**Shweta Sharma**

**Big Data ETL Test Engineer**

**+91- 8800911134| shweta.sharma0718@gmail.com**

# Professional Summary

* A highly dedicated and dynamic **IT** professional with over 3 years of experience in BFS Domain as a Big Data Engineer in Hadoop Framework. Well versed in agile methodology and working in small to large teams.

# Skills Summary

|  |  |  |  |
| --- | --- | --- | --- |
| * HADOOP (HDFS) | * RDBMS/SQL | * PYTHON | * MAPREDUCE |
| * JIRA | * SQOOP | * UNIX/BASH | * API TESTING |
| * OOZIE * PODIUM/FM * HIVE | * ORACLE * KAFKA * HCATALOG | * TAC * ALM * PIG | * SHELL SCRIPTING * ETL |

* **Hadoop**- Good knowledge of Hadoop architecture.
* **RDBMS/SQL-** Good experience in writing intermediate to complex SQL queries.
* **KAFKA-** Experience in testing of kafka streaming application End-to-End.
* **UNIX Shell Scripting-** Strong working knowledge of Unix Shell Scripting.
* **Debugging –** Strong analytical skills to debug and do the RCA.
* **Big Data, Hadoop, Spark & Hive-** More than 2 years of experience in Big Data and Hadoop Ecosystem. Experience in writing Hive QL queries using Spark shell to load and process data in HDFS. Experience in creating Hive tables, loading data and writing hive queries using HQL.Good experience in understanding Hadoop Architecture. Hands on experience in troubleshooting error in shell, Hive, MapReduce, kafka and HCatalog.
* **Hive QL-** Experience in writing Hive QL(Queries) to check the logic of the Data.
* **SQOOP-** Experience in importing data from different databases to HDFS, hive(Historical and incremental).
* **Unix-** Developed automated scripts using Unix shell for performing ingestion of 2300+ files.
* **QC/ALM :** Proficient in QC-ALM in uploading test cases, logging defects, etc. Also, to generate the daily dashboard report for client.
* **JIRA :** Good understanding of raising tickets on JIRA.
* **TAC :** Hands-on experience on Talend Administrator Controller to run the jobs.
* **Podium/FM :** Proficient in these tools which are used to ingest data in the cluster.
* **Structured and Semi-Structured Data Sources-** Experience in working with different data sources like Flat Files, JSON Files and databases.
* **Data Model**- Good understanding of Data Modelling (Schema Design, handling medium complexity scenarios)
* **Documentation** : Good hands on experience in creating all the required documents.
* **ETL**: Experience and knowledge in functional, technical and design aspects of ETL.
* Experience in loading historical,incremental and full truncate and load data through different layers using hive.
* **PIG :** Use of HCatalog to read and write from hive interface.

# Experience

* **ClairvoyantSoft Pvt. Ltd.**: Big Data Quality Analyst | 1st June 2018 to Present | Paypal Client
* **Wipro Technologies**: Big Data Quality Analyst | Feb 2016 to 30th May 2018 | TD Bank Client

# Key Responsibilities and Project Description

* **CLAIRVOYANTSOFT Pvt. Ltd.:**

**Project 1: Kafka Implementation of Cross-Streaming-Architecture** : **Kafka Postman**

* Data is extracted from the System of Record (SOR) data sources using the data extraction framework and the data is posted on to a Kafka Cluster
* Kafka message queueing is used as intermediate storage on the Cross Property HRZ before getting moved to client environment, and segregate Class 2 and Class 3 data into different topics in Kafka to be processed separately.
* Kafka Messages are in AVRO data format and we will defining the schema of the payload in Avro.
* There are different types of validations done during the streaming process like checks for Class-2 elements, file creation, message decryption, message desearlisation and datalake processing and load testing.
* Good experience in analysing log files and creating a report which is understandable by the stakeholders.
* Good experience in functional API testing, integration testing and UI testing.
* Good data analytical skills to check the quality and mapping of data.

**Project 2: Class 2 Big Data Project** **from June 1st to Present**

**Environment: Hadoop Hive**

* This project deals Class 2 Data which cannot be stored in the data lake. Class 2 data needs to extracted and brought into the data lake via the Class 2 Data Pipeline after all the class 2 data elements are replaced with UUIDs from the Class 2 Data Store.
* A batch application(Data Transfer API) is designed to process the PGP files and generate the txt files in which sensitive encrypted data is converted to UUID(Unique Identifiers).
* Then, Data needs to be ingested in a secured fashion.
* **WIPRO TECHNOLOGIES:**

**Project 1: Business Data Warehouse(Medium/Large Scale) from August 2017 to 30th May 2018**

**Environment: Hadoop Hive TAC SHARE HDFS**

* A large scale project which involved the Talend Administrator Controller to run the Talend Jobs.
* Podium Publish used to transfer data from the Parsed zone to SHARE – zone.
* Validations performed between both the zones.
* After the data is transferred to SHARE-IN, data is transferred to SHARE-OUT zone. This is done to establish security.
* Logic-based queries are written for the SHARE-IN 🡪 SHARE-OUT data transfer.
* Unit testing of code and create reports.
* Automated testing of data flow record count using shell.

**Project 2: (Complex/Small Scale) Projects from May 2017 to July 2017**

**Environment: Hive HDFS Hadoop Oracle Sqoop**

* This was a sqoop-out(Extraction) project.
* The ingested data was hierarchical data through which complex query generation was done.
* The Validation was done between between the sqoop-out data and the data generated from the QA Hive queries.
* Appropriate documents were created during the project phase.

# Education

|  |  |  |  |
| --- | --- | --- | --- |
| **S.no.** | **Degree** | **University/Board** | **Marks** |
| 1 | BTECH | University of Delhi | 83.2% |
| 2 | 12th | CBSE | 85.8% |
| 3 | 10th | CBSE | 87.4% |