



# SATYASOBHAN PUJARI

## PROFILE

Having strong background and working knowledge of Hadoop and Big Data Technologies stack, Data Science, Machine Learning along with the strong experience and skills to work in java/j2ee/Scala/Python Technologies with excellent verbal and written communication and interpersonal skills

## CAREER INTEREST

- ☞ Data Science
- ☞ Machine Learning
- ☞ Big Data Architecture
- ☞ Technical Architecture
- ☞ Solution Designing
- ☞ Engineering Practices

## PERSONAL DETAILS



+91 9000 692 782



Satya.pujari@gmail.com



A 905, Aeropolis, Porwal  
Road, Dhanori, Pune 411047



11 | Aug | 1978

## PROFESSIONAL SUMMARY

- ✓ Over 15 years of experience in Software industry involving in complex, high-profile projects including 6+ years of experience in **BigData** and **BigData analytics** in **Banking Domain** using **Hadoop** Ecosystem.
- ✓ Strong experience in building **Statistical Model using Machine learning algorithm, Natural Language Processing and Data Mining** using **Python**
- ✓ Hands on experience in designing and developing Big Data Application using **Spark, Spark Streaming, Kafka, Flume, HBase**
- ✓ Experience in developing proof of concept models. Design and developing data pipeline for various big data projects
- ✓ Good exposure with Devops toolset e.g. **Jenkins, Maven, Selenium, GitHub, Nexus, Jira, TeamCity**
- ✓ Experience in participation on both in-house and open hackathons
- ✓ Exposure to cloud platform using **AWS**

## AREA OF STRENGTH

- ✓ Hands on experience in Hadoop and its ecosystem (HDFS, MapReduce, HIVE, Sqoop, Flume, HBase ) Apache Spark , Kafka, Python
- ✓ Full life cycle experience in building machine learning models including Data gathering, Data Cleaning, Model Building, Testing and Deployment
- ✓ Expertise in J2EE technologies with experience in e-Business technologies and various Web and Integration tools, Implementation of Client/Server, SOA application, Internet/ Intranet based Database Applications and Legacy Applications, Distributed Application development.
- ✓ Experience in Banking and Financial Domain.
- ✓ Strong technical and Functional skills and effectively handling of complex situations.
- ✓ Deep knowledge of object oriented programming and software development life cycle. Profound ability to work on server side java programming.
- ✓ Confidence in learning and implementing any new application and languages.
- ✓ Leading the team of technical professional during the development stages.
- ✓ Enthusiastic in Big Data Analytics, Machine Learning, Data Visualization.

