|  |  |
| --- | --- |
| **Name: Sreeja**  **Email: sm.dataeng@gmail.com**  **Phone: 312-776-9279**  **linkedin.com/in/smdataeng** |  |

**Summary:**

* 13+ years of experience as a Certified Data Engineer, Data Analyst and Microsoft SQL Server Suite of Products– T-SQL, SSIS, SSRS, Power BI with expertise in **Data Mapping, Data Validation**, and dealing with **statistical data analysis** such as transforming business requirements into analytical models, and strategic solutions that scale across massive volumes of data.
* Proficient in Azure Big data technologies such as **Azure Data Lake Analytics, Azure Data Lake Store, and Azure Data Factory**, **Blob storage, Delta Lake Storage, Azure Key Vault, Azure SQL, Logics Apps and Synapse Analytics** with hands-on experience in creating Proof of Concepts (POCs) for data migration and processing**.**
* I have extensive experience extracting and loading data from relational databases such as Teradata, Oracle, and DB2 into Azure Data Lake storage via the Azure Data Factory.
* Experienced in developing production-ready Spark applications using Spark components such as Spark SQL, MLlib, Graphix, Data Frames, Datasets, Spark-ML, and Spark Streaming.
* Strong experience in writing scripts using **Python API, and Spark API** for analyzing the data; Hands-On experience with **Spark Core, Spark SQL, Spark Streaming, Rest API** and creating the Data Frames handle in SPARK with Scala.
* Hands-on experience in **AVRO and Parquet** file format, Dynamic Partitions, Bucketing for best Practice and Performance improvement.
* Experience with source control systems such as **Git Hub, Bit bucket, and Jenkins** and in **CI/CD** deployments.
* Developed Data pipelines to synchronize and aggregate data from various data sources. Build the query data platform and provide guidelines for partner team to use aggregated data. Publish standard data sets & schema.
* Strong experience in data engineering and building **ETL pipelines** on batch and streaming data using **PySpark** and **Spark SQL**.
* In-depth knowledge in normalization/de-normalization techniques, logical and physical data modeling for relational databases and dimensional modeling for data warehouses.
* Expert in creating database objects such as tables, views, stored procedures, triggers, joins, indexes, user-defined data-types and functions.
* Demonstrated experience in developing interactive reports using Power BI and SQL Server Reporting Services(SSRS) from various data sources including OLTP databases and OLAP cubes
* In depth knowledge in Power BI advance analysis features like Actions, Calculations, Parameters and Trend lines.
* Highly Proficient in creating SQL Server Report(SSRS) including drill down, drill through, Parameterized, Cascading, Table, Matrix, Chart, line, Ad-hoc reports, Sub Reports and Snapshots.
* Developed ETL process using SSIS with various control flow, data flow tasks and store procedures for work order validation process.
* Experience in creating jobs, alerts, SQL mail agent, and scheduling DTS packages.
* Experience in using sql Profiler, Index Tuning Wizard and Performance Monitor to monitor SQL Server for dead locks and long running queries.
* Ability to work independently or as part of a team to accomplish critical business objectives and to make decisions under pressure.
* Solid analytical ability and creativity in analyzing and implementing timely solutions to requirements.

# Education:

* Bachelor of Technology, J.N.T.U, Hyderabad,2006, India.
* Master of Science, West Virginia University, Morgantown, WV,2009, US.

**Certifications:**

* Microsoft Certification: Microsoft SQL Server – Implementation and Maintenance
* DP 203: Microsoft Certified: Azure Data Engineer Associate
* PL 300: Microsoft Certified: Power BI Data Analyst Associate

