**SHIVA SRINIVAAS**

**Sr. Data Engineer**

+1 (972) 256 8725

sivadataengineer40@gmail.com

**Professional Summary:**

* Around 8+ years of IT experience as on Azure Cloud. Experience on Migrating SQL database to Azure data Lake, Visualizing with Power BI, 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 Azure Data factory.
* Experience in Developing Spark applications using Spark - SQL in Databricks for data extraction, transformation, and aggregation from multiple file formats for analyzing & transforming the data to uncover insights into the customer usage patterns.
* Good understanding of Spark Architecture including Spark Core, Spark SQL, Data Frames, Spark Streaming, Driver Node, Worker Node, Stages, Executors and Tasks.
* Experience with MS SQL Server Integration Services (SSIS), T-SQL skills, stored procedures, triggers.
* Strong experience in migrating other databases to Snowflake.
* Participates in the development improvement and maintenance of snowflake database applications.
* Design and 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.
* Design and implement database solutions in Azure SQL Data Warehouse, Azure SQL
* Experience in Database Design and development with Business Intelligence using SQL Server 2014/2016, Integration Services (SSIS), DTS Packages, SQL Server Analysis Services (SSAS), DAX, OLAP Cubes, Star Schema and Snowflake Schema.
* Excellent communication skills with excellent work ethics and a proactive team player with a positive attitude.
* Created Azure SQL database, performed monitoring and restoring of Azure SQL database. Performed migration of Microsoft SQL server to Azure SQL database.
* Have designed and developed ETL mapping for data collection from various data feeds using REST API.
* Expertise in various phases of project life cycles (Design, Analysis, Implementation, and testing).
* Proficiency in handling software development sprints, test, and delivery cycle for the development teams.

**Technical skills:**

|  |  |  |
| --- | --- | --- |
| ETL | : | Azure Data Factory, SAP BODS. Azure Data Bricks. Talend, DBT, DAX, Data Modelling. |
| Cloud Data Engineering | : | Azure, S3, Glue, Spark, Data Pipelines, Airflow, EC2, Azure, ADF, Azure Data Bricks, Azure synapse, Azure SQL (TSQL), IAM, Docker |
| Data Warehouse | : | Redshift, Azure SQL |
| Scrum & Version Tools | : | Jira, Git, Bitbucket |
| Database | : | MS SQL Azure, MS SQL Server, DB2, Oracle |
| Operating Systems | : | Windows 8, Windows 10, Windows Server 10, Linux |
| Programming | : | Python, SQL, PySpark, SSIS, T-SQL |
| Visualization | : | Power BI, Tableau |

**Professional Experience:**

**Aramco, Houston, TX March 2022 - Present**

**Sr. Azure Data Engineer**

**Responsibilities:**

* Analyze, design, and build Modern data solutions using Azure PaaS service to support visualization of data. Understand current Production state of application and determine the impact of new implementation on existing business processes.
* Superior involvement in developing reports and dashboards from multiple data sources such as on-premises and cloud using Power BI dektop .
* Creating reports from importing, direct query and live data from several data sources.
* Experience with, power pivot and dashboard designer.
* Troubleshoot and support reports in Power Bl workspaces whenever there are issues in production
* Creating pipelines, data flows and complex data transformations and manipulations using Azure Data Factory (ADF) and PySpark with Databricks.
* 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. Data Ingestion to one or more Azure Services - (Azure Data Lake, Azure Storage, Azure SQL, Azure DW) and processing the data in In Azure Databricks.
* REST APIs to retrieve analytics data from different data feeds.
* Developed HQL, Mappings, tables, external tables in Hive for analysis across different banners and worked on partitioning, optimization, compilation, and execution.
* Created Pipelines in ADF using Linked Services/Datasets/Pipeline to perform ETL on data from different sources like Azure SQL, Blob storage, Azure SQL Data warehouse, write-back tool and backwards.
* Developed 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.
* Responsible for estimating the cluster size, monitoring, troubleshooting of the Spark data bricks cluster.
* Design, develop and maintain SSIS packages to consume web services data for Analysis and Reporting of APS Decision Making System.
* Developed data pulls from Cosmos using Scope Scripts.
* Worked with the Microsoft COSMOS DB to generate distributed systems technology. All the datasets.
* Architect & implement medium to large scale BI solutions on Azure using Azure Data Platform services (Azure Data Lake, Data Factory, Data Lake Analytics, Stream Analytics, Azure SQL DW, HDInsight/Databricks, NoSQL DB).
* Created SQL scripts to migrate data from SQL server to Cosmos DB.
* Extensively involved in writing T-SQL Stored Procedures, Functions to convert complex business logic into backend.
* Extensive role in fixing performance issues by tuning slow running queries and run-away queries using different methods, by evaluating statics, joins, indexes, and code changes.
* Designed, Developed, Tested, Published, and maintained PowerBI Functional reports, Dashboards for the management teams to make healthy decisions.
* Created maintained SQL Agent Jobs for automating and scheduling ETL process and Database maintenance tasks.
* Develop Databricks Python notebooks to Join, filter, pre-aggregate, and process the files stored in Azure data lake storage.
* **Environment**: Azure Data Factory (ADF v2), Power BI,DAX, Spark - Python, Hive, Apache NiFi 1.8.0, Spark Streaming, Docker Containers, PostgreSQL, RabbitMQ, Celery, Flask, ELK Stack, MS-Azure, Azure SQL Database, Azure functions Apps, Azure Data Lake, BLOB Storage, SQL server, Data bricks, ADLS Gen 2, Cosmos DB, Azure Event Hub, Machine Learning, HDP 2.6, Jenkins, Kafka, HBase, Sqoop, Kerberos, Sentry, AWS, Data Modelling

