

HARSHALI SINGH

Available from Jan 2017

(617) 373-0528 | singh.ha@husky.neu.edu | 42 St Germain St, Boston, MA 02115 | github.com/harshalisingh

EDUCATION

Northeastern University, Boston, MA

Expected Graduation: Dec 2016

Candidate for a Master of Science in Computer Science, GPA: 3.6

Courses: Algorithms, Parallel Data Processing in MapReduce, Statistics and Data Analysis in R, Data Visualization

Teaching Assistant: Software Development (Spring 2016), Information Retrieval (Spring 2015)

University of Mumbai, Mumbai, India

Bachelor of Engineering in Computer Engineering, GPA: 3.79

May 2012

TECHNICAL SKILLS

Languages: Java, C#, Python, Scala

Cloud Services: Amazon EC2, S3, EMR, Microsoft Azure

Platforms: Linux, Mac OS, Windows

Databases: SQL Server, MySQL, Elasticsearch

Statistical Tools: R, Tableau, Pandas, Scikit-Learn, D3.js

Big Data: Hadoop, Spark, MapReduce

WORK EXPERIENCE

Red Hat, Raleigh

May 2016 – Aug 2016

Application Platform Intern

- Scraped job websites of Red Hat's top 50 Middleware customers using ParseHub scraping tool. Performed text mining in R on around 20,000 IT job descriptions to identify open source technology and language trends.
- Performed Hierarchical Clustering on the data and plotted dendrograms to identify term similarities.
- Created interactive visualizations in Plotly to represent geographical distribution of programming language hiring trends in USA using Choropleth maps.

Boston Consulting Group, Boston

June 2015 – Dec 2015

Software Developer Co-op

- Developed the UI and business layer of a website using ASP.NET MVC, JQuery and NHibernate to manage invoices and employee's finance reports subscriptions.
- Developed a console application in C# to run as a weekly scheduled job to retrieve the differences in journal entries between two financial accounting (General Ledger) tables in Oracle.

Microsoft, Bangalore, India

July 2012 – April 2014

Technical Consultant

- Delivered high quality service to Microsoft Premier customers by solving business critical technical problems on ASP.NET web sites. Engaged with solution architects to perform risk analysis on large web server farms.
- Developed a diagnostic engine in C# that automated collection and analysis of server's configuration data.
- Analyzed and debugged web application performance issues (high CPU, memory leak and application crash).

ACADEMIC PROJECTS

- Parallel Data Processing Framework (Java, EC2, S3, Shell):** Designed and developed own lightweight version of MapReduce framework in Java which runs in parallel across a distributed cluster of AWS EC2 instances or on a single computer using all of its available CPU cores. Achieved EC2 inter-node communication, cluster management, end-to-end automation and fault-tolerance.
- Distributed Sorting (Java, EC2, AWS CLI, Shell):** Implemented Sample Sort algorithm to sort 6TB of weather dataset on Amazon EC2 instances. Automated creation, destruction of clusters and execution of sort jobs.
- Big Data Analytics (Java, Hadoop, EMR, Weka):** Analyzed 27 years of Bureau of Transport Statistics' On-time Performance dataset on EMR clusters using Hadoop. Built a prediction model using Naïve Bayes Classifier and Weka API to predict flight delays and suggest optimal two-flight routes.
- Missed Connections (Spark, Scala):** Computed missed connections for each airline in Scala using Apache Spark.
- Search Engine (Python):** Developed a Search Engine using Elasticsearch Index and Okapi BM25 ranking model. Queried a corpus of 85000 stemmed documents and returned ranked matching documents.
- Twitter Sentiment Analyzer (Python, Pandas, Scikit-Learn):** Analyzed tweets and performed sentiment analysis on 2016 GOP Presidential Debate data as part of Kaggle competition.