NISANTH TUMU

Data Engineer/ Cloud Data Engineer
Peterborough, ON |Canada
+1 (249) 688 2395| nisanthtumu9@gmail.com

Website
GitHub
LinkedIn

Skills

- Microsoft Azure (ADLSGen2, Data Factory, Databricks, Synapse Analytics)
- Databricks (Auto loader, Delta Lake, Unity Catalog, Declarative Pipelines AKA Delta Live Tables)
- Microsoft Power BI, Tableau and Excel
- Data Warehousing - Data Modeling (Star/Snowflake Schema), Data Governance and SCD Types 1 and 2.
- Data Build Tool (DBT), Version Control (Git)
- Python (Pandas, NumPy, SciPy, Matplotlib, OOPs)
- PySpark, Spark Batch and Stream processing.

Work Experience

➢ **Senior Data Engineer –** Remote: Confidential Client – Ontario, Canada **Feb 2025 – Present**

- Built foundational **data pipelines** using **Azure Data Factory**, **Databricks Autoloader**, and **Structured Streaming**, enabling both batch and near real-time ingestion and monitored alerts and notifications with **Azure Logic Apps.**
- Led the initial rollout of **Delta Live Table workflows** using **Medallion architecture**, implementing **SCD Types 1 and 2**, schema evolution, and time-travel-enabled audit tracking.
- Piloted the adoption of **Unity Catalog** for centralized access control and governance across multi-workspace environments, improving compliance and cross-team collaboration.
- Developed modular, parameterized pipelines with **integrated Databricks Workflows**, enabling seamless CI/CD deployments and reducing release cycle times by 43% using **Azure DevOps.**
- Modeled curated datasets in **Azure Synapse** exposing unified semantic layers to **analytical teams** for executive and operational reporting.
- Partnered with business analysts to redesign and modernize legacy **SQL-based Power BI reports**, tuned performance using **Z-Ordering, partition pruning, and caching strategies** to reduce data latency and boost refresh speeds across streaming dashboards.

➢ **Data Engineer –** Infosys PVT Limited – Hyderabad, India **Aug 2021 – Aug 2023**

- Designed and deployed scalable ETL pipelines using **Azure Data Factory** and **ADLS Gen2**, with dynamic parameterization and triggers to automate data integration from multiple sources (csv, Json, xml, https, and Rest API).
- **Migrated the on-premise SQL Server data** into **Azure Data Lake**, using **Self-hosted Integration Runtime**, incremental loads with watermarking, and schema mapping techniques to ensure secure and reliable cloud integration.
- Improved data quality and pipeline resiliency through structured logging, robust error handling, and modular control flows using ADF activities like **CopyData**, **ForEach**, **If Condition**, **GetMetaData** and **Lookup**.
- Developed end-to-end transformation logic in **Azure Databricks (PySpark)** using **Delta Lake**, implementing **SCD Types**, schema evolution, and optimized partitioning within **multi-layered architecture**.
- Created and managed **dimension and fact tables** in **Azure Synapse Analytics** using **CETAS**, stored procedures, and **star/snowflake schema**, enabling efficient Power BI reporting for CRM analytical down streams.

Key Projects

➢ **End-to-End Azure Data Engineering Projects – E-Commerce & Netflix Data** **Source Codes: (e-Com) – (Netflix)**

- Architected **Azure-based data pipelines** with the **Medallion Architecture** using **ADF** and **Databricks**.
- Ingested data from sources - **GitHub** and **MongoDB**, handling **batch and real-time** using **Autoloader** and **Spark Streaming**.
- Configured **Azure Data Lake Storage Gen2 (ADLS)** for raw and processed data, ensuring scalable and secure storage.
- Utilized **Databricks Notebooks** and **PySpark** for data cleansing, transformation, and enrichment.
- Applied **Databricks Utilities** for dynamic parameterization and **Workflows** for orchestration of multi-stage pipelines.
- Integrated **Databricks Delta Live Tables (DLT)** for declarative data transformations and incremental processing.
- Leveraged **Unity Catalog** for fine-grained access control and unified governance across data assets.
- Implemented **SQL-based data modeling** and transformations for analytics-ready tables in **Azure Synapse Analytics**.

➢ **Databricks Declarative pipelines (DLT)** (Source-Code)
➢ **Azure Databricks Project** (Source-Code)
➢ **Azure Data Factory Project** (Source-Code)
➢ **Azure Data Warehousing Project** (Source-Code)
➢ **Basic Azure Data Engineering Project** (Source-Code)

Education

MASTER OF SCIENCE– Trent University – Peterborough, Canada. **Sep 2023 – Dec 2024**
Specialization: **Big Data Engineering** (GPA: 3.8/4)