

# PRAVEEN PANDEY

☎ +1 (857) 544-9465 | ✉ praveenpdy88@gmail.com | 🔗 linkedin.com/in/praveenpdy | 🏠 South Plainfield, NJ

## PROFESSIONAL SUMMARY

Data Engineer with 3+ years of experience in **cloud-native, data-driven solutions** across Azure, AWS, and GCP. Proven record of **leading end-to-end implementations**, reducing ETL runtimes by **85%**, and delivering **real-time analytics pipelines** for faster executive decision-making. Strong client-facing consultant skilled in **stakeholder engagement, project management, and risk mitigation**, with certifications in Microsoft Azure Data Engineer and Databricks.

## TECHNICAL SKILLS

**Programming & Scripting:** Python (4+ years), SQL, PySpark, Scala

**Cloud & Big Data:** Azure (Data Factory, Databricks, Synapse Analytics, ADLS), AWS, GCP, Apache Spark, Kafka, Event Hubs, Delta Lake, Flink, Hadoop

**Databases & Warehousing:** Snowflake, Azure Synapse, SQL Server, Oracle, Teradata, Delta Lake, Neo4j, Cosmos DB

**Orchestration & ETL:** Apache Airflow, Azure Data Factory, dbt

**CI/CD & DevOps:** Jenkins, Git, Docker, Kubernetes, Azure DevOps, Bitbucket

**Data Processing:** Pandas, NumPy

**Visualization & BI:** Power BI, Tableau

**Project Management:** JIRA, GitHub, Jupyter Notebooks

## PROFESSIONAL EXPERIENCE

### Data Engineer

Mar 2022 – Present

*Tata Consultancy Services / Kenvue (Former Johnson & Johnson Inc)*

*Skillman, NJ*

- **Architected high-throughput ETL pipelines** using Azure Data Factory and Databricks Workflows, orchestrating daily ingestion and real-time transformation of **500M+ records from 20+ sources**, enabling reliable, on-demand analytics at enterprise scale.
- **Designed Databricks Lakehouse architectures** using medallion (bronze, silver, gold) frameworks, leveraging Databricks SQL, Spark Structured Streaming, and Delta Live Tables for scalable analytics.
- **Architected & optimized semantic layer loads** (35+ mins → 4 mins), improving query efficiency by **25%** across SQL Server, Oracle, and Teradata through advanced performance tuning and indexing strategies.
- **Improved processing efficiency by 30%** through advanced PySpark and Spark SQL solutions, including dynamic XML parsing, deep-nested JSON extraction, and parallelized API-based data ingestion, establishing new standards for scalable data engineering.
- **Reduced batch data latency by 40%** by profiling and optimizing legacy scripts, automating workflow orchestration, and streamlining data pipelines, accelerating downstream processing and improving daily reporting timelines.
- **Implemented enterprise-grade data governance**, compliance controls, data lineage, and risk mitigation strategies using **Databricks Unity Catalog and cloud-native security frameworks**.
- **Migrated 5,000+ Parquet tables to Delta Lake format** by designing and executing a scalable framework in collaboration with senior engineers, unlocking performance gains, schema evolution support, and advanced data governance for multi-tenant data lakes.
- **Led cross-functional SAP S/4HANA and SAP ECC integration**, spearheading technical enhancements to silver and gold data layers through dimensional modeling to ensure KPI continuity, regulatory compliance, and scalable lineage tracking across hybrid ecosystems.
- **Drove architectural decision-making** by leading a Neo4j vs. Cosmos DB proof-of-concept, delivering a scalable graph database solution that enhanced supply chain insights and stakeholder visibility.
- **Enabled continuous agile deployments** by integrating CI/CD tools and Jenkins automation, supporting frequent updates, rapid rollback capabilities, and urgent off-cycle releases while maintaining **99.9% system reliability**.

### Research Assistant

Dec 2019 – Dec 2021

*Northeastern University*

*Boston, MA*

- Developed **ETL workflows for healthcare datasets**, accelerating reporting and improving decision-making speed by **25%**.
- Built **Python-based workflows** for multi-source time-series data processing, ensuring data quality and compliance with healthcare standards.
- Designed **SQL queries and BI dashboards (Power BI, Tableau)**, cutting analysis time by **30%** for clinical researchers.

## EDUCATION

### Northeastern University

Boston, MA

*Master of Professional Studies in Artificial Intelligence*

*Sep 2019 – Dec 2021*

## CERTIFICATIONS

- |  |          |
|--|----------|
| • Deep Learning: Image Recognition             | Jan 2021 |
| • Neo4j Certified Professional                 | Aug 2023 |
| • Microsoft Azure AI Fundamentals (AI-900)     | Jan 2024 |
| • Microsoft Fabric Analytics Engineer (DP-600) | Sep 2025 |