

Professional Summary:

- **3.3 years of Experience** in different domains in IT industry including Big Data Hadoop Development.
 - Have hands on experience Spark Architecture, SparkCore, SparkSql.
 - Understanding of Partitioning, Bucketing concepts in Hive and designed both Internal and External tables in Hive to optimize performance.
 - Experience in creating tables, partitioning, bucketing, loading and aggregating data using Hive.
 - Uploaded and processed gigabyte, terabytes of data from various structured sources into Hive.
 - Understanding the requirements of input to output transformations.
 - Also having knowledge of scheduler Airflow.
 - To sound with Oozie, Git, Postgre, SQL server.
 - Have experience writing data extraction logic in pyspark.
 - Written Hive Queries for Data Analysis to meet the business requirements.
 - Implement Data Validation, Quality Checks, Profiling.
-

ORGANISATIONAL EXPERIENCE:

Organization: Larsen & Toubro Infotech pvt. Ltd.Pune
Designation: Senior Data Engineer
Address: LTI - Hinjewadi VI,Pune-7th Floor,IT-9,Pune
Star Date : 18th April 2022 to 3rd August 2022

Organization: Genex Corporate Services Pvt. Ltd
Designation: Storage & Backup Management L1
Client Site: VodafoneIdea Ltd., Kalyani Nagar, Pune.
Duration: April 2021 to December 2021

Organization: EC Mobility Pvt. Ltd. Aurangabad
Designation: Data Engineer
Address: A-2, Ground Floor, Aurangabad Business Centre,
Adalat Road, Aurangabad – 431005
Duration: 22 October 2018 to 2 April 2021

Project # 1 : Telecom Data Analysis**Experience: 1.2 year****Technologies Skills :** Spark,Python,PostgressSql,AWS RDS, Hbase, Hive, AWS S3,EMR Cluster for Hbase and Hive.

The Project purpose is telecom data ingestion from AWS cloud and stored it into S3 bucket. Lots of source from data gather from Apps, website, network. The rapid rise in the use of smartphones and other connected mobile devices has triggered a spurt in the volume of data flowing through the networks of telecom operators.

Responsibilities:

- Understand the AWS RDS source and work with business cases.
 - Written pyspark jobs to extract records from AWS RDS source.
 - Involve in creating Hive tables, loading with data and writing Hive queries.
 - Created HBase tables to load large sets of structured, semi-structured and unstructured data coming from RDS source
 - Extract data from AWS RDS to Hbase by using Spark JDBC driver.
 - Raise JIRA for Infrastructure, Platform issues.
 - Extract data from Hbase to Hive by using transformation and stored into Hive.
 - Written Hive Queries for Data Analysis to meet the business requirements.
 - Designed both Managed and External Hive Tables and Defined static and dynamic partitions as per requirement for optimized performance on production datasets.
 - Checks duplicate record in Hive and perform Deduplication Job for remove duplicate record.
 - Create pyspark script for Generate report as per daily wise, weekly wise and monthly wise and stored into S3.
-

Project #2: Data Ingestion Project**Experience: 1.2 year****Role: Data Engineer****Technologies Skills:** Hive, Sqoop, Hadoop , MapReduce

In this project we have migrated number of source data of our client into Data Lake, because our client is having thousands of data source and dataset are there, for that like we build a data ingestion framework, our data ingestion framework support three ingestion pattern database ingestion, file based ingestion and incremental data load.

Responsibilities:-

- In depth understanding/knowledge of Hadoop Architecture and various components such as HDFS, Application Master, Node Manager, Resource Manager, Name Node, Data Node and MapReduce components.
- History creation process of data warehouse application and perform aggregation operations on Hadoop Big Data platform.

- Imported required tables from RDBMS (Oracle, SQL Server) to HDFS using Sqoop.
- Involved in creating Hive tables, loading with data and writing hive queries that will run internally in MapReduce way.
- Used Hive to analyze the partitioned and bucketed data and compute various metrics for reporting.
- Importing and exporting data into HDFS and Hive using Sqoop.
- Performance Implement Data Validation, Quality Checks, Profiling
- Ingested Flat Files like Delimited, Fixed Length into Hive Warehouse.
- Written Hive Queries for Data Analysis to meet the business requirements.
- Designed both Managed and External Hive Tables and Defined static and dynamic partitions as per requirement for optimized performance on production data sets.

ACADEMIC RECORD:

Course	University	Year of passing	Percentage /Grade
B.E (CSE)	Shavitribai Phule Pune University	2017	63.66%
Diploma(IT)	MSBTE	2014	65.94%
S.S.C	Maharashtra Board	2011	70.55%

PERSONAL DETAILS:

Address : Nere Grampachayat Building,Near Dattwadi,Nere,Pune

Date of Birth : 01/01/1995

Nationality : Indian

Marital status : Married

Language Known: English, Marathi, Hindi

DECLARATION

I hereby declare that the particulars of information and facts stated here in above are true, correct and complete to the best of my knowledge and belief.

PLACE:

DATE:

Rahul Dabhade