**Common Spirit Health, Phoenix, AZ May 2020 – Feb 2022**

**Azure Data Engineer**

**Responsibilities:**

* Expertise in data transformations such as adding calculated columns, manage relationships, create different measures, remove rows, replace values, split column, date & time column.
* Experience in combining data from different sources by using of power query.
* Design and implement multiple ETL solutions with various data sources by extensive SQL Scripting, ETL tools, Python, Shell Scripting, and scheduling tools. Data profiling and data wrangling of XML, Web feeds and file handling using python, Unix, and SQL
* Loading data from different sources to a data warehouse to perform some data aggregations for business Intelligence using Python.
* Design and build data structures on Azure Data Lake and data processing using PySpark, to provide efficient reporting and analytics capability.
* Developed Spark applications using PySpark and Spark-SQL for data extraction, transformation, and aggregation from multiple file formats.
* Used Spark Streaming to receive real-time data from the Kafka and store the stream data to Data Lake using Python.
* Collected data using Spark Streaming from Azure storage account in near-real-time and performs necessary Transformations and Aggregation on the fly to build the common learner data model and persists the data in Data lakes.
* Authoring Python (PySpark) Scripts for custom UDF's for Row/ Column manipulations, merges, aggregations, stacking, data labeling and for all Cleaning and conforming tasks.
* Design and implement database solutions in Azure SQL Data Warehouse, Azure SQL.
* Architect & implement medium to large scale BI solutions on Azure using Azure Data Platform services (Azure Data Lake, Data Factory, Data Lake Analytics, Stream Analytics, Azure SQL DW, HDInsight/Databricks, NoSQL DB).
* Design Setup maintain Administrator the Azure SQL Database, Azure Analysis Service, Azure SQL Data warehouse, Azure Data Factory, Azure SQL Data warehouse.
* Performed ETL operation using SSIS and loaded the data into Secure DB.
* Good hands-on experience in Data Vault concepts, data models, well versed understanding, and implementation on Data warehousing concepts/Data Vault.
* Designed, reviewed, and created primary objects such as views, indexes based on logical design models, user requirements and physical constraints.
* Worked with stored procedures for data set results for use in Reporting Services to reduce report complexity and to optimize the run time. Exported reports into various formats (PDF, Excel) and resolved formatting issues.
* Worked on Sqoop to import and export data from database to HDFS and Data Lake on AWS.
* Designed the packages to extract data from SQL DB, flat files and loaded into Oracle database.
* Performed data migrations from on-prem to Azure Data Factory and Azure Data Lake.
* Responsible for defining cloud network architecture using Azure virtual networks, VPN, and express route to establish connectivity between on premise and cloud.
* **Environment**: Azure Data Factory (ADF v2),Power BI, Data Modelling, DAX, Spark (Python/Scala), Hive, Jenkins, Kafka, Spark Streaming, Docker Containers, PostgreSQL, RabbitMQ, Celery, Flask, ELK Stack, MS-Azure, Azure SQL Database, Azure functions Apps, Azure Data Lake, BLOB Storage, SQL server, UNIX Shell Scripting, AZURE PowerShell, Data bricks, ADLS Gen 2, Azure Cosmos DB, Azure Event Hub.

