SOMESH TIWARI | INFOSYS

Email id: someshkumartiwari96@gmail.com   **Current Location**: Pune, India

Contact No: +91-9007309076   **Role Designation:** Technology Analyst

---

Professional Summary

- **3.5 years** of experience as a **Big Data Developer** to design and develop applications/services to cater client needs using in **Hadoop**, **HDFS**, **PySpark** , **Hive**, **Sqoop**, **Oozie**.
- Currently working as, a **Big Data Developer** to design and develop applications/services to cater client needs.
- Understands the complex processing needs of big data and has experience developing code and modules to address those needs.
- Implemented Proofs of Concept on Hadoop stack and different big data analytic tools.
- Migrated a high volume and time consuming batch process from Mainframe to HDFS using Sqoop, PySpark, Oozie.
- Have re-engineered Legacy Mainframe Applications into Big Data Technology.
- Good knowledge of SDLC processes like waterfall model and agile technology.
- Good understanding of technical documents and requirements.
- Coordination with onshore / offshore team and client.

TECHNICAL SKILLS

Business Domain	☐ Banking – Cards and Payments, Anti Money Laundering
Primary Skills	☐ Big Data: Pyspark, python, Hadoop, Hive, Sqoop, HDFS, SQL, Bit Bucket, GitHub, Basic Shell scripting.
Languages	☐ Python
Schedulers	☐ Oozie
Database	☐ Oracle , MySQL, Hive, DB2
Tools/Utilities	☐ Putty, WinSCP, Hive, Pyspark, HDFS, Pycharm

Project Details

- ❖ **Project**   *Anti-Money laundering* solutions for a leading finance client
  Timeline – April'2019 to till Date

- ➢ Client Name   o   American Express
- ➢ Project Name   o   Cadence-CRR
- ➢ Tools &Technologies   o   Spark, Sqoop, HDFS, Unix,Oozie
- ➢ Role Played   o   BigData Developer
- ➢ Team Size   o   6

- ➢ Project Overview: Worked as a Big Data Engineer in the Risk Management Team where the bank wanted to store process & manage the huge amount of data in day to day operations collected from various sources. The system majorly checks the credibility of the customer & looks for the credit risks.
- ➢ Approach/Responsibilities:
  - Performed requirement gathering by engaging with onshore team and client for Designing, creating, testing and maintaining the complete data management & processing systems over various daily, weekly and monthly calls.
  - Analysed data for different data source systems, involving transactional, account and customer data.
  - Transformed daily data from DB2 source systems by using **Sqoop job.**
  - Developed Spark code in **Pyspark using Spark SQL & Data Frames** for faster aggregation and Created Oozie workflow jobs for scheduling queries & actions in a 10-nodes Hadoop cluster.
  - Used different operations of CONCAT, TRIM, SUBSTR, REGEX_REPLACE, GREP and other Unix/Hadoop commands to slice and dice the aggregated data.
  - Was responsible for the creation of the necessary Functional and Technical Specification document for the changes that were made as part of the project. Also ensured that the documents followed the specified guidelines as maintained by the client.
  - Helping team members to achieve targets.
  - Acting as a single technical contact point and selecting candidates for project.

- ❖ **Project**   *CADENCE*   *Aug'2018* to April'19

- ➢ Client Name   o   American Express
- ➢ Project Name   o   Cadence-OLOB
- ➢ Tools &Technologies   o   PySpark, HDFS, Hive, Oozie
- ➢ Role Played   o   Spark Developer
- ➢ Team Size   o   4
                                        .
- ➢ Project Overview:

    The purpose of the project is to offer a service to predicate offenses to money laundering. Currently this project is getting implemented to multiple regions. This project aims at helping the client to perform the AML with improved efficiency and reduced cost.

- ➢ Approach/Responsibilities:
  - Designed, developed and deployed various components for Cadence implementation with the help of multiple teams to improved efficiency and reduced cost.
  - Provided technical leadership and expertise within the Hadoop environment using technologies such as **HDFS**, **Hive** and other Big Data technologies
  - Using Different methods and operators of Python to slice and dice the aggregated data and store in **Parquet** file format for highest compression and fastest execution of data.
  - Used Hive Query Language for inserting data into Hive tables and used HDFS in loading the tables present at the Local or HDFS Server to the Hive External and Internal Tables and to process them with Linux scripting
  - Managing & querying the external tables and using **Map join** and **Broadcast Join** to optimise the Hive query.
  - Implemented **PySpark** and **SparkSQL** for faster processing of data.
  - Using **TEZ** execution engine for fast processing of data in Hive.
  - Developing Spark programs using PySpark to compare the performance of Spark with Hive queries.
  - Helping team members to understand requirement and achieve targets.

Education

Passout Year	Degree	University	Institution
2017	B.Tech	Maulana Abul Kalam Azad University of Technology, West Bengal	Institute of Engineering & Management, Kolkata
2012	Higher Secondary	West Bengal Council Of Higher Secondary Education	Gyan Bharati Vidyapeeth, Kolkata
2010	High School	Uttar Pradesh Board	Balmeeki Inter College

Other Particulars:

▪ **Date of Birth:** 24 September, 1996
▪ **Father's Name:** Pramod Tiwari
▪ **Languages Known**: English, Hindi