

PRADEEP RAJA MOHAN

Chicago, IL | pmohan3@hawk.iit.edu | Phone: +1 (312) 721-1940

LinkedIn: [linkedin.com/in/mpradeep1994](https://www.linkedin.com/in/mpradeep1994) | github.com/mpradeep1994 | pradeepraja.me

EDUCATION

Illinois institute of technology – Master’s, Information Technology & Management (GPA: 3.7/4.0) Dec - 2017

Anna University, Chennai, India -Bachelor of Technology, Information Technology (CGPA: 8.1/10.0) May - 2015

TECHNICAL SKILLS

Programming Languages: JAVA, Python, C++
Database: MS-SQL, MySQL, Oracle, MongoDB, HBase
Distributed System Computing: MapReduce, SQOOP, Hive, Elastic Search (ELK), Spark, Kafka
ETL Tools: SSRS, SSIS, Pentaho, Data Integration & Pipelining
Cloud Services: EC2, Lambda, Dynamo Db, S3, RedShift
Version control: Git, SVN

EXPERIENCE

Application Developer Intern at Federal Home Loan Bank of Chicago (Chicago, IL) May 2017 - Aug 2017

- Created and Implemented a system design which involves automation of AWS instance - shutdown/restart by lambda function (Python) with a UI designed in MVC framework that lists features for all instances to customize work timings
- Automated the detection and turning down AWS instances with less utilization, saved \$200k / Month to the organization
- Designed scripts to write/read Instance tags, which is used as storage fields for instance work timings, instead of a data store
- Implemented an ETL process in SSIS (SQL, C#) to estimate the customers pledged collateral values by pipelining from CUSIPS generation in staging tables to transferring data to pricing vendors (IHS, Bloomberg, IDC)
- Successful job creation, repository maintenance and troubleshooting performance issues of ETL pipeline is carried out

Software Data Engineer Intern at Metarvrse Technologies (Chennai, India) Mar 2015 - Dec 2015

- Developed a data life cycle pipeline to migrate virtual user experience data from VR device to MySQL data store for analysis
- Carried out the Extract, Transform and Load operation to support the data scientists for modeling operations from different VR-AR devices based on business requirements and cleaning to create a consistent target data mart after transformation
- Successfully carried out data imputation operations to maintain consistency using dummy variables
- Successfully resolved inconsistencies across 90% of the usable data and achieved 100% success rate in avoiding selection bias
- Created an automatic Data Quality Report for data, which improved efficiency and accuracy based on data transformed

PROJECTS

REAL-TIME ANALYSIS OF TAXI PRICE SURGE (Aug 2016) GitHub Link - <https://goo.gl/xaPZCW>

- Built a Real-Time Big Data Processing Pipeline prototype using Kafka for data ingestion and Spark for analyzing taxi surge price
- Surge price for sample locations is calculated by analyzing real-time pricing data through UBER API in Python by processing in spark and designed a database model for this Big Data in HBase
- Successfully processed the data using Spark RDDs and visualized the surge behaviors on Tableau

WEB LOG ANALYSIS USING HADOOP AND ITS TOOLS (Jan 2017) GitHub Link - <https://goo.gl/gZLmpq>

- Created an analysis on Historical web log in Hadoop using MapReduce, SQOOP, HIVE and PIG
- Ingested data into MySQL and HDFS using SQOOP. Insights like most visited URL by each unique IP address, most occurred error code and its reason analysis and summarized the page content using MapReduce, HIVE and PIG

SENTIMENT AND CLUSTER ANALYSIS THROUGH TWITTER API (Nov 2016) GitHub Link - <https://goo.gl/cDiy76>

- Developed a python application which identifies the twitter user who uses the current trending hash tags and builds a graph between them. Used Girvan Newman cluster analysis algorithm, clusters among them are identified
- Tweets are classified using Machine learning models - SVM, Naïve bayes and logistic regression for predicting sentiment. Computed testing accuracy for each model using Cross validation

PERSONAL MEDICAL TRACKING APPLICATION (Jan 2016) GitHub Link - <https://goo.gl/JRYYYC>

- Created an end to end product for family medical tracking using core Java in Object-Oriented Design and modeled database
- Functionality to keep track of medicines of family members, doctors use the medical history of family & setting up appointments
- Implemented CRUD functionalities following MVC design pattern for modeling user interface to its underlying data models
- XML Parsers (DOM and SAX) are created to load the data from XML to MySQL Database

BIO-METRIC BASED AUTHENTICATION IN ATM (Jan 2015) GitHub Link - <https://goo.gl/bo5GYK>

- Implemented a highly-secured ATM application using Iris image of individual user to authenticate transactions securely
- Digital image processing algorithms like canny edge detection and Hough’s circle detection algorithm for parsing the iris image for authentication with the use of Public Key Infrastructure (PKI), MD5 hashing for client and server-side validations

CERTIFICATIONS

Hadoop certification FITA Services | EMC² Certified Academic Associate, Cloud Infrastructure and Service. | Microsoft Certified Solution Associate | ITIL v3 Foundation certificate on IT Service Management. | Oracle Certified Java SE6 Developer

PUBLICATION

- Published paper titled “Secure Banking Based on Machine Learning in IRIS Pattern Recognition” in IJCTA. ISSN: 0974-5572
- Presented and published paper titled “Identifying and Optimizing Data Duplication by Efficient Memory allocation in Repository by Single Instance Storage” in IRAJ. ISBN: 978-93-85465-30-7