**SUBHADEEP MONDAL**

**Email:** [subhadeepmondal.0604@gmail.com](mailto:subhadeepmondal.0604@gmail.com)

**Phone: +91-8763900434; +91-9477424067**

A collaborative engineering professional with substantial experience designing and executing solutions for complex business problems involving large scale **data warehousing, real-time analytics and reporting solutions**

**Location Preference: Hyderabad, Pune, Kolkata**

**Permanent Address:** Sarada Sarani, Ramkrishna Palli, P.O. - Hridaypur, North 24 PGS, Kolkata – 700127, West Bengal.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Experience Summary:**

* Having 3 years of experience in **Infosys Big Data COE Team**.
* Good knowledge of **Object Oriented Programming** concepts.
* Trained in **Business Intelligent (BI)** Stream in training period.
* Experience and knowledge in **Big Data Technologies (Hadoop, Hive, Hbase)**.
* Having knowledge and experience on **Spark Java API (Batch/ Streaming/SQL)**.
* Experience in development of **Real Time POCs on Apache Storm**.
* Having knowledge and experience on **Apache Kafka**.
* Experience and knowledge in **Unix/Linux shell scripting**.
* Has the motivation to take independent responsibility as well as ability to contribute and be a productive team player.
* Good team member, positive attitude and result oriented Self-motivated, quick learner, willing to adapt to new challenges & new technologies.

**Achievements:**

* Got 84% in **INFY** **Big Data** **certification**.
* Has been awarded with **Insta** award for technical aptitude to scale up on technologies in **Hbase, Kafka and Spark Streaming** in the FY 2014.
* Got **appreciation letter** from SPM regarding my good work.

**Experience Profile:**

* Working as Senior System Engineer for **Infosys Limited** with 3 years of experience.

**Academic Profile:**

* Bachelor of Technology (B-TECH, Electronics and Communication Engineering) from Haldia Institute of Technology, West Bengal 2012 with DGPA 7.73
* Class XII from Barasat Mahatma Gandhi Memorial High School, West Bengal 2008 with 73.2%.
* Class X from Barasat Mahatma Gandhi Memorial High School, West Bengal 2006 with 83%.

**Technical Profile:**

|  |  |
| --- | --- |
| **SKILLS** | **TECHNOLOGIES** |
| **Big Data Technologies** | Hadoop, HBase, MapReduce, Hive, Spark Java API(Batch/ Streaming/ SQL),Apache Storm, Apache Phoenix, Kafka, CDH 5 installation Using Cloudera manager |
| **Languages** | JAVA 6.0,7.0, PL-SQL |
| **Operating Systems** | Windows 7, Linux, Unix(Shell Scripting) |
| **Database Software** | Oracle 10g, MySQL 5.0 |
| **JSE Technologies** | Core Java, JDBC |
| **Servers** | Tomcat |
| **Logging Tool** | Log 4J |
| **Build Tool** | Maven |

**Project Profile:**

**Project #1**

Entity resolution

**Company Name : Infosys**

**Client** **: Internal**

**Role** **: Senior System Engineer**

**Technologies Used : Storm, Kafka, Core Java, Hbase**

**Operating System : UNIX, Windows 7**

**Tools** **: Eclipse, Putty, Winscp**

**Duration** **: Apr 2016 to Till**

**Project Description**

This POC was build for finding records in a data set that refer to the same entity across different data sources. Reference data was stored in Oracle database, and input data has landed in Kafka topic from where Storm starts processing. Matching reference records are being pulled based on blocking keys of each input record and then applying fuzzy scoring and score calculation according to business rules after that final output is pushed into Hbase table in Json format.

**Project #2**

Data Quality Migration in Spark

**Company Name : Infosys**

**Client** **: Internal**

**Role** **: Senior System Engineer**

**Technologies Used : Spark (SQL & Batch), Core Java, HDFS**

**Operating System : UNIX, Windows 7**

**Tools** **: Eclipse, Putty, Winscp**

**Duration** **: Jan 2016 to Apr 2016**

**Project Description**

This project was aimed at migrating Data Quality solution in Apache Spark. It has four layers, Data Validation, Data duplication, Data Lookup & Data Enrichment. The last module involves run time rule Ingestion using Drools implementation.

