



HARSH VARDHAN

DATA ENGINEER

+91-9065420367 vardhan76547@gmail.com Bengaluru 4 Years of experience

PROFILE SUMMARY

Driven, analytical, and proactive data engineer with 4 years of Experience. Enjoying designing and building systems for collecting, storing, and analyzing data. Have experience on Domains like Healthcare, Retail (Fashion & Lifestyle) , Learning & Development

KEYSKILLS

- Python
- Pyspark
- ETL
- DevOps
- Azure Data Lake
- JIRA
- Microsoft Fabric
- Visual Studio
- SQL
- Spark
- Data Warehousing
- Machine Learning
- Data Lake
- Big Data
- Artificial Intelligence
- Microsoft Azure
- Power BI
- Ci - Cd
- Azure Databricks
- Azure Data Factory
- SQL Server
- Copilot

CERTIFICATIONS

- Microsoft Certified: Azure Data Engineer Associate
- Microsoft Certified Azure Fundamentals
- Microsoft certified: azure ai fundamentals
- Microsoft Certified Power BI data analyst
- Microsoft certified: azure data fundamentals

WORK EXPERIENCE

Data Engineer - Infinite Computer Solutions

Aug 2024 - Present

Infinite Computer Solutions

- Developed a robust ETL pipeline utilizing Azure Data Factory, efficiently extracting, transforming, and loading data from Salesforce into an Azure SQL Database.
- Implemented Slowly Changing Dimensions (SCD) Type 2 using PySpark and SQL in Databricks, minimizing data discrepancies by 90%.
- Designed and created a comprehensive Power BI dashboards as per Business & Project Requirement.
- Automated data quality checks and implemented error handling in Databricks notebooks, decreasing manual intervention by 85% and significantly improving data reliability for downstream analytics.
- Leveraged Temporal Tables in SQL Server for automatic change data capture, streamlining data management processes.
- Utilized Azure DevOps and AutoSys for improved orchestration and scheduling of data workflows.
- Enhanced data processing efficiency through effective collaboration within cross-functional teams.

Data Engineer - Happiest Minds Technologies

Aug 2022 - Aug 2024

Happiest Minds Technologies

- Developed and maintained scalable ETL/data pipelines utilizing Azure Data Factory (ADF), Azure Databricks, PySpark, and Spark SQL, ensuring efficient enterprise-level data processing.
- Built and optimized over 50 high-performance data integration pipelines in ADF to handle structured and semi-structured data from diverse sources and formats, including Parquet, JSON, and CSV.
- Automated complex ETL workflows across billions of rows of data, achieving a 67% reduction in manual effort through innovative use of ADF, Azure Databricks, PySpark, T-SQL, and Spark SQL.
- Designed and implemented robust data ingestion and transformation processes for batch data pipelines within cloud-based environments, enhancing overall data management.
- Developed and fine-tuned Spark SQL queries and Databricks notebooks for extensive data exploration, transformation, validation, and file format conversion.

PROJECTS

Duals_Availity (Molina Healthcare)

3 Months

- Developed and managed project stories for Azure Data Factory (ADF) at Molina Healthcare, boosting data integration across multiple states.
- Collaborated with cross-functional teams to design and implement Databricks solutions, enhancing data processing and analytics workflows.

- Conducted thorough testing and validation of data pipelines, ensuring accuracy and reliability within healthcare data processing systems.
- Engaged in continuous improvement initiatives, identifying optimization opportunities that led to a 35% reduction in costs and a 33% increase in processing speed.
- Successfully deployed the data pipeline in production following unit testing, QA sign-off, and PO review.
- Utilized Autosys for efficient scheduling of pipelines, ensuring timely availability of critical data.
- Leveraged Linked Servers in SQL Server Management Studio (SSMS) to effectively load data into Availability tables, improving data accessibility.

ANG Data Migration (Molina Healthcare)

5 Months

- Led the ANG Data Migration project for Molina Healthcare across multiple states, including MI, CA, MA, AZ, ID, OH, and IL.
- Developed Databricks Notebooks utilizing PySpark to fulfill complex business requirements efficiently.
- Designed and implemented Azure Data Factory (ADF) pipelines based on detailed Jira Stories to enhance data integration processes.
- Achieved a 33% optimization of ADF pipelines and improved processing speed by 35%.
- Reduced defects by 75% through rigorous unit and negative testing, assisting in QA validation.
- Managed RFC pipeline deployment, Notebook deployment, and DB script deployment along with creating and activating Autosys jobs.
- Conducted internal reviews with team leads and facilitated business code reviews, ensuring alignment with project goals.
- Ensured all pipelines and notebooks were successfully running in production via Scheduler Autosys.

CIO Mandate (Molina Healthcare)

7 Months

- Executed the migration from SSIS to Azure Data Factory (ADF) for Molina Healthcare, significantly enhancing data processing efficiency.
- Developed and optimized data pipelines in ADF, ensuring seamless data flow and integration across multiple systems.
- Utilized SQL to code and create stored procedures, effectively supporting complex data transformation processes. - Leveraged Autosys for efficient scheduling of ADF pipelines, improving operational time management.
- Achieved a 50% reduction in defects through rigorous testing and continuous optimization of the migration process.
- Reduced operational costs by 30% and increased processing speed by 40% by optimizing ADF pipelines through parallel execution techniques.
- Successfully migrated 22 SSIS packages to ADF, deploying them in a phased manner while ensuring minimal defects.
- Implemented comprehensive unit test documentation, resulting in all 22 packages being successfully operational in production.

UM Data Migration CT (Molina Healthcare)

5 Months

- Led the successful UM Data Migration project for Molina Healthcare, integrating QNXT with Emblem from June 2025 to October 2025.
- Developed and managed project stories within Azure Data Factory (ADF), enhancing data integration across multiple states.
- Collaborated with cross-functional teams to design and implement Databricks solutions, optimizing data processing and analytics workflows.
- Conducted comprehensive testing and validation of data pipelines to ensure accuracy and reliability in healthcare data processing systems.
- Identified and implemented continuous improvement initiatives, achieving a 35% reduction in costs and a 33% increase in processing speed.
- Successfully deployed the data pipeline in production after thorough unit testing, QA sign-off, and PO review. - Utilized Autosys for efficient scheduling of data pipelines, ensuring timely access to critical data.
- Leveraged Linked Servers in SQL Server Management for enhanced database connectivity and performance.

Duals (Molina Healthcare)

8 Months

- Developed over 30 Azure Data Factory (ADF) pipelines for Molina Healthcare, enhancing data integration across multiple regions for Duals States.
- Managed project stories within ADF, streamlining data flow and fostering collaboration among cross-functional teams.
- Designed and implemented optimized Databricks solutions, significantly improving data processing and analytics workflows.
- Conducted comprehensive testing and validation of data pipelines, ensuring accuracy and reliability in healthcare data systems.
- Identified and executed cost-saving and performance initiatives, leading to a 38% reduction in costs and a 32% increase in processing speed.
- Successfully deployed data pipelines in production, adhering to rigorous unit testing, QA sign-off, and PO review processes.
- Utilized Autosys for efficient scheduling of data pipelines, ensuring timely access to critical datasets.
- Achieved 100% data match accuracy through meticulous data handling and validation.

QNXT Upgrade (Molina Healthcare)

5 Months

- Led the QNXT Upgrade project for Molina Healthcare, enhancing system performance and functionality.
- Developed and optimized stored procedures to meet diverse business requirements, improving data retrieval efficiency.
- Diagnosed and resolved defects within SSIS packages, ensuring seamless data integration processes.
- Utilized SQL and SSIS technologies, demonstrating strong technical skills throughout the project lifecycle.

Inovalon (Molina Healthcare)

6 Months

- Developed and maintained ETL pipelines leveraging PySpark, SQL, and Python within the Azure Databricks environment.
- Conducted data ingestion from multiple source systems into the Databricks File System (DBFS) for improved data access.
- Implemented comprehensive data transformations across Curated, Staging, and Consumption layers, enhancing overall data integrity.
- Ensured high standards of data quality through meticulous cleansing and validation processes for reporting and analytics.
- Managed large-scale schema enhancements, increasing the number of columns from 200 to over 1300, significantly improving data richness.
- Oversaw code deployment processes from DEV/UAT to PROD environments, utilizing Azure DevOps pipelines for optimized workflows.
- Created and maintained analytical notebooks and workflows to streamline data processing and support efficient reporting tasks in Databricks.
- Executed defect tracking, testing, and issue resolution using Autosys and Jira.

Ascend Learning

12 Months

- Developed 40+ Azure Data Factory (ADF) pipelines, significantly enhancing data ingestion and processing workflows for Ascend Learning.
- Executed complex SQL queries and optimized stored procedures using SQL Server Management Studio (SSMS), leading to improved overall database performance.
- Created customized Databricks notebooks to meet specific business requirements, facilitating efficient data processing operations.
- Conducted ETL transformations and developed insightful dashboards in Power BI, utilizing DAX queries for dynamic reporting capabilities.
- Enhanced data processing efficiency by implementing PySpark code for data extraction and transformation tasks.
- Utilized Python for parsing XML files and performing data transformations with Pandas, ensuring 100% data accuracy and integrity.
- Achieved a 30% optimization in pipeline performance and a 20% cost reduction through strategic improvements.
- Implemented parallel processing techniques, reducing processing time by 33%.

Forever New

12 Months

- Conducted a comprehensive analysis of SQL Agent Jobs, evaluating schedule time, job steps, packages, tables, and stored procedures utilizing SQL Server Management Studio on Remote Desktop.
- Developed and deployed SSIS packages within the solution using Visual Studio, emphasizing data capture from tables, stored procedures, and data flows.
- Created and managed ingestion pipelines in Azure Data Factory, facilitating effective data ingestion from on-premises systems.
- Built orchestration pipelines in ADF, leveraging insights from SSMS and SSIS to enhance data processing efficiency.
- Designed and implemented Databricks Notebooks utilizing a Multi-Hop Architecture (Bronze, Silver, Gold) with distinct environments for Development and Production, optimizing data transformation and business logic application.
- Validated Databricks Notebooks against SSMS outputs, reinforcing data integrity and consistency.
- Developed Views which was later used for PowerBI dashboard creation.

EDUCATION

2022

B.Tech / B.E. | Computer Science and Engineering (CSE)

Biju Patnaik University of Technology (BPUT)

ADDITIONAL INFORMATION

Languages: English, Hindi