

**Somanath Nanda**

E-mail : [somanath\\_nanda@rediffmail.com](mailto:somanath_nanda@rediffmail.com)

Github : <https://github.com/babasomanath/>



Contact : +353-899768305

---

• **Senior Bigdata Consultant and ML enthusiast**

---

**SUMMARY :**

- 7 years and 3 months of experience in Software Design & Development , analytics product and service development across various domains such as Manufacturing, HealthCare, Supply-Chain management, Inventory & Logistics Management , Media and Entertainment.
- I have worked on several use cases using Hadoop and Bigdata related technologies. Hands on experience with components of Hadoop ecosystem : **HDFS (Architecture), MapReduce Programming, YARN, HIVE , HBASE, AWS Kinesis, AWS DynamoDB, EMR and Apache Spark.**
- Have worked on various programming languages and technologies like **JAVA, JDBC, Servlet 3.0, JSP, Struts 1.3 , Spring 3.0 IOC & Web MVC, Spring Security, Spring JMS, JASIG-CAS Single Sign On Implementation in JBOSS AS 7.1 for Amazon AWS, Hibernate 3.x and iBatis ORM** with good knowledge of **design patterns (Façade Design Pattern, Filter design Pattern, Singleton Design Pattern, Factory and Abstract Factory Design Pattern).**
- Experienced in the design and implementation of dynamic Web-based Applications using **JAVA EE Server side** programming and implementation on the **Web server** and **Application Servers** like **IBM WAS 7.0.0.19 , IBM WAS 7.0.0.27 (JPA 2.0 integration), JBOSS 5.1 , JBOSS AS 7.1 and Tomcat 7.0** respectively.
- Developed applications for Distributed environment using **Hadoop , JAVA and Bigdata** technologies (Spark). Expertise in **debugging and troubleshooting** MR Jobs, Spark Apps on a hadoop cluster with query optimization and cluster parameters fine tuning to enhance the performance..
- Energetic self-starter with excellent analytical and organizational skills. Outstanding team player with an aptitude to learn, achieve goals, objectives in an accurate and consistent manner.

**CERTIFICATION :**

- **Oracle Certified Java Professional – Version 1.6 – December 2012**
- **Cloudera Certified Developer for Apache Hadoop – CDH4 (CCD-410) – October 2013**

**COURSES :**

- SPARK and Machine Learning (edX) - Certified
- Regression and Analysis (Coursera)

**TECHNICAL SKILLS :**

<b>Programming languages/Scripting :</b>	C, Java, Scala, Unix Shell, Bash and AWK Script, JavaScript
<b>Databases :</b>	Oracle 10 G (SQL , PL/SQL) , HBASE (NOSQL)
<b>Big data :</b>	HDFS, MapReduce, Hive, YARN, HBase, Apache Spark, Oozie
<b>Cloud Environments :</b>	AMAZON EC2, Dynamo DB, EMR, Kinesis, S3, IAM, DynamoDB
<b>Operating Systems :</b>	Windows XP, GNU/Linux- Ubuntu 11.10
<b>IDEs :</b>	Eclipse (Helios,Indigo & Juno), Netbeans IDA
<b>Data Visualization :</b>	Microsoft Excel, R
<b>Queue Implementation :</b>	JBOSS-HORNET-Q

**EDUCATION :**

**B.Tech.** in Electronics and Telecommunication from Biju Pattanaik University of Technology in the year 2010

**INTERESTS :**

Learning new technologies like Bigdata Concepts and Analysis , Design and Analysis of Algorithms, ML techniques and business implementations, Design Patterns and Linux Kernel

## PROFESSIONAL EXPERIENCE :

**Virtusa Polaris**

**Location : Dublin**

**Sep 2017 – Till date**

---

### **Project # 1 :**

**Title :** Predict the Foreign Exchange

**Description :** The client want to predict the Foreign exchanges involved to a particular transaction for a particular user and based on this there will be a rebate given by the Sales & Marketing Team. My daily activities involve the followings:

- To design and build the ML model based on 14 different features.
- To predict the foreign exchange involved for a particular transaction for a particular client.
- I have developed two Spark UDFs as well to correct the miss classified values based on two other features.
- The ML model is built using the Random Forest Classification model.

