



Harsha Buddana

harshavardhanabuddana@gmail.com

469-325-3754

PROFESSIONAL SUMMARY:

- ❖ 8+ year Experience with Azure transformation projects and Azure architecture decision - making. Strong development skills with Azure Data Lake, Azure Data Factory, SQL Data Warehouse Azure Blob, Azure Storage Explorer
- ❖ Hands on experience in data integration, data management, data governance, data analytics and business intelligence projects.
- ❖ Develop Power BI reports & effective dashboards after gathering and translating end-user requirements.
- ❖ Recreating existing application logic and functionality in the Azure Data Lake, Data Factory, SQL Database, and SQL Data warehouse environment
- ❖ Propose architectures considering cost/spend in Azure and develop recommendations to right-size data infrastructure.
- ❖ Data center Migration, Azure Data Services have a strong virtualization experience.
- ❖ Experience in troubleshooting and resolving architecture problems including database and storage, network, security and applications.
- ❖ I do have experience as Azure Cloud Data Engineer in Microsoft Azure Cloud technologies including Azure Data Factory(ADF), Azure Data Lake Storage(ADLS), Azure Synapse Analytics(SQL Data warehouse), Azure SQL Database, Azure Analytical services, Polybase, Azure Cosmos NoSQL DB, Azure Key vaults, Azure Devops, Azure HDInsight BigData Technologies like Hadoop, Apache Spark and Azure Data bricks.
- ❖ Experience managing Big Data platform deployed in Azure Cloud.
- ❖ Implemented Copy activity, Custom Azure Data Factory Pipeline Activities for On-cloud ETL processing.
- ❖ Experience in Monitoring and Tuning SQL Server Performance.
- ❖ Experience in configuration of report server and report manager for job scheduling, giving permissions to a different level of users in SQL Server Reporting Services (SSRS).
- ❖ Experience of planning and implementation of Dynamics AX system quality tests to ensure that
- ❖ business requirements, functionality, configuration and data meet the expectation.
- ❖ Excellent knowledge on integrating Azure Data Factory V2/V1 with variety of ... UNIX Shell Scripting, AZURE PowerShell, Data bricks, Python, ADLS Gen 2.0
- ❖ Experience with different cloud-based storage systems like S3, Azure Blob Storage, Azure Data Lake Storage, Gen2, Gen, D365, Azure Integration

- ❖ Data Ingestion to one or more Azure Services - (Azure Data Lake, Azure Storage, Azure SQL, Azure DW) and processing the data in In Azure Data bricks
- ❖ Optimized Azure Data Factory (ADF) Pipelines with Various Partitioning Techniques

TECHNICAL SKILLS:

Tools: SSMS, Power BI, MS Office (Excel, SharePoint, Visio), Adobe Analytics, Azure Data Factory, Azure Synapse, Azure Data Lake, Azure SQL Databases, Azure Management tools, MSBI (SSIS, SSAS, SSRS) Azure Databricks, Azure SQL database, Azure SQL Datawarehouse

Version Control: GIT

Programming Languages: .NET, C#, JSON, U-SQL, Scala, Python, Spark SQL

Version Control: SVN, CVS, TFS, GIT, Bit Bucket

Platforms: Windows, Linux

Databases: SQL Server, MySQL, Oracle

PROFESSIONAL EXPERIENCE:

Client: Humana, CA

April 2022 to Present

Role: Azure Data Engineer

Responsibilities:

- ❖ Design and implement database solutions in Azure SQL Data Warehouse, Azure SQL.
- ❖ Migrate data from traditional database systems to Azure databases
- ❖ Co-ordination with external team members and other stake holders to understand the impact of their
- ❖ changes to complete release work with comfort. That helps a lot to avoid any integration issues in
- ❖ Explore.MS application.
- ❖ 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
- ❖ Co-ordination with external team members and other stake holders to understand the impact of their
- ❖ changes to complete release work with comfort. That helps a lot to avoid any integration issues in VLin-Box application.
- ❖ Responsible for estimating the cluster size, monitoring and troubleshooting of the Spark databricks cluster.
- ❖ Experienced in performance tuning of Spark Applications for setting right Batch Interval time, correct level of Parallelism and memory tuning.
- ❖ Performed ETL using Azure Data Bricks. Migrated on-premise Oracle ETL process to Azure Synapse Analytics
- ❖ To meet specific business requirements wrote UDF's in Scala and Pyspark

