Vishal Vallapu

+1 (940)-758-4780 | vishalvallapu@gmail.com | LinkedIn | GitHub | Portfolio |

PROFESSIONAL SUMMARY

Innovative data engineer and problem-solver with around 4 years of experience in distributed software development, data processing, data transformation and visualisation. Demonstrated expertise in big data technologies, machine learning. Proven track record of thriving in fast-paced, global environments while delivering creative solutions through cross-functional collaboration.

EDUCATION

Masters in Science, Advanced Data Analytics-University of North Texas,Texas,USA-GPA-3.8/4   **August 2022-December 2023**
Bachelors in Technology ,Computer Science - J.N.T.U Hyderabad, India-GPA-3.3/4   **August 2014-May 2018**

SKILL SET

Programming Languages: Python

Technical Skills: Risk Management, Big Data Analytics, Data Engineering, Distributed Systems, Security Implementation, Data Visualization, Machine Learning, Data Curation, ETL/ELT Processes, Pyspark, databricks, CI/CD Pipelines, Data Transformation
Data Visualization: Power BI, Tableau
Cloud Platforms: Google Cloud Platform GCP(BigQuery, Compute Engine, Dataproc, Looker Studio), Azure (ADF, Databricks, Virtual Machines, Azure Storages and Databases), Amazon Web ServicesAWS(Redhsift,Lambda,Kinesis,EMR)
Database Management: SSIS ,SSRS, SSMS, MySQL, PostgreSQL,Cosmos DB, Vector databases(Pinecone)

PROFESSIONAL EXPERIENCE

Technology Crest Corporation, USA   **February2024-Present**
Data Engineer
- Built a robust data ingestion system using Python and Cloud Functions to fetch data from APIs, CSVs, and databases, reducing manual data handling.
- Worked on RCM and building KPI's for patients payback to the healthcare firm by analyzing EMR, Client's data.
- Innovated and implemented distributed software solutions for data access and delivery, improving system efficiency by 30%
- Led cross-functional teams in developing creative solutions for complex data engineering challenges
- Designed secure data pipelines incorporating risk management protocols and security paradigms
- Collaborated with global teams to create innovative mobile-first data visualization solutions
- Created stored procedures and optimized SQL queries data structures to streamline data transformation processes, enhancing efficiency and reducing execution times.
- Utilized Azure Databricks, CI/CD, and GitHub for rigorous data validatiion, testing, and collaborative code management in Scrum sprints.

Winfocus Solutions Pvt Ltd   **October 2018-July2022**
Software Engineer
- Spearheaded the SLAM (Software for Locomotive Asset Management) project for the Ministry of Railways, optimizing locomotive shed operations and reducing maintenance time by 20%, while enhancing locomotive availability by 25%. Leveraged advanced data engineering techniques with Python, SQL, and Tableau to extract, clean, and transform data, enabling precise trend identification and informed decision-making.
- Conducted data quality assessments and rigorous testing on diverse data sources, ensuring Data Management accuracy in data migration, integration for real-time analytics.
- Developed and implemented ETL Pipelines to stream continuous data from sensors, ensuring real-time data collection and designed and scheduled data processing workflows to handle large volumes of streaming data efficiently. Stored processed data in a relational database, ensuring data integrity and accessibility for analysis.

CERTIFICATIONS

- Certified by **Microsoft** : Azure Data Engineer Associate
- Certifications by **AWS** : Fundamentals of Analytics on AWS, Amazon EMR, Amazon Kinesis, Data Analyst Quest Essentials on AWS, Planning on large data migration to AWS
- Certifications by **Deeplearning AI** : Building Gen AI applications with Gradio, Streamlit, Building and Evaluating advanced RAG
- Certifications by **Google** : Introduction to Generative AI, Large Language Model, Responsible AI by Google, Image generation, Attention mechanism, Transformers and BERT model, Encoder-Decoder architecture

---

KEY PROJECTS

- **Virtual Internships:** Price Water house Coopers PWC
  - ➤ Power BI(Analyzing Customer behavior, Customer Analytics, Customer Churn analysis)
  - ➤ Completed a job simulation where I strengthened my PowerBI skills to better understand clients and their data visualization to improve business performance.
  - ➤ Demonstrated expertise in data visualization through the creation of Power BI dashboards that effectively convey KPIs Accenture in Data Analytics and Visualization(Data preprocessing,analysis and visualization) by Forage.

- **Azure end-to-end Data engineering project :** The use case for this project is building an end to end solution by ingesting the tables from an on-premises SQL Server database using Azure Data Factory ADF and then storing the data in Azure Data Lake ADLS. Then Azure databricks is used to transform the RAW data to the most cleanest form of data and then ASA to load the clean data creating data pipelines and finally using Microsoft Power BI to integrate with Azure synapse analytics to build an interactive dashboard. Azure Active Directory (AAD) Azure Key Vault for the monitoring and governance.

- **Netflix Data analysis**: Created an interactive dashboard using Netflix dataset, analyzing top genres, descriptions, duration, release year, and various other variables in Tableau.

- **Azure Datafactory pipeline challenge**: Azure Data Factory ADF Pipelines to prepare log data for analysis in Azure Synapse Analytics.The log data must be available in a specific storage account, blob container, and folder for Azure Synapse Analytics.

- **Baggage Handling System Optimization for American Airlines**
  - ➤ Aimed to address the challenges faced by an airline company with a shortage of personnel and equipment in its BHS.
  - ➤ Predict and analyze the number of baggage items that will arrive for any given outbound flight in Python.

- **Apache Airflow end-to-end data pipeline**
- **End to end data engineering with DBT**
- **End to end Pyspark data engineering project**