

MANIKANDAN MUTHIAH

Senior Data Engineer

✉ pcbzmani@gmail.com

☎ +91 9500327540 / +974 51173177

📍 Doha, Qatar

🌐 [LinkedIn](#)

🌐 [Portfolio](#)

PROFESSIONAL SUMMARY

Senior Data Engineer with 9.5+ years of experience in architecting and scaling high-performance data platforms across Azure (Databricks, ADF) and GCP (BigQuery, Cloud Composer). Specialist in transitioning legacy environments to modern cloud-native architectures, with expertise in PySpark, Delta Lake, and performance tuning (Z-Ordering, Partitioning). Proven track record of integrating cutting-edge GenAI solutions, including RAG pipelines and LLM-powered insights, to drive business value in the Healthcare, Aviation, and Retail sectors. Expert in bridging complex data engineering with business intelligence, leveraging CI/CD (Terraform, Azure DevOps) to ensure 99.9% reliability.

EDUCATION

Degree — Bachelor of Engineering in Electrical and Electronics, National Engineering College — 91%

Aug 2012 - Apr 2016

GPA: 9.13

SKILLS

Languages: PySpark, Python, SQL (PostgreSQL, MS SQL, Oracle, DB2), COBOL (Legacy Migration)

Azure Stack: Azure Data Factory (ADF), Databricks, Synapse, Delta Lake, Azure DevOps, Terraform

GCP Stack: BigQuery, Dataproc, Cloud Composer (Airflow), Vertex AI, GCP Agents

Data Engineering: Medallion Architecture, ETL/ELT, Data Modeling, Z-Ordering, Partitioning, Data Skew Management

AI/ML: GenAI, Retrieval-Augmented Generation (RAG), LLMs, Model Context Protocol (MCP), Agentic AI

Orchestration & CI/CD: Airflow, Jenkins, Bitbucket, Git, Terraform (IaC)

EXPERIENCE

Senior Data Engineer / Senior Software Engineer — EMAKINA (EPAM Systems)

Sep 2023 – Present

- Architected a multi-cloud GCP data ingestion framework, automating the migration of 100TB+ legacy on-premise datasets to BigQuery to support global aviation retail operations.
- Pioneered GenAI integration by engineering Retrieval-Augmented Generation (RAG) pipelines and LLM-powered features, reducing internal operational team response times by 40%.
- Modernized BigQuery infrastructure using advanced clustering, partitioning strategies, and materialized views, resulting in a 35% reduction in job latency and a 25% decrease in cloud compute spend.
- Orchestrated complex, production-grade workflows using Cloud Composer (Airflow), achieving 99.9% pipeline uptime and ensuring strict data governance across multi-stage ETL processes.

Senior Data Engineer / Senior Software Engineer — EPAM Systems

Apr 2022 – Sep 2023

- Deployed highly scalable cloud-native ingestion frameworks on Azure, enabling real-time global inventory analytics for high-velocity data streams.
- Architected Medallion (Bronze/Silver/Gold) data structures in Azure Databricks to power 'Flight 360' KPIs and predictive booking models, enhancing downstream data science accuracy.
- Optimized Delta Lake performance via Z-Ordering, data skipping, and file compaction, delivering a 50% query speed improvement for enterprise-level analytics.

Application Development Senior Analyst (Data Engineer) — Accenture International

Apr 2021 – Apr 2022

- Engineered automated ETL/ELT pipelines in Azure Data Factory (ADF) to centralize disparate retail and distribution data from PostgreSQL, Oracle, and MS SQL into a unified Data Lake.
- Debugged and resolved critical Spark data skew issues by implementing broadcast joins and custom salting techniques, maximizing cluster resource utilization.
- Standardized CI/CD pipelines using Jenkins and Bitbucket, accelerating product deployment cycles by 20% while maintaining rigorous code quality standards.

IT Analyst / Data Engineer — Tata Consultancy Services (TCS)

Jul 2016 – Apr 2021

- Modernized legacy healthcare data systems by refactoring COBOL logic into optimized PySpark and Hive workflows, reducing processing windows by 60%.
- Implemented HDFS storage strategies, including partitioning and bucketing, to ensure high-performance data migration to AWS Redshift ecosystems.
- Delivered high-impact data solutions for Mortgage Insurance and Healthcare portfolios within an Agile/Scrum framework, consistently exceeding sprint velocity targets.

PROJECTS

Duty Free Business Intelligence (Aviation)

- Developed a real-time duty-free forecasting engine in BigQuery, integrating high-volume luxury retail and travel data for actionable aviation insights.
- Built Model Context Protocol (MCP) servers and RAG pipelines to securely connect AI agents with internal aviation datasets, enhancing data accessibility.
- Engineered PowerBI reporting layers to optimize food waste management and inventory turnover for stakeholders.

Enterprise Data Factory (Retail & Distribution)

- Designed a cross-region Data Factory solution using Databricks and Delta Lake to manage multi-terabyte retail inventory datasets.
- Optimized Spark executors and memory configurations to eliminate processing bottlenecks during peak load periods.

Healthcare Data Modernization

- Led a complex migration of IBM DB2 legacy data to a modern HDFS/Hive architecture.
- Enhanced query response times by 40% through Hive bucketing and Spark DataFrame optimization techniques.

CERTIFICATIONS

- Certified in Azure Data Engineer (DP-203)
- Certified in Databricks — Data Engineer Associate and Data Analyst
- Certified in Google Professional Data Engineer Associate
- Certified Partner associate Gemini Enterprise deployment

EXTRACURRICULAR ACTIVITIES

- Published a app in Playstore - PanamKasu personal finance app

LANGUAGES

English — Fluent

ADDITIONAL INFORMATION

Nationality: Indian