**Thomson Reuters, Eagan, MN July 2017 – March 2020**

**Data Engineer**

**Responsibilities:**

* Used update strategy to effectively migrate data from source to target.
* Moved the mappings from development environment to test environment.
* Designed ETL Process using Informatica to load data from Flat Files, and Excel Files to target Oracle Data Warehouse database.
* Interacted with the business community and database administrators to identify the Business requirements and data realties.
* Experienced in Agile Methodologies and SCRUM Process.
* Built various graphs for business decision making using Python matplotlib library.
* Worked in development of applications especially in UNIX environment and familiar with all its commands.
* Implement code in Python to retrieve and manipulate data.
* Reviewed basic SQL queries and edited inner, left, &rightjoins in Tableau Desktop by connecting live/dynamic and static datasets.
* Reported and created dashboards for Global Services & Technical Services using SSRS, Oracle BI, and Excel. Deployed Excel VLOOKUP, PivotTable, and Access Query functionalities to research data issues.
* Created conceptual, logical and physical relational models for integration and base layer; created logical and physical dimensional models for presentation layer and dim layer for a dimensional data warehouse in Power Designer.
* Involved in reviewing business requirements and analyzing data sources form Excel, Oracle SQL Server for design, development, testing, and production rollover of reporting and analysis projects.
* Analyzing, designing, developing, implementing, and maintaining ETL jobs using IBM Info sphere Data stage and Netezza.
* Designed Data Flow Diagrams, E/R Diagrams and enforced all referential integrity constraints.
* Extensively worked in Client-Server application development using Oracle 10g, Teradata 14, SQL, PL/SQL, Oracle Import and Export Utilities.
* Coordinated with DB2 on database build and table normalizations and de-normalizations.
* Conducted brainstorming sessions with application developers and DBAs to discuss about various de-normalization, partitioning and indexing schemes for Physical Model.
* Involved in several facets of MDM implementations including Data Profiling, metadata acquisition and data migration.
* Extensively used SQL Loader to load data from the Legacy systems into Oracle databases using control files and used Oracle External Tables feature to read the data from flat files into Oracle staging tables.
* Involved in extensive Data validation by writing several complex SQL queries and involved in back-end testing and worked with data quality issues.
* Used SSIS to create ETL packages to validate, extract, transform and load data to data warehouse databases, data mart databases, and process SSAS cubes to store data to OLAP databases
* Created ETL packages using OLTP data sources (SQL Server, Flat files, Excel source files, Oracle) and loaded the data into target tables by performing different kinds of transformations using SSIS.
* Migrated SQL server 2008 to SQL Server 2008 R2 in Microsoft Windows Server 2008 R2 Enterprise Edition.
* Developing reusable objects like PL/SQL program units and libraries, database procedures and functions, database triggers to be used by the team and satisfying the business rules.
* Performed data validation on the flat files that were generated in UNIX environment using UNIX commands as necessary.
* Strong understanding of Data Modeling (Relational, dimensional, Star and Snowflake Schema), Data analysis, implementations of Data warehousing using Windows and UNIX
* Used Model Mart of Erwin for effective model management of sharing, dividing and reusing model information and design for productivity improvement and identified and tracked the slowly changing dimensions and determined the hierarchies in dimensions.
* Designed high level ETL architecture for overall data transfer using SSIS from the source server to the Warehouse and defined various facts and dimensions in the data mart including Fact less Facts and designed the Data Mart defining Entities, Attributes and relationships between them.
* **Environment**: Azure Data Factory, Power BI, Data Modelling, DAX, Power Designer, Teradata, Oracle, PL/SQL, MDM, SQL Server 2008, ETL,DB2, SSIS, SSRS, SAS, SPSS, Datastage, Informatica, SQL, T-SQL, UNIX, Netezza, Aginity, SQL assistance etc.

**Flipkart, Bangalore, India June 2015 – March 2017**

**SQl Server Developer**

**Responsibilities:**

* Provide support and fix issues to all ETL/SSIS schedule jobs and maintain compliance to same and develop and maintain various standards to perform ETL codes and maintain an effective project life cycle on all ETL processes.
* Design, build and deploy ETL/SSIS packages with different data sources (SQL Server, Flat Files, Excel source files, Share Point etc.) and then loaded the data into destination tables by performing complex transformations using SSIS packages.
* Design, build and deploy ETL/SSIS packages with different data sources (SQL Server, Flat Files, Excel source files, Share Point etc.) and then loaded the data into destination tables by performing complex transformations using SSIS packages. Developed ETL Packages Involving transformations like derived column, conditional split, lookup, data conversion to load data into SQL Data Warehouse.
* Involved in various Transformation and data cleansing activities using various Control flow and data flow tasks in SSIS packages during data migration.
* Worked on all activities related to the development, implementation, administration and support of ETL processes.
* Used SSIS to create ETL packages to validate, extract, transform and load data into data warehouse databases.
* Used advanced SSIS tools (Derived Columns, Conditional Split, Look-up Transformation, Aggregations, Multicasting).
* Performed efficient tuning of SQL source queries for data load/Usage of stored procedures for optimized performance.
* Performance tuning of Queries and Stored Procedures.
* Performance tuning on existing ETL packages
* **Environment**: MS Azure, MSBI, SQL Server, SQL Azure, Tableau, Azure Data bricks, Data Lake, Cosmos DB

**Education:**

Bachelor’s in Information Technology - 2015