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

Avinash Mishra

+81-70-2188-8241
avinash.mishra.2388@gmail.com
Nissincho 23-01, #911

- IT Professional with over 4.5 years of experience in the domain of Distributed Systems, Big data technologies and Data Science.
- Team leading experience, consisting of very diverse team members and motivated engineers.
- Master of Technology in computer science and Engineering.
- Proven ability to manage multiple projects while meeting challenging deadlines.
- Love to play with data.

EXPERIENCE

Data Scientist/Team Lead	*Paradigmshift.io, Japan*	*Sept. 2016 ~*

Repchecker **[**see project**]**

- Repchecker is a B2B cloud based reputation checker of modern hotel industry.
- I am responsible for designing the search component of it. Customer can search and monitor their competitors on selected parameters. I am working on it from scratch, designed the architecture which involves technologies like Java, Cassandra multi-node cluster, Solr, Hadoop, Nginx and Jenkins.
- Managing AWS instances of Repchecker server.

Valuechain **[**link**]**

- Valuechain is an NLP tool to process review data from OTAs.
- I am leading this project and we are team of 3 members.
- I have redesigned the architecture and we perform a lot of NLP techniques like Noise removal, Lexicon Normalization, Object standardization, Entity Parsing, Statistical featuring and text classification and matching etc.
- We recently started doing decision tree based supervised machine learning to predict customer's revisit.
- Tools and technologies used : NLTK, Kuromoji, Apache Lucene, Java, Python, Maven.
- Managing AWS instances of Valuechain server.

Distributed web crawle **[**link**]**

- I am also leading this project. We are a team of 4 engineers for it.
- I designed distributed web crawling architecture. It has centralized controller, distributed spiders and distributed storage in Cassandra cluster.
- Tools and technologies used: Apache Spark, Apache Kafka, Scrapy web crawling framework, Apache zookeeper, Django, Python, bash.

Price RecSys **[**link**]**

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

Data Lake

- It is a centralized repository hosted on Digitalocean which stores all structured, unstructured and semi structure 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.

Systems Engineer	*Infosys Ltd., India*	*Mar. 2011 - Aug. 2012*

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.

Teaching assistant	*VJTI Mumbai, India*	*Sept. 2013 - June 2014*

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

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

- *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 hadoop 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.

TECHNICAL EXPERIENCE

Fun Projects

- *Content based image retrieval system*: Used Java to implement exact match algorithm for comparing images and based on Euclidean distance method retrieved optimal match.
- *Cassandra snapshot Backup and restore* **[**see project**]**: Using nodetool snapshot logic I created a cassandra snapshot backup as well as restore program. I am using ansible and python to do this task.

ADDITIONAL EXPERIENCES/ AWARDS/ HOBBIES

- Rakuten Project Award for GSP.
- HackerRank : Among top java developers in Japan. avinash_mishra02 *Current rank is 1*
- My profile link : avinash-mishra.github.io
- 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.
- Best business Idea recognition to increase traffic in public chat of Viber app.

LANGUAGE AND TECHNOLOGIES

- Language : C , 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, Laravel, VueJs
- OS : Ubuntu, CentOs, Amazon AMI
- Data Science : Scikit-learn, NLP, NLTK, KuromoJi, Pandas, sentiment-analysis.
- Data Visualization tools : Tableau, DOMO, incubator-superset, matplotlib
- Cloud Platform: AWS, Digitalocean, Amazon Lightsail

LANGUAGES