**Role :** Senior Bigdata Consultant

**Duration :** Apr 2018 – Till Date

**Technologies used :** Spark 2.1.0 and Spark 2.3.0 (including the ML libraries), Hive, Oozie

**Team Size :** 2

### **Project # 2 :**

**Title :** Analytics Product

**Description :** My work area relates to design and development of Scala code to be run on Cloudera clusters. We have three different Cloudera versions used at the moment for Dev, UAT and Prod environment. My daily activities involve the followings:

- To design and architect the different business modules based on the requirement.
- To develop efficient scala codes to be run on Cloudera Clusters.
- Writing oozie workflows using forks and joins to run spark actions concurrently.
- Managing YARN queues using the Fair Scheduler for the best utilization of the resources in different environments such as DEV, UAT and PROD.
- Development of utility modules using scala and register them as hive UDF wherever needed.
- Finding the best possible way to make the spark applications performant.
- Design and Maintenance of data lake which receives data from the upstream such as RDBMS, online streamed data and web-apps data from the servers. I have implemented data profiling and data security using different tools and concepts to provide the transformed useful data for business end users and App consumers.

**Role :** Senior Bigdata Consultant

**Duration :** Sep 2017 – Apr 2018

**Technologies used :** Cloudera CDH 5.7.5 cluster (UAT) , Cloudera CDH 5.9.1 cluster (DEV), Spark, Hive, Oozie

**Team Size :** 9 members

**Amazon Data Service Ireland Limited**

**Location : Dublin**

**May 2015 – Sep 2017**

---

### **Project # 1 :**

**Title :** Developing and Troubleshooting issues on EMR (DevOps, SysAdmin, Development)

**Description :** My work area relates to development of JAVA code for big data users on EMR, using MR and SPARK, writing hive queries, tuning hive queries, troubleshooting performance issues for running jobs, configuration of EMR environments, working on capacity schedulers and fair schedulers for different users for

different pools/queues on YARN and Apache Spark. My job role involved consulting, development and troubleshooting (DevOps and SysAdmin) in Linux environment.

**Role :** Bigdata Cloud Support Engineer II

**Duration :** 11 May 2015 – 15 Sep 2017

**Technologies used :** Amazon EC2, S3, EMR, DynamoDB, Kinesis, IAM

**Team Size :** 20 members

---

**Cognizant Technology Solutions**

**Location : India**

**January 2014 – February 2015**

---

**Project # 1 :**

**Title :** Product Development for Sales Force Admin and Analytics in Amazon Cloud

**Description :** This Product is meant for analyzing business based on different geography and customers assigned to a particular geographic location. I was an individual contributor to develop a query processor on various unstructured data .

**Role :** Programmer Analyst

**Duration :** January 2014 – Till Date

**Technologies used :** Spring 3.x, Hibernate 3.x, JPA 2.0, Spring Security, JBOSS AS 7.1, Single Sign On using JASIG-CAS, AMAZON AWS Cloud, JSON, Multi-tenancy on Cloud Server

**Database used :** My SQL

**Team Size :** 20 members

---

**Tech Mahindra**

**Location : India**

**August 2011 – December 2013**

---

**Project # 1 : (Duration : August 2013 – December 2013)**

**A.**

**Title :** I had to find out the top 10 mobile phones , on which we are getting more number of tweets.

**Role and Description :** People are tweeting their experience on different brands of mobile phones. I had written the JAVA code using Twitter4J API to collect the data from twitter. Then the collected data was processed in a MAPREDUCE program to make the data structured and to find out the counts of different tweets and only top ten record counts were saved.

**Role :** Hadoop Developer

**Technologies used :** HDFS, MapReduce Programming, WebServices, twitter4j

**Responsibilities :**

- Wrote Java code to extract data (tweets, status updates) from Twitter.
- Wrote efficient MapReduce programs for filtering out the unstructured data.
- Wrote Map Reduce program as per the requirement.

**B.**

**Title :** There are 30Lakh records given by one of the health care giant. They wanted to know the top 10 medicines which are sold in the year 2012 - 2013.

**Role and Description :** To solve this objective, I have Written a MapReduce program using JAVA and found out the sum of total amount spent on a particular medicine in the year 2012-2013. Then I sorted out the top 10 amounts to find out the medicine names using TreeMap in JAVA. This took maximum 2 minutes to process

all the 30L records on a 3 node cluster environment ; whereas it was taking more than 7 minutes to compute using RDBMS.

