Manthan Thakar

617-371-7788 | thakar.ma@husky.neu.edu | manthan.io | github.com/Manthan787

EDUCATION

Northeastern University, Boston, MA
College of Computer and Information Science, **GPA**: 3.7/4.0 Jan 2016 - present
Candidate for Master of Science in Computer Science Expected graduation: May 2018
Related Courses: Natural Language Processing, Machine Learning, Information Retrieval, Large Scale Data Processing
Activities: Master's Teaching Assistant, CS6120 Natural Language Processing

Sardar Vallabhbhai Patel Institute of Technology, Vasad, India
Bachelor of Engineering in Computer Engineering, **CGPA**: 8.16/10 Apr 2015

TECHNICAL KNOWLEDGE

Languages: C, Go, Java, Python, R, Scala, Javascript, Racket, HTML, CSS
Cloud / Devops: AWS EC2, AWS EMR, Docker
Software: git, Apache Spark, Hadoop, GraphQL, Scikit-learn, NLTK, NetworkX, Pandas, Django, Flask, Hive, Flume, Kafka, Akka
Databases: Postgres, Oracle, SQL Server, MySQL, MongoDB, DynamoDB

WORK EXPERIENCE

Programming Research Lab, Northeastern University, Boston, MA Dec 2017 - present
Research Assistant
- Analyzing all Javascript packages on npm using dynamic analysis and dependency analysis and developing toolchain in Node js to perform experiments

Lexumo, Burlington, MA Jan 2017 - Aug 2017
Software Engineer Co-op
- Built a crawler to crawl github for 100K open source Python software projects, analyzed programming patterns using ASTs and built a source code similarity model using TF-IDF, Tree Edit Distance and Characteristic Vectors
- Designed and developed Lexumo's licensing feature for C/C++ and Java source code, responsible for extracting license information from source code using String Matching and Information Retrieval based methods
- Designed ETL pipeline to collect metadata about all open-source projects written in Java, Javascript & C

Northeastern University, Boston, MA May 2016 - Dec 2016
HPC Graduate Assistant – Research Computing, Information Technology Services
- Automated tasks using Python and Bash scripts for monitoring of Northeastern's Discovery HPC cluster
- Assisted researchers in automating and parallelizing jobs by designing SLURM and IBM LSF batch scripts

Acharya, Vadodara, India Dec 2014 – Dec 2015
Software Engineer
- Designed a single page SaaS application for educational institutes using AngularJS and jQuery
- Designed and developed REST API using Laravel framework and deployed on AWS

PROJECTS

Join optimization of RDDs Oct 2017 – Dec 2017
- Improved Apache Spark's RDD **join operation execution time by 24%** and **reduced network I/O by 89%,** by building a **Scala compiler plugin**
- Modified Spark's join implementation to incorporate Broadcast joins and **improved performance by 50%**

PyDO: Python Distributed Objects Nov 2017 – Dec 2017
- Designed and implemented a distributed object system in Python that allows **distributed object-oriented programming** and keeps object distribution, caching, placement and migration transparent to programmers

Distributed Key Value Store Nov 2017
- Designed and Built a fault-tolerant distributed key value store with proxy server in Go and node servers in Python

Clustering Music Artists using Apache Spark Oct 2017
- Implemented agglomerative hierarchical clustering and k-means clustering from scratch in Scala on top of Apache Spark to cluster music artists and deployed application on Amazon EMR

Flight Suggestions Sep 2017
- Implemented a flight suggestion algorithm to suggest two-hop flights that minimize the chance of missing a connection in R by processing historical flight data using Hadoop (Java)

Hitch - *A cross-platform Airdrop* Aug 2017 - Sep 2017
- Built a Desktop and Android application for file transfer on local network using multicast DNS protocol for service discovery, TCP connections for data transfer, electron javascript framework for Desktop App and Java for Android App

Search Engine May 2016 – Aug 2016
- Developed a multi-threaded crawler using python with distributed frontier management to crawl 200K web pages, indexed to a distributed Elasticsearch index and implemented PageRank and HITS algorithms to rank websites
- Trained a Linear Regression model to increase average precision of IR system by 156%