

# NAVEEN NATARAJAN

Email: naveen.natz@gmail.com

Website: <http://www.nnatz.com/>

Phone: 443-562-8451

## PROFESSIONAL EXPERIENCE

### Rock Central (Quicken Loans)

**Position:** Senior Big Data Engineer

*March 2017 - Present*

- Designed and built data domain(Mesh) to manage the data effectively and for quick retrieval for data products
- Built an API on DynamoDB to enable faster access to data.
- Configured the deployment environment on AWS EMR and enabled it to run effectively for large data.
- Developed the GAM data product and API using config driven approach using pyspark.
- Built privacy by design application used by many different teams to enable compliance with CCPA regulations.
- Part of team which created data lake and moved data from on prem to cloud.
- Deploy and maintain the existing applications on AWS and Cloudera clusters.

### IMS Health

*June 2015 – March 2017*

**Position:** Software Engineer – Big Data

- Designed and architected a scalable and reliable system to gather data from various social platforms.
- Built the system using tools like Hadoop, Nutch, Solr, Kafka, Storm, Cassandra, Flume and MySQL.
- Improved query turn-around time by more than 60% using Spark Dataframes.
- Built a data api layer to query and retrieve data from the system.

### Dow Jones

*July 2013 - June 2015*

**Position:** Application Developer

- Develop and update the existing customer self-service site that is used by all the WSJ & Barron's customers.
- Built a fault-tolerant abstraction layer that would handle request from all the customer service systems.
- Built many proof of concepts using Python, NodeJS and Java.

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

*June - Aug 2012*

**Position:** Database Developer Intern.

## TECHNICAL SKILLS

- **Languages:** Python, Java
- **Big Data Tools:** Spark , Hadoop, Kafka, NiFi, Phoenix, S3, SageMaker, EventBridge, EMR, Glue, Athena
- **Databases:** DynamoDB, Hive, Hbase
- **Frameworks:** Map Reduce, Spark, Amazon Web Services(AWS), Hibernate, Java Spring
- **Technologies:** Shell, Linux, SQL, Git, Oozie, ETL, Cloudera, CICD

## EDUCATION

### Johns Hopkins University, Baltimore, Maryland

*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)

## PROJECTS

### Data Domain(Mesh) Platform

**Technology Used:** Python, Spark, AWS EMR, CICD, Athena, Glue ETL, Oozie, Hive

Designed, architected and built a reusable and config driven application to gather data from various sources. The data is extracted and loaded to the pre processed layer from the data lake. The data is then transformed, quality checked and survivorship rules applied before storing in conformed layer. The data in the conformed layer is cataloged in Glue and can be queried in Athena.

### Get A Mortgage

**Technology Used:** Python, Spark, Hbase, Hive, Java, AWS API Gateway

Developed a pyspark application which generates data which is used by different applications to populate client information. The data is gathered from various sources using a hive query, then passed through the ETL where quality checks and survivorship rules are applied and finally written to a Hbase table. A java spring based API is used to query the data. The API is deployed on AWS API gateway.

### Privacy By Design

**Technology Used:** Python, Spark, AWS S3, Hive

Designed and implemented a pyspark application which removes the client information who requested to be removed to adhere to CCPA regulations. The application takes the input data and scans it to check if any details of the client to be removed is present and removes that data.