# Professional Experience:

**Mesirow, Chicago, IL Nov2022 – Present**

# Sr. Azure Data Engineer

# Responsibilities:

* Employed Azure Data Factory orchestration in conjunction with **Azure Data Lake Storage Gen 2** to seamlessly ingest, transform, and load transactional data sets, encompassing historical performance data and loan portfolio data sourced from disparate sources resulting in a notable **17.3%** decrease in processing time, streamlining operations, and enhancing overall efficiency.
* Engineered and deployed scalable data ingestion solutions utilizing **Azure Data Factory** to bolster real-time and batch data processing capabilities within the financial domain. This implementation enables near-real-time analytics and reporting functionalities, facilitating swift decision-making, and enhancing operational agility.
* Leveraged Azure Data Factory's robust pipeline orchestration features, harnessing the potential of Data Flow for task execution and monitoring purposes.
* Skillfully orchestrated complex data workflows, resulting in anotable **13% improvement in workflow** efficiency.
* Leveraged Databricks to analyse data from Azure data storages, utilizing **Spark cluster** capabilities for deriving insights.
* Involved in extracting, transforming, and loading data from source systems to **Azure Synapse** using **Azure Data Factory, Databricks, Spark, Spark SQL, and U-SQL Azure Data Lake Analytics**.
* Created pipelines in Azure Data Factory (ADF) for ETL/ELT operations involving **Azure SQL, Blob storage, and Azure synapse.**
* Integrated data from both on-prem **(MYSQL, Cassandra) and cloud (Blob storage, Azure SQL DB)** using **Azure Data Factory** and applied transformations for loading into **Snowflake.**
* Deployed streaming ETL pipelines using Spark Streaming for extracting, transforming, and loading real-time data into Azure Synapse Analytics.
* Utilized Medallion architecture for data integration, migration, and ingestion.
* Efficiently loaded data into Azure SQL Synapse Analytics (DW) using **Azure BLOB and Data Lake storage**.
* Developed data ingestion pipelines on **Azure HDInsight Spark cluster** through **ADF** and **Spark SQL.**
* Implemented data partitioning and indexing strategies**, reducing data query time by 21% and optimizing overall data storage costs by 35%.**
* Proficiently used PySpark in both **HDInsight and Databricks** within the Hadoop ecosystem.
* Designed and implemented real-time data streaming solutions utilizing **Azure EventHub**.
* Involved in performance tuning and optimization for **Azure Logic Apps** and the related data processing pipelines.
* Collaborated with DevOps engineers to develop **automated CI/CD and test-driven development pipelines** using Azure, meeting client requirements.
* Developed Spark Streaming applications for processing real-time data from sources such as **Kafka and Azure Event Hubs.**
* Developed **ELT/ETL pipelines** for data movement to and from Snowflake data store using **Python and Snowflake Snow SQL.**
* **Created custom alerts** using Azure Data Factory, SQLDB, and Logic Apps.
* Contributed to the development, improvement, and maintenance of Snowflake database applications.
* Worked on Snowflake modelling, incorporating data warehousing techniques, data cleansing, Slowly Changing Dimension phenomenon, surrogate key assignment, and change data capture.
* **Developed logical and physical data models for Snowflake**, implementing partitioning, bucketing, and optimization through parameters.
* Configured and managed multi-cluster warehouses in Snowflake to handle high-concurrency workloads effectively.
* Defined roles and access privileges in Snowflake to ensure proper data security and governance.
* **Implemented Snowflake caching** mechanisms to improve query performance and reduce data transfer costs.
* Utilized Snow pipe for real-time data ingestion into Snowflake, ensuring continuous data availability and automated data loading processes.
* Monitored Spark clusters using Log Analytics and **Ambari Web UI** and Extensively utilized Spark core, Spark SQL, and Spark Streaming for real-time data processing.
* Conducted continuous review of business-critical databases to proactively identify space issues and performance tuning challenges.
* Developed Spark core and Spark SQL scripts using Scala for enhanced data processing speed.
* Executed all phases of software engineering, encompassing requirements analysis, application design, and code development and testing.

**Environment:** Azure Cloud Services, Azure Data Factory, Azure Databricks, Azure Synapse, Blob, ADLS, Bigdata, Hadoop, PySpark, Spark, Python scripts, SQL, Tableau, Data Model, Hive, Sqoop, Snowflake

