

## Data Engineer

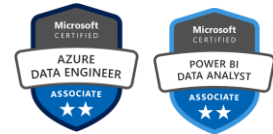
Sri Thanmaie Nuthalapati

[nthanmaie23@gmail.com](mailto:nthanmaie23@gmail.com)

9373563611

### PROFESSIONAL SUMMARY

- Having 5+ years of experience in MSBI Tool(SSIS), MS-SQL Server, Azure cloud technologies such as Azure Data Factory, Azure SQL, Azure data lake storage Gen1 & Gen2, Azure data studio, Azure data explorer, Azure Synapse analytics, Azure Blob Storage and Azure data Bricks.
- Working in Microsoft Azure Data Factory, Azure SQL, and Azure.
- Design and deploy customized monitoring solutions to monitor infrastructure, Operating system, database, and applications.
- Designing solutions in Azure Data Factory (ADF) using ADLS & Data Bricks.
- Building complex SQL Queries, Joins, Constraints, Stored procedures, DML to implement the business logic.
- Designed and developed end-to-end data integration solutions using Azure Data Factory, orchestrating data movement and transformation workflows across on-premises and cloud-based data sources.
- Designed and implemented data integration solutions using Azure Logic Apps to connect various data sources and destinations, including databases, APIs, and cloud services.
- Having experience in deploying the code across the environment by using CI/CD.
- Having experience in creating notebooks in Azure Data Bricks.
- Configured error handling and retry mechanisms within Azure Logic Apps to ensure data integrity and reliability, minimizing downtime and data loss.
- Designed and implemented messaging solutions using Azure Service Bus to enable asynchronous communication and event-driven architectures between distributed components.
- Create Notebooks in Data bricks for loading data from source to target ADLS.
- Experience on unified data analytics with Databricks, Databricks workspace user interface, managing Databricks Notebooks, Delta Lake with python, Delta Lake with Spark SQL.
- Worked on Databricks environment and used spark structured streaming to consume the data and perform merge operations on delta lake tables.
- Developed spark applications Pyspark and spark SQL for data extraction, transformation and aggregation from various.
- Worked with spark core, Pyspark and spark SQL modules of Spark.
- Involved in converting SQL queries into Spark transformations using APIs like Spark SQL, Data Frames and python.
- Strong exposure on Logic apps to send Email notifications, write blob to store, generate excel outputs, to get log files etc. based on pipeline events.
- Written Unit Test cases for Web API's and Azure functions.
- Developed Azure function apps as API services to communicate DB.
- Transform data from various data sources like Excel, and SQL Server Database using OLEDB connection by creating various SSIS Packages.
- Design packages Using Control Flow Tasks like For Loop Container, For Each Loop Container, Sequential Container, Execute SQL Task, and Data Flow Task.
- Extract data from sources and transformed the data using different transformations like data conversion, derived columns, look up, Conditional Split, Aggregate, Union all, merge join and multi cast transformations.
- Extensive knowledge Python Data manipulation libraries Pandas, NumPy etc.
- Analyzed the SQL scripts and designed the solution to Developed a custom ETL pipeline by using Python.
- Analyze data using python libraries Pandas, NumPy etc.
- Experience with Snowflake utilities, SnowSQL, SnowPipe, Big Data model techniques using Python.
- Worked on complex SQL queries on Snowflake.
- Experience and understanding of architecting, designing and operationalization of large-scale data and analytics solutions on snowflake Cloud Data Warehouse.



- Experience in building and migrating large scale data platforms using Snowflake
- Involved in migration of on-premises history data to Snowflake cloud data warehouse.
- Created Snowflake stage, file formats, schema's, sequences, views, tables.
- Created internal /external stage and transformed data during load Worked on Data Sharing in Snowflake, Database, Schema and Table structures, Automatic Clustering.
- Having good exposure on Architecture of Azure synapse warehouse.
- Expressions and Tabular Model. Skills in Power BI visualization, high quality Power BI reports and Dashboards.
- Experience in developing Power BI reports and dashboards from various Data Sources like databases, Excel files, flat files.
- Used Power BI to prepare dashboards to create attractive and interactive visualizations.
- Created reports in Power BI preview portal utilizing the SSAS Tabular using Power BI Gateway.

