



HARI JANARDHAN VANKA

[GitHub](#) | [LinkedIn](#) | [SalesForce Trailblazer](#) | <https://www.harijanardhanvanka.info/>

1614 15th Ave S, Apt G, Birmingham, Alabama – 35205 | harijanardhanvanka@gmail.com | +1 205-370-8266

SUMMARY

- A Software Engineer with 3 years of experience working on Data Analytics principles and Big Data/Hadoop Eco System.
- Extensive experience in developing stored procedures, functions, Views and Triggers, Complex queries using SQL Server, TSQL and Oracle PL/SQL.
- Experience on analysing services on hospital claims, insurances data, and expertise in using ROS (Revenue Optimization Service) to analyse, process and provide required data.
- Developed Tableau reports & effective dashboards after gathering and translating end user requirements.
- Expertise in building and maintaining ETL Batch and Kafka streaming pipelines using Pyspark, Python and SQL.
- Extracted data from SQL Server Database, copied into HDFS File system and used Hadoop tools such as Hive to retrieve the data required for building models.
- Worked on pre-processing the data from ISA to Parquet, build jar files which runs the model cluster on AWS.
- Certified Cloud Infrastructure Associate with profound understanding of data architecture in the Cloud.
- Experience in developing data views using CRON Jobs to generate reports and build dashboards using BI tools.
- Expertise in writing SQL Queries and procedures to generate DML Scripts that modified database objects dynamically based on inputs. Created SSIS package to import and export data from various CSV files, Flat files, Excel spread sheets and SQL Server.

EDUCATIONAL DETAILS

Master's in Computer & Information Sciences

University of Alabama at Birmingham

[Jan 22 – May 23]

Bachelor's in Computer Science & Engineering

Jawaharlal Nehru Technological University Hyderabad

[Aug 16 – Sep 20]

WORK EXPERIENCE

Data Engineer

Jan 2023

SnowStack

- Ingested data from different data sources using a combination of SQL, Pyspark, and Python and created views and temp tables to be consumed by BI Tools
- Functioned complex SQL Queries, Stored Procedure, Triggers, Views & Indexes using DML, DDL commands and user defined functions to implement the business logic.
- Design dashboards and data visualization using the DAX command to communicate meaningful metrics using Power BI.
- Built ETL pipelines and performed ETL operations over the raw data by including data extraction, transformation, and manipulation.
- Optimize, and create functions or processes to load the data from Pyspark data frames into the hive or Impala tables.

Graduate Student Assistant

Aug 2022

University Of Alabama at Birmingham

- Functioned as a Teaching Assistant for the Capstone Project Course, upskill the students with Python, SQL.
- Resolve queries from students who are working on projects with whole usage and following SDLC process.

Software Engineer

Mar 2021

Innova Solutions of ACS (American Cyber Systems) Group

- Project is to analyse services on hospital claims, where insurances are underpaid it, where ROS (Revenue Optimization Service) team Pitches in to analyse, process and provide required data to hospital decline by insurance. ROS collects all the documents and try to show required reports to get back funds from insurance.
- As a Software Engineer implementing Data Engineer principles in healthcare project working with Revenue Optimization service pipeline.
- Stream documents to AWS s3 bucket using services like EMR for data processing using Apache Spark, glue's crawler.
- Migrate data from ISA to HDFS, apply transformation techniques and save the data to a parquet file.
- Used Teradata SQL to develop metrics to derive business solutions and for budgeting and forecasting.
- Automate Data retrieval process from Oracle DB using FTP and SQOOP.

- Implemented AWS lambda functions to accelerate and improve performance of the data upstream and notify using AWS SNS service.
- Developed ETL Pipelines and performed regression and unit testing, to handle redundancies in the pipeline.

ETL Developer (Salesforce)

September 2020

Trailhead systems

- Experience on using Salesforce.com setup, Configuration, Customization, Administration, Data Migration, Mapping, Data Modelling designing of Custom objects, Custom components, Custom fields, Role based page layouts, Custom Tabs & Custom Application.
- Migrate and integrate data with databases using Data loader for development and customizations of Salesforce.com applications using Visualforce, Apex classes, Apex triggers, Apex testing, Salesforce web-based integrated development environment (IDE).
- Created Dimension Tables and Fact Tables based on the warehouse design.
- Wrote Triggers and Stored Procedures using PL/SQL for Incremental updates.
- Integrated various sources into the Staging area in Data warehouse to Integrate and for cleansing data. Involved in building of the Data Warehouse which includes the Design of Data mart using Star Schema.

INTERNSHIPS

IOT Developer - Internship on Internet of Things (IOT)

May 2019

The SmartBridge Services Pvt Ltd

- Intelligent Access Control System for Safety Critical Areas in Industries.
In some industries it is necessary for the workers to wear safety helmets and shoes while working. So, to check whether workers are taking safety precautions or not we are proposing this system.

PROJECTS / IMPLEMENTATIONS

Data Science and Machine Learning Implementations

Stock Market Real Time data Analysis using Kafka_ → [GitHub](#)

- Implemented a streaming pipeline using Apache Kafka to stream the whole data into the Aws bucket in real time scenario and then used AWS Glue, Crawler and Athena to query and manipulate the data.

YouTube Data Analysis → [GitHub](#)

- Streamline YouTube Analysis on structured and semi-structured YouTube Videos data from data ingestion to building ETL pipeline to visualize data in QuickSight.

Credit Risk Analysis_ → [GitHub](#)

- Implemented with risk dataset, where to analyse whether person's loan can be approved or rejected classifying using Random Forest Classifier.

DonorChoose.Org Application Screening → [GitHub](#)

- Donorchoose.org receives thousands of project proposals each year for classroom project in need of funding, where we automate the screening processes so that the results will be posted as quickly and efficiently as possible.

Data Science Process Pipeline to Solve Employee Attrition & Job Performance & Predicting with AI → [GitHub](#)

- This project deals with predicting whether an employee would continue to work in an organization or leave due to various reasons by using data science methodology. By analysing the result, try to understand what the reason and we could be then take steps to rectify them. **Algorithms** – Random Forest classification, Logistic Regression, Support Vector Machine, Multi-layer Perceptron Classifier, Data Science process pipeline. Libraries – Pandas, NumPy, Sci-kit learn, Matplotlib.

Annotations → [GitHub](#)

- **Uber Trajectory** – Annotation is done by using Georeferencer.com to recover the trip information and applied Data Analysis.
- **Floods** - Annotating the satellite images to train ML models for flood extent mapping using segmentation based Topological Data Science and elevated guided BFS concepts.

TABLEAU IMPLEMENTATIONS

Bank Analysis - Data given below is from an imaginary bank operating in the UK. Analysed how the different regions are represented in the customer base. → [Tableau Public](#) , [GitHub](#)

Sales Insights - Atliq Hardware is a company which supplies hardware peripherals to various clients like stores and gains insights from regional branch performances and decides for improvement. → [Tableau Public](#) , [GitHub](#)

Full Stack Development → [GitHub](#)

- **Developed Web Application Called Foodie** where users can order food online. Included features like from ordering food, sign in/up, rate/review, payment methods, order history, favorite food.
Technologies used – Python-Django, PostgreSQL, Html, CSS, Bootstrap, jQuery, JavaScript.

- **Developed Web Application called My Profile** where I used AWS services for deployment, to store structured or unstructured data and send notifications about the data.
Technologies used - AWS {S3 buckets, EC2 Instances, RDS, Lambda, SNS.}, Python, Html, CSS Bootstrap, JavaScript.

TECHNICAL SKILLS

Programming Languages - Python, Java. **Web** - Html, CSS, JavaScript, Django, Flask

Query Languages - MySQL, SQL, SOQL, SOSL. **Databases** - MySQL, Mango DB, SQL server, PostgreSQL, MariaDB

Big Data Technologies - Apache Spark, Spark SQL, Hadoop, Apache Hive, Apache Kafka, Apache Airflow, HDFS, SBT, Postman, MapRGit, Scala, AWS, Teradata, Looker, BigQuery.

Cloud Technologies - AWS - {S3, DynamoDB, RDS, EMR, Athena, Glue, Redshift}, AZURE - Data Lake, Data Bricks, GCP

Visualisation Tools – Tableau, Power BI, SAS. **SalesForce Clouds CRM** - Administrator, Developer

Operating Systems - Windows – [xp, 2007, 10, 11] and Linux

Methodology - Agile

IDE'S - Sublime Text, PyCharm, IntelliJ, Node.js, Anaconda, VS code. **Computational Skills** - MS Office

CERTIFICATIONS

- **Oracle Cloud Data Management, Infrastructure Foundations Associate – Oracle**
- **Databricks Lakehouse Fundamentals, Generic AI Fundamentals.**
- **ML Observability Fundamentals - ARIZE University**
- **SQL & Programming Essentials – SAS**
- **Cognitive Apache Architecture Class – IBM**
- **Machine Learning** course authorized by **Stanford University**.
- **Platform Developer (PD1) – SalesForce Certified**
- **Microsoft Technology Associate Certification** - Python Programming