**Adhav Kiran Babasaheb**

# Educational Qualification:

* Bachelor of Engineering in Computer engineer from university of Pune - 2015.

Professional Summary:

* Total 5.9 years of IT experience in the field of Data Engineering, Data Lake, Big Data & ML .
* 3 years of relevant work experience on GCP, Data Science with Python, Machine Learning, Kafka, Elastic search, Hadoop Big Data, Spark, SQL and Java project development.
* Meeting the clients and understanding their requirements, presenting data to leadership team.
* Proficient at working as an individual / team player with an aim of contributing towards the success of the overall project & organization.

# Technical skill set:

Languages: Python, Core Java.

Data Science: Machine Learning, Computer Vision

Hadoop/Big Data: HDFS, Hive, Hadoop distribution of Cloudera. SPARK, Kafka, PySpark

Operating Systems: Windows 10, Linux

Programming Languages: Python, Java, J2EE, Spring.

Tools: Jupyter Notebook, PyCharm,Google Colab

RDBMS Technology: JDBC, MySQL, Mongo DB, PostgreSQL

## Project # 1

**Project details: Data Ingestion (News Corp Australia Client)**

**Duration: Sep 21- Dec 2022**

**Role: Data Engineer**

## Roles and Responsibilities:

* Data transfer from Crowd Tangle to Big Query using ASCEND tool using PySpark.
* Read the URL from Big query table using pyspark. Read fields as per client requirement and store in big query table for final Business process.
* Created queries using final data as per client requirements.
* Created cloud composer jobs.
* Created Data flow batch processing.

# Tools and Technology: Google Cloud Platform (GCP), Big Query Ascend.io, PySpark, Data Flow, Cloud Composer.

## Project # 2

**Project details: Data Migration (US Client)**

**Duration: Jan 2021 – Sep 2021**

**Role: Associate Cloud Engineer**

**Roles and Responsibilities:**

* Understanding the specification given by client.
* Create VM in GCP as per requirement and installing s/w.
* Done Couchbase setup on GCP VM(Linux)
* Done Cassandra setup on GCP VM (Linux)
* Create Cluster on Confluent Kafka cloud and peering with vpc in GCP.
* Add brokerid in DNS as per requirements.
* Create Topics, Producer, Consumer, and Schema Registry in Confluent Kafka.
* Connect Confluent Kafka with Elastic Search.

# Tools and Technology: Google Cloud Platform (GCP), Confluent Kafka, Cassandra, Couchbase DB.

## Project # 3

## Project details: Warehouse Inventory Management using Machine Learning.

## Duration: Aug 2020 – Dec 2020

**Role: Data Engineer**

## Roles and Responsibilities:

* To help client in achieving inventory auditing producer i.e., Cycle count with minimal Human Resources and with time saving prediction.
* Proposed project is to identify cycle count error.
* Machine Learning Model built for prediction of classification cycle count error.
* Classification model is employed.
* ETL activities performed by using Azure SQL and Azure Data Factory.
* Pipelines are created using Azure ML Studio.
* Worked for 3 different warehouses (Chemical, Life Science, Retails)
* Worked for One Infosys internal project on cycle count error prediction using ML for business purpose (Involved in END-END solution)
* Understanding the specification given by client and transforming and computing the data to meet the business specification.
* Performed ETL activities on Azure SQL databases and load data into ML pipeline using Azure datastore.
* Responsible for creating ETL pipeline in Azure Data Factory.
* Responsible for EDA, Feature Engineering and Training model.

**Tools and Technologies:** Python, Machine Learning, Azure Data Factory, Azure Data Studio, Azure ML studio, Azure python notebook.

## Project # 4

## Project details: Diabetes Prediction (Training Project into Infosys)

## Duration: June 2020 – July 2020

## Role: Data Engineer

## Roles and Responsibilities:

* Exploratory Data Analysis.
* Data preprocessing.
* Model Selection and application.
* Create Classification model
* Hyper Parameter

**Tools and Technologies:** Python, Jupyter, Machine Learning.

## Project # 5

## Project details: Real time attendance system (Costa Rica Client)

## Duration: April 2019 – May2020

## Role: Data Engineer

## Roles and Responsibilities:

* To create Spring API as per requirements.
* To work on Spark java and Kafka for reading data from MSSQL server and inserting into PostgreSQL.
* Working on Stream sets and creating pipeline for attendance process.
* Doing manual testing.

**Tools and Technologies:** Spring, Spark, Kafka, PostgreSQL, MSSQL Server, Stream Sets.

## Project # 6

**Project details: Twitter online data streaming using Spring Rest API, Spark and Kafka and Store in Mongo DB.**

**Duration: May2018 –June2018**

**Role: Software Engineer**

## Roles and Responsibilities:

* To create Spring Rest API and configure Kafka and spark.
* To generate Twitter keys online on twitter website.
* To store data successful in Mongo DB.

## Tools and Technologies: Spring, Spark, Kafka and Mongo DB.

## Project # 7

## Project details: Stock Market Prediction (This is my personal project development) In Stock Market Prediction to calculate last 5-year company data and calculate prediction and assume that current stock market company shares values.

## Duration: Feb 2018 – April 2018

## Role: Software Engineer

## Roles and Responsibilities:

* To create Spring Rest API as per client requirements.
* To work on Spark and Hive. Create CSV data files and upload in hive and create hive tables queries.
* Using spark to read and write data and configure beans spring and spark inXML.
* Manual testing.
* To configure project setup on client system.

**Tools and Technologies:** Spring, Spark, Mongo DB, Hive.

## Project # 8

**Project details: Anti-money laundering (AML) is a term mainly used in the financial and legal industries to describe the legal controls that require financial institutions and other regulated entities to prevent, detect, and report money laundering activities.**

**Duration: June 2017 –Jan 2018**

**Role: Software Engineer**

## Roles and Responsibilities:

* To create Spring Rest API as per client requirements.
* To create JSON files as per requirements. E.g. To create Maths Formulas in JSON using Hives Math Functions.
* To work on Spark and Hive. Create CSV data files and upload in hive and create hive tables queries.
* Using spark to read and write data and configure beans spring and spark inXML.
* Manual testing.
* To configure project setup on client system.

**Tools and Technologies:** Spring, Spark, Mongo DB, Hive.

**Project # 9**

**Project details: Worked as a Trainee Software Engineer for “Financial Application Framework” Project. The Framework project use for representing in the financial statements the effect of an entity’s risk management activities when they use financial instruments to manage exposures arising from particular risk and those risks could affect profit or loss.**

**Duration: June 2016 – May 2017**

**Role: Software Engineer**

**Roles and Responsibilities:**

* To create Spring Rest API as per client requirements.
* To create JSON files as per requirements.
* To work on Spark and Hive. Create CSV data files and upload in hive and create hive tables queries.
* Manual testing.
* To configure project setup on client system.

**Tools and Technologies:** Spring, Spark, Mongo DB, Hive.