

Ashutosh Modi

Senior Data Engineer

Jacksonville, FL · contact@modiashu.dev · modiashu.dev · linkedin.com/in/modi-ashutosh · github.com/modiashu · U.S. work authorization — no sponsorship required

SUMMARY

Senior data engineering leader with 14+ years of experience building petabyte-scale data platforms, identity resolution systems, and self-serve analytics for cross-functional stakeholders. Proven track record leading 0-to-1 platform builds using Python, SQL, Spark, Airflow, BigQuery, Databricks, and Snowflake. Combines deep technical execution with strategic communication to executive leadership. U.S. Patent holder and published AI thought leader on agent communication patterns and RAG evaluation strategies.

EXPERIENCE

Senior Data Engineer — Sam's Club (Walmart)

Jacksonville, FL · 2022 – Present

- Architected a privacy-first media data platform using Databricks Clean Rooms and Unity Catalog, enabling secure multi-party data sharing for audience onboarding and campaign measurement across marketing stakeholders.
- Engineered a scalable identity-matching platform on Delta Lake supporting 200M+ identities and 2,000+ attributes, serving as the canonical source for high-volume marketing analytics workloads.
- Designed and operated a petabyte-scale BigQuery data warehouse managing 25+ TB of data and 110+ billion records, serving as the foundational self-serve analytics platform for business stakeholders.
- Built and maintained scalable ETL/ELT pipelines using Python, SQL, Spark, Airflow, and GCP, enabling reliable ingestion, transformation, and governed access to critical business datasets.
- Optimized query performance and warehouse costs through advanced partitioning, clustering, and query tuning — reducing analytics query costs by 99% and improving insight latency for end users.
- Secured U.S. Patent 12141166B2 — manifest-based incremental synchronization framework across Azure Delta Lake and GCP BigQuery, delivering 75% time savings and 90% cost reduction (\$30K+ savings in Year 1, with compounding savings ongoing).
- Led a team of data engineers, providing technical mentorship, managing task allocation via Agile roadmaps, and establishing engineering best practices that scaled across the organization.
- Directed full project lifecycle for high-visibility media initiatives, driving cross-functional alignment across Media, Product, and Analytics under tight deadlines.
- Partnered with cross-functional business and analytics stakeholders to translate complex requirements into governed, production-ready data products; established enterprise-wide data quality validation standards adopted across teams.
- Designed regression dashboards on materialized views to accelerate UAT and KPI validation — doubling deployment frequency and reducing manual effort by 20% per sprint.
- Established CI/CD pipelines and access controls for data workflows via GitHub, improving deployment reliability and enforcing secure engineering practices.
- Authored AI thought leadership published internally at Walmart DX: agent communication patterns and RAG evaluation strategies — featured in Walmart CTO's weekly communication.
- Published reusable AI Skills to the Walmart Meta Registry (git-eval-prep, git-user-contributions, git-project-summary) to automate engineering productivity audits at scale.

Software Engineer III — Sam's Club (Walmart)

Bentonville, AR · 2018 – 2022

- Designed and implemented a Neural Machine Translation (NMT) system using NLP, RNN-based Seq2Seq models, attention mechanisms, and fastText embeddings — evaluated with BLEU scores to improve product description quality at scale.
- Implemented enterprise-grade data governance frameworks including lineage tracking, access controls, and privacy enforcement across business and analytics teams.
- Built image collection reporting dashboards in Power BI on Azure for operational visibility.

Research Assistant — University of South Florida

Tampa, FL · 2016 – 2018

- Developed a traffic yield-rate calculation algorithm combining pedestrian and vehicle detection at signalized intersections, improving accuracy of safety and flow analysis.
- Engineered pedestrian tracking system using OpenCV, GPU acceleration, and C++, enabling reliable detection and identification across video frames.

Software Engineering Intern — Optym

Gainesville, FL · 2017

- Developed an incremental learning prediction model to forecast shipment tonnages, enabling the optimization engine to execute three hours earlier and accelerating daily decision-making.

Earlier career: Senior Software Engineer at iGate/Capgemini (2012–2016) — automation prototypes for Rio Tinto mining operations. Project Engineer at Wipro Technologies (2010–2012) — telecom protocol development for Avaya systems. Software Engineer at Dell (2010) — enterprise deployment packaging.

PATENT

U.S. Patent 12141166B2 — "System and Method for Service Synchronization" (Walmart Apollo LLC, Nov 2024). Manifest-based incremental synchronization framework across Azure Delta Lake and GCP BigQuery.

SKILLS

Languages & Frameworks	Python (PySpark), SQL, C++, Shell, Apache Spark, Apache Airflow, Apache Kafka, Spark Structured Streaming, dbt
Data Platforms	BigQuery, Snowflake, Databricks (Delta Lake, Unity Catalog), GCP, Azure, DuckDB, Lakehouse Architecture
Infrastructure & DevOps	Terraform, Docker, CI/CD, GitHub Actions, Distributed Systems
Analytics Engineering	Dimensional Modeling, Metric Definition, Data Governance, Data Quality, Privacy-First Design, Identity Resolution
AI / ML	LangChain, Ollama, Hugging Face, scikit-learn, PyTorch, RAG Pipelines, Prompt Engineering
Visualization	Looker Studio, Power BI, Streamlit
Leadership & Delivery	Team Leadership, Technical Mentorship, Stakeholder Management, Agile Roadmaps, Engineering Best Practices

PUBLICATIONS & THOUGHT LEADERSHIP

- AI thought leadership (Walmart DX): "Prompts vs. Instructions vs. Skills", "Text Summarization Evaluation Paradox", and "Turning Git History Into Engineering Insights with Reusable Skills" — all featured in Walmart CTO's weekly communication.
- Research: Hinduja, S., Modi, A., et al. "Enhancing Pedestrian Safety at Signalized Intersection" — submitted to IEEE ICIP 2018.
- Personal technical writing on Medium (medium.com/@modiashutoshd), including "From Hello World to Streaming Insights: What I Learned Building with Local LLMs".

EDUCATION

- **M.S., Computer Science** — University of South Florida, Tampa, FL (GPA 3.96/4.0)
- **B.E., Electronics and Telecommunication** — PES Modern College of Engineering, Pune, India (First Class with Distinction)

AWARDS

- Two-time Wall of Fame quarterly award at Capgemini/iGATE (team of 150) for successful project delivery.
- Winner, data mining prediction competition (USF).
- IEEE Technical Paper Presentation — IEEE Systems, Man and Cybernetics Society, Bombay Section (2009).