maintained reference files in forecast optimization models with live SQL queries against inventory tables, improving accuracy and removing a manual data-refresh step.
- Led a cross-functional initiative to automate a manual demand-planning update process — designing the data schema and coordinating with backend and UI engineers to replace an untracked manual process with an automated, trackable pipeline surfaced in a business-facing planning UI.

Cost Optimization & Security Automation (Total \$140K+ Annual Savings)

- Spearheaded a company-wide cost optimization initiative, analyzing CPU usage across streaming pipelines and BigQuery cron jobs, and redeploying workloads to achieve \$40K+ in annual compute savings.
- Built an Azure data purge automation script that removed 500TB of unused Azure Data Lake storage, saving \$100K+ annually.
- Tracked cost savings in detail and provided daily updates to leadership on optimization progress.
- Integrated Akeyless secrets management into BigQuery cron jobs, removing hardcoded certificates and improving security.

Data Pipeline Framework & Infrastructure Onboarding

- Enhanced a batch data pipeline framework where jobs are orchestrated via Apache Airflow (Astronomer) and executed on Google Cloud Dataproc, moving data from BigQuery to Databricks — onboarded 10+ tables with stable production operation.
- Onboarded a new Apache Flume agent and a new data center into the company's ingestion framework, with full documentation and knowledge-transfer sessions.
- Found and permanently fixed a critical Flume pipeline bug related to zstd compression by upgrading the Flume version and handling edge cases — the fix was adopted company-wide.

System Reliability & Framework Development

- Diagnosed and resolved critical defects in a core supply-chain data pipeline, strengthening system reliability; performed root-cause analysis on pipeline job failures and shared RCA reports with leadership and business stakeholders.
- Built a data reconciliation framework on the company's existing codebase with DAG-based scheduling and one-click execution for streamlined operations.
- Developed a CI/CD governance framework enabling configuration changes through pull-request-based approvals, improving change management.
- Contributed to a proof-of-concept for dynamic configuration management across streaming applications using Kafka Connect.

Agile Leadership & Team Collaboration

- Served as Scrum Master for the first two quarters, leading sprint retrospectives, sprint demos, and daily standups, and helped set up a rotational Scrum Owner model.
- Owned the full development and onboarding of GCP pipelines and a data center rollout, mentoring team members along the way.
- Provided after-hours and weekend production support, demonstrating strong ownership.

Innovation & Professional Development

- Built an AI on-call support agent (company hackathon project, still being actively improved) that uses LangGraph to orchestrate multi-step incident triage — classifying alerts, looking up likely root causes, and suggesting remediation steps — to help reduce on-call response time.
- Participated in a 3-day company hackathon, building an AI chatbot to handle supply chain container arrival questions.
- Submitted "Remy – Personal Cuisine Companion," an AI chatbot, to an internal innovation challenge.
- Attended the Databricks Conference and shared key learnings with the team.

OTHER PROFESSIONAL EXPERIENCE

Axxess, Dallas, TX, US — Software Engineering Intern

Jan 2022 – Aug 2022

- Extended key functionalities for new clients and resolved 50+ production bugs in the Axxess Home Care solution, using Entity Framework Core and Dapper ORM to improve code efficiency by 10%.

- Analyzed production logs via Kibana for data-driven debugging and performance improvements.

Times Internet, Gurugram, India — Sr Software Development Engineer In Test**Sept 2020 – Jan 2021**

- Built a fully automated CI/CD pipeline in Jenkins (30% time savings) and implemented microservices for the Longwalks journaling app using MongoDB and Neo4j.
- Designed and optimized database schemas and Elasticsearch indices, improving system performance and availability by 50%.

Airtel Digital Limited, Gurugram, India — Sr Software Engineer**Nov 2018 – Aug 2020**

- Built and maintained 5+ high-performance microservices with a Cassandra NoSQL database serving millions of users for a large-scale music streaming application.
- Designed distributed data models and optimized query patterns for Cassandra, improving read/write performance by 30%.
- Developed an end-to-end functional testing automation framework, contributing to 20% customer base growth across 3 projects.

Paytm, Noida, India — QA Engineer**July 2018 – Nov 2018**

- Implemented REST API testing automation for the payment processing module and built a log aggregation web service, improving testing team efficiency by 10%.

Accenture Solutions Pvt Ltd, Gurugram, India — Test Engineering Analyst**Oct 2014 – July 2018**

- Developed data-driven Selenium automation frameworks and VB scripting tools for HP ALM integration. Received Self Starter and Client Champion awards.

LEADERSHIP & OTHER EXPERIENCE

New Joinees Mentor and Project Leader

- Mentored 5+ new joinees, overseeing orientation and knowledge transfer, and guided them on decision-making and planning. Led the Hello Tunes project at Airtel.
- Participated in HackDFW 2022 and won 3rd place in the sponsor challenge from TIAA.

TECHNICAL SKILLS

Data Engineering: PySpark, Apache Spark, Apache Kafka, Kafka Connect, Apache Airflow, Apache Flume, Apache Storm, ETL/ELT Pipelines

Cloud & Big Data: Google BigQuery, Google Cloud Dataproc, Azure Data Lake, Databricks, Databricks Autoloader, Google Cloud Platform (GCP), Azure, Multi-Cloud Architecture

ML / Forecasting: CatBoost, Ensemble Modeling, SciPy Linear Optimization, Forecast Pipeline Engineering

Agentic AI: LangChain, LangGraph, Prompt Engineering, AI Agent/Plugin Development

Databases & Storage: SQL, MySQL, MongoDB, Cassandra, Elasticsearch, Data Warehousing, Data Modeling

Programming: Python, Java, Shell Scripting, SQL, PySpark, JavaScript

DevOps & Tools: Git, CI/CD (Jenkins), Docker, Linux, Postman, Astronomer, Data Pipeline Orchestration, nbconvert/nbformat

EDUCATION

University of Texas at Dallas, Dallas, TX, US**Jan 2021 – Dec 2022**

- Master of Science in Computer Science; GPA 3.8

Dr. Akhilesh Das Gupta Institute of Technology and Management, Delhi, India**Sept 2010 – July 2014**

- Bachelor of Technology in Computer Science and Engineering; GPA 3.3