- ❖ Developed JSON Scripts for deploying the Pipeline in Azure Data Factory (ADF) that process the data using the Sql Activity.
- ❖ Created Build and Release for multiple projects (modules) in production environment using Visual Studio Team Services (VSTS)
- ❖ Design and implement database solutions in Azure SQL Data Warehouse, Azure SQL.
- ❖ Propose architectures considering cost/spend in Azure and develop recommendations to right-size data infrastructure.
- ❖ Setup and maintain the Azure SQL Database, Azure Analysis Service, Azure SQL Data warehouse, Azure Data Factory, Azure SQL Data warehouse
- ❖ Develop conceptual solutions & create proofs-of-concept to demonstrate viability of solutions.
- ❖ Implement Copy activity, Custom Azure Data Factory Pipeline Activities
- ❖ Primarily involved in Data Migration using SQL, SQL Azure, Azure storage, and Azure Data Factory, SSIS, PowerShell. Develop dashboards and visualizations to help business users analyze data as well as providing data insight to upper management with a focus on Microsoft products like SQL Server Reporting Services (SSRS) and Power BI.
- ❖ Responsible for creating Requirements Documentation for various projects.

Environment: Azure SQL, Azure Storage Explorer, Azure Storage, Azure Blob Storage, Azure Backup, Azure Files, Azure Data Lake Storage, SQL Server Management Studio, Visual Studio, VSTS, Azure Blob, Power BI, PowerShell, C# .Net, SSIS, DataGrid, ETL Extract Transformation and Load, Business Intelligence (BI).

Client: PNC Bank, PA.

Nov 2020 to May 2022

Role: Azure Data Engineer

Responsibilities:

- ❖ Designed and executed pipelines, data flows, and intricate data transformations and manipulations utilizing Azure Data Factory and PySpark on Databricks.
- ❖ Implemented data ingestion into one or more Azure Services, such as Azure Data Lake, Azure Storage, Azure SQL, and Azure DW, with data migration processes in Azure Databricks.
- ❖ Utilized Data Flow Debug to construct efficient Azure Data Factory data flow pipelines, and improved performance by implementing optimization options, including effective partition utilization during transformations.
- ❖ Constructed intricate ETL pipelines in Azure Data Factory using mapping data flows with multiple input and output transformations.
- ❖ Experienced in working with Azure BLOB and Data Lake storage, as well as loading data into Azure SQL Synapse analytics.
- ❖ Familiar with the Azure SQL Database Import and Export Service.
- ❖ Incorporated Azure Key vault as the central storage for secrets, referencing them in both Azure Data Factory and Databricks notebooks.

- ❖ Developed a common SFTP download/upload framework using Azure Data Factory and Databricks.
- ❖ Designed ETL pipelines in Databricks using notebooks, Spark DataFrames, Spark SQL, and Python scripting.
- ❖ Created Databricks job workflows that extract data from SQL Server and upload files to SFTP using PySpark and Python.
- ❖ Built the logical and physical data models for Snowflake, based on required changes.
- ❖ Determined virtual warehouse sizing for Snowflake for various types of workloads.
- ❖ Implemented Spark applications with PySpark and Spark SQL for data extraction, transformation, and aggregation from multiple file formats.
- ❖ Accountable for estimating cluster size, monitoring, and troubleshooting Spark Databricks clusters.
- ❖ Wrote Spark code in Scala and Spark SQL/Streaming for faster data processing.
- ❖ Implemented Spark Streaming to break down streaming data into batches for processing in Spark Engine.
- ❖ Worked on ETL pipelines outside of the data warehouse using a combination of Python and Snowflake Snow SQL, writing SQL queries against Snowflake.
- ❖ Participated in branching, tagging, and release activities on GitHub version control.

