

Tushar Gupta

Mobile: + 91-7838900186

Email-ID: tushargupta93.tg@gmail.com

1) Experience Summary:

- Working as a Consultant with **Deloitte Consulting India Pvt. Ltd.** for 1 year 6 months.
- Worked as System Engineer with **Tata Consultancy Services** for 3 years and 8 months.
- Completed training on variety of Big Data tools and Technologies.
- Strong technical knowledge in developing Hadoop solutions, Big data using components of the ecosystem like Spark, Spark Scala, Hive, Pig, Map Reduce, Sqoop, Impala, Airflow, Oozie and working with TALEND and AWS.
- Extensive experience in Database Management & Big Data, Oracle SQL, SQL, Powercenter Informatica.
- Hive joins, performance tuning in joins.
- Spark Scala programs using Dataframes and RDDs.
- Sqoop tools for data ingestion, import data into Hadoop ecosystem.
- Experience of working on GIT and SVN for version control.
- Proven abilities in problem solving and resolving complex software issues.
- Experience of working on the Agile methodology for project delivery.
- Successfully led a team of 4 people to achieve desired goals.
- Good communication & interpersonal skills.

2) Technologies and Tools:

- **BIG Data:** Hadoop, Map Reduce, Hive, Pig, Talend, Spark, Spark SQL, Sqoop, Impala, Airflow, Oozie, AWS
- **Database:** Teradata, Oracle, My SQL, AWS S3
- **BI Reporting tool:** MS-Excel
- **ETL Tools:** Talend, Data Stage, Powercenter Informatica
- **Operating Systems:** MacOS, Windows XP/7

3) Professional Experience:

Deloitte Employer (May2019-Current):

3.1 Employer: Deloitte-Client: Will disclose if required -RTW Data ingestion-Hadoop Developer - November 2019- Current

Environment:

Hadoop, Big Data, MS SQL, Spark, Spark SQL, Spark Scala, Oozie, Shell scripting, AWS S3, AWS EMR, AWS Athena.

Description: The Project involves designing an ingestion framework in Spark to fetch data from the MS SQL source server into data lake, use various transformations in Spark to cleanse data and make it available for the visualization team.

Responsibilities:

- Responsible for designing, coding, testing and deploying the new ingestion framework using Spark Scala.
- Ingested data from client MS SQL server to the processed layer in AWS S3 by establishing the JDBC connection through the framework.
- Used the properties file and table list file to parameterize the code to read the values of the source database, target database, load type and tables to be ingested from the external files.
- Optimized the code for proper utilization of the resources available on the AWS EMR cluster.
- Used the savemode functionality of spark to segregate the jobs for truncate and load and incremental load.
- Created the oozie job to schedule the job to run at a specified time in a day.

3.2 Employer: Deloitte-Client: Will disclose if required -CPV Data Transformation-Hadoop Developer - May 2019- November 2019.

Environment:

Hadoop, Big Data, Sqoop, Oracle, MS SQL, Impala, Airflow, Shell scripting, Python, AWS S3, Spotfire

Description: The Project is about Designing an automated process to ingest the data from varied sources and transforming the data using Impala scripting and visualizing the data through Spotfire Client.

Responsibilities:

- Worked on designing, coding, testing and deploying the New model in Hadoop.
- Ingested files from client server to intermediate server and then finally into the AWS S3.
- Built Sqoop jobs to transfer data from Oracle and MS SQL into the cloud storage.
- Made python scripts to convert the excel into csv and applying the column level transformation on certain columns.
- Created Impala scripts to transform the raw data as per the requirement to load into the final distribution table.
- Used the parquet tables to enhance the performance.
- Used Airflow to build the scheduled DAGs to automate the workflow.

TCS Employer (Sep 2015-May2019):

3.4 Employer: TCS-Client: Will disclose if required - TDD Framework- Test Driven Development - Hadoop Developer -Oct 2018- Sep 2019

Environment:

Hadoop, Map-Reduce, Hive, Java, Big Data, Spark, Spark Scala, GIT

Description: The project is about creating automated test cases in spark framework using Spark Scala and implementing them on the already ingested data in the HDFS.

