



**SUBHASH CHANDRA REDDY**  
BigData Consultant

Highly organized and independent professional who is able to effectively coordinate tasks to accomplish projects with timeliness and creativity.

09545 666120 hadoopsubhash@gmail.com

## Profile Summary

A result-oriented, professionally qualified Software Engineer with proven experience in Big Data technologies combined with excellent business, communication, analytical & problem solving skills. Quick to learn new Technologies and able to work alone under pressure and manage conflicting priorities to consistently meet strict deadlines. Motivated by Client satisfaction and can effectively reach stakeholder expectations.

## Roles and Responsibilities

- Proficient in Building data systems and data pipelines that extract, classify, merge, and deliver new insights.
- Good experience with Hadoop, Spark, Python technologies.
- Skilled in designing and building data analysis systems for large data sets.
- Expert in Building high-performance algorithms, prototypes, predictive models and proof of concepts.
- Create custom software components and analytics applications
- Extensive knowledge of designing and building end to end big data applications.
- Excellent oral and written communication skills
- Experience in mentoring junior members of the team.

## IT Skills



## Certifications

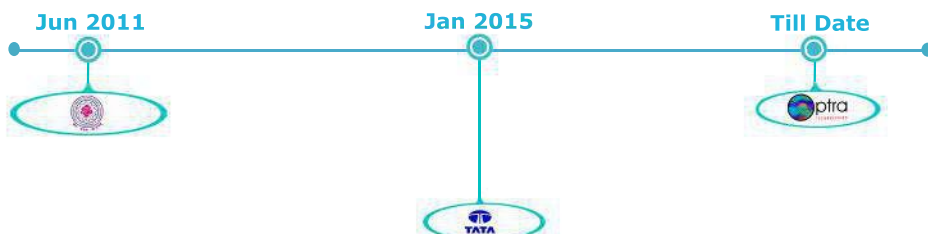
BU level certified Bigdata Administrator

BU level certified Bigdata Developer

## Rewards and Recognitions

- ★ Innovative Team member award
- ★ Inter University badminton winner
- ★ Won many tournaments in badminton in India.

## Career Timeline

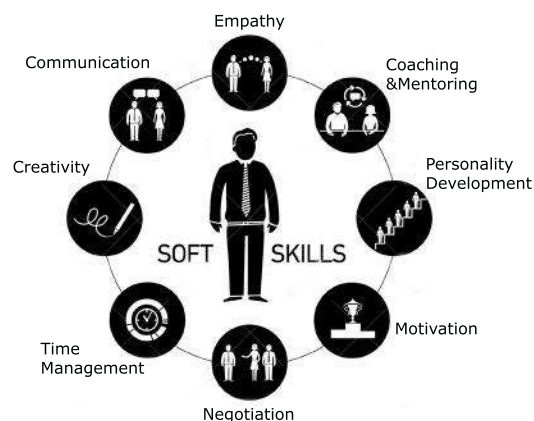


## Education

2011 B.Tech:Computer Science  
JNTU Hyderabad,India.



## Soft Skills



## Core Competencies

Spark  
Hadoop  
Python  
Hadoop Admin

## Online Profile



### Subhash Reddy

Hadoop Developer at optra  
Pune, Maharashtra, India | Information Technology and Services

Current Optra Systems, Inc  
Previous TCS  
Education Jawaharlal Nehru Technological University

## QR Code Business card



Scan QR code to get my contact details

## Project Details:

### PROJECT1:

Project Name : Multi-Policy Predictive Analytics  
Project Role : Hadoop Developer  
Environment : HDFS, Apache Hadoop, Apache Spark, Hive, Apache Nifi  
Duration : Nov 2015 to Till Date

**Description:** Predict existing customer's interest in opting for multiple insurance policies.

### **Roles and Responsibilities:**

- Ingest Insurance policy data into the Hdfs data lake using Apache Nifi.
- Create an Apache Nifi data flow using in built processors i.e. GetFTP to fetch datasets from an ftp server and put the files into Hdfs using PutHDFS processor
- Ingest Customer and Sales data using Apache &pop into the Hdfs.
- Create External Hive tables and load the data into the hive tables from the datasets in the data lake.
- Developed Spark core algorithms in Python using RDD's and Datagrams to normalize the data.
- Implement data normalization using spark core API which includes many different processing models such as graph processing, SparkSQL and all of these models can be unified using DAGs

### PROJECT2:

Project Name : I-Phronesis  
Project Role : Hadoop Developer  
Environment : HDFS, Apache Pig, Hbase, GATK, SAMTOOLS  
Duration : Sep 2015 to Till Date  
Client : GE

### **Description:**

Genome analysis is used to analyze human genome data using Hadoop. A single human genome contains about 3 billion base pairs. This is less than 1 gigabyte of data but the intermediate data produced by a DNA sequencer, required to produce a sequenced human genome, is many hundreds of times larger. Beyond the huge storage requirement, deep genomic analysis across large populations of humans requires enormous computational capacity as well. Efforts exist for adapting existing genomics data structures to Hadoop, but these don't support the full range of analytic requirements. Our approach is to implement an end-to-end analysis pipeline based on GATK and running on Hadoop.

