

# Praveen Ganapathy Ravi

Email: [erpraveenganapathy@gmail.com](mailto:erpraveenganapathy@gmail.com)

LinkedIn: [linkedin.com/in/pravi01](https://www.linkedin.com/in/pravi01)

**Professional Summary:**

- Data Engineer with 6+ years of experience architecting scalable ETL pipelines, cloud-native data solutions, and enterprise data warehouses on the Azure Stack.
- CSM-certified professional with a proven track record of leading critical migrations of ~4TB of production data for Fortune 500 and Telecom clients (NTT Data, TalkTalk).
- Extensive expertise in Azure Data Factory, Azure Synapse, and SQL Managed Instances, ensuring zero data loss and 99.9% availability during complex cutovers.
- Engineered high-throughput data ingestion pipelines to feed Generative AI models (LLMs) and RAG architectures, bridging the gap between data engineering and AI product strategy.
- Highly proficient in Python (Advanced), SQL (T-SQL/PL-SQL), and Apache Spark, optimizing backend performance and query speeds by up to 40%.
- Designed and maintained CI/CD pipelines using Azure DevOps and GitLab, reducing release deployment times by ~40% through automated testing and "Infrastructure as Code."
- Managed complex ETL workflows for major UK Telecom clients, processing sensitive transactional data in strict compliance with GDPR and regulatory standards.
- Spearheaded legacy application modernization, transitioning monolithic PHP backends and on-premise servers to microservices-based, cloud-native Azure architectures.
- Developed and standardized RESTful APIs and data dictionaries, creating repeatable content frameworks that accelerated developer adoption by 30%.
- Identified inefficiencies in cloud resource allocation, implementing auto-scaling strategies that generated over \$50K in operational savings for clients.
- Skilled in predictive modeling and data analytics, enabling real-time business intelligence through optimized Power BI and Tableau executive dashboards.
- Authored comprehensive technical guides, SOPs, and audit-ready documentation to support internal knowledge transfer and ensure engineering consistency.
- Built and executed rigorous Data Quality Frameworks (DQC) to detect anomalies in real-time, reducing data discrepancy resolution time by 50%.
- Partnered with cross-functional teams (Product, IT, QA) in fast-paced Agile/Scrum environments to translate complex business requirements into functional data specifications.
- Led SEO domain migration and infrastructure consolidation projects, maintaining over 99% data integrity and preserving search rankings during platform shifts.
- Demonstrated leadership by mentoring junior analysts and facilitating training sessions on Azure best practices, SQL tuning, and version control.

**Technical Skills & Tools:**

| Category         | Skills / Tools                                                                                     |
|------------------|----------------------------------------------------------------------------------------------------|
| Cloud & ETL      | Azure Data Factory, Azure Synapse, SQL Managed Instances, AWS (S3, Lambda, EMR), Databricks, Jedox |
| Languages        | Python (Advanced), SQL (T-SQL, PL/SQL), Scala, JavaScript, Shell Scripting                         |
| Data Engineering | ETL/ELT Design, Data Warehousing, Apache Spark, Airflow, Data Modeling                             |
| DevOps & Tools   | Git, CI/CD (Azure DevOps, GitLab), Docker, Kubernetes, Terraform                                   |
| AI & Analytics   | GenAI Integration, RAG Architectures, Predictive Modeling, Power BI, Tableau                       |
| Methodologies    | Agile, Scrum, CI/CD, Legacy Modernization                                                          |

**Data Engineer****Feb 2026 – Present****World Bank Group – Washington DC, US****Responsibilities:**

- Designed and built 20+ Azure Data Factory ETL/ELT pipelines and data flows integrating financial, budget, and HR data from ODBC sources, REST APIs, SFTP, and SAP (FI, CO, HR) into the World Bank's new Budget Planning Systems.
- Developed core pipelines and transformations for Budget, CapEx (assets, depreciation, cost share, fund movements, investments), Commitment Items, Fund Center Hierarchy, Exchange Rate, and consolidated FICO/RMRF (incl. secure variants), loading curated data into ADLS.
- Built and configured Jedox ETL integrators to connect planning and budgeting data flows between Jedox and upstream Azure/SAP sources, enabling accurate, automated data refreshes for enterprise budget planning.
- Set up and managed 20+ scheduled triggers to orchestrate daily pipeline execution across budget, CapEx, commitment, fund hierarchy, and SAP integration workflows, ensuring reliable, timely data availability.
- Configured linked services, integration runtimes, and connections for secure data movement across on-prem and cloud sources.