**Project #3**

CRT (Customer Reporting Technology) letters

**Company Name : Infosys**

**Client** **: Charles Schwab**

**Role** **: Senior System Engineer**

**Technologies Used : Spark (SQL & Batch), Mongo DB, Core Java, HDFS**

**Operating System : UNIX, Windows 7**

**Tools** **: Eclipse, Putty, Winscp**

**Duration** **: Nov 2015 to Jan 2016**

**Project Description**

This project was aimed at exploring and employing Big Data technologies for the purpose of generation of customer letters for US based banking client. The technology stack chosen was Mongo DB as the source and the intermediate database with Apache Spark as the processing engine. The source for the Hadoop team was Mongo DB documents which were processed, joined, blended and then converted to XML documents and passed on the downstream systems.

**Project #4**

Data Ingestion Tool in Spark

**Company Name : Infosys**

**Client** **: Internal**

**Role** **: Senior System Engineer**

**Technologies Used : Core Java, MySQL, Spark (Streaming & Batch), Kafka, Hbase, HDFS**

**Application Server : Node.js**

**Operating System : UNIX, Windows 7**

**Tools** **: Eclipse (Indigo/ Kepler), Putty, Winscp**

**Duration** **: June 2015 to Nov 2013**

**Project Description**

This project was to build a Data Ingestion Framework in Spark, where you can choose Spark Batch and Streaming processing and can push Text, Csv, Json, Xml and MySQL data. The tool could egress and store data into HDFS, Kafka, MySQL, Hbase. A web UI wrapper was created to provide a GUI base tool.

**Project #5**

NLP (Natural Language Processing)

**Company Name : Infosys**

**Client** **: Infosys Internal Product**

**Role** **: Senior System Engineer**

**Technologies Used : Core Java, Spark (Streaming), Hbase, Kafka, Apache Solr**

**Application Test : Manual Testing (White Box Testing)**

**Tools** **: Eclipse, MS Excel, Maven**

**Duration** **: Jan 2014 to May 2014**

**Product Description**

This project was about building a near real time **NLP** application working on **Apache Spark**. Raw web crawler data was coming in as xml format that landed in Kafka topic, was processed by Spark Streaming in near real time through a pipe line of data filtering, NLP and indexing.

**Project #6**

Storm Hive POC for Wells Fargo

**Company Name : Infosys**

**Client** **: Wells Fargo**

**Role** **: Storm Developer as Senior System Engineer**

**Technologies Used : Core Java, Hive, Storm, Apache Active MQ, Memcache**

**Application Server : Apache Tomcat**

**Operating System : UNIX**

**Tools** **: Eclipse, WinSCP, Putt**

**Duration** **: May 2015 to June 2015**

**Project Description**

This POC purpose was, creating a web UI that can query on Hive table, which data was populated in real time through Apache Storm. Raw data was in xml format, in the storm bolt, this xml was parsed and the data was stored into hive table using hive-thrift server. There was another module where hiveql was written by end user and this hiveql was going to fetch data form hive table in a real time manner and show the result on web user interface.

**Project #7**

SCD (Slow Changing Dimension) Type 2 in Hbase

**Company Name : Infosys**

**Client** **: Internal**

**Role** **: Hadoop Developer as System Engineer**

**Technologies Used : Core Java, Hbase**

**Operating System : UNIX, Windows 7**

**Tools** **: Eclipse, WinSCP, Putty**

**Duration** **: Sept 2014 to Sept 2014**

**Product Description**

The purpose of this proof of concept (POC) was creating a Slow Changing Dimension type 2 table in Hbase (No SQL). The code base was used in Amex project. Whole code was written in java. This code was then changed into Map-reduce codebase for scale up processing speed and volume.

**Project #8**

Assets under Management POC

**Company Name : Infosys**

**Client** **: Wells Fargo**

**Role** **: Hadoop Developer as System Engineer**

**Technologies Used : HDFS, Storm, Apache Active MQ, Core Java, Hbase, Memcache, Node.js**

**Operating System : UNIX, Windows 7**

**Tools** **: Eclipse, WinSCP, Putty**

**Duration** **: Oct 2013 to Feb 2014**

**Project Description**

The purpose of the POC was to show a helpful dashboard for every assets manager. The real time data calculation is done by Apache Storm after reading data from Apache Active MQ. After calculating assets and other revenue, result was stored into Hbase table for backup as well as send into the Memcache. End user dashboard was made by Node.js and collect data form memcache to show the result. There was the fault tolerant and data lost backup mechanism in storm.

**Personal Details:**

Father’s Name : Subhas Chandra Mondal

Strengths : Optimistic, Leadership Qualities, Team Player, Multi- tasking.

Date of Birth : 06th April 1990

Languages Known : English, Hindi, Bengali.

Hobbies : Painting, Origami, Reading Story Books, Playing Computer Games, Cooking.

Place : Bhubaneswar