# Skills:

|  |  |  |
| --- | --- | --- |
| AWS/Azure Data Lake | Snowflake Database | MS SQL Server |
| Python / PySpark | ETL / ELT/ ITL | Azure Data Factory / SSIS |
| Databricks | Power BI | Data warehouse |

**Work History**

|  |  |
| --- | --- |
| **Data Engineer**  US Farm Data, Mt. Juliet, TN | **Aug 2019 – Present** |

Designing, building, and maintaining data warehouses, databases, and pipelines is the central focus of this role. To ensure data availability, accuracy, and scalability, they involve data scientists and analysts. Data engineers apply data modeling, develop, and enhance ETL processes, Data Integrations from different sources and Motivated / Supporting Team members. The position entails Senior Snowflake Data Engineer role like follows business trends, Client Requirement, Data Synchronization, Data Integration, Implementing Solutions, assessing Technology, and making sure that data infrastructure meets the demands of the Customer Service / needs or Business requirements.

* Implement DWH solution consisting of ETLs, **On Premise to Cloud Migration** and expertise building and deploying batch and streaming data pipelines on cloud environment.

# Snowflake concepts - Snow Pipe, Tasks, Streams, Zero- copy cloning, Time Travel, Resource Monitors, Role Based Access Controls, Data Sharing, Multi -Cluster Virtual Warehouse Sizing, Query Performance Tuning, Data Masking, UDF, Materialized Views for continuous load data.

* Create and manage RDS SQL server instance to deliver the **Million-dollar data analytics**

project.

* Created Pipelines in **ADF** using **Linked Services** / **Datasets** / **Pipeline** to Extract, Transform, and load data from different sources like Azure SQL, Blob storage, Azure SQL DWH.
* Implemented complex data transformations using **ADF** Data Flow activities, leveraging transformations like **Filter, Derived Column, Lookup, Conditional Split, Aggregate, and Mapping Data Flow**s to optimize data processing and ensure the accuracy of the data warehouse.
* Experience on Migrating SQL database to Azure data Lake, Azure data lake Analytics, Azure SQL Database, **Data Bricks** and Azure SQL Data warehouse and controlling and granting database access and Migrating On premise databases to Azure Data Lake store using **ADF**.
* Writing in **Python, PySpark** Code for **Data Manipulation, ETL Tools**, utilizing libraries such as Pandas and NumPy to clean, transform, and analyze large datasets.
* Develop Spark applications using **PySpark** and **Spark-SQL** for data extraction, transformation, and aggregation from multiple file formats for analyzing, transforming the data to uncover insights into the customer usage patterns.
* Extract files from a **Data lake - S3** bucket using AWS SDK or API, and efficiently processed and transformed the data at a Snowflake warehouse, resulting in actionable and usable insights.
* Good understanding of **Spark Architecture** including **Spark Core, Spark SQL, Data Frames, Spark Streaming, Driver Node, Worker Node, Stages, Executors** and Tasks.
* Implemented **Replication** and **Failover** strategies to ensure high availability and data redundancy in critical systems.
* Develop and manage a data warehousing for the sales and marketing teams to track the status of the process flow, make data-driven decisions, and significantly reduce **their workload by up to 20%,** resulting in enhanced products / productivity and revenue growth.
* Extensive experience in Healthcare / Banking / Finance **Database Design**, **Normalization**, Selective **De-normalization**, **Forward/ Reverse Engineering** using Erwin.
* Azure Data tools, handling storage (**Data Lake, SQL Database**), as well as processing and orchestrating data efficiently in the Azure ecosystem.
* Created, provisioned different **Databricks clusters** needed for batch and continuous streaming data processing and installed the required libraries for the **clusters**.
* Knowledge in Analysis Business Requirements, Data Migration Projects to New Technologies, RBAC, Architecture, Resource Monitor, Data Ingestion, JSON, CSV, XML, Data Processing.
* Communication with CEOs, CTOs, and other Organizational leaders to understand and interpret data requirements, align data engineering initiatives with strategic objectives, and ensure the eﬀective implementation of data-driven solutions and New Concepts.