## EDUCATION SUMMARY

Master's in Computer Science from University of Dayton

## TECHNICAL SKILLS

Languages	SQL, T-SQL, Python, Pyspark
Databases	MS SQL Server, Azure SQL, Snowflake
ETL Tools	SQL Server Integration Services (SSIS)
Reporting Tools	Power-Bi, QlikView
Operating System	Windows, Linux
Microsoft Azure Technologies	Azure SQL Server, Azure Data Factory, Azure Data Lake Storage, Azure Blob, Azure Data Bricks, Azure Synapse Analytics, Azure data studio, Azure data explorer

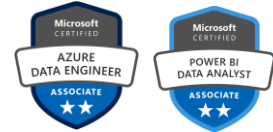
## PROJECT SUMMARY

Crowe, Chicago, IL  
Sr. Data Engineer

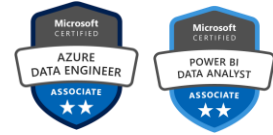
Jan 2021 – Present

### Responsibilities:

- Create Azure Data Factory pipelines to copy the data from source to target systems.
- Migrating the data from sources to the destination with the help of ADF.
- Create Azure Data Factory pipelines with extensive usage of activities like Copy, Metadata, For Each, If, Stored Proc, Mapping Data Flow etc., and Transformations in Data flow like Lookup, Derived column, Select, Aggregate, Union all, Join, Alter row, Conditional Split etc.
- Create Linked services on the source and as well for the destination servers.
- Create automated workflows with help of triggers.
- Scheduled jobs in workflows and ADF pipelines.
- Involve in deployment activities and moved the code across different environments using CI/CD.
- Load the source data into Azure synapse warehouse tables by creating some external tables (polybase mechanism).
- Leveraged .NET Core for cross-platform development of data processing services and microservices, ensuring compatibility and scalability across different operating systems and environments.
- Integrated Azure Logic Apps with Azure Data Factory pipelines to orchestrate end-to-end data workflows, enabling streamlined data movement and processing at scale.



- Implemented monitoring and logging solutions for Azure Logic Apps using Azure Monitor and Azure Log Analytics, enabling real-time visibility into workflow performance and health.
- Connect to SharePoint and placed file onto BLOB using Logic Apps.
- Created and configured pipelines in Azure Data Factory to extract data from various sources, including Azure Blob Storage, Azure SQL Database, SQL Server, and REST APIs, ensuring seamless data ingestion.
- Integrated .NET applications with cloud-based data services such as Azure SQL Database, Azure Cosmos DB, and Azure Blob Storage, enabling seamless data interaction and synchronization.
- Written Unit Test cases for Web API's and Azure functions.
- Configured queues, topics, and subscriptions within Azure Service Bus to decouple application components and facilitate reliable message delivery with support for publish-subscribe messaging patterns.
- Implemented data transformation activities using Azure Data Factory's built-in transformations, such as mapping, filtering, and aggregating, to cleanse and prepare data for analytics and reporting purposes.
- Implemented message filtering, dead-lettering, and retry policies within Azure Service Bus to handle message processing errors and ensure message reliability and durability.
- Developed stateful and stateless services on Azure Service Fabric, leveraging reliable collections and actor models to manage distributed state and orchestrate complex workflows.
- Developed Azure function apps as API services to communicate DB.
- Create Notebooks in Data bricks for loading data from source to target ADLS.
- Monitoring daily running Azure pipelines for different applications and supporting multiple applications.
- Collaborated with cross-functional teams to gather requirements, design data integration solutions, and troubleshoot issues within Azure Logic Apps, ensuring alignment with business objectives.
- Development of Data Bricks Notebooks using SQL.
- Utilized Azure Service Fabric's rolling upgrades and auto-scaling capabilities to seamlessly update and scale services, minimizing downtime and maximizing resource utilization.
- Implementation of logics in Data Bricks as per client requirement.
- Experience on unified data analytics with Databricks, Databricks workspace user interface, managing Databricks Notebooks, Delta Lake with python, Delta Lake with Spark SQL.
- Worked on Databricks environment and used spark structured streaming to consume the data and perform merge operations on delta lake tables.
- Deployment of Azure Data bricks and ADF using Azure DevOps.
- Developed spark applications Pyspark and spark SQL for data extraction, transformation and aggregation from various.
- Stayed updated with the latest features and best practices in Azure Logic Apps, continuously optimizing data integration processes and enhancing overall system performance
- Worked with spark core, Pyspark and spark SQL modules of Spark.
- Involved in converting SQL queries into Spark transformations using APIs like Spark SQL, Data Frames and python.
- Create Cosmos scripts which pulls data from upstream structured streams and include business logic & transformations to meet the requirements.
- Create Indexes like Clustered/Non-Clustered Indexes to speed up the data retrieval.
- Written complex stored procedures and user defined functions using T-SQL.
- Worked on Azure Key vaults to use it in linked services.
- Create complex pipelines to extract the data from SQL server into different csv files.
- Create event-based triggers in Data factory to capture data from ADLS.
- Work on Azure SQL Synapse analytics.
- Creating the linked service for source and target connectivity Based on the requirement.
- Create highly complex SSIS packages using various Data transformations like Derived Column, Conditional Split, Lookup, Multi Cast and Copy Column.



