

Shashank Puli

(484) 278 - 1189 | pulishashank9@gmail.com | [LinkedIn](#) | [GitHub](#)

SUMMARY

Data Analytical Engineer with 4+ years of experience architecting scalable cloud infrastructure and delivering high-impact business intelligence across Fintech, Healthcare, and Marketing sectors. Expert in fusing AWS and GCP engineering with statistical modeling to drive \$8M in savings, optimize marketing ROI, and ensure regulatory compliance. Proven ability to transform raw data into executive insights, bridging the technical gap between backend pipelines and strategic decision-making for top-tier global organizations.

TECHNICAL SKILLS

- **Cloud:** AWS (S3, Glue, Lambda, Redshift, EMR, Data Lake), Azure (ADF, Databricks, Data Lake, Azure SQL), GCP (BigQuery, Dataflow, Composer)
- **Data Engineering:** Snowflake, Redshift, PySpark, Airflow, dbt, Kafka, Apache Beam, Apache Spark, DynamoDB, MongoDB, ETL/ELT Design, Spark Streaming
- **Analytics & BI:** Power BI, Tableau, QuickSight, Looker Studio, Excel, SSRS, Alteryx, A/B Testing
- **Programming & Scripting:** Python (Pandas, NumPy, Scikit-learn), R Programming, SQL (Advanced), Java, Scala, JavaScript, Shell Scripting, Statistics, Financial Modeling
- **Governance & Compliance:** HIPAA, SOX/SEC Reporting, FHIR/HL7 Standards, Data Lineage (Collibra concepts)
- **Development:** Kubernetes, Docker, GitHub Actions, Rest APIs, FastAPI

PROFESSIONAL EXPERIENCE

Crowe LLP

Data Engineer Analytics

Jun 2024 - Present

Tampa FL

- Architected Snowflake Data Warehouse for real-time financial analytics, migrating 12TB using AWS Glue and dbt; implemented dimensional modeling with SCD Type 2, achieving 35% query latency reduction and enabling sub-second compliance reporting.
- Streamlined PySpark pipelines handling 1.2 billion monthly financial events from Apache Kafka to Redshift, applying broadcast joins and improved partitioning to reduce execution time by 27% (from 52 to 38 minutes) and save \$40K in infrastructure costs annually.
- Engineered automated reconciliation system between Core Banking and General Ledger (GL) using SQL window functions; processed \$75B+ daily transactions at 99.9% accuracy, eliminating \$8M in annual discrepancies and reducing manual audit effort by 87%.
- Built modular AWS Glue job templates and dbt macro libraries, standardizing 80+ data pipelines; reduced new pipeline deployment time from 4 days to 1 day, enabling faster market response for emerging compliance requirements.
- Developed and implemented a comprehensive Data Governance framework for the organization, documented over 150 data flow maps and 200 metadata standards, cut SOX audit findings by 60%, and set up a central data stewardship council to ensure ongoing compliance.
- Conducted quarterly Variance Analysis on \$500M+ financial portfolios using Python and SQL; identified critical budget drivers in less than 24 hours versus 5 days manually, enabling \$15M quarterly capital reallocation and 40 %faster executive decision-making.
- Built statistical anomaly detection model on \$75B+ daily transactions using isolation forests; flagged suspicious patterns with 97 %precision, reducing downstream fraud investigation time by 80 % and preventing \$2.5M+ in potential losses.

GP InfoTech

Data Engineer

Mar 2021 - Jul 2023

Hyderabad, India

- Developed HIPAA-compliant Python ETL pipelines using GCP Dataflow with Apache Beam to process 4TB+ monthly electronic health records; implemented FHIR standard transformations for HL7 data integration, improving clinical data accuracy from 94% to 99.8% and eliminating compliance violations.
- Designed Pub/Sub + Cloud Composer streaming architecture for 800K daily prescription records to BigQuery. Optimized schema and partitioning, reducing pharmacy analytics query latency by 45-60%.
- Optimized GCP Dataproc PySpark jobs analyzing 1.5M patient encounters. Broadcast joins and partition tuning reduced job latency from 35 to 23 minutes (34% improvement).
- Created Power BI dashboards that tracked 60+ clinical KPIs for 200+ staff, leading to targeted actions that cut 30-day hospital readmissions by 12%, saved 240 patient-days, boosted patient satisfaction by 18%, and delivered \$1.2M in savings.
- Performed pharmacy cost variance analysis across 8 hospitals using ARIMA time-series forecasting models; identified \$500K in annual inventory waste through pattern recognition and demand forecasting, driving hospital-wide procurement optimization that reduced pharmacy costs by 8% (\$400K savings)
- Implemented automated data quality validation framework using Great Expectations for 1.5M+ patient encounters daily; flagged pre-production anomalies in real-time, achieving 99.2% accuracy, preventing HIPAA compliance violations, and maintaining zero audit exceptions across eight consecutive external audits.

Carborundum Universal

Data Analyst

Feb 2020 - Mar 2021

Chennai, India (remote)

- Developed a Power BI platform that combined 12+ KPIs from 5 different systems and automated daily reports, cutting manual Excel work by 85% and delivering real-time dashboards for executives.

- Created Power BI dashboards analyzing 150K customer records with SQL queries, delivering sales team insights that improved conversion rates 22% through customer segmentation analysis
- Designed Power BI dashboards with RFM customer segmentation (150K records). Delivered segment-specific insights to sales leadership for targeted outreach campaigns.
- Collaborated with finance, operations, and C-suite executives to translate business requirements into actionable dashboards and KPI scorecards; delivered 4 executive-level reports with interactive drill-down capabilities, reducing decision-making cycle time by 60% and improving forecast accuracy for quarterly planning
- Analyzed A/B test results for 8+ marketing campaigns (email, landing page, pricing), helping leadership accelerate feature rollout decisions by 35% and boost conversion rates by 12%.

EDUCATION

University of South Florida

Master of Science, Computer Science

CERTIFICATIONS

- AWS Certified Data Engineer - Associate
- Google Data Analytics Professional
- Microsoft Power BI Data Analyst
- AWS Solutions Architect - Associate