

PRADEEP RAJA MOHAN

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

LinkedIn: <http://www.linkedin.com/in/mpradeep1994> | Portfolio: <http://www.pradeeppraja.me>

EDUCATION

Illinois institute of technology – Master's in Information Technology and Management (GPA: 3.8/4.0) Dec - 2017

► Advance Database Management | Distributed computing | Online Social Network Analysis | Data Analytics | OO Development

TECHNICAL SKILLS

Programming Languages: JAVA, Python, C++

Big Data: Hadoop, MapReduce, SQOOP, Hive, Elastic Search (ELK), Spark, Kafka, Tableau

ETL: SSIS, Pentaho, Data Integration & Pipelining

Database Technologies: MySQL, Oracle, MongoDB, HBase, DynamoDB.

Version control: SVN, Git, JIRA

AWS Services: EC2, Lambda, Dynamo Db, S3, RedShift

Machine Learning: Linear, Logistic Regression, Decision Tree, K-Means Clustering, Support Vector Machine

EXPERIENCE

Federal Home Loan Bank of Chicago (Chicago, IL)

May 2017 - Aug 2017

IT Application Engineer

- Created and Implemented a system application 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 them to pricing vendors (IHS, Bloomberg, IDC) the lowest price is feed into STOC.
- Successful job creation, repository maintenance and troubleshooting performance issues of ETL pipeline is carried out.

Metarvrse Technologies (Chennai, India)

Mar 2015 - Dec - 2015

Software Engineer

- Involved in entire data life cycle from data collection, mining, filtering, migration, analysis and reporting.
- Processed and Migrated around 10000 raw records every day from different VR-AR device usage stats sources by filtering bases on business need and cleaning to create a consistent target data mart after transformation.
- Created an automatic Data Quality Report for data, which improved efficiency and accuracy on data transformed.
- Worked in both Waterfall and Agile methodologies of the SDLC.

PROJECTS

github.com/mpradeep1994

REAL-TIME DATA PIPELINE TO ANALYZE UBER PRICE SURGE

Aug – 2016

- Built Real-Time Streaming Data Processing Pipeline using Kafka for data ingestion & Spark for analyzing behavior in price.
- Price surge is calculated by analyzing with historical price data with the real-time price based on location collected by batch processing and visualized using Tableau.
- Data scraping and transformation was done on important fields to extract only the fields of interest using Spark.

WEB LOG ANALYSIS USING HADOOP AND ITS TOOLS

Jan – 2017

- Created an analysis on Historical web log in Hadoop using MapReduce, SQOOP, HIVE and PIG.
- Imported 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

- Developed an application in Python that streams tweets from twitter, users who has used the hashtags and builds a relationship graph between users Friend's suggestion for the users has been made.
- Parsed the JSON twitter data filtered and Classified tweets using Machine learning models like K-Means and Girvan Newman cluster analysis algorithm to create clusters among the users.
- Logistic Regression, SVM and naïve based is used to predict sentiments for tweets, computed test accuracy using Cross validation.

SALARY PREDICTION MODEL FOR EMPLOYEES IN CHICAGO

May - 2016

- Implemented a salary prediction model, consuming department and job description to predict salary using Logistics regression, Linear regression and Decision Tree by tuning parameters to accuracy benchmark is performed.
- Data set has been pre-processed and SMOTE algorithm is used to correct data imbalance.

BANKING CONSOLE WITH CUSTOMER ANALYSIS

Jan - 2016

- Created a banking application using core Java and provided visualization using JavaFx.
- Implementation CRUD functionalities in MVC architecture for modelling user roles and its features.
- XML Parsers (DOM and SAX) are created to load the data to MySQL Database from XML.
- Using the Historical Data, Business oriented analytics has been performed to identify ideal customer for providing credit cards.

CERTIFICATIONS

Hadoop certification FITA Services | **Oracle Certified Java SE6 Developer** | **IBM Certified DB2 Academic Associate** | **EMC² Certified Academic Associate, Cloud Infrastructure and Service** | **ITIL v3 Foundation** certificate on IT Service Management.

PUBLICATION - Presented and published paper titled "Identifying and Optimizing Data Duplication by Efficient Memory allocation in Data Repository by Single Instance Storage" in IRAJ. **ISBN: 978-93-85465-30-7.**