

# NAVEEN NATARAJAN

Email: naveen.natz@gmail.com

Website: <http://cs.jhu.edu/~nnatara2/>

Phone: 443-562-8451

---

## EDUCATION

### Johns Hopkins University, Baltimore, MD

Aug 2011 - May 2013

- Master of Science & Engineering in Computer Science (MSE)

### Anna University, India

Aug 2007 - Apr 2011

- Bachelor of Technology in Information Technology (B.Tech)
- Won inter college programming contests.

## EXPERIENCE

### Dow Jones

July 2013 - Present

- **Position:** Product Technologist- Application Development and Infrastructure

### Children's Hospital of Philadelphia (CHOP)

June - Aug 2012

- **Position:** Summer Intern in Database Services Team

## PROJECTS

### Enhancing Customer Support Website

**Technology Used:** Java, JavaScript

Included a time out to lock the screen and display password prompt after a certain time interval. Lock the screen and display the password prompt when certain features are clicked. Display and hide certain features for different kinds of users.

### Generating Chat Reports

**Technology Used:** Pentaho, Java, JavaScript, Oracle

Designed pentaho process to generate hourly and daily chat reports.

### Molecular Dynamics Database Project (Mddb)

**Technology Used:** Python, Greenplum/PostgreSQL, ETL process, Query Processing & Optimization

This is an adaptive control system used for scientific simulations. I designed a common parser to parse different energy file formats using python scripts. This involved implementing ETL process in which I extracted data from different kind of files, cleaned the data, transformed and loaded it into a Greenplum database. Designed a schema for the database and ways to efficiently load and manage the data in the database which is in the order of terabytes. Came up with new queries and designed queries to query the data efficiently and analyze them. I also worked on the optimization of these queries.

### Comparison of Iterative Processing of Large Data Sets in Hadoop and HaLoop

**Technology Used:** Java, Hadoop and HaLoop frameworks

The Hadoop platform lacks built-in support for iterative programs, which arise naturally in many applications including data mining, web ranking, graph analysis etc. HaLoop extends MapReduce with programming support for iterative applications. I evaluated both the frameworks on real queries and real datasets such as Twitter tweets, mutual friends and web ranking. The limitation and advantages of both the frameworks were analyzed in detail.

### Comparison of Parallel DBMS(MPP) and Distributed DBMS (NoSQL)

**Technology Used:** Greenplum, Cassandra and HBase

Parallel and distributed DBMS are used to manage large amounts of data. In this I compared if massively parallel database management system (MPP) can offer the same or comparable or better performance and scalability than a distributed database system (NoSQL) on different datasets. I compared and benchmarked these three databases on performance, I/O operations, throughput and scalability. The limitation and advantages of these different systems were analyzed in detail.

### Paper Bot – Distributed Information Retrieval System

**Technology Used:** Python, Web Scraping, Web Robots, HTML

Paper Bot is a meta-search engine that blends the top web search results from Google, Microsoft and CiteSeer. Distributed information retrieval comes into play when a user wants to get information from different sources in parallel. It utilizes collection fusion algorithms to compile results from many of the Web's major search properties, delivering more relevant and comprehensive results every time a search is done.

### Jay List – An Android Application

**Technology Used:** Java, Android Programming

Jay List is an android application which is used to help Hopkins students to post and search about housing, buy and sell things within the Hopkins community. This uses the REST architecture. I was involved in both the server side development and the development of Android side functions.

## SKILLS

- **Programming Languages:** Java, Python, C++, JavaScript
- **Database:** MySQL, PostgreSQL, Cassandra, Oracle
- **Operating Systems:** Microsoft Windows, Linux
- Hadoop, Pentaho, HBase, Hive, Sqoop, Amazon Web Services, Web Robots