

## SUMMARY

Having Professional Experience in Big Data/Hadoop and Spark Development, having a strong background with File distribution system in a big data arena. Understands the complex processing needs of big data and has experience developing codes and modules to address these needs

## HIGH-LEVEL SUMMARY OF MY RESUME

- Having **3.2** years of Experience in **Big Data and Hadoop**.
- Extensive working experience on **Hadoop** eco-system and its components **Spark, HDFS, Map Reduce, Hive, Sqoop, PySpark , HBase**.
- Experience in building **Hive** and **Pyspark** scripts.
- Designing and creating **Hive external tables** using shared meta-store instead of derby with partitioning, dynamic partitioning and buckets.
- Experience in writing **HiveQL** queries to store processed data into **Hive tables** for analysis.
- Experience in developing scripts in Python and **Spark-SQL** for faster testing and processing of data. Import the data from different sources like **HDFS/HBase** into **SparkRDD**.
- Having good exposer on **Spark (RDD)** and its working functionality.
- Experience with all flavor of **Hadoop distributions, including Cloudera, Horton works**.
- Experience in importing and exporting data using **Apache Sqoop** from **Relational Data Base Systems** to **HDFS** and vice-versa.
- Experience in writing **HiveQL** queries to store processed data into **Hive tables** for analysis.

## TECHNICAL SKILLS SET

Operating Systems:

Languages Known:

Database:

Hadoop Tools:

## TOOLS USED

Windows, UNIX , Linux

Python, SQL, Scala, Java

MySQL, Oracle,

SPARK, SCALA, HDFS, HIVE , Map Reduce, Hbase ,  
Sqoop, PySpark .

## PROFESSIONAL WORK EXPERIENCE

**Company Name-** ADROSONIC IT Consultancy Services, Mumbai

**Designation-** Data Engineer from Oct 2019 to till date

### PROJECTS HANDLED:

#### PROJECT SUMMARY

##### Project # 1

Domain Type : Retail

Roles and Responsibilities:

- Understand the requirements and clarify the doubts with BA.
- Working in a team of Hadoop developers for getting data from MySQL to Hadoop using PySpark.
- Capturing data from existing databases that provide interfaces using Python.
- Used PySpark to Code the Spark Application.
- Used Python Module To Connect and Obtain Data From MySql Data Source.
- Involves developing shell scripts for automation, creation of hive queries for data transformation, providing hive optimization.
- Techniques, deploying solution in production. Development of Spark jobs for data transformations.
- Writing the script files for processing data and loading to HDFS.
- Completely involved in the requirement analysis phase.

##### Project # 2

Domain Type : Banking

Roles and Responsibilities:

- Involved in gathering the requirements, designing, development and testing.
- Writing the script files for processing data and loading to HDFS.
- Loading files to HDFS and writing hive queries to process required data.
- Involved in Partitions of Hive tables. Creating Hive tables to store the processed data in tables.
- Setup Hive with MySQL as a Remote Meta store.
- Moved all log/text files generated by various products into HDFS location.
- Investigate and analyze alternative solutions to data storing, processing etc. to ensure most streamlined approaches are implemented.
- Designed both Managed and External Hive Tables and Defined static and dynamic partitions as per requirement for optimized performance on production datasets.

## EDUCATION

Degree & Date	University	Specialization	Year
Bachelor Of Engineering [B.E]	Dr. Babasaheb Ambedkar Marathwada University, Aurangabad	Computer Science and Engineering	2014

## PERSONAL DETAILS

Date of Birth : 27<sup>th</sup> June 1993

Sex : Male

Marital Status : Unmarried

Languages Known : English, Hindi and Marathi.

Address : Nehru Nagar, Kanjurmarg East ,Mumbai - 400042

PLACE: Mumbai

DATE: 19/12/2022