

# PRUDHVISAI REDDY CHADUVULA

Little Elm, TX | +1 (940) 843-9399 | [prudhvisaireddy1243@gmail.com](mailto:prudhvisaireddy1243@gmail.com) | [LinkedIn](#) | [Portfolio](#)

---

## PROFESSIONAL SUMMARY

Data Engineer with 3+ years of experience building batch data pipelines, SQL transformation layers, and analytics-ready datasets in cloud environments. Proficient in Azure Data Factory, Azure Databricks, Azure Synapse Analytics, Python, PySpark, SQL, and Delta Lake to ingest and transform structured data from APIs, files, and databases. Skilled in incremental data processing, data quality validation, and delivering curated fact and dimension tables that improve data reliability. Collaborative team member supporting SQL-based reporting, analysis, and analytics workflows in U.S. enterprise environments.

---

## TECHNICAL SKILLS

**Data Engineering:** Data Pipeline Development, ETL Processing, Batch Data Ingestion, Data Transformation, Basic Data Modeling, Analytics-Ready Dataset Preparation, Data Quality Checks, Incremental Data Loads.

**Cloud Data Platforms:** Azure Data Factory, Azure Databricks, Azure Synapse Analytics, Delta Lake, Unity Catalog.

**Programming and Data Processing:** Python, SQL, PySpark.

**Databases and Data Stores:** Azure Synapse SQL, SQL Server, PostgreSQL, MySQL, IBM DB2.

**Data Integration and Formats:** REST APIs, JSON, CSV, Parquet, Cloud Storage Integration.

**Analytics Enablement:** SQL Transformations, Fact and Dimension Tables, Curated Data Layers for Reporting.

**Data Operations:** Pipeline Monitoring, Logging, Error Handling, Performance Tuning (Basic).

**Operating Systems and Scripting:** Linux, Shell Scripting.

---

## PROFESSIONAL EXPERIENCE

### Data Engineer

July 2025 - Present

#### DA2Technologies LLC

- Engineered batch ETL pipelines using Azure Data Factory and Python to ingest structured data from REST APIs and cloud storage, reducing manual data preparation effort for analytics teams by 40%.
- Developed PySpark transformation workflows in Azure Databricks to cleanse, normalize, and standardize raw JSON and CSV datasets into curated Delta tables for reliable analytical consumption.
- Modeled fact and dimension tables in Azure Synapse SQL with optimized joins and schemas, enabling faster analytical queries and lowering average report query time by 30%.
- Implemented incremental data load logic with Delta Lake to process only new and updated records, cutting daily pipeline runtimes by 25% while maintaining historical accuracy.
- Validated data quality through schema checks, null thresholds, and record-count comparisons within Databricks transformation layers, supporting governed access and metadata visibility using Unity Catalog and decreasing downstream data defects by 35%.
- Integrated multi-format datasets including JSON, CSV, and Parquet into a unified data model, improving consistency and usability of data consumed across reporting and analysis workflows.
- Monitored pipeline executions and failure logs across Azure Data Factory and Databricks, collaborating with cross-functional teams to resolve issues and maintain 99% successful pipeline completion rates.
- Partnered with analysts and business stakeholders to translate reporting requirements into SQL-based transformations and analytics-ready datasets aligned with ongoing business needs.

### Data Integration Engineer

Feb 2021 - July 2023

#### Tecnic's Integration Technologies Pvt Ltd

- Automated data-dependent operational tasks using shell scripting on Linux systems, improving reliability of scheduled data jobs and reducing manual intervention by 20% across production environments.
- Supported SQL Server and IBM DB2 databases by validating data availability, integrity, and consistency for application workflows and internal reporting needs used by multiple teams.
- Analyzed data discrepancies through targeted SQL queries to identify record-level issues, collaborating with engineering teams to reduce investigation and resolution cycles by 30%.
- Optimized frequently executed SQL queries by reviewing execution plans and access patterns, contributing to more stable data retrieval during peak operational workloads.
- Investigated data flow issues between application layers and backend databases in partnership with cross-functional teams, improving accuracy and completeness of production data outputs.

- Assisted with performance tuning efforts by reviewing database behavior and system logs, helping stabilize data processing during high-volume transaction periods.
- Documented SQL queries, data validation steps, and operational procedures to support long-term maintainability and smoother knowledge transfer for data-related workflows.
- Coordinated with developers and QA teams to validate data changes during releases, ensuring downstream data consumers and reporting processes were not impacted by production updates.

---

## PROJECTS

### Data Operations and Reporting Support

- Consolidated operational data from multiple SQL Server sources using structured SQL queries and joins, improving dataset consistency by 35% and reducing reporting discrepancies across teams.
- Transformed raw transactional records into structured reporting tables through SQL-based filtering and aggregation logic, cutting manual reconciliation effort for business teams by 30%.
- Validated dataset completeness through record-count checks and targeted query analysis, increasing accuracy of recurring reports and audit datasets by 25%.

### Data Processing and Quality Enhancement

- Processed export-related datasets by applying SQL transformations to standardize records across upstream systems, improving consistency of export tracking data by 40%.
- Implemented data validation logic using SQL queries to identify missing values and format issues, reducing data-related errors in export analytics outputs by 28%.
- Supported analytics consumption by preparing cleaned and structured tables, decreasing additional data preparation time for analysis teams by 32%.

### Data Reliability and Transformation Project

- Extracted structured operational data from backend databases using optimized SQL queries, improving data availability for internal analysis and reporting workflows by 30%.
- Refined datasets through SQL-based filtering, aggregation, and deduplication logic, enhancing usability and clarity of shared data assets by 27%.
- Assisted in maintaining data reliability by reviewing query outputs and collaborating with stakeholders to resolve discrepancies, lowering recurring data issues by 22%.

---

## EDUCATION

### Masters in Computer and Information Sciences

University of North Texas | TX, USA

Aug 2023 - May 2025

### Bachelor of Technology in Computer Science and Engineering

Gitam University | India

Jun 2016 - Jul 2020

---

## CERTIFICATIONS

- AWS Certified Cloud Practitioner - **AWS**
- Microsoft Certified Azure Fundamentals - **Microsoft**
- Google Data Engineering Professional Certificate - **Coursera**
- IBM Data Engineering Professional Certificate - **Coursera**
- Databricks Lakehouse Fundamentals - **Coursera**
- SQL for Data Science - **Coursera**