

# Shubham Malpani

Shubham.malpani@gmail.com | 315-303-6468 | LinkedIn

## PROFESSIONAL SUMMARY

AI-focused Data Engineer with 5 years of experience building the data infrastructure that makes LLMs production-ready. Shipped 7 MCP servers connecting Claude to 200+ warehouse tables with multi-layer security and observability, surfacing \$7.6M in unrealized capacity that traditional dashboards had missed. Focused on bridging the gap between raw enterprise data and AI systems that business users can actually rely on.

## EXPERIENCE

### Vital Impact Group (formerly Arabella Advisors), New York

February 2022 - Present

#### Data Architect and AI Engineer

December 2025 - Present

- Developed 7 production MCP servers in Python (~8,200 lines, 65 tools) connecting Claude to the enterprise data warehouse across 200+ tables, enabling natural-language queries that identified \$7.6M in unrealized billable capacity across 160 staff.
- Launched AI-powered data cubes that let non-technical staff query warehouse data in natural language, reducing turnaround on routine data questions from days to minutes across finance, HR, and operations.
- Integrated a pipeline diagnostics agent into the MCP infrastructure that correlates failure logs with source notebook context, identifies root causes, and recommends fixes across 125 production pipelines, cutting mean time to resolution from hours to minutes.
- Deployed AI-powered entity matching across enterprise CRM (Salesforce), ERP (NetSuite), HRIS (Paycor), and project management (OpenAir) systems where no shared keys existed, matching customers, projects, and employees across platforms with approved matches feeding into the next pipeline run automatically, reducing data row rejections from 25% to 5%.
- Designed and delivered the enterprise data warehouse on Microsoft Fabric, consolidating 21+ enterprise source systems (ERP, CRM, HRIS, billing, project management) into a governed platform serving 400+ users.
- Pioneered the data engineering function, recruiting and leading a team of 8 contractors that accelerated delivery pace 5x.
- Designed an intelligent query routing system with 8-layer security validation, 3-tier observability logging, and a prompt engineering layer that guides LLM query selection across 5 database layers, achieving 94.5% success rate across 1,300+ weekly analytical queries.

### Assoc. Director, BI and Analytics

January 2023 - November 2025

- Devised a revenue reconciliation engine across 5 enterprise systems (ERP, CRM, billing) and 3 service lines, then expanded it with Power Apps, Logic Apps, and Dataverse to serve 200+ additional users.
- Designed a 3-phase integration project replacing a broken email-based provisioning system, constructing API-driven Logic Apps connecting HRIS (Paycor), Active Directory, and downstream enterprise systems so onboarding, offboarding, and role changes flow end-to-end automatically, cutting provisioning time by 93%.
- Automated manual month-end reporting for 15 clients with data pipelines, saving 300+ analyst hours per month at 99% accuracy.
- Presented at the CDATA data integration conference as an invited industry speaker (60+ attendees), demonstrating how the team standardized client onboarding across QuickBooks, Sage, and 10+ source systems using unified connectors instead of one-off API builds.

### Data Analyst

February 2022 - December 2022

- Consolidated reporting across 3 acquired companies into a single data model, cutting processing time by 40% and saving \$75K/year.
- Architected the company's first gross profit report across 150 engagements, identifying 30+ unprofitable projects and giving leadership a project-level view of revenue vs. cost that triggered pricing and staffing corrections.

## EDUCATION

M.S. in AI Systems Engineering, Harrisburg University (Expected Oct 2027)

Harrisburg, PA

M.S. in Information Systems, Syracuse University (Dec 2022)

Syracuse, NY

B.E. in Electronics & Telecommunication, University of Mumbai (May 2020)

Mumbai, India

## TECHNICAL SKILLS

**AI & LLM Integration:** Model Context Protocol (MCP), FastMCP, Claude API & Code, LLM-to-database integration, AI pipeline diagnostics, prompt engineering, LLM observability, agentic AI

**Data Platforms:** Microsoft Fabric, Delta Lake, Snowflake, AWS (S3, Glue, Redshift), PostgreSQL, MongoDB

**Languages & Engineering:** Python, T-SQL, Spark SQL, DAX | ETL/ELT, medallion architecture, dimensional modeling (Kimball), SCD Type 4, REST APIs

**DevOps & Tools:** Azure DevOps (CI/CD, Git, Pipelines), Power BI, Fivetran, JIRA, Agile/Scrum