**Data & AI Software Engineer****Aug 2025 – Dec 2025****Goodie Bag - US, Remote****Responsibilities:**

- GenAI Pipeline Architecture: Engineered high-throughput data ingestion pipelines to feed Generative AI models (LLMs), translating complex AI research concepts into scalable, client-facing product features.
- API Standardization: Designed and standardized robust RESTful APIs and data dictionaries, creating a repeatable content framework that accelerated internal developer adoption by 30%.
- Backend Optimization: Collaborated directly with the CTO in a fast-paced startup environment to optimize backend query performance and structure data workflows within a CI/CD ecosystem.
- Technical Documentation: Spearheaded the internal adoption of new data tools by authoring comprehensive API documentation and technical guides, ensuring 100% consistency across engineering deliverables.
- Data Strategy: Established data governance protocols for early-stage AI products, ensuring clean, structured datasets were available for RAG (Retrieval-Augmented Generation) implementations.
- Product Integration: Bridged the gap between data engineering and product strategy, rapidly prototyping and deploying data features that enabled dynamic content generation for users.

**Senior Data Engineer (Azure Stack)****Jun 2021 – Jan 2024****Infosys Limited – Chennai, India****Responsibilities:**

- Cloud Migration (NTT Data): Led the end-to-end migration of ~4TB of mission-critical data from SQL Azure VMs to Managed Instances, ensuring zero data loss and 99.9% uptime during complex cutovers.
- ETL & Compliance (TalkTalk): Managed high-volume ETL/ELT pipelines using Azure Data Factory for telecom data, implementing row-level security and masking to ensure strict GDPR compliance for sensitive PII.
- Performance Tuning: Optimized T-SQL stored procedures and Apache Spark/Databricks jobs, utilizing partitioning and broadcast variables to reduce execution time by 40% and lower compute costs.
- DevOps & CI/CD: Built unified CI/CD pipelines using Azure DevOps, GitLab, and Bamboo, reducing release cycles by ~40% through "Infrastructure as Code" (IaC) and automated regression testing.
- Legacy Modernization: Spearheaded the transition of monolithic PHP stacks and on-premise servers to cloud-native Azure PaaS architectures, maintaining over 99% data integrity during infrastructure consolidation.
- Cost & Quality: Developed Python/SQL-based Data Quality Frameworks (DQC) to cut discrepancy resolution time by 50% and implemented auto-scaling strategies achieving \$50K+ in operational savings.
- Data Modeling & Architecture: Translated complex telecom business requirements into scalable Star Schema and dimensional data models, optimizing storage efficiency and query logic for downstream reporting.
- Data Security Governance: Implemented Row-Level Security (RLS) and Dynamic Data Masking within ETL workflows, ensuring granular access control and protection of sensitive PII across multi-tenant environments.
- Release Stability: Engineered automated rollback mechanisms and integrated static code analysis into deployment pipelines, preventing production failures and ensuring consistent release quality during agile sprints.
- Analytics Support: Enabled real-time business intelligence by building optimized data marts that supported high-performance Power BI and Tableau reporting for C-level executives.
- Leadership & Awards: Recognized with multiple Infosys INSTA Awards (2022, 2023) for delivery excellence and mentored junior engineers on Azure best practices and SQL performance tuning.
- SEO & Data Integrity: Executed critical SEO domain migrations and URL redirection strategies during platform upgrades, preserving search engine rankings and minimizing service disruption.

**Data Engineer**  
**Wipro Limited – Chennai, India**  
**Responsibilities:**

**Feb 2020 – May 2021**

- **Python ETL Development:** Developed and deployed efficient Python-based ETL scripts (using Pandas/NumPy) for large-scale data validation, automating manual entry processes and reducing error rates.
- **Automated Quality Assurance:** Built and executed rigorous Data Quality Checks (DQC) and automated testing scripts for web application modules, ensuring high fidelity of production data.
- **Performance Tuning:** Analyzed and optimized backend data operations, achieving measurable improvements in API response times and application latency benchmarks.
- **Backend Operations:** Completed intensive training in backend data architecture and immediately applied best practices to production pipelines, ensuring data integrity across validation modules.

**Education:** Master's degree in data science, University of Massachusetts Dartmouth

**Certification:** Certified Scrum Master, Microsoft Certified: Fabric Data Engineer Associate (DP-700)