

## **PAVAN KUMAR REDDY KATTAMEDDI**

BigData Engineer

**Mobile No:** - 7997224662

**Email:** - kpkreddy404@gmail.com

### **Career Objective:**

To achieve high career growth through a continuous learning process and keep myself dynamic and inspiration for others and competitive with the changing scenario of the world.

### **Experience Summary:**

- I Have 4 years of professional experience in IT. Handling Data Warehouse tools.
- Around 2+ years of experience in Hadoop and spark frameworks.
- Good exposure to Software Development Life Cycle following Agile methodologies in Requirement gathering, Design, Application Development in IT systems.
- Good knowledge of installation, configuration, management and deployment of Big Data solutions and the underlying infrastructure of Hadoop cluster.
- In depth knowledge of Hadoop Architecture and various components such as HDFS and Map Reduce components.
- Experience in Pyspark and Spark SQL for processing data in a Python environment.
- Experience in importing data using SQOOP from RDBMS to HDFS and vice-versa.
- Knowledge on SQL Databases like MYSQL.
- Knowledge on job scheduler tools like Oozie etc.
- Experience in Designing, developing and implementing connectivity products that allow efficient exchange of data between our core database engine and Hadoop ecosystem.
- Hands on experience working with IDE, build tools such as Eclipse, putty.
- Experience on Cloudera Hadoop Distribution.
- Hands on Experience with ETL.
- Capable of processing large sets of structured, semi-structured and supporting systems.
- Proficient in client interaction and gathering information related to projects.
- High interest to work in BigData Technologies and Data Analytics.
- Good exposure on BigData solutions in both Hadoop and spark framework.
- Quick learner, keen in learning new technologies and quickly adapting to new environments.

### **Work Experience:**

- Currently working with HTC Global Service as L2 Engineer.
- Previously worked with Netfinity Technologies Pvt Ltd.

## Technical Skills:

BigData : Hive, Sqoop, HDFS, Python, Oozie, Scala, Hadoop, Pyspark and snowflake  
Database : Oracle, MYSQL.  
Other Tools : Putty, Eclipse.  
Hadoop Distribution : Cludera.

## Education:

B.Tech in Computer Science Engineering College at Sree Vidyanikethan Engineering College in 2018.

## Project Details:

### #Project 3:

Project : Transportation Analytics.  
Client : State form  
Role : L2 Engineer.  
Technologies : Python, Hive, Spark, and Oozie, Sqoop ,SQL.  
Duration : Dec 2021 to Aug 2022

## Description:

The purpose of this project is to develop Cost to Deliver capability for the Consumer segment within the region leveraging the Cost to Deliver platform with the goal to create visibility to opportunity areas to further analyze and generate cost savings benefits as well as create visibility for customer and product segmentation.

## Responsibilities:

- Involved in extracting data from various data sources into Hadoop HDFS .
- Worked on creating Hive managed and external tables based on the requirements.
- Provided Customer, Delivery, Delivery line and Product data to downstream system based on the queries.
- Performed data preprocessing validations on the file based on the requirement.
- Generated Exception reports using pandas

## #Project 2:

Project : Internet Banking – Module.  
Client : MUFG Bank, Cayman Islands (Canada).  
Role : Hadoop Developer.  
Technologies : Hadoop, Apache pig, Hive, SQOOP, and Oozie.  
Duration : Oct 2020 to Sep 2021.

### Description:

This is an internet banking application that provides very useful features to bank customers who have opted for such a facility. It is a free registration facility provided by the bank. Once the customers register for an online banking facility they will be provided user id and password to carry out the online transactions in a secure manner.

### Responsibilities:

- Involved in extracting data from various data sources into Hadoop HDFS by using SQOOP.
- Worked on creating Hive managed and external tables based on the requirements.
- Implemented partitioning and Bucketing in Hive Table for better performance.
- Worked on various file formats like Parquet, JSON and ORC.
- Used for Spark-SQL to process the data and to run on the Spark engine.
- Worked on Spark for improving and optimization of existing algorithms in Hadoop using Spark-SQL and Python.
- Experienced in handling large datasets using partition's, Spark in memory capabilities, Effective efficient joins, Transformations and others during the ingestion process itself.
- Worked on Oozie to develop workflows to automate the ETL data pipeline.
- Troubleshoot various software issues using the debugging process and coding techniques.
- Supported code/design analysis strategy development and project.
- Planning Responsible for managing data coming from different sources.
- Monitor multiple tasks on the Hadoop cluster environment.

## #Project 1:

Project : Health Care, Data Migration.  
Role : Developer, Support.  
Environment : Hive SQOOP, HDFS, Pyspark.  
Client : Medplus  
Duration : sep 2018 to Aug 2020

## **Description:**

In this Project, Previously data was in a Relational Database like MYSQL. But now because of handling large volumes of data. We are migrating those data to HDFS location through SQOOP. Sometimes by using Hive, we have to analyze those data and provide them with the different kinds of insurance coverage which suits the patents. Based on the patient's needs, they choose the appropriate coverage for them.

## **Responsibilities:**

- Analyzing Json data with Hive.
- Worked in Partitioning, Bucketing, Join optimizations and Query optimization in Hive.
- Worked on Data processing with Pyspark.
- Handled importing other enterprise data from different sources into HDFS using Sqoop.
- Exported the analyzed data to a relational database using SQOOP.
- Worked with Hadoop cluster using Cloudera Distribution to store data on HDFS.

## **Declaration**

I hereby declare that all the particulars given above are true to the best of my Knowledge

**Pavan Kumar Reddy Kattameddi**