**GE Digital, Bengaluru, India Mar 2019 – Oct 2022**

# Sr. Data Engineer

# Responsibilities:

* Understand Business requirements, analysis and translate into Application and operational requirements.
* Designed a one-time load strategy for moving large databases to Azure SQL DWH.
* Data Ingestion to one or more **Azure Services** - (Azure Data Lake, Azure Storage, Azure SQL, Azure DW) and processing the data in **Azure Data bricks.**
* Implemented Copy activity, Custom **Azure Data Factory** Pipeline Activities. Analyzed existing systems and proposed improvements in processes and systems for usage of modern scheduling tools like Airflow and migrating the legacy systems into an Enterprise data lake built on **Azure Cloud**.
* Experience in building and architecting multiple Data pipelines, end to end ETL and ELT process for Data ingestion and transformation in **Azure**.
* Extract Transform and Load data from Sources Systems to **Azure Data Storage services** using a combination of **Azure Data Factory, T-SQL, Spark SQL, and U-SQL Azure Data Lake Analytics.**
* Create and maintain optimal data pipeline architecture in cloud Microsoft Azure using **Data Factory and Azure Databricks.** Working Experience on **Azure Databricks** cloud to organize the data into notebooks and making it easy to visualize data using **dashboards**.
* Prepared **ETL** design document which consists of the database structure, change data capture, Error handling, restart and refresh strategies.
* Provided application support via phone. Developed and tested Windows command files and SQL Server queries for Production database monitoring in 24/7 support.
* Configured and implemented the **Azure Data Factory** Triggers and scheduled the Pipelines; monitored the scheduled ADF pipelines and configured the alerts to get notification of failure pipelines.
* Maintained **NoSQL** database to handle **unstructured data**, clean the data by removing invalidate data, unifying the format and rearranging the structure and load following steps.
* Wrote **Python scripts** to parse **XML** documents and load the data in database.
* Worked on **Data load** using **Azure Data factory** using external table approach. Automated recurring reports using **SQL** and **Python.**Developed purging scripts and routines to purge data on **Azure SQL Server** and **Azure Blob storage.**
* Resolved the **data type inconsistencies** between the source systems and the **target system** using the **Mapping Documents.**
* Wrote **T-SQL procedures** to generate **DML scripts** that modified database objects dynamically based on user inputs.
* Created Row Level Security with Power BI and integrated it into Power BI service Portal.
* Implemented row level security to restrict data for the business groups using analysis service data models.
* Developed interactive visual reports, dashboards, and KPI scorecards using Microsoft Power BI desktop. Create high-quality, interactive dashboards and reports using Power BI, incorporating best practices for data visualization and user experience.
* Integrated data from diverse sources, transforming and cleaning data as needed to ensure consistency and reliability for reporting purposes.
* Utilized Data Analysis Expressions (DAX) to implement sophisticated calculations and measures for in-depth analysis within Power BI.
* Optimized the performance of Power BI reports and dashboards, addressing issues related to data refresh, query performance, and overall responsiveness.
* Work closely with data engineering teams to ensure the availability and quality of data necessary for Power BI reporting.
* Pulling data from several data sources like Azure SQL Server, Snowflake, Azure Synapse, Power BI data set into Power BI and creating robust data model for reporting.
* Created Slicers and Filters to accurately depict information in visualizations according to the user's needs.
* Designed complex data intensive reports and visualizations in Power BI such as Table, Tree Maps, gauge, funnel, line etc for better business analysis.

**Environment:** Azure Cloud, Azure Data Factory (ADF v2), Azure functions Apps, Azure Data Lake, BLOB Storage, Visual Studio 2012/2016, Microsoft SQL Server 2012/2016, SSIS 2012/2016, Power BI Desktop, Teradata Utilities, Windows remote desktop, UNIX Shell Scripting, AZURE PowerShell, Data bricks, Python, Erwin Data Modelling Tool, Azure Cosmos DB, Azure Stream Analytics, Azure Event Hub.

**Algonomy Software PvtLtd., Bengaluru, India Feb2017 –Jan 2019**

**SQL Lead**

# Responsibilities:

* Gathered Data requirements from all Modules in the Project-WebApp/UI, Data Science and PowerBI and Prepared ETL Specification Documents.
* Gathered Business and Technical requirements from the stake holders.
* Created ER diagrams and designed data mappings.
* Designed and developed new Target Database implementing business rules and technical rules.
* Designed SSIS package to move data from Staging to Target DB utilizing Data flow components.
* Created Views, Stored Procedures, Triggers to support web application.
* Creating Database objects–Tables, Views, and Functions and developing procedures for implementing application functionality at the database side for performance improvement.
* Performing query plans and making sure each query is using appropriate useful Indexes.
* Creating and Modifying T-SQL Stored Procedures/Triggers for validating the integrity.
* Used Performance Monitor and SQL Profiler for Monitoring memory, processor, SQL Queries etc.
* Deployed the created SSIS packages in Development and Testing environments.
* Wrote/utilized T-SQL objects Complex joins, correlated subqueries, CTEs to match (data matching) current and historical data for reporting.
* Provided source to target mappings to perform initial, full, and incremental loads into the target data mart.
* Utilized many system functions and user functions to cleanse and transform data as per the ad-hoc

Reporting needs.