Responsibilities:

- Responsible for creating the TDD framework to automate unit testing using Spark Scala.
- Reading the data in the Dataframes and RDD from the hive table.
- Implemented various transformation logics using Spark SQL on the datasets available in the Dataframes.
- Used file-based operations to read data from hive tables and write the results back to the HDFS.
- Using GIT to track changes and maintain versioning of the code and then finally submitting the job to Spark-submit.

3.5 Employer: TCS-Client: Will disclose if required -RX Purge-Hadoop Developer -Sep 2017- Oct 2018

Environment:

Hadoop, Map-Reduce, Hive, Java, Big Data, Sqoop, Oracle, Teradata, Talend DI , Talend with Big Data Description: The Project is about Designing an automated process to purge the Pharmacy Data from Datawarehouse (Teradata) and Hadoop Cluster using Talend as Integration tool. The requirements were driven by complex business legal compliance rules and regulations.

Responsibilities:

- Responsible for rigorous data analysis to formulate the Design approach for the process.
- Created detailed flow diagrams and Designs to take the customer through all the existing Data anomalies.
- Used Talend as an Integration Tool with Datawarehouse built in Teradata.
- Developed Talend Generic jobs to handle 200+ TD tables leading to major cost saving for the customer.
- Built Talend generic joblets for error handling and logging.
- Using Talend with Big Data to handle the process for Hadoop cluster.
- Built complex HIVE and Teradata queries for handling business rules and data discrepancies.
- Optimized Talend jobs for performance and fine tuning.
- Used custom talend components for checkpointing and restartability.

3.6 Employer: TCS-Client: Will disclose if required -RX-Offload- Hadoop Developer -Jul 2016-Aug 2017

Environment: Hadoop, Map-Reduce, Pig, Hive, Java, Big-Data, Sqoop, Oracle, Teradata, Datastage, Abnitio.

Description: Project is about Converting Existing Data Model from Teradata, Abnitio and Data Stage to Hadoop using different Hadoop components.

Modules:

- Load Ready
- Noise Reduction
- Raw Ingestion (Hadoop Scripts)

Responsibilities:

- Responsible for designing, coding, testing of New Model in Hadoop.
- Ingested file from a third server to main server and then to HDFS using a Raw Ingestion Framework.
- Built Hive scripts and did reformation with actual data Using Pig and finally inserted data back to Hive tables with dynamic partitioning.
- Assisted in upgrading, configuration of Hive and Pig scripts to make the performance better.
- Assisted In designing Data Quality framework through which we can handle ^M like character in hadoop as these characters behave as Line breaker by default.
- Built Sqoop for transferring of data from hive tables to Teradata and vice-versa.
- Successfully implemented Noise Reduction logic using Apache Hive and Pig.
- Used special ORC Format tables with UTF-8 encoding so that all the special characters are handled perfectly even in Hive tables.
- Developed Pig and hive scripts which will load the data into final hive tables
- Successfully implemented Data Compaction logic using Apache Hive and Pig

3.7 Employer: TCS-Client: IVR data load -Informatica and SQL Developer -Sept 2015 -Jun 2016

Environment: Powercenter Informatica, Oracle SQL, Unviewer Console UNIX.

Description: Project is about loading the IVR (Interactive Voice Response) data in the form of files to the fact table using informatica and SQL scripts.

Responsibilities:

- Responsible for designing, coding, testing of IVR data load.
- Loaded the IVR data in form of files to the staging tables using Informatica mappings.
- Moving the data obtained in the staging tables to the intermediate schema after applying the required transformations through informatica.
- Merging the data to the fact table where data from other sources of interactions is also available using the update else insert SQL scripts.
- Performance tuning of the SQL scripts.
- Scheduling the incremental loading of the IVR data job using Unviewer console (Dollar Universe).

4) Personal Details:

First Name: Tushar

Last Name: Gupta

Address: S-42, Vijay Vihar, Uttam Nagar, New Delhi 110059

City: New Delhi | **State:** Delhi | **Country:** India

5) Declaration:

I hereby declare that the above particulars are true & correct to the best of my knowledge and belief. In the event of any Information being found false or incorrect, my candidature will be liable to be cancelled.

Place: Bengaluru

Date: