



RAHUL VARMA ALLURI

Dearborn, MI (Open to Relocate) | (586) 788-4133 | alluri@umich.edu | [LinkedIn](#) | [GitHub](#)



SUMMARY

Microsoft Certified Fabric Data Engineer (DP-700) with over 5 years of experience creating and scaling data systems in healthcare, manufacturing, and pharma. Skilled in building complete Medallion lakehouse architectures, real-time streaming pipelines, and dimensional models. Has a proven record of improving data warehouse queries by 60% and lowering pipeline latency for high-throughput IoT and LIMS systems. Combines strong SQL and Python skills with modern cloud analytics to provide reliable data platforms.

TECHNICAL SKILLS

- **Microsoft Fabric Ecosystem:** OneLake, Fabric Lakehouse & Data Warehouse, Fabric Data Factory (Pipelines & Dataflows Gen2), Real-Time Intelligence (Eventstreams, KQL), Data Activator, Power BI (Direct Lake mode)
- **Cloud & Azure Legacy:** Azure Data Factory (ADF), Synapse Analytics, Azure Databricks, ADLS Gen2, Event Hubs, AWS (S3, Lambda, Glue, SageMaker)
- **Data Architecture & Modeling:** Medallion Architecture (Bronze/Silver/Gold), Micro-services, Star/Snowflake schemas, Dimensional Modeling, Slowly Changing Dimensions (SCD), Change Data Capture, Partitioning, Indexing
- **Languages & Querying:** SQL, T-SQL, PL/SQL, Advanced CTEs, Window Functions, Python, PySpark, pandas, NumPy, SQLAlchemy, Spark SQL, C#/.NET
- **Data Orchestration & DevOps:** Airflow, Hangfire, CICD, Jenkins, Azure DevOps, Octopus Deploy, Git, Docker

CERTIFICATIONS

- **Microsoft Certified: Azure Fabric Data Engineer Associate (DP-700)**, Data Factory Pipelines, Lakehouse, Synapse, Databricks, ADLS Gen2, Spark, Event Hubs, partitioning & incremental load patterns
- **Microsoft Certified: Power Platform Developer (PL-400)**,
- **Microsoft Certified: Azure Data Fundamentals (DP-900)**
- **Salesforce Agentforce Specialist**, agentic AI workflows; daily use of GitHub Copilot and Claude Code for AI-assisted development

PROFESSIONAL EXPERIENCE

Bytebridge Technologies(Client: Jhonson & Jhonson)

Contract | Data Engineer - Digital Manufacturing & Agentic AI

Remote, US

Sept 2025 - Present

- Architected an end-to-end real-time ingestion pipeline in **Microsoft Fabric**, utilizing **Fabric Eventstreams** to capture high-throughput PLC and sensor telemetry over MQTT/OPC UA, reducing data latency from ~8s to under 1.5s
- Designed a **Fabric Lakehouse** using a **Medallion architecture** (Bronze/Silver/Gold), processing streaming edge data into Delta Parquet format via Fabric Notebooks (PySpark) with structured retry and dead-letter queue patterns
- Engineered scheduled and event-driven orchestration workflows using **Fabric Data Factory** to automate validation report generation and log parsing, **achieving a 92% automated task success rate**
- Developed real-time OEE and anomaly-detection dashboards utilizing **Power BI in Direct Lake mode**, bypassing traditional import/DQ refresh cycles to surface recurring equipment failure patterns directly from **OneLake**

Eurofins IT Solutions India Pvt Ltd

Senior Software Engineer - Healthcare LIMS Data Platform

Bangalore, India

Apr 2021 - Aug 2023

- Designed and deployed a snowflake-schema data warehouse on SQL Server for laboratory analytics serving **200K+** users globally, centralizing global sample trends that previously required manual extracts
- Accelerated consumer-facing report **performance by 60%** by implementing date-based partitioning on fact tables, applying composite indexes, and refactoring legacy nested queries into optimized CTEs
- Built high-performance .NET ETL workflows transforming complex XML/JSON lab feeds with **schema validation** and **ActiveMQ** messaging, reducing pipeline **processing time by 38%** and preventing data loss
- Migrated a monolithic LIMS database architecture into **6 decoupled** domain-driven data stores (Results, Reports, Master Data), orchestrating parallel run states and executing staged data migrations with **zero downtime**
- Delivered automated analytics dashboards (Tableau, Power BI, Angular) tracking operational KPIs, replacing 6 manually maintained legacy Excel reports

- Delivered Tableau-style analytics dashboards (Angular + Apache ECharts) for sample-trend, contract-performance, and turnaround-time metrics, with PNG/SVG/PDF/CSV/Excel export and lazy-loaded chart modules, replacing 6 manually-maintained Excel reports
- Established **CI/CD pipelines** (Jenkins, Octopus, Git) with **TDD** (XUnit, **95%+ coverage**), peer-review gates, and runbook documentation; reduced production incidents **35%**, new-hire ramp-up time **30%** across a 5-engineer team

Pratian Technologies

Bangalore, India

Software Engineer - Data Pipelines & Automation

Apr 2019 - Mar 2021

- Built production **ETL pipelines** transforming raw XML/JSON ERP data into normalized relational models, reducing manual data entry workloads by 75% and **cutting processing times by 38%**
- Developed metadata-driven ingestion validators and implemented asynchronous data processing patterns, **achieving 99.7% uptime** on customer-facing ingestion modules
- Integrated automated testing gates within Jenkins CI/CD pipelines, reducing regression **defects by 35%** and streamlining production deployments
- Achieved **99.7% uptime** on customer-facing modules through async patterns (RxJS), state management (NgRx), and structured retry/circuit-breaker policies (Polly)

PROJECTS

Healthcare Analytics Engineering Platform (dbt + Prefect)

Present

- Built a modern data stack portfolio project orchestrating Medicare/healthcare datasets from raw ingestion to reporting-ready gold tables
- Implemented **dbt Core** for modular SQL transformations, enforcing data quality tests, schema validations, and generating automated documentation DAGs
- Orchestrated the end-to-end ELT lifecycle using **Prefect**, managing failure alerts and dependency mapping

Cloud-Based Diabetes Prediction System (Healthcare ML on AWS)

Jan 2025 - Apr 2025

Cloud Computing

- Architected end-to-end serverless healthcare ML pipeline on AWS (S3, API Gateway, Lambda, SageMaker XGBoost endpoint), achieving **85% accuracy and 80% recall** on PIMA diabetes dataset with sub-second real-time inference
- Built CI/CD pipeline with AWS CodeBuild for automated model and infrastructure deployment; resolved CORS, IAM, and API Gateway authorization configurations to enable secure cross-origin health-metric submission

Formula 1 World Championship, Relational Data Warehouse

Aug 2023 - Dec 2023

Database Systems

- Designed normalized RDBMS schema across 8 entities (drivers, races, qualifying, results, constructors) to analyse qualifying-to-final-race correlation and maintenance-issue impact across 70+ years of championship data
- Tuned query performance **20%** through composite indexing on driver-race-circuit join paths and execution-plan analysis; built Tableau dashboard for circuit-level driver performance comparison

DATA OPERATIONS & GOVERNANCE

- **DataOps & CI/CD:** Proficient in setting up environment-backed deployment pipelines, utilizing Fabric's native Git integration with Azure DevOps/GitHub to manage seamless code promotion between Dev, QA and Prod
- **Data Quality & Monitoring:** Experienced in configuring automated data quality alerts, schema drift detection, and operational alerts using Fabric Data Activator and Prefect automation hooks
- **Compliance & Security:** Well-versed in implementing Row-Level Security (RLS) and Column-Level Security (CLS) within Direct Lake Power BI models to uphold strict data privacy standards (SOC 2, HIPAA)

EDUCATION

University of Michigan - Dearborn

Dearborn, MI, US

Master of Science, Data Science

Aug 2023 - May 2025

Coursework: Database Systems, Big Data & Visualization, Advanced Database Management, Cloud Computing

GITAM University

Visakhapatnam, MI, US

Bachelors of Technologies, Electronics and Communications

Apr 2014 - Jun 2018