### **Roles and Responsibilities:**

- Configuring Hortonworks Hadoop cluster setup with 5 nodes using ambari.
- Writing pig scripts to process the Genome data.
- Writing shell scripts to run the GATK commands.
- Writing the script files for processing data and loading to HDFS.

### PROJECT 3:

Project Name : Avarx

Project Role : Hadoop Developer  
Environment : HDFS, Apache Pig, Hbase, Javas, Sqoop.  
Duration : May 2015 to Nov 2015  
Client : Torrey Insights

### **Description:**

AvaRx is an online tool that expands on the Available Pharmaceutical Products List. Previously they used PHP and MySQL to develop this tool now they want to change this tool with Java and Hbase. So they want move all the data from MySQL to Hbase.

### **Roles and Responsibilities:**

- Moving all their data from MySQL to Hdfs using sqoop scripts.
- Writing the Pig scripts to process the data.
- Storing all processed data into Hbase using Hbase loader.
- Involved in gathering the requirements, designing, development and testing
- Analyzing the requirement to setup a cluster
- Configuring Hadoop cluster setup with ten nodes.

### **PROJECT 4:**

Project Name : PKI  
Project Role : Hadoop Developer  
Environment : HDFS, Apache Pig, Hive, spotfire, R  
Duration : Feb 2015 to May 2015  
Client : Janssen

### **Description:**

This Project is all about the storing the medical products data into hdfs after that connecting hive to spotfire and analyze the data using spotfire.processing data using pig and storing reports using hive. Moving all hive data into Hbase using Hive Hbase handler.

### **Roles and Responsibilities:**

- Moved all medical products data to HDFS for further processing
- Written the Apache PIG scripts to process the HDFS data.
- Created Hive tables to store the processed results in a tabular format.
- Creating connection between hive and spotfire to generate visualization In Spotfire
- Configuring ads in spotfire.

### **PROJECT 5:**

Project Name : 4 Medica Clinical Data Integration  
Project Role : Hadoop Developer  
Environment : HDFS, Apache Pig, Hive, SQOOP, Java, UNIX, SQL  
Duration : Nov 2013 to Nov 2014  
Client : 4 Medica

### Description:

4medica is the nation's leading provider of cloud-based clinical data exchange, which provides clinicians with a unified, real-time view of patient information across disparate care locations. The company's flagship clinical integration platform, Integrated Electronic Health Record (iEHR), builds upon organizations' existing technologies to supply the exact level of health connectivity needed to address meaningful use requirements, from basic health information exchange to integration with existing electronic health records (EHRs), practice management systems and other healthcare applications.

### Roles and Responsibilities:

- Written the Apache PIG scripts to process the HDFS data.
- Created Hive tables to store the processed results in a tabular format.
- Developed the sqoop scripts in order to make the interaction between Pig and SQL Database.
- Involved in gathering the requirements, designing, development and testing
- Writing the script files for processing data and loading to HDFS
- Writing CLI commands using HDFS.
- Developed the UNIX shell scripts for creating the reports from Hive data.
- Completely involved in the requirement analysis phase.
- Analyzing the requirement to setup a cluster
- Created two different users (hduser for performing hdfs operations and map red user for performing map reduce operations only)
- Setting Password less Hadoop
- Setting up cron job to delete hadoop logs/local old job files/cluster temp files
- Setup Hive with SQL as a Remote Metastore
- Created External Hive Table on top of parsed data.

### Project 6:

**Hadoop Framework** : HDFS, HIVE, Hbase, Sqoop, PIG  
**Project Role** : Hadoop Developer  
**Scripting** : Python, UNIX  
**Hardware** : Ubuntu 11.10 operating system  
**Duration** : Jan 2013 to Nov2013

### Description:

Maintaining the customer member details and rewards points transaction are very difficult in terms of storage and processing. Member loyalty management system is replacing the existing reward management system which is developed as a web service provider with the help of database sharing. Aim of this system is to reduce the response time of web service. This system is designed with Hbase storage handler and later planning to remove some BI reports generation using Hive. The solution is based on the open source Big Data s/w Hadoop. The data will be stored in Hadoop file system and processed using Map/Reduce jobs.

### Project Responsibilities:

- Application installation of Hadoop, Hive , Map Reduce & Sqoop
- HDFS support and maintenance and Adding/Removing a Node, Data Rebalancing.
- Involved in developing the Pig scripts
- Involved in developing the Hive Reports.

### Project 7:

**Project** : Prepaid system  
**Project Role** : Sql Developer  
**Client** : Tata Tele Services Limited  
**Period** : Jun2011 to Aug2012

**Prepaid Accounting System:** The main aim of this project is to provide the prepaid revenue accounting entries to SAPERP by gathering information from all the source systems which impact the prepaid customer.

**Roles & Responsibilities:**

- Team member as Developer.
- Contributed to project maintenance and enhancement phases.
- Work mainly involves requirements gathering, design, coding, testing and implementation.
- Played a Key role in developing procedures to efficiently handle and manipulating huge volumes of data.

**Solution Environment:**

- Oracle 10g and Solaris

**Declaration:**

I hereby declare that the information furnished here is true to the best of my knowledge. I bear the responsibility for the correctness of the above mentioned particulars.

Place: Pune

Subhash.T