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 May 18
Master of Science in Computer Science
Related Courses: Natural Language Processing, Machine Learning, Information Retrieval, Large Scale Data Processing
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 15

TECHNICAL KNOWLEDGE

Languages: C, Go, Java, Python, R, Scala, Javascript, Racket, HTML, CSS
Cloud: AWS EC2, AWS EMR, Docker, gitlab-ci, Ansible
Software: Apache Spark, Hadoop, GraphQL, Scikit-learn, Tensorflow, Django, Flask, Kafka, Akka, Celery, Lucene, Elasticsearch, dplyr
Databases: Postgres, Oracle, SQL Server, MySQL, MongoDB, DynamoDB

WORK EXPERIENCE

Programming Research Lab, Northeastern University, Boston, MA Dec 17 - present
Research Assistant

- Analyze all Javascript packages on npm using dynamic analysis and dependency analysis; Developed toolchain in NodeJS to perform experiments

Lexumo, Burlington, MA Jan 17 - Aug 17
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 Licex, a license extractor tool to extract licenses from 300K open-source Java and C projects using string matching and information retrieval methods
- Implemented an efficient streaming YAML parser in Python that decreased memory pressure by 200%
- Designed ETL pipeline to collect metadata about all open-source projects written in Java, Javascript & C

Northeastern University, Boston, MA May 16 - Dec 16
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 14 – Dec 15
Software Engineer

- Designed a single page SaaS application for educational institutes using AngularJS and jQuery; Built the REST API using Laravel framework and deployed on AWS

SELECTED PROJECTS

Fiddler: Composing music using Neural Networks [Project: https://git.io/vxD43] Feb 18 – Current

- Implemented a configurable Recurrent Neural Network architecture using Tensorflow to train a model that learns to compose music from sequential music sheet data and generates novel compositions

Join optimization of RDDs [Project: https://git.io/vxD4W] Oct 17 - Dec 17

- Built Scala Compiler Plugin and Improved Apache Spark's RDD join operation execution time by 24% & reduced network I/O by 89%; Modified Spark's join implementation to incorporate Broadcast joins and improved performance by 50%

PyDO: Python Distributed Objects Nov 17 – Dec 17

- 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 [Project: https://git.io/vNEnV] Nov 17

- 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 [Project: https://git.io/vNEcr] Oct 17

- 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

Hitch - *A cross-platform Airdrop* [Project: https://git.io/vNEcb] Aug 17 - Sep 17

- 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.