- Develop the SSIS Packages to make the data available from various relational databases, Excel, Flat files etc. for analysis and reporting purposes.
- Utilize multiple transformations provided by SSIS such as Data Conversion, Conditional Split, Lookup, Merge, Union all etc.

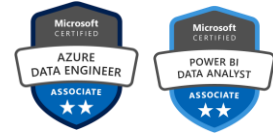
Environment: SQL, SSIS, Data Factory, Data Bricks, Data Lake Gen2, Blob, Python, NumPy, Pandas, DevOps, T-SQL, Synapse, Snowflake, Git, Jira, Agile.

Teradata, San Diego, CA  
Data Engineer

Aug 2019-Dec 2020

#### Responsibilities:

- Designed AWS architecture, Cloud migration, AWS EMR, DynamoDB, Redshift and event processing using lambda function
- Implemented usage of Amazon EMR for processing Big Data across a Hadoop Cluster of virtual servers on Amazon Elastic Compute Cloud (EC2) and Amazon Simple Storage Service (S3).
- Utilized AWS services with focus on big data analytics, enterprise data warehouse and business intelligence solutions to ensure optimal architecture, scalability, flexibility
- Experienced in Importing and exporting data into HDFS and Hive using Sqoop.
- Participated in development/implementation of Cloudera Hadoop environment.
- Experienced in running query-using Impala and used BI tools to run ad-hoc queries directly on Hadoop.
- Using Bash and Python included Boto3 to supplement automation provided by Ansible and Terraform for tasks such as encrypting EBS volumes backing AMIs.
- Involved in using Terraform migrate legacy and monolithic systems to Amazon Web Services.
- Wrote Lambda function code and set Cloud Watch Event as trigger with Cron job Expression.
- Validate Scoop jobs, Shell scripts & perform data validation to check if data is loaded correctly without any discrepancy. Perform migration and testing of static data and transaction data from one core system to another.
- Worked on creating and running Docker images with multiple micro – services and Docker container orchestration using ECS, ALB and lambda.
- Developed Spark scripts by writing custom RDDs in Scala for data transformations and perform actions on RDDs.
- Created Metric tables, End user views in Snowflake to feed data for Tableau refresh.
- Generated Custom SQL to verify the dependency for the daily, Weekly, Monthly jobs.
- Implemented Kafka producers create custom partitions, configured brokers, and implemented high level consumers to implement data platform.
- Developed best practice, processes, and standards for effectively carrying out data migration activities. Work across multiple functional projects to understand data usage and implications for data migration.
- Prepared data migration plans including migration risk, milestones, quality and business sign-off details.
- Performed advanced procedures like text analytics and processing, using the in-memory computing capabilities of Spark using Scala.
- Worked on migrating MapReduce programs into Spark transformations using Scala.
- Developed spark code and spark-SQL/streaming for faster testing and processing of data.
- Wrote Python modules to extract data from the MySQL source database.
- Worked on Cloudera distribution and deployed on AWS EC2 Instances.
- Deployed the project on Amazon EMR with S3 connectivity for setting a backup storage.
- Created Jenkins jobs for CI/CD using git, Maven and Bash scripting
- Built regression test suite in CI/CD pipeline with Data setup, test case execution and tear down using Cucumber- Gherkin, Java, Spring DAO, PostgreSQL.
- Extracted and loaded CSV files, Json files data from AWS S3 to Snowflake Cloud Data Warehouse.