* Modified existing SSIS packages to meet the change requests from the users.
* Used SSIS to create ETL Packages to validate, extract, transform and load data to data warehouse databases, data mart databases.
* Generated packages using different Transformations like Look ups, Derived Column, Merge Join, Fuzzy Lookup, ForLoop, ForEachLoop, Conditional Split, Union all, Script component etc.
* Experience in providing Logging, Error handling by using EventHandler, and Custom Logging for SSIS Packages.
* Imported Data from SAS to SQL Server using SSIS packages.
* Created reports in Power BI Desktop and exported them to be utilized in a Hybrid dashboard.
* Designed and Implemented Power BI reports and effective dashboards after gathering and translating end user requirement.
* Experience in creating calculated measures and columns with DAX in MS POWER BI DESKTOP.
* Experience in creating advanced calculations, groups, dashboards, presentations, and graphs.
* Created different visualizations like Line, Bar, Matrix, Scatter, Water Fall, Tree Maps etc.

**Environment:** SQLServer2019, VisualStudio2017, SSIS, SAS 9.4, PowerBI, DAX, T-SQL, SSAS, Agile, SVN.

**Provant Health Solutions, East Greenwich, RI Apr 2013–Jun 2016**

**Sr. SQL Business Intelligence Developer**

**Responsibilities:**

* Created tables and Indexes. Implemented SQL Constraints like Unique, Check and Default to enforce Business Rules.
* Created stored procedures with input and output parameters to be utilized by micro services.
* Developed SQL Queries to fetch complex data from different tables in remote databases using joins, and database links.
* Developed and created stored procedures, advanced queries, triggers, views, indexes, and functions for databases.
* Designed dynamic SSIS Packages to transfer data crossing different platforms, validate data during transferring, and archived data files for different DBMS.
* Responsible for ETL operation using SSIS and Worked with users in training and support and also have experience in incremental data loading.
* Developed SSIS Packages to extract data from various data sources including Access database, Excel spreadsheets, and flat files into SQL server.
* Developed stored procedures to implement business logic to process raw data and to be used in SSIS packages and to be called from portal.
* Performed maintenance of Stored Procedures to improve **Performance** of different front end Applications.
* Created and maintained SQL objects including stored procedures, partitioned and indexed views, complex joins, CTEs, Triggers, temporary tables, table variables.
* Performed various data massaging activities like treating null values, cleaning data, sorting, data conversions by using the Power Query Editor.
* Created interactions of Highlight action, Filter action add interactivity to the visualizations. Involved in advanced calculations, creating Parameters, slicers and Interactions.
* Used Aggregate Functions, Filter Functions, Time & Date Functions and Logical Functions in Data Analysis Expressions (DAX) for various fact calculations and efficient data visualization in Power BI.
* Designed high level ETL architecture for overall data transfer from the OLTP to OLAP with the help of SSIS. Improved data load performance by analyzing database design and query plans.
* Implemented parallel execution of multiple data flows, tasks and packages in SSIS by modifying max current executables and engine thread properties.
* Developed high performance packages to process large data sets/files by identifying and fixing ETL bottlenecks, implementing multithreading, modulo function and conditional splitting the data into multiple processes.
* Extensively worked on Data Analysis, Data Modeling, Data Cleansing, Transformation, Integration, Data import and export.
* Created and analyzed source to target matrix, mapping specification and data dictionary documents.
* Deployed SSIS package into production environment and used package configuration to export various package properties to make package environment independent.
* Tested, cleaned and standardized data meeting the business standards using fuzzy /exact lookups using SSIS tasks.
* Involved in creating SSIS jobs to automate the reports generation and cube refresh packages.
* Designed and developed scorecards and dash boards using SharePoint Server. Created interactive reports with an ability to drill down into details.
* Monitored **SQL Server logs** to ensure the backup operations, batch commands, or other scripts and processes have completed successfully.
* Performed various database development activities including Index Optimizations, Table Partitions, Sub Partitions, Views and Temporary Tables.
* Optimize the database by creating various clustered, non-clustered indexes and indexed views.
* Knowledgeable in HL7 Transformations, ICD-10 and 999 rejections.
* Completed workshop on HIPAA.

