NAVEEN NATARAJAN

Email: naveen.natz@gmail.com Website: http://www.nnatz.com/ Phone: 443-562-8451

PROFESSIONAL EXPERIENCE

IMS Health *June 2015 - Present*

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.
- Developed an archival system as a proof of concept to archive and analyze the old data using Hive.
- Deploy, maintain and expand the cluster on Amazon Web Services.

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 a coupon manager system using Django.
- Developed an encryption and decryption algorithm in Java for storing documents in Nuxeo.
- Built many proof of concepts using Python, NodeJS and Java.

Children's Hospital of Philadelphia (CHOP) *June - Aug 2012*

Position: Database Developer Intern

- Twelve weeks of summer internship in the Database Services Team as a Database Administrator.
- Worked on Oracle, Microsoft SQL Server and Epic Cache Databases.

TECHNICAL SKILLS

- **Languages:** Java, Python
- **Big Data Tools:** Hadoop, Kafka, Storm, Flume, Solr, Spark
- **Databases:** NoSQL(Cassandra, Hive), Relational(MySQL, PostgreSQL)
- **Frameworks:** Map Reduce, Spark, Amazon Web Services(AWS)
- **Technologies:** SQL, Git, Jenkins, HUE, Oozie, Eclipse, ETL, Cloudera, Django, Agile
- **Operating Systems:** Linux, Windows

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

Nexxus Social Big Data Search Platform

Technology Used: Java, Hadoop, Flume, Nutch, Solr, Cassandra, Storm, Kafka, AWS, MySQL

Designed, architected and built a system to gather data from various social media platforms like Twitter, Facebook, news sites, blogs, forums, etc. The system is designed to be reliable, scalable and fault tolerant. Wrote multiple map reduce jobs to gather data, clean it, tag the data based on the keywords and write it to Cassandra. The data in Cassandra is indexed by Solr so it could be searched. The data is analyzed, useful and relevant information is displayed on the dashboard based on the keyword searched by the user.

Query performance improvement using Spark

Technology Used: Python, Spark, Oracle

Wrote a Python Spark program which used dataframes to reduce query execution time by more than 60%. The program reads data from Oracle database, executes the queries on Spark using dataframes and writes the result back to the database.

Data API Layer

Technology Used: Java, PHP, Solr

Wrote a data api layer to retrieve data from the existing search platform. This can be used by other teams or users to retrieve data from the search platform or add new keywords or url's or pages to be searched by the search platