

# HARI JANARDHAN VANKA

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

1614 15<sup>th</sup> Ave S, Apt G, Birmingham, Alabama – 35205 | [harijanardhanvanka@gmail.com](mailto:harijanardhanvanka@gmail.com) | +1 205-370-8266

## SUMMARY

- Professional experience of 2 years on Automating ETL process by building ETL pipelines to make easier data wrangling.
- Built and maintained real-time data processing using Apache Kafka streaming pipelines ingesting data using s3, python.
- Equipped with Monte Carlo simulation methods, Mathematical and differential equations, Statistics.
- Well Equipped with Credit Risk Management objectives, quality control analysis, credit modelling, credit risk mitigation strategies.
- I had experience with SAS interface by working on SAS procedural SQL, Programming. Also completed multiple SAS certifications
- Experience in using Salesforce CRM cloud with administrative and development works and Ingested data from data sources using combination of SQL, Salesforce API using python to create data views to be used in BI tools.
- Experience on pre-processing the data from ISA to parquet and creating jar files, which runs the model cluster on AWS.

## WORK EXPERIENCE

### SnowStack

Jan 2023

#### Data Engineer

- 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.
- 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 transactional data by including data extraction, transformation and manipulation.
- Extracting data from Splunk API and performing data cleaning and loading into the tables in the desired format.
- Optimize, and create functions or processes to load the data from Pyspark data frames into the hive or Impala tables.

### Teaching Assistant

Aug 2022

#### University Of Alabama at Birmingham

- Working as a Teaching Assistant for the Capstone Project course.
- 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

- Worked as a Software Engineer following **Data Engineer** principles in healthcare project on working with Revenue Optimization service pipeline where it collects all the documents and stream it to AWS s3 using services like EMR for data processing using Apache Spark, glue's crawler to query the data and run code with various services also resolve the data issues and fix bugs.
- Use to pre-process the data from ISA to HDFS and apply transformation technique to save as parquet file, ingest parquet and performed extraction by normalization.
- Retrieve data from client through FTP or Sqoop job the data from oracle db.
- Implemented AWS lambda functions to accelerate and improve performance of the upload and notification alert using AWS S3 buckets in AWS cloud environment.
- Worked on ETL processing and SQL query to get optimized and accelerate the process.
- Performed troubleshooting data-related issues and resolve redundancies in the data.
- Supported technical design process in analysis of technical application requirements.

## INTERSHIPS

### The SmartBridge Services Pvt Ltd

May 2019

#### Internship on Internet of Things (IOT) using IBM Cloud → [GitHub](#)

- 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.

## 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.

## 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.

## 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

**Query Languages** - MySQL, SQL, SOQL, SOSL

**Databases** - MySQL, Mango DB, SQL server, PostgreSQL, MariaDB

**Big Data Technologies** - Apache Spark, Spark SQL, Hadoop, HDFS, SBT, Postman, MapRGit, Scala, AWS

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

**Visualisation Tools** – Tableau, Power BI.

**Web Development Languages** - Html, CSS, JavaScript, Django, Flask

**SalesForce Cloud 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

## EDUCATIONAL DETAILS

---

**Master's in Computer science - Data Science**

[Jan 2022 – May 2023]

University of Alabama at Birmingham

**Bachelor's in Computer Science & Engineering**

[August 2016 – May 2020]

Jawaharlal Nehru Technological University Hyderabad

## CERTIFICATIONS

---

- **ML Observability Fundamentals** - ARIZE University
- **SQL & Programming Essentials** – SAS
- **Cognitive Apache Architecture Class** – IBM
- **Machine Learning** (MOOCS) course authorized by **Stanford University** and offered by **Coursera**.
- **Platform Developer (PD1)** – **SalesForce Certified**
- **Microsoft Technology Associate Certification** - Introduction to Programming using **Python**.