

Technical Skill Set: Programming Languages Apache Hadoop, Python, shell scripting, SQL Technologies Hadoop  
B.E. Electronics & Communication Indore, Madhya Pradesh Medi-caps Institute of Technology & Management  
Hadoop developer

hadoop,hive,sqoop,flume,pig,mapreduce,python,impala,spark,scala,sql,unix.

#### Skill Details

APACHE HADOOP SQOOP- Experience - 31 months

Hadoop- Experience - 31 months

HADOOP- Experience - 31 months

Hive- Experience - 31 months

SQOOP- Experience - 31 months

python- Experience - Less than 1 year months

hdfs- Experience - Less than 1 year months

unix- Experience - Less than 1 year months

impala- Experience - Less than 1 year months

pig- Experience - Less than 1 year months

unravel- Experience - Less than 1 year months

mapreduce- Experience - Less than 1 year months

dataiku- Experience - Less than 1 year months

company - Tata Consultancy Services

description - Project Description

Data warehouse division has multiple products for injecting, storing, analysing and presenting data. The Data

Responsibility 1. Developed proactive Health Check utility for Data Lake. The utility proactively checks the s

2. Getting the data in different formats and processing the data in Hadoop ecosystem after filtering the data

3. Developed data pipeline utility to ingest data from RDBMS database to Hive external tables using Sqoop

4. Developed and automated various cluster health check, usage, capacity related reports using Unix shell s

5. Optimization of hive queries in order to increase the performance and minimize the Hadoop resource utiliz

6. Creating flume agents to process the data to Hadoop ecosystem side.

7. Performed benchmark testing on the Hive Queries and impala queries.

8. Involved in setting up the cluster and its components like edge node and HA implementation of the service

9. Filtering the required data from available data using different technologies like pig, regex Serde etc.

10. Dataiku benchmark testing on top of impala and hive in compare to Greenplum database.

11. Moving the data from Greenplum database to Hadoop side with help of Sqoop pipeline, process the data

12. Dealing with the Hadoop ecosystem related issues in order to provide stability to WM Hadoop ecosystem

13. Rescheduling of job from autosys job hosting to TWS job hosting for better performance.

#### Declaration:

I hereby declare that the above mentioned information is authentic to the best of my knowledge

company - Tata Consultancy Services

description - Clients: 1. Barclays 2. Union bank of California (UBC) 3. Morgan Stanley (MS)

#### KEY PROJECTS HANDLED

Project Name ABSA- Reconciliations, UBC and WMDATALAKE COE

company - Tata Consultancy Services

description - Project Description

Migration of data from RDBMS database to Hive (Hadoop ecosystem) . Hadoop platform ability with strong c

Responsibility 1. Optimising the SQL queries for those data which were not required to move from RDBMS t

2. Writing the Hive queries and logic to move the data from RDBMS to Hadoop ecosystem.

3. Writing the hive queries to analyse the required data as per the business requirements.

4. Optimization of hive queries in order to increase the performance and minimize the Hadoop resource utilization.
5. Writing the sqoop commands and scripts to move the data from RDBMS to Hadoop side.

company - Tata Consultancy Services

description - Project Description

Create recs and migrating static setup of reconciliations from 8.1 version to 9.1 version of the environment In

Responsibility 1. Have worked on extracting business requirements, analyzing and implementing them in de

3. Done the back end work where most of the things were related to writing the sql queries and provide the c

Project Name PSO