Environment: Azure, ADF, Azure Data Lake Gen2, PySpark, dataflow jobs, copy activity, lookup activity, Data Flow, linked services, logic apps, Event-Hub, Scala, Snowflake, Streaming, Agile methods GIT.

Client: Macy's, New York, NY

April 2019 to Oct 2020

Role: Data Engineer

Responsibilities:

- ❖ Migrated the existing data from Teradata/SQL Server to Hadoop and perform ETL operations on it.
- ❖ Responsible for loading structured, unstructured, and semi-structured data into Hadoop by creating static and dynamic partitions.
- ❖ Worked on different data formats such as JSON and performed machine learning algorithms in Python.
- ❖ Created a task scheduling application to run in an EC2 environment on multiple servers.
- ❖ Strong knowledge of various Data warehousing methodologies and Data modeling concepts.
- ❖ Created Hive partitioned tables using Parquet&Avro format to improve query performance and efficient space utilization.
- ❖ Exposure to Microsoft Azure in the processing of moving the on-prem data to azure cloud.
- ❖ Developed Sqoop scripts for writing the processed data into HBase tables which helps BI team for the data visualization.

- ❖ Used Azure DevOps to build and release different versions of code in different environments.
- ❖ Creation and maintenance of Informatica users and privileges.
- ❖ Exported the aggregated data into RDBMS using Sqoop for creating dashboards in the Tableau and developed trend analysis using statistical features.
- ❖ Experience managing Azure Data Lakes (ADLS) and Data Lake Analytics and an understanding of how to integrate with other Azure Services. Knowledge of USQL and how it can be used for data transformation as part of a cloud data integration strategy
- ❖ Implemented spark and Beam machine learning algorithms and successfully deployed to production.
- ❖ Used CI/CD tools Jenkins, Git/GitLab, Jira and Docker registry/daemon for configuration management and automation using Ansible.
- ❖ Created and Managed Storage accounts in Azure Portal and also pipelines, links, datasets in Azure Data Factory (ADF)
- ❖ Created Containers in Docker. Moving ETL pipelines from SQL server to Hadoop Environment and worked on GLBA.
- ❖ Used advanced SQL methods to code, test, debug, and document complex database queries.
- ❖ Report at daily Scrum of Scrum meetings also ensured the development teams are practicing the Agile principles.
- ❖ Design relational database models for small and large applications.
- ❖ Designed and developed Scala workflows for data pull from cloud-based systems and applying transformations on it.
- ❖ Worked with integration with marketing cloud for user access management.
- ❖ The ability to develop reliable, maintainable, efficient code in most of SQL, Linux shell, and Python.
- ❖ Implemented Apache-spark code to read multiple tables from the real-time records and filter the data based on the requirement.
- ❖ Stored final computation result to Cassandra tables and used Spark-SQL, spark-dataset to perform data computation.
- ❖ Created Informatica Folders and assigning the permissions to the Folders.
- ❖ Used Spark for data analysis and store final computation result to HBase tables.
- ❖ Experienced on Agile processes and facilitated planning meetings and retrospectives.
- ❖ Troubleshoot and resolve complex production issues while providing data analysis and data validation. Participated in the Scrum master team meetings.

Client: Zen3, Hyderabad, INDIA

May 2016 to Dec 2018

Role: Data Engineer

Responsibilities:

- ❖ Created and presented models for potential holdings to fund managers. Achieved 20% better returns v/s historical performance and Predicted stock price 25% better than traditional figures.