**Role :** Hadoop Developer

**Technologies used :** HDFS, MapReduce Programming

## **Project # 2 :**

**Title :** Social network Analytics

**Description :** This Project is meant to analyze the sentiments of the viewers based on the data available on various social networking sites such as twitter, facebook, youtube etc. The purpose of this project was to load data from the social networking sites using the java program into HDFS file system. The data obtained is in unstructured format, MapReduce Programming is used to filter out the obtained data and the filtered data is stored in HBase. The data stored in HBase is feed to the R-Tool, which is a statistical computing and graphics tool and used to analyze the sentiments of the user. Previously conversion of unstructured data into filtered structured data was done through java program and data is feed to R-tool which makes the operation slower; with the introduction of Hadoop and its components data is stored and formatted (filtered) efficiently and thus makes the operation faster.

**Role :** Hadoop Developer

**Duration :** June 2013 – August 2013

**Technologies used :** HDFS, MapReduce Programming, WebServices, restfb, twitter4j, g-data api, imdb api, R-tool

**Database used :** HBase

**Team Size :** 4 members

**Responsibilities :**

- Wrote Java code to extract data (movie reviews, tweets, comments, status updates) from various social media websites (Facebook, YouTube & Twitter and IMDB).
- Wrote efficient MapReduce programs for filtering out the unstructured data.
- Wrote program using HBase API to stored data in to HBase.
- Integrated Hadoop system with R-tool.

## **Project # 3 :**

**Title :** Concern Report and Counter Measure

**Description :** Nissan was operating from Sunderland, U.K . To manage the escalation level for different raised issues the employees go for various level of approvals to solve and manage. Previously they were following the traditional pen-paper methodologies. To make the process automated and reporting , this application is developed. (Still Continuous development and enhancement is going on).

**Role :** JAVA – J2EE Developer

**Duration :** April 2013 – June 2013

**Technologies used :** Spring 3.x , Hibernate 3.x, JSP, Javascript

**Database used :** Oracle 11G

**Application Server used :** IBM WAS 7.0.0.27

**Team Size :** 5 members

**Responsibilities :**

- Developing Action Classes using Spring 3.x
- Developing Oracle SQL Queries and PL-SQL Procedures, functions and triggers

- Designing workflows using Spring for Various levels of approvals

#### **Project # 4 :**

**Title :** Russian Plant Project

**Description :** Nissan is currently operating from St. Petesburg , Russia known as Nissan Motor Manufacturing which is based on Complete Knock Down business rules. Along with that two more plants named as NMGR-Shikyu and NMGR-DKD are also operating to support and enhance the business. Development of applications which automate the complete ordering system. Along with that maintaining the parts in warehouses and in different shops related to car manufacturing industry is a crucial business aspect of the project. This project is completely based on Inventory & Logistic Management and Supply Chain Management.

**Role :** JAVA – J2EE Developer

**Duration :** January 2012 – April 2013

**Technologies used :** Struts 1.3 , iBatis, JSP, Javascript

**Database used :** Oracle 11G

**Application Server used :** IBM WAS 7.0.0.19

**Team Size :** 5 members

#### **Responsibilities :**

- Developing Action Classes and JSP pages
- Developing Oracle SQL Queries and PL-SQL Procedures, functions and triggers
- Analyzing issues and solving them as per the requirement of the Business
- Supports Applications Developed Using JAVA and ORACLE.
- Develops Web Based applications using different frameworks of Java
- Develops Code for Oracle Stored procedures, triggers and cursors as per the business requirement.
- Accomplishes the relation between the technology and the business requirement as per the feasibility of the system.
- Improves the productivity of my esteemed customer providing new business solutions which can be implemented through technology.

#### **PERSONAL DETAILS :**

<b>Present Address</b>	: 35D, Belarmine Square, Belarmine, Stepside, Dublin-18, Ireland
<b>Nationality</b>	: Indian
<b>Gender</b>	: Male
<b>Passport No.</b>	: J0399214
<b>PPS No.</b>	: 1598973BA
<b>GNIB No.</b>	: 692871 (Stamp IV)
<b>Current Location</b>	: Dublin

#### **DECLARATION :**

I, hereby, declare that the information given above is true to the best of my knowledge and belief

**Place :**

**Date :**

**Signature**