**Environment:** AWS, Snowflake, Python, synapse, Data bricks, PySpark, Azure, Power BI, SSIS, SSAS, ER-Win, Pl\SQL, Microsoft SQL server 2017/2019, SQL Server Profiler, T-SQL, Snow SQL, ANSI SQL, PowerShell, Redshift, Jira, GitHub

|  |  |
| --- | --- |
| **BI Developer**  US Farm Data, Madurai, India | **Jun 2015 – Aug 2019** |

* Created both **Tabular** and **Multi-dimensional SSAS** cubes from relational databases, utilizing Data Modelling and aggregation techniques to optimize query performance. Generated comprehensive reports based on these cubes.
* Developed efficient **ETL pipelines** using SSIS to seamlessly Data Extractions from diverse sources, perform necessary Data transformation, and successfully load it into the designated Database Systems.
* Leveraged Business Intelligence tools to **analyze relationships** between farmers, sellers, and consumers within the USDA dataset, identifying missing values and inconsistencies.
* Comprehensive knowledge and experience in process improvement, **Normalization / De- Normalization**, Data Extraction, Data Cleansing, Data Governance
* Experience using **ER diagram**, Dimensional data modeling, **Star Schema** modeling, **Snow- flake** modeling using Erwin tool and implement in existing data warehouse & operational database.
* Designed, implemented, and maintained **Disaster Recovery** (DR) and **High Availability (HA) solutions**, including Log Shipping, Database Mirroring, and Replication, ensuring the resilience and availability of critical systems.
* Engaged proactively with business users and customers to gather requirements, designed visually appealing and interactive visualizations, and comprehensive training on self-service BI tools, empowering users to explore and analyze data independently.
* Developed **Logical / Physical dimensional** data models using **Erwin**.
* Took responsibility for capturing diverse database, server, and application metrics, analyzing trends, and utilizing insights for long-term capacity planning and analysis eﬀorts.
* Expertly acquired diverse types of data files from multiple sources and meticulously validated and cleansed the data, employing advanced techniques to ensure PII **data quality** and **reliability**.
* written complex Queries using **Stored Procedures**, Merge, common Table expressions (CTEs), Temporary Tables to support Design Standards in Power BI and SSRS Reporting.
* Provide guidance on following **Agile / Waterfall Methodology**, including Sprints planning, daily stand-ups, and retrospectives, to promote continuous improvement, boost project efficiency, and facilitate seamless collaboration within the team and follows corporate protocol.

|  |  |
| --- | --- |
| **Database Administrator**  GoLeads, Omaha, NE | **Jul 2011 – Jun 2015** |

* Created the **ﬁrst comprehensive farm database** in the US, stitching together a dynamic dataset from multiple federal and agriculture university real-time data feeds.
* Manage and executed synchronization processes between **production, staging, testing,** and **development environments**, promoting data consistency and facilitating seamless collaboration among teams for reliable testing and development activities.
* **Schedule** and **meticulously** maintain **Backup and other packages** on a daily, weekly, and monthly basis using SQL Server Agent within SQL Server Management Studio (**SSMS**), ensuring the timely execution of data processes, and maintaining the integrity and reliability of data operations.
* Created comprehensive logical and **physical modeling, including data structures, primary keys,** and **foreign keys**, to support the implementation of Large Volume databases, resulting in efficient data organization and optimal performance based on Operational requirements.
* Implemented robust database security measures, such as **Access Controls**, **user authentication,** and **Data Encryption**, resolve Performance Issues to safeguard sensitive data and protect against unauthorized access.
* Develop and implement T-SQL objects and scripts, including **stored procedures, triggers,** and **functions**, to ensure consistent **data integrity** across multiple databases within database servers, SQL Queries, software update, networking solutions, Data Transfer, Data Formats, Performance Analysis, Manage Multiple Accounts / Modes, Data Mapping, Fault Tolerance, Data Classification, Domain Areas, Master Data Management, Quality Management, Metadata, Tolerance, Data Flow, System Requirements, Business Process, Project Implementation, DB Maintenance, software development, user administration, and enhancing data management efficiency and overall database reliability.

**EDUCATION**

**MBA** – Madurai Kamaraj University, Madurai - 2019

**BE –** Anna university, Chennai – 2011

**CERTIFICATIONS**

* + **Snowflake** Certified Snow Pro Core
  + **AWS Certiﬁed** Cloud Practitioner Certificate
  + **Microsoft Certiﬁed** Azure Database Administrator Associate
  + **Cisco Certiﬁed** Network Associate - CCNA
  + **Microsoft Certiﬁed MCSA** - Windows Server 2012
  + **Microsoft Certiﬁed** specialist on Server Virtualization