- ❖ Facilitated User Acceptance Testing and identified KPIs in the Reference data to predict failure of client's payments.
- ❖ Fashioned with machine learning algorithms like linear regression, KNN and decision tree for trading problems, estimated machine learning algorithm's performance for time series data (stock price data).
- ❖ Calculated and presented data visualization report for daily returns, cumulative returns, simple moving averages, Sharpe ratio, portfolio value for stock performance optimization.
- ❖ Analyzed large data sets using pandas and used regression models using SciPy to predict future data and visualized them and used SQL to manipulate data
- ❖ Proficient in importing/exporting large amounts of data from files to Teradata and Created Teradata specific physical data models which included primary indexes, secondary indexes and joined indexes.
- ❖ Implemented Dimensional Data Modeling to deliver Multi-Dimensional STAR schemas and Developed Snowflake Schemas by normalizing the dimension tables as appropriate.
- ❖ Worked with data compliance teams, Data governance team to maintain data models, Metadata, Data Dictionaries; define source fields and its definitions.
- ❖ Used HIVE to do transformations, event joins and some pre-aggregations before storing the data onto HDFS.
- ❖ Used Redshift, S3 within AWS along with Informatica Cloud Services to load data into s3 bucket.
- ❖ Worked with HDFS file formats like Avro, Sequence File and various compression formats like Snappy.
- ❖ Collaborated with Data engineers and operation team to implement ETL process, Snowflake models, wrote and optimized SQL queries to perform data extraction to fit the analytical requirements.
- ❖ Written SQL Scripts and PL/SQL Scripts to extract data from Database and for Testing Purposes.
- ❖ Created data pipeline package to move data from amazon S3 bucket to MYSQL database and executed MySQL stored procedure using events to load data into tables.
- ❖ Designed and developed Tableau graphical and visualization solutions based on business requirement and expert with the backend data retrieval team, data mart team to guide the proper structuring of data for Tableau reporting.
- ❖ Worked on metadata in Tableau desktop to alter data types, assign roles, rename fields, joins, refresh data and change extract data options to prepare data sources to be utilized by users.

Environment: Hadoop, ETL operations, Data Warehousing, Data Modelling, Cassandra, AWS Cloud computing architecture, EC2, S3, Advanced SQL methods, NiFi, Python, Linux, Apache Spark, Scala, Spark-SQL, HBase

Client: PayPal- Hyderabad, IND

Mar 2014– May 2016

Role: Data Analyst

Responsibilities:

- ❖ Analyzed business requirements, facilitating planning and implementation phases of the OLAP model in Team meetings.
- ❖ Participated in Team meetings to ensure a mutual understanding with business, development and test teams.
- ❖ Encapsulated frequently executed SQL statements into stored procedures to reduce the query execution times
- ❖ Created SSIS packages to implement error/failure handling with event handlers, row redirects, and loggings.
- ❖ Managed packages the in SSISDB catalog with environments; automated deployment and execution with SQL agent jobs
- ❖ Worked on design, development, customization, Implementation, Integration and administration of MS
- ❖ Dynamics 365 solutions.
- ❖ Involved in the design of Data-warehouse using Star-Schema methodology and converted data from various sources to Sql tables
- ❖ Obtained user approvals from the client for the collected requirements to ensure similar understanding between development team and business.
- ❖ Extensively utilized SSIS packages to create complete ETL process and load data into database which was to be used by Reporting Services
- ❖ Designed and developed various SSIS packages (ETL) to extract and transform data and involved in Scheduling SSIS Packages.
- ❖ Implemented, customized and maintained Dynamics Field Service for First-line workers and trained
- ❖ users to use Microsoft Field Service and PowerApps mobile apps
- ❖ Migrated data from Dynamics CRM On-premises to Dynamics 365 - Customer Engagement Plan
- ❖ Created ETL metadata reports using SSRS, reports include like execution times for the SSIS packages, Failure reports with error description.
- ❖ Identified the dimension, fact tables and designed the data warehouse using star schema
- ❖ Developed Multi-Dimensional Objects (Cubes, Dimensions) using SQL Server Analysis Services (SSAS).

Environment: MS Office 2007, MS Excel, SharePoint, PowerPoint, MS Project, SQL Server, Erwin, Business Objects, MS Outlook