

---

# Madhavi Mukesh Singh

Email: [madhvims11@gmail.com](mailto:madhvims11@gmail.com)

Ph.no: 8796693893

Location: Gurugram, Haryana

## SUMMARY

---

1.2 years of Experience in different domains in IT industry including **Big Data – Hadoop Development** and **SQL**. Loading data from different datasets and deciding on which file format is efficient for a task. Understanding the requirements of input to output transformations. Defining Hadoop Job Flows. Have strong knowledge on Apache Sqoop. Hands on experience in various Big Data application phases like data ingestion, data analytics and data lake. Experience in using Hadoop distribution like Cloudera. Experience in transferring data from RDBMS to HDFS and Hive table using Sqoop. Experience in creating tables, partitioning, bucketing, loading and aggregating data using Hive.

## TECHNICAL SKILLS

---

- **Platforms:** Windows, Linux
- **Programming Languages:** Scala, Python, .Net with MVC
- **Big-Data Ecosystem:** HDFS, YARN, Map-Reduce, Sqoop, Hive, Spark
- **Database:** SQL, Oracle
- **Other:** Azure Cloud, PowerBI, ML

## EXPERIENCE

---

**Software Engineer (January 2022 – February 2023)**

**Capgemini, Pune**

**Project: HDFS Data Ingestion Project**

**Technologies used:** Hive, Sqoop, Hadoop, MapReduce

**Role: Data Engineer**

In this project we migrated large number of our client's data sources and datasets into a Data Lake. This involved building a data ingestion framework that supports three ingestion patterns: Database ingestion, file-based ingestion and incremental data load ingestion.

**Roles and Responsibilities:**

- Possess an in-depth knowledge of Hadoop Architecture and its key components, including HDFS, Application Master, Node Manager, Resource Manager, Name Node, Data Node, and MapReduce components.
- Import essential tables from relational databases (RDBMS) such as Oracle and SQL Server into HDFS utilizing Sqoop.
- Involved in the creation of Hive tables, data loading, and crafting Hive queries designed to run efficiently in a MapReduce environment.
- Utilize Hive for the analysis of partitioned and bucketed data, contributing to insightful data-driven decision-making. Proficiently import and export data between HDFS and Hive, leveraging Sqoop for seamless data transfer.
- Work with various HDFS file formats like Avro, Sequence File, and employ compression formats such as Snappy for efficient data storage and retrieval.
- Set up the SFTP protocol between the source server and the edge node/landing zone, ensuring secure and reliable data transfer.

## EDUCATION

---

- **2017-2020 - B.E (Information Technology)** - PVG'sCOE, Nashik.
- **2014-2017 - Diploma (Information Technology)** - GGSP, Nashik.
- **2013-2014 - 10<sup>th</sup> (SSC)** - Sukhdev Madhyamik Vidyamandir, Nashik.

## Certification

---

- Microsoft Certified: Azure developer associate.
- Machine Learning (Internshala)
- AI (Dell)