- Moved data from S3 bucket to Snowflake data warehouse for generating the reports.
- Managed Roles and Grant management on Snowflake.
- Hands on experience running ETL data pipelines on snowflake.
- Connected Redshift to Tableau for creating a dynamic dashboard for the analytics team.
- Conducted ETL Data Integration, Cleansing, and Transformations using AWS glue Spark script.

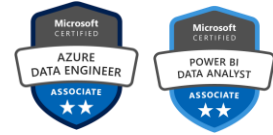
**Environment:** Redshift, DynamoDB, Pyspark, Snowflake, EC2, EMR, Glue, S3, Java, Kafka, IAM, PostgreSQL, Jenkins, Maven, AWSCLI, Shell Scripting, Git.

Deloitte, Dallas, TX  
Role: Azure Data Engineer

Nov 2018- Aug 2019

**Responsibilities:**

- Worked in different applications with both development and support project activities.
- Created Azure Data Factory pipelines to copy the data from source to target systems.
- Migrate the data from sources to the destination with the help of ADF.
- Create Azure Data Factory pipelines with extensive usage of activities like Copy, Metadata, For Each, If, Stored Proc, Mapping Data Flow etc., and Transformations in Data flow like Lookup, Derived column, Select, Aggregate, Union all, Join, Alter row, Conditional Split etc.
- Involved in deployment activities and moved the code across different environments.
- Migrated data from Client data scope to Azure SQL server and Azure Data Lake Gen2.
- Loaded the source data into Azure synapse warehouse tables by creating some external tables.
- Creating the linked service for source and target connectivity Based on the requirement.
- Connected to SharePoint and placed file onto BLOB using Logic Apps.
- Written Unit Test cases for Web API's and Azure functions.
- Implemented monitoring and logging solutions for Azure Data Factory pipelines using Azure Monitor and Azure Log Analytics, enabling real-time visibility into pipeline execution status and performance metrics.
- Developed Azure function apps as API services to communicate DB.
- Created Notebooks in data bricks for loading data from source to target ADLS.
- Provided technical guidance and support to junior team members on Azure Data Factory best practices and design patterns, fostering skill development and collaboration within the team.
- Monitoring daily running Azure pipelines for different applications and supporting multiple applications.
- Collaborated with cross-functional teams to design and implement scalable and cost-effective messaging solutions using Azure Service Bus, aligning with business requirements and objectives.
- Development of Data Bricks Notebooks using SQL.
- Implementation of logics in Data Bricks as per client requirement.
- Developed spark applications Pyspark and spark SQL for data extraction, transformation and aggregation from various.
- Worked with spark core, Pyspark and spark SQL modules of Spark.
- Involved in converting SQL queries into Spark transformations using APIs like Spark SQL, Data Frames and python.
- Developed multiple POCs using Spark and deployed on the Yarn cluster, compared the performance of Spark, with SQL
- Deployment of Azure Data Bricks and ADF using Azure DevOps.
- Written complex stored procedures and user defined functions using T-SQL.
- Worked on Azure Key vaults to use it in linked services.
- Create complex pipelines to extract the data from SQL server into different csv files.
- Create event-based triggers in Data factory to capture data from ADLS.



- Create scheduling for pipelines based on the requirement.
- Given support and fixing the issues during the test phases.
- Expressions and Tabular Model. Skills in Power BI visualization, high quality Power BI reports and Dashboards.
- Experience in developing Power BI reports and dashboards from various Data Sources like Databases, Excel files, flat files.
- Applied transform rules Data Conversion, Data Cleansing, Data Aggregation, Data Merging and Data split.
- Created Custom columns and Conditional Columns in the Power BI using Power Query.
- Created various measures and columns in Power BI Modelling using Data Analysis Expression.
- Created data visualizations using various visual objects (Line chart, Bar chart, Tree map, Single card, KPI, Map, Pie chart, custom visuals .... etc.). Published reports to the Power BI Online.
- Used Power BI to prepare dashboards to create attractive and interactive visualizations.
- Created reports in Power BI preview portal utilizing the SSAS Tabular using Power BI Gateway.

Environment: SQL, ADF, Data Bricks, ADLS Gen2, Blob, Logic Apps, DevOps, T-SQL, Git, Synapse, Power BI.