## EDUCATION & CERTIFICATION



Master in computer application (MCA)  
MK University,  
TamilNadu



Bachelor of  
Commerce  
Sambalpur University,  
Odisha



International School of  
Engineering,  
Hyderabad Certified  
Data Scientist



Cloudera Certified  
Apache Hadoop  
Developer

## TOOLS & TECHNOLOGY

### Languages:

**JAVA | SCALA | PYTHON**

### BigData Technologies:

**HADOOP | MAPREDUCE | HIVE | KAFKA |  
STORM | SQOOP | FLUME | SPARK**

### J2EE/Web Technologies:

**JSP | SERVLETS | ENTERPRISE JAVABEANS |  
JMS | XML | WEBSERVICES | MICROSERVICES  
| REST | HTML 5 | CSS 3 | FLASK**

### ML Library:

**SCIPY | NUMPY | SPACY | NLTK |  
TENSORFLOW | SKLEARN | XGBOOST**

### JavaScript Library:

**JQUERY | NODE.JS | D3 | ANGULAR JS |  
BOOTSTRAP**

### NoSql DB:

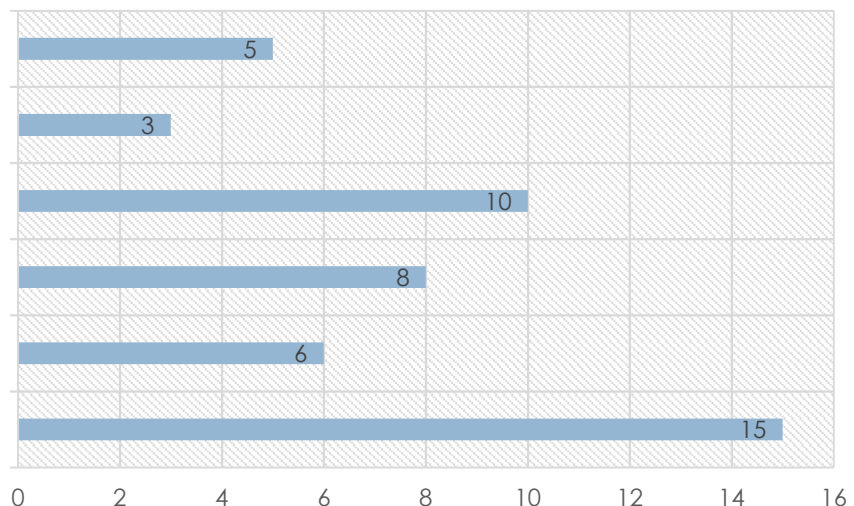
**MONGODB | HBASE**

### Devops Tools:

**GIT | MAVEN | TEAMCITY | JENKINS | JIRA**

## Years of Experience

Team Management And Delivery Management



Machine Learning Model Building

Agile Methodology

Solution Design And Architecture

BigData Application Development

Java/J2EE Application Development

## CAREER TIMELINE



Aug-2003 –  
Jan 2005

RBI (Reserve Bank of India) As Research Associate



Jan, 2006 -  
June 2007

Synechron Technologies, Pune As Sr. Software Engineer



July, 2007 -  
Jan 2010

Satyam Computers Service Ltd, Pune, As a Tech Lead



Jan 2010 –  
Nov 2012

HSBC GIt, Pune (India) As a Consultant Specialist



Nov 2012 –  
Nov 2015

Barclays Technology Centre India, Pune As Technical Delivery Manager



Nov 2015 -  
July 2017

IBM India Pvt. Ltd As BigData Consultant



July 2017 –  
Till Date

Barclays Global Service Centre, Pune

## ACHIVEMENTS

### ROBUST PROCESSING SYSTEM DELIVERED AT SCALE

Designing and implementing robust spark streaming system to handle around 2 billion records daily and process batch of 3.6 million of records within timeframe of 5 mins for Barclays equities platform.

Successfully delivered reporting system using BigData technologies to handle petabyte of records for TIAA. Clean & Clone receive best delivered bigdata project for TIAA

### HACKATHON

Winner of Barclays RFT hackathon to delivered a quick solution using BigData on AWS cloud.

Stand in top ten of leaderboard for MachineHack hackathon for designing predictive modelling solution

## PROJECT DETAILS

**MachineHack Hackathon**  
**Project: Predicting The Costs Of Used Cars**

**Technology:** Python, Predictive Modeling using Ensemble of Gradient Boosting, XGBoost, AdaBoost, Regressor

**Description:** predict the price of a used car based on a given set of features. Size of training set: 6,019 records, Size of test set: 1,234 records

**Details:**

- ☞ Implementing data preprocessing
- ☞ Encoding categorical variables
- ☞ Imputing missing values
- ☞ Scaling the dataset
- ☞ Model designing

Got an accuracy of 94% and stand in top ten of leaderboard.

**Barclays**  
**Project: TradeStore / QuoteStore**  
**Role: Individual Contributor**

**Technology:** Hadoop cdh5, Hive, Impala, Flume, Spark Streaming, Jdk 1.7, Scala

**Description:** TradeStore/QuoteStore is a centralized cross-asset platform responsible for providing analytical representation of Trade data ingested from various upstream system into Hadoop/Hbase in various format. Upstream sends intraday/live trades that are ingested into Hadoop as soon as the trade is booked. It also provides a near real time web based view generation engine to see live and historical trades.

**Details:** Responsibilities include:

- ☞ Design and implementing and data pipelines.
- ☞ Implementing Spark Streaming jobs to receive daily Trades and Quotes from Kafka Cluster.
- ☞ Develop various autosys and Shell Script job to process Snap and eod files to Hadoop cluster.
- ☞ Developed TradeStore refactoring engine to provide customize view of TradeData to downstream application.

**IBM**  
**Project: HADOOP ESP, Clean & Clone**  
**Client: TIAA**  
**Role: Consultant**

**Technology:** Hadoop cdh5, Hive, SQOOP, Spark, Python, Java

**Description:** HADOOP ESP project is to create a data-lake by transferring existing TIAA asset data from various staging platform to Hadoop cluster and applying various adjustment and balance calculation by joining multiple dataset.

**Clean and Clone** Project to build deliver Balance calculation of the Market Product of Guaranteed and Non Guaranteed on Number of Shared allotted to/bought by customers and matching the expected rate of interest and Fund Balance as per current Share rate for each Plan. The Design build on PySpark and Hive and to ensure the data get processed on daily basis.

**Details:** Responsibilities include:

- ☞ Worked as a BigData Consultant for implementing and designing data pipelines.
- ☞ Develop Shell Script/SQOOP job to transfer traditional data to Hadoop cluster.
- ☞ Develop Hive/Spark script to collect data from various source system for balance adjustment calculation
- ☞ Developed ETL using PySpark for over multi billion records history data for one time processing of 7 years of financial records
- ☞ Solution building and Development in PySpark for processing 3 Million incremental run
- ☞ Designed Hive tables for quick data response by implementing Buckets, Partition and Parquet/Avro format
- ☞ Migrate some HIVE/PIG jobs to SPARK to achieve low latency and better performance.

## DECLARATION

All the information mentioned in the resume are correct to the best of my knowledge & belief.

**Satyasobhan Pujari**