

Charitha Y

Data Engineer

Irving, Texas-75038 | +1 972-850-6542 | charitha.y@outlook.com

SUMMARY

- 4+ years of experience as Data Engineer with expertise spanning multiple cloud platforms, specializing in Azure and AWS environments.
- Around 4+ years of IT experience as on Azure Cloud. 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 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.
- Azure Data Factory (ADF), Integration Run Time (IR), File System Data Ingestion, Relational Data Ingestion.
- Created Azure SQL database, performed monitoring and restoring of Azure SQL database. Performed migration of Microsoft SQL server to Azure SQL database.
- 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.
- 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.
- Experience includes troubleshooting Power BI dashboards, performing Windows VM patching on Azure, and utilizing Azure Databricks for data validations.
- Involved designing and setting up an Enterprise Data Lake using AWS services.
- Proficient in orchestrating AWS infrastructure, creating Glue ETL jobs, and establishing CICD tools for efficient code deployment.
- Extensive experience in analyzing data using Hadoop Ecosystems including HDFS, MapReduce, Hive & PIG.
- Contributed by developing ETL solutions using Spark, PySpark, and Azure Data Factory.
- In-depth experience in data modeling, ETL processes, and optimization solutions for enhanced performance.
- Technical competencies include Python, PySpark, SQL, T-SQL, and familiarity with tools such as Power BI, Git, and Jenkins.
- Solid understanding of Data Warehousing concepts, Star Schema modeling, and Snowflake modeling.
- Demonstrated ability to collaborate with cross-functional teams, ensuring alignment with business goals.
- Strong analytical and problem-solving skills applied to diverse projects, resulting in successful project delivery.
- Continuously updated skills in a dynamic technological landscape, staying current with industry best practices.
- Seeking to bring a wealth of technical expertise and a history of successful project contributions to a new and challenging role in data engineering.
- Excellent communication skills with excellent work ethics and a proactive team player with a positive attitude.

KEY COMPETENCIES:

- Azure Data Lake, Data factory
- Azure Databricks,
- Azure SQL database
- Azure SQL Datawarehouse
- SQL Server 2019/17/16/14
- SQL, MSBI (SSIS, SSAS, SSRS)
- Data Visualization
- Data Migration
- Azure, AWS
- Cosmos DB, Redshift, RDS
- Python, Scala, PySpark, Spark SQL, T-SQL
- Power BI, Git, Jenkins, Terraform
- Star Schema, Snowflake Schema

WORK EXPERIENCE

Cardinal Health, Austin, TX | Data Engineer

Jan 2023 – Present

- 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.
- 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.
- Implemented Proof of concepts for SOAP & REST APIs
- REST APIs to retrieve analytics data from different data feeds
- 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 Data warehouse, write-back tool and backwards.
- Building ADF pipelines to extract and manipulate data from Azure Blob storage/Azure Data Lake/Cosmos DB/SQL Server on cloud.
- Building pipelines using Azure logical apps to extract and manipulate data from SharePoint.
- Developed scope script using U-SQL to generate daily, monthly structured streams/ORC files in Cosmos.
- Enhancing Multi-Dimensional Revenue Data Cube from ORC files generated in cosmos using TITAN for Business solutions.
- Maintaining and troubleshooting existing Power BI dashboards in case of data refresh failures which are owned by our team.
- Responsible for Windows VM patching on AZURE.
- Performing research to identify source and nature of data required for ETL solutions using Azure Databricks.
- Extensively used Azure Databricks for data validations and analysis on Cosmos structured steams.
- Developing job monitoring alerts for job failures and latency using scope script.
- Extensively used Dragonfly Job Scheduler for scheduling, monitoring the Scope jobs on daily/monthly basis.
- Experience working with Talend ETL and Snowflake cloud database systems.
- Performing Data Validation between HDInsight's cluster & Target SQL Server using Hive SQL.
- Highly knowledgeable in concepts related to Data Warehouses schemas, Star Schema modeling and Snowflake modeling and processes.
- Developing custom stored procedures for delta loads, functions, triggers using SQL, T-SQL on cloud SQL server.

Environment: Spark, Spark SQL, Hive, Pig, HDFS, Azure, Data Factory, Data Lake, Data Lake Analytics, Databricks, Data Factory, SQL, PL/SQL, ETL, Git

Fisker, India | Data Engineer

Feb 2020 – Mar 2022

- Design and implement database solutions in Azure SQL Data Warehouse, Azure SQL.
- Implemented 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 & implement migration strategies for traditional systems on Azure (Lift and shift/Azure Migrate, other third-party tools).
- Engage with business users to gather requirements, design visualizations and provide training to use self-service BI tools.
- Used various sources to pull data into Power BI such as SQL Server, Excel, Oracle, SQL Azure etc.
- Propose architectures considering cost/spend in Azure and develop recommendations to right-size data infrastructure.

- Build Complex distributed systems involving huge amount data handling, collecting metrics building data pipeline, and Analytics.
- Designed and set up Enterprise Data Lake to provide support for various use cases including Storing, processing, Analytics and Reporting of voluminous, rapidly changing data by using various AWS Services.
- Used various AWS services including S3, EC2, AWS Glue, RedShift, EMR, SNS, SQS, DMS
- Responsible for provisioning key AWS Cloud services and configure them for scalability, flexibility, and cost optimization
- Created VPCs, subnets including private and public, NAT gateways in a multi- region, multi-zone infrastructure landscape to manage its worldwide operation.
- Managed Amazon Web Services (AWS) infrastructure with orchestration tools such as CFT, Terraform and Jenkins Pipeline
- Created Terraform scripts to automate deployment of EC2 Instance, S3, EFS, EBS, IAM Roles, Snapshots and Jenkins Server
- Extracted data from multiple source systems S3, Redshift, RDS and Created multiple tables/databases in Glue Catalogue by creating Glue Crawlers.
- Created AWS Glue crawlers for crawling the source data in S3 and RDS.
- Created multiple Glue ETL jobs in Glue Studio and then processed the data by using different transformations and then loaded into S3, Redshift and RDS.
- Created multiple Recipes in Glue Data Brew and then used in various Glue ETL Jobs.
- Designed and Developed ETL Processes in AWS Glue to migrate data from external sources like S3, Parquet/Text Files into AWS Redshift.
- Build Cloud data stores in S3 storage with logical layers built for Raw, Curated and transformed data management
- Created data ingestion modules using AWS Glue for loading data in various layers in S3 and reporting using Athena and Quick Sight.
- Created manage bucket policies and lifecycle for S3 storage as per organizations and compliance guidelines
- Created parameters and SSM documents using AWS Systems Manager
- Established CICD tools such as Jenkins and GitBucket for code repository, build and deployment of the python code base.
- Used Lambda functions and Step Functions to trigger Glue Jobs and orchestrate the data pipeline
- Used PyCharm IDE for Python/PySpark development and Git for version control and repository management.

Environment: AWS Glue, S3, IAM, EC2, RDS, Redshift, EC2, Lambda, Boto3, DynamoDB, Athena, VPC, EBS, ELB, CloudFormation, Lambda, GIT, Glue, Athena Python and PySpark.

Infosys, India | SQL/SSIS Developer

Jun 2018 – Jan 2020

- Designed, reviewed, and created primary objects (views, indexes, etc.) based on logical design models, user requirements and physical constraints.
- Created Stored Procedures, Triggers, Functions, Indexes, Tables, Views using T-SQL code and SQL joins for applications.
- Wrote stored procedures and User Define scalar Functions (UDFs) to be used in the SSIS packages and SQL scripts.
- Involved in Normalization and De-Normalization of existing tables for faster query retrieval.
- Used DDL and DML for writing triggers, stored procedures to check the data entry and payment verification.
- Responsible for optimizing all indexes, SQL queries, stored procedures to improve the quality of software.
- Created SSIS Packages to transfer data from various sources like Text Files, SQL Server, Excel and Access to SQL Server.
- Configured Error and Event Handling: Precedence Constraints, Break Points, Check Points, Logging in SSIS and TRY and CATCH Blocks in TSQL.

- Creating Star and Snow Flake Schema design using Visio, defining relations between tables
- Designed, modified and troubleshoot SSIS packages.
- Utilized SSIS Checkpoint, Loading Dimension with the SSIS slowly changing Dimension Wizard and used different SSIS Control Flows.
- Created ETL in DTS packages by using Data Transform Task, VBScript task, SQL task and send Email task to migrate data from different functional databases into SQL Server 2016/14.
- Created notifications, alerts to let support team know in case there are any failures.
- Designed and created Report templates, bar graphs and pie charts based on the financial data using SSRS 2016/14.
- Developed various types of reports like Drill down, Drill through and parameterized reports using SQL Server Reporting Services.
- Defined Key Performance Indicator metrics to create scorecards.
- Generated Sub-Reports, Drill down reports, Drill through reports and parameterized reports using SSRS 2016/14.
- Created reports to retrieve data using Stored Procedures that accepts parameters.
- Developed complex reports using multiple data providers, user defined objects, aggregate aware objects, charts, and synchronized queries.
- Administered interface to organize reports and data sources, schedule report execution and delivery, and track reporting history using SSRS 2012/2008.
- Scheduled Cube Processing from Staging Database Tables using SQL Server Agent.
- Used SQL Server profiler for auditing and analyzing the events, which occurred during a particular time horizon and stored in script.
- Used "SQL Profiler TSQL Duration" template for tracking execution time of TSQL Statements.

Environment: MS SQL Server 2016/14, Management Studio (SSMS), ETL, ASP.NET, Crystal Reports9.0, Data Modeling, OLAP, Data warehousing SSIS, C#, VB.net T-SQL, DB2, SSAS, SSIS, SSRS, IBM, UNIX scripts

Education

Masters- Southern New Hampshire University (Business Analytics)