NAVEEN SRIVATHSAV

DATA ENGINEER | DATABRICKS CERTIFIED
Chennai, India | 9840249713 | nsrivathsav@gmail.com

PROFESSIONAL SUMMARY

Data Engineer with 4+ years of experience specializing in robust ETL/ELT pipelines, data warehousing and BI reporting. Currently a **freelance Data Engineer delivering Azure Databricks and Azure Data Factory pipelines for an enterprise data platform**; previously a Data Analyst at Stats Perform. Certified in Databricks and Power BI, with advanced expertise in PySpark, Delta Lake, Microsoft Azure and SQL. Proven track record of designing medallion-architecture data lakes and automated ingestion workflows that turn complex operational data into actionable business insight. BCA (Data Science) graduate, currently pursuing an MBA in Data Science and AI.

CORE TECHNICAL SKILLS

Data Engineering & Big Data: Databricks, PySpark, Apache Spark, Delta Lake, Unity Catalog, Microsoft Azure, Azure Data Factory, dbt, ETL / ELT Pipelines, Data Lakes, Medallion Architecture

Programming & Databases: Advanced SQL, Python, PostgreSQL, Oracle, REST APIs, Git, Azure DevOps

Data Ingestion & Integration: Metadata-driven pipelines, incremental / delta loads, parquet landing zones, relational database and API sources, job scheduling and run logging

BI & Analytics: Power BI, Tableau, DAX, KPI Dashboards, Automated Reporting, Performance Tuning

Warehousing & Quality: Dimensional Modeling, Star Schema, Source-to-Target Mapping, Data Validation, dbt Tests, Data Quality

PROFESSIONAL EXPERIENCE

Data Engineer (Freelance) | Remote **Jul 2026 - Present**

- Build and maintain **Azure Databricks (PySpark, Delta Lake, SQL) pipelines** that move sales-order, billing and delivery data through **Bronze, Silver and Gold medallion layers** as part of an enterprise data platform migration.
- Orchestrate ingestion with **Azure Data Factory**: metadata-driven pipelines, incremental delta-load extraction, parquet landing zones and scheduled Databricks jobs with run logging and data-quality audit tables.
- Validate Databricks gold views against source-system reports, comparing records and measures row by row and tracing each mismatch to its root cause.
- Log, triage and document defects in **Azure DevOps** with reproducible SQL evidence; write fix proposals that include backups and rollback scripts before any live view or table is changed.
- Design consumption views in Unity Catalog for BI and reporting users, including a combined history-plus-delta delivery view that unifies two source layouts.
- Work in a config-driven pipeline framework (YAML definitions, code generation and automated checks) under Git and Azure DevOps, and write pipeline architecture and run-history documentation used for team handover.

Python Intern | DLK Software Solutions **Sep 2025 - Nov 2025**

- Completed a three-month Python internship alongside final-year degree studies.

Data Analyst | Stats Perform, Chennai, India **Mar 2020 - Feb 2024**

- Architected and optimized end-to-end ETL/ELT pipelines and data integration workflows using SQL, PySpark and automated scripts, reducing manual processing time by **25%**.
- Processed, transformed and validated multi-sport operational datasets into analytics-ready tables with **98%+ accuracy**.
- Built Power BI and Tableau dashboards for KPIs such as pass completion rate, shot accuracy and reporting SLA performance.
- Developed automated reporting data layers with SQL and data pipelines, cutting recurring extraction effort by **30%**.
- Collaborated with global data engineering, QA and software teams to troubleshoot database feed bottlenecks, tune queries and maintain strict data processing SLAs.

PROJECTS

- **Databricks Medallion Lakehouse Pipeline:** Designed data pipelines with Databricks, PySpark and Microsoft Azure that take large datasets through Bronze, Silver and Gold layers.
- **dbt & SQL Data Transformation Layer:** Built modular dbt models with automated tests, schema validations and referential-integrity checks for clean downstream consumption.
- **Electricity Bill Prediction Pipeline:** ML pipeline with feature engineering, missing-data handling and SVR / Random Forest models.
- **Sports Performance Analytics & Power BI Platform:** Scalable SQL models and Power BI dashboards with custom DAX measures for real-time, self-service insights.

EDUCATION & CERTIFICATIONS

MBA, Data Science and AI | SRM University **Expected 2027**

BCA (Data Science), First Class, CGPA 8.22 | SRM Institute of Science and Technology **2023 - 2026**

Coursework: Programming for Analytics, Data Mining & Warehousing, NoSQL, Deep Learning, Cloud Application Development.

Certifications: Databricks Associate | Microsoft Power BI Associate