NISANTH TUMU

Data Engineer/Cloud Data Engineer

+1 (249) 688 2395| nisanthtumu9@gmail.com
Peterborough, ON |Canada
https://www.linkedin.com/in/nisanth-t-486a4b192/

Skills

- Microsoft Azure Data Engineering (ADLSGen2, Data Factory, Databricks, Synapse Analytics)
- Microsoft Power BI, Tableau
- SQL (SQL Server, MySQL)
- Data Build Tool (DBT)
- Microsoft Fabric (One Lake, Data Factory, DataFlowGen2, Data Engineering, Data Warehousing).
- Apache PySpark
- Python (Pandas, NumPy, SciPy, Matplotlib)

Work Experience

Data Engineer – Infosys PVT Limited – Hyderabad, India

- Developed and maintained data pipelines to support CRM analytics by integrating structured and semi-structured data.
- Automated and optimized ETL processes using SQL and Python, improving data processing efficiency significantly.
- Enhanced accuracy of data migration using rule-based validations and logging mechanisms ensuring 99% error free transfers.
- Contributed and supported reporting needs by preparing clean and structured datasets for downstream consumption.

Projects

Azure Data Engineering Project With E-Com Data

• Setting up Azure Data Factory • Configuring ADLS Gen2 storage • Understanding Medallion Architecture • Implementing data ingestion • HTTP endpoints (GitHub) • SQL Tables • Creating your first data pipeline • Real-world implementation scenarios • Performance optimization techniques • Azure Databricks implementation • Data transformation techniques • MongoDB integration for data enrichment • Synapse Analytics setup • Visualization with Power BI/Tableau/Fabric • Performance optimization.

Azure Data Engineering Project with Netflix Data

• Azure Data Architecture • Data Understanding • Azure Data Fundamentals • Azure Data Factory • ETL Pipelines with ADF • ADF Real Time Scenarios • Databricks Unity Catalog • Databricks Spark Cluster • Incremental Loading with Autoloader • Spark Streaming • Data Ingestion using PySpark • Parameters using Databricks Utilities • Data Orchestration with Databricks Workflows • Data Transformation using PySpark • Big Data Analytics with Apache Spark • Databricks Delta Live Tables • End to End Pipeline in Databricks.

Azure Databricks Project

• Data Understanding • Creating Azure Resources • Databricks Overview • Unity Catalog • Data Ingestion• Autoloader with parquet files • Spark Structured Streaming • ETL Jobs • PySpark Functions • Python OOP with PySpark • PySpark Advanced Functions • Slowly Changing Dimension • Delta Live Tables • Star Schema • Databricks End-To-End Pipeline.

Azure Data Factory Project

• Create Azure resources• Azure Data Lake (ADLSGen2) • Create ADF Workspace • Dataset and Linked Service • Data Ingestion in Data Lake • Copy Activity • Copy Data Using REST API • Get Metadata Activity • IF Condition • For Each Activity • Expression Builder • Parameterized Pipeline • Data Flow • Data Transformation • Schedule Trigger • Set Variable Activity • Storage Event Trigger • Real Time Scenarios • Delete Activity • Debugging Pipeline • Execute Pipeline Activity • End-To-End Data Pipeline.

Azure Data Warehousing Project

• Creating Synapse Analytics Resource • ETL Pipelines• Incremental Data Loading • Data Transformation using PySpark Data Flows • Serverless SQL Datawarehouse• External Data Sources• Openrowset() Function • External Tables in Synapse • Schema Handling • End-To-End Data Pipeline • Dimensional Data Modeling • STAR schema vs SNOWFLAKE schema • Creating Dimension Tables using CETAS • Creating Fact Table • STAR schema • Slowly Changing Dimensions.

Basic Azure Data Engineering Project

• Data Understanding (API) • Creating Azure Resources • Azure Data Lake Gen2 • Data Ingestion (Bronze Layer) • ETL Pipelines with ADF • Real-Time Scenarios with ADF with parameterization • Databricks Cluster & Overview • Service Principle • Data Transformation • Apache Spark (Silver Layer) • Pyspark for advanced transformations. • Big Data Analytics with Pyspark • Azure Synapse Analytics (Gold Layer) • Openrowset () function • External Tables in Synapse • Integrate Data Warehouse with PBI.

Education

MASTER OF SCIENCE IN BIG DATA ANALYTICS – Trent University – Peterborough, Canada.
Majors: Big Data Analytics (GPA: 3.8/4)