

## Professional Summary

- Having **4.5 Years** of experience in the IT Industry, including **3.5 years** of exclusive experience in **Hadoop** and its components like **HDFS, Hive** and **SPARK CORE, SPARK SQL**.
- Extensive experience in **Hadoop/Spark** ingestion, storage, querying, processing and analysis of Big Data.
- I have experience in **Snowflake** (Snow SQL, data loading, and unloading), Apache **NiFi** (data routing, transformation, and migration), and Apache **Airflow** (workflow scheduler).
- Experience in designing and developing Scripts in **PySpark** for loading the data in Cloudera and better processing performance.
- Involved in loading and transforming large sets of structured, semi-structured data and analyzed them by running Hive queries.
- Good Experience in importing and exporting data from HDFS to Relational Database Systems (**RDBMS**).
- Experience in implementing Spark Core and Spark SQL for faster testing and processing of data.
- Expert in working with Hive data warehouse and Developing the HiveQL queries.
- Knowledge in PYTHON Programming Language.
- Basic knowledge on NOSQL databases like HBase, Cassandra.

## Professional Experience

- Worked as a **Data Engineer** in **Tata Consultancy Services (TCS)** Chennai, India from Oct 2021 to Nov 2022.
- Worked as a **Hadoop Developer** in **Intesome Solutions Pvt. Ltd.** Pune, India from Aug 2019 to Sept 2021.

## Academic Profile

Completed Bachelor of computer science from Dr. Babasaheb Ambedkar Marathwada University, Aurangabad (2011 - 2014).

## Software Skills

- |                         |  |
|-------------------------|--|
| • Big Data & Eco-System | : HDFS, Hive, Sqoop, Spark, SQL, Kafka, Nifi, Airflow. |
| • NoSQL Databases       | : HBase, Cassandra.                                    |
| • Database Languages    | : Snowflake, Oracle Database, MySQL                    |
| • Operating Systems     | : Linux (Ubuntu), Windows                              |
| • Programming Languages | : Python.  |

## Project Details

### Project -1 (TCS)

Project Name : Netezza Migration  
Client : Marriott International, USA  
Environment : Oracle, Snowflake, Nifi, Airflow, AWS, Spark.  
Role : Data Engineer

#### Description:

This Project Involves migration of data from SQL database (Netezza, Oracle) to Snowflake.  
Project is divided into multiple parts.

- 1) Ingestion of data from existing databases (Netezza, Oracle) or fileserver to Amazon S3 bucket using Apache Nifi /Apache Kafka.
- 2) Copying data from Amazon S3 to Snowflake Stage.
- 3) Converting Netezza SQL /Oracle SQL to Snowflake.
- 4) Create a DAG using Airflow to periodically insert data in the Snowflake table.

#### Roles and Responsibilities:

- Ingesting data from SQL database (Netezza / Oracle) or File server to Amazon S3 bucket.
- using Apache NiFi, Copying data from Amazon S3 bucket to Snowflake Stage.
- Writing a script to call a Snow SQL file.
- Schedule the JOB using Apache Airflow.
- Do end to end data validation for migrated tables.
- Creates the data mapping document.

### Project -2 (Intesome Solutions Pvt. Ltd)

Project Name : JSK-Business Analytics  
Client : JSK LOGISTICS LLC, Metuchen, NJ, USA.  
Environment : Hadoop, Spark, Hive, UNIX, MySQL.  
Role : Hadoop and Spark Developer

#### Description:

The purpose of this project is to address the challenge faced by logistics providers in managing the vast amount of data generated while handling the movement of goods and people. With millions of shipments being processed worldwide each day, various types of information such as location, content, size, weight, origin and destination need to be tracked and gathered across global delivery and transport networks. However, traditional databases are not efficient in handling and scaling this large volume of data.

To overcome this challenge, JSK Logistics aims to re-platform their current data warehouse system to a Hadoop solution, which offers a cost-effective solution to manage and scale this large volume of diverse data.

## **Roles and Responsibilities:**

- Import all data from RDBMS to HDFS for further processing.
- Writing the script files for processing data and loading to HDFS.
- Develop the Sqoop commands for import to the HDFS data.
- Created Hive tables to store the processed results in a tabular format.
- Batch processing of data sources using Spark.
- Implemented Spark RDD transformations, actions to implement business analysis.
- Implemented partitioning, dynamic partitions and buckets in HIVE.
- Exported hive tables data into MySQL database through the spark SQL.

Nitin Patil