

Nissincho 23-01, #911  
Kawasaki City, Tokyo, Japan

**Avinash Mishra**

+81-70-2188-8241  
[avinash.mishra.2388@gmail.com](mailto:avinash.mishra.2388@gmail.com)



## EXPERIENCE

*Data Scientist/Team Lead*

*Paradigmshift.io, Japan*

*Sept. 2016 ~*

### Valuechain [\[link\]](#) [\[see project\]](#)

- Valuechain is a machine learning tool which provide deep insights of OTAs reviews.
- I have designed the architecture and we perform various NLP techniques like Noise removal, Lexicon Normalization, Object standardization, Entity Parsing, Statistical featuring, text classification and matching etc.
- Tools and technologies used : NLTK, MeCab, ML-Classifer, Python, flask.
- Managing AWS instances of Valuechain server.

### Price RecSys [\[link\]](#)

- This is a price recommendation engine for OTA price master. We are using Hybrid filtering with Spark ALS to recommend price.

### Smart web crawler [\[link\]](#)

- Designed distributed web crawling architecture. It has centralized controller, distributed spiders and distributed storage in Cassandra cluster.
- Tools and technologies used: Apache Kafka, Scrapy web crawling framework, Apache zookeeper, Django, Python, bash.

### Data Lake

- It is a centralized repository hosted on Digitalocean which stores all structured, unstructured and semi-structured data of the company. We use HDP for data storage, Ambari for cluster management. System is powered with bulk Read api and writing is done by talend.

*Software Engineer(Big Data  
Department)*

*Rakuten, Japan*

*Oct. 2014 - Sept. 2016*

### GSP [\[see project\]](#)

- GSP is Global search platform which is being used for almost all of the Rakuten services.
- Created automated testing framework (Ngauto) for distributed search services involving systems like Solr, Zookeeper, Cassandra and Hadoop.
- Created search peripheral components like dictionary compiler and word extractor to support search sub-functions like “did you mean”, “related-words”, “spell-check” and “auto-completion”.
- DevOps : In-charge of CI/CD . Used docker, chef , Jenkins and some shell scripts for OS-provisioning and delivery pipeline.

### Survey Panda

- This was our training project we as a team designed Survey Panda for PC support help desk feedback. Project was evaluated as the best training project of October 2016 batch and product was chosen to be used at PC support help desk.
- I was part of backend team working on Spring boot (J2EE). My main task was to design survey database and validate survey form and post feedback form.

### Dynamic Search UI [\[see project\]](#)

- Initial team member responsible for improving dynamic search UI of Rakuten ichiba.

<i>Systems Engineer</i>	<i>Infosys Ltd., India</i>	<i>Mar. 2011 - Aug. 2012</i>
-------------------------	----------------------------	------------------------------

### Molina Healthcare management

- Molina was a client of infosys and they wanted to upgrade their health care product. It was a pretty big project and we were team of 20+ engineers. I was responsible to manage users' health related information in database. I designed the DB in SQL server and performing stored procedures, triggers and functions to optimize it.
- Technologies : Java,Python, SQL server.

### Employee SWAP portal

- India with big demographic region sometimes people gets posting at a place where they don't want to live. So my team and I have designed one web based swap portal for employees who are interested to swap their posting locations. It made transfer process easy.
- Technologies : Java, Jsp, Servlet.

<i>Teaching assistant</i>	<i>VJTI Mumbai, India</i>	<i>Sept. 2013 - June 2014</i>
---------------------------	---------------------------	-------------------------------

During my Master of Technology course I was working as a Teaching assistant in VJTI for various courses mentioned below.

Courses: Data structures and algorithms, Software engineering, Mathematical foundation for computer science.

## EDUCATION

<i>Master of Technology</i>	<i>Software Engineering, Department of CS (VJTI), Mumbai , India</i>	<i>Sept. 2012 - June 2014</i>
<i>Bachelor of Engineering</i>	<i>Computer Science &amp; Engineering (RGPV), Bhopal, India</i>	<i>July 2006 - June 2010</i>

- *Post Graduate Coursework:* Human computer interaction; Design patterns; Artificial Intelligence; Computational Theory; Network analysis, Advanced Data Mining;

- *Undergraduate Coursework:* Operating Systems; Databases; Algorithms; Digital electronics, Programming Languages; Comp. Architecture; Engineering Mathematics, Data Mining.

*Thesis work :*

#### **Data on the cloud (2014)**

- Provided different ways of putting analyzed data into cloud and performed a theoretical as well as practical comparison between two famous storage formats HDFS and Amazon S3 on various parameters like Scalability, Durability, Persistence, Price, Performance, Security, Limitations etc.

## **FUN PROJECTS**

---

*For my Fun Projects I create github repositories to convey some vertical ideas.*

- *NLP-with-NLTK* [[see project](#)]
- *Cassandra snapshot Backup and restore* [[see project](#)]

## **ADDITIONAL EXPERIENCES/ AWARDS/ HOBBIES**

---

- [Rakuten Project Award](#) for GSP.
- Event organizer in Rakuten Technology conference 2014.
- Placement coordinator of software engineering branch during 2013-2014
- **Hobbies** : Solving puzzles, Reading about International affairs, Playing Badminton.

## **LANGUAGE AND TECHNOLOGIES**

---

- Language : Java , Python, SQL , bash
- Server : Apache tomcat, Jetty, Nginx
- Big data technologies : HDFS (HDP), Yarn, MapReduce, Spark, Kafka, Zookeeper, Ambari, talend
- Search technologies: Apache lucene, Solr
- Database : Mysql, Cassandra (Apache, DSE), MongoDB, MariaDB, SQL server, Redis
- DevOps/CI : Jenkins, chef, docker, Vagrant, Apache Maven
- IDE/Editors: IntelliJ Idea , PyCharm, Atom, Vim, Jupyter, franchise
- Tools : github, Jira, Confluence, Bitbucket.
- Web Frameworks: Django, Flask, Laravel, VueJs
- OS : Ubuntu, CentOS, Amazon AMI
- Data Science : NLP, NLTK, aspect-analysis, sentiment-analysis, Recommendation Engine
- Data Visualization tools : Tableau, DOMO, incubator-superset, matplotlib
- Cloud Platform: AWS, Digitalocean, Amazon Lightsail