

Sushant Sawant  
Contact No. : +91 7021415633

Email:sushantsawant007@yahoo.com

### **Academic Credentials**

- BEIT from Mumbai University in 2014 - **68%**
- Executive Post Graduation Program in Data Science, Business Analytics and Big Data in association with IBM by Aegis School of Business, Data Science & Telecommunication.

### **IT Credentials**

**(Started career with a BigData Database in Product Startup! We build our own database.)**

**Over all 4.7 years of experience in developer role.**

- Programming Language - Java, Scala, Erlang, R, SQL, C, C++, Python (Cpython, Ironpython), Nodejs, Golang.
- Database - Oracle 2012, MYSQL, Cassandra, MongoDB, Apache Solr, Hadoop, Elasticsearch, Redis.
- Big data processing - Apache Flink, Apache Spark, Apache Thrift, Apache Airflow, Kubeflow.
- Queuing Technology- Apache Kafka, NSQ
- BI tools: IBM Cognos, QlikSense, Tableau.
- Any technology most welcomed. **Also an language agnostic.**

**You could visit me on**

**Github Account:- <https://github.com/saint7007>**

**Stackoverflow :- <http://stackoverflow.com/users/7504347/saint>**

**Kaggle :- <https://www.kaggle.com/sushantsawant>**

**Tech blog:- <https://saint007code.blogspot.in>**

**Hackathons. Incase you host it. Let me know.**

### **Organisations-**

**Amsys Innovative Solution(Startup) as Senior Software Engineer(Analytics)  
Sep 2019 - Current**

**Project - Data pipeline using Airflow for OCR on medical documents**

**Team Size : 1**

**Description:** Developed OCR pipeline with cocktail of custom **Opencv** python, **AWS textract** and **Google vision api** and dumping result of image to text in **Elasticsearch** for visualization.

**Project - Data pipeline using Kubeflow for OCR on medical documents**

**Team Size : 1**

**Description:** Developed OCR pipeline on **Kubeflow** for **handwritten documents** to text using **Google vision**.

**Project - Reactive Web Api for Push Notification**

**Team Size : 1**

**Description:** Developed super concurrent push notification based on reactive framework known as **Vertx** and used **Scala** as language. Will update the bench marking soon.

**Project - Backtrading for Bitcoin data**

**Team Size : 2**

**Description:** Developing **python Flask** based back trade analysis for traders using python library called **Backtrader**. Used **Arctic db** as backend wrapper on **Mongodb**.

**Bprise Analytics(Startup) as Senior Software Engineer(Analytics)  
Aug 2018 - Sep 2019**

**Project - Real time Streaming Analytics on advertisement (Real Time Bidding)**

**Team Size : 1**

**Description:** Post Real time bidding, the successful bids were logged into kafka and so the impression and click, calculation of realtime statistical numbers. Achieved throughput of **4100/sec throughput** during kafka stream enrichment using Apache Flink. And data database storage was **Elasticsearch, Redis, Cassandra, ES-hadoop with 4 node cluster**. Solved this :<https://www.slideshare.net/FlinkForward/flink-forward-berlin-2018-lasse-nedergaard-our-successful-journey-with-flink> slide 23, faced same issue.

#### **Project - Web analytics**

**Team Size : 1**

**Description:** Analytics on click stream data and calculating various parameters, such as CTR, Average session duration etc with respect to day, week and month. On similar lines of google analytics for our client. Technology side have used **Elasticsearch, Redis, Apache Flink jobs and ES-hadoop with 4 node cluster**.

#### **Project - Retarget marketing**

**Team Size : 1**

**Description:** Showing interested users the advertisement using cookie sync using Spring boot, Redis, Cassandra and bit of javascript. Cassandra table modeling for reads. User base classification to increase conversion rate based on previous clicks and behaviours using **LogisticRegressionWithLBFGS, Pyspark with 2 node cluster**.

### **Headway.ai(Stealth mode startup) as Application Developer February, 2018 - Aug 2018(0.7 years)**

#### **Project - CV Feature Extraction**

**Team Size : 1**

**Description:** Feature extraction using **Spacy** and **Stanford NLP** and **language modelling** using n-gram. Used grammar for extraction of subject verb object by **POS tagger**. Using **Naive Bayes** classifier for further classification of text using spark mllib for topic classification using Logistic Regression. The engineering layer includes **Apache Thrift**, and **Apache spark**, more specifically, **Pyspark**.

#### **Project - Data Pipeline for Asynchronous processing.**

**Team Size : 1**

**Description:** Using **Apache Kafka 1.1.1** version, and wrote **python producer** and **nodejs consumer** client of Kafka, built an data pipeline for ML engine for CV feature extraction project.

#### **Project - Multi-tenant HR web application with recommendation engine**

**Team Size : 1**

**Description:** Built an architecture for this system, and technology stack involved around Big data ecosystem which includes **Python Django, Nodejs, Postgres, MongoDB, Apache Spark**.

### **WorkJapan.jp (Japanese startup) as Developer/Lead December, 2017 - February 2018(Contract Based)**

#### **Project - Job Recommendation engine(Apache Spark,Python)**

**Team Size : 1**

**Description:** Job feed according to user's demo graphics using collaborative filtering technique using ML lib Pyspark.

#### **Project - User behaviour analytics**

**Team Size : 1**

**Description:** Patterns from URL visits by customer i.e click events. Implemented GBIDE algorithm which works on tree pruning based search. Clustering on successful leads and unsuccessful leads with K-means clustering.

**Tata Consultancy Services**  
**(on pay roll of C-edge an TCS-SBI joint venture)**  
**October 2015 - Nov 2017 (2.1 Year)**

**Awarded trophy and certificate for performance from TCS head.**

**Project - Credit Risk Modelling**

**Team Size : 1**

**Description:** Developed an python flask web application and was given R code of calculating credit risk score by Third party company at Client, since R being not production ready converted same into python code.

**Project - CBS client in Spring MVC**

**Team Size : 8**

**Description:** Lead this team and developed this web application using Spring Java and Oracle. All transactions like NEFT,RTGS etc from bank with Core Banking Server where handled.

**Project - Automating National Pension Scheme/Atal Pension Yojna**

**Team Size : 4**

**Description:** Developed an java based web application (Struts, Oracle) with Rest full webservice interacting with government organization NSDL.

**Project - Social Media Banking(Middleware)**

**Team Size : 1**

**Description:** Developed an middle ware for PIN code encryption/decryption using Triple Des and RSA for mobile app and CBS for largest financial institution in India.

**Blobcity iSolutions Pvt. Ltd as Java Developer**  
**October 2014 to October 2015 (1 Year)**

Blobcity iSolution is one of startup organisation in big data space, making web application in various domains Trading, Social etc.

**finstein.com**

**Description:** Developed a trading platform with NSE live data 6 minutes delay. (**Scala, Java,Cassandra,Mongodb**)

Usual user mapped actions application used to interact with mongoDb, cassandra for storing click data.

**droover.com**

**Description:** Developed a search engine with classification in category. (**Scala,Java,Cassandra,Mongodb**)

Usual user mapped actions application used to interact with mongoDb, cassandra for storing click data.

**XMPP based Chat Server in Erlang**

**Description:** Same language as Whatsapp server and same architecture.

- More added feature like saving chats on server, save file during transfer on server.

**mpensystems.com**

**Description:** Developed web based search document using **Apache solr** and flat file based custom built database, for an Energy Science related company.

**Mail dump in Hadoop**

**Description:** Developed a project for reading mail from mail server from POP3 protocol, using java mail reader API and dumping in into **hadoop**.

**Visualization Charts**

**Description:**

- For demo purpose and data visualization web pages were built using **D3JS** library.
- Bench marking read write operation between Mongo, Cassandra and our own database.