**Environment:** SQL Server 2012/2014, SSDT, SSIS, T-SQL, MDX, ETL, VB.Net, C#, Oracle 12c, TFS, Erwin, HL7.

**Memorial HealthCare System, Miramar, FL Jan 2012 - Apr 2013**

**SQL Developer/ SSIS Developer**

**Responsibilities:**

* + Designed the Functional Requirements and Mapping Technical Specifications on the basis of Business Requirements.
  + Creating Database objects – Tables, Views, and Functions and developing procedures for implementing application functionality at the database side for performance improvement.
  + Performing query plans and making sure each and every query is using appropriate useful Indexes.
  + Creating and Modifying T-SQL Stored Procedures/Triggers for validating the integrity.
  + Used Performance Monitor and SQL Profiler for Monitoring memory, processor, SQL Queries etc.
  + Rebuilding Indexes and Tables as part of performance tuning exercise.
  + Deployed the createdSSIS packages in Development and Testing environments.
  + Wrote/utilized T-SQL objects Complex joins, correlated sub queries, CTEs to match (data matching) current and historical data for reporting.
  + Provided source to target mappings to perform initial, full and incremental loads into the target data mart.
  + Studied the existing OLTP system(s) and Created facts, dimensions and star schema representation for the data mart.
  + OLAP model based on Dimension and FACTS for efficient loads of data based on Star Schema structure on levels of reports.
  + Converted the data mart from Logical design to Physical design, defined data types, Constraints, Indexes, generated Schema in the Database.
  + Analyzed data and created new data match columns/keys to identify and match data from different departments to reconcile cash transactions. Improved the reconciliation process.
  + Utilized many system functions and user functions to cleanse and transform data as per the ad-hoc reporting needs.
  + Built packages and handled slowly changing dimensions to maintain the historical data.
  + Modified existing SSIS packages to meet the change requests from the users.
  + Used SSIS to create ETL Packages to validate, extract, transform and load data to data warehouse databases, data mart databases.
  + Generated packages using different Transformations like Lookups, Derived Column, Merge Join, Fuzzy Lookup, For Loop, For Each Loop, Conditional Split, Union all, Script component etc.
  + Experience in providing Logging, Error handling by using Event Handler, and Custom Logging for SSIS Packages.

**Environment:** SQL Server 2012/2008 R2, Tableau Desktop, SSAS, SSIS, SSRS, T-SQL, TFS, Epic Clarity, Lawson.

**Nationwide Financial, Columbus, OH Feb 2010 – Jan 2012**

# SQL Developer/ SSRS/SSIS Developer

Nationwide Financial Services (NFS) is the holding company for Nationwide Life Insurance and is itself held by insurance giant Nationwide Mutual. It offers mortgage and investment services to personal bankers and small businesses. The insurance segment is into selling different types of insurance policies like automobile insurance, home insurance, health insurance, life insurance etc. Finance segment is into lending different types of loans like home loans, auto loans etc.

**Responsibilities:**

* Involved in User Requirement Gathering and Technical Specification Documentation.
* Generated various Report models on Views (due to security issues) and deployed them on to the production reporting server and generated various Report models and ad-hoc reports.
* Developed daily, weekly, monthly reports and delivered reports to the required users.
* Maintained the table performance by following the tuning tips like normalization, creating indexes and collect statistics.
* Responsible for Debugging reports using Stored Procedures.
* Built effective queries for high-performance reporting and rendered them to HTML, XML, PDF and Excel formats using Reporting Services.
* Identified and defined the Datasets for the report generation and wrote the queries for the drill down reports,sub reports.
* Created a SQL 2008 Reporting Services (SSRS) and SSIS Infrastructure that pulls data from data sources and provides a simple to use reporting environment.
* Created SSRS reports using Report Parameters, Drop-Down Parameters, Multi-Valued Parameters Debugging Parameter Issues Matrix Reports and Charts.
* Created Ad-Hoc Reporting with Report Builder and Programming Reporting Services.
* Used SSRS Report Catalog to store the Configuration, Security and Caching Information for the Operation of the Report Server.
* Deployed the generated reports on to the Reportserver and fine-tuned the database objects and server to ensure efficient data retrieval.
* Developed SQLSSIS Packages to extract data from various data sources including Access database, Excel spreadsheets, and flat files into SQLserver 2008.
* Performed Data Analysis and Reporting by using multiple transformations provided by SSIS such as data conversion, conditional split, bulk insert, merge and union all.
* Worked with different methods of logging in SSIS.
* Retrieved logins and passwords using Configuration Manager.
* Extensively used transformations, control low tasks, data flow tasks, containers and event handlers.
* Experience in upgrading SQLServer software to new versions and applying service packs and hot fixes and unattended Installations.
* Extensively used SSIS Import/Export Wizard, for performing the ETL operations.
* Performed database transfers and queries tune-up, integrity verification, data cleansing, analysis and interpretation. Developed, monitored and deployed SSIS packages.
* Optimized SQL Queries using Query Analyzer.
* Designed and created views for security purposes, implemented rules, defaults, and user defined data types.
* Designed and implemented data access stored procedures and triggers for automating tasks.

**Environment:** SQL Server 2008, SSIS, SSRS, SSAS, Visual Studio 2010, Microsoft Visual Basic, Microsoft Visio, Windows 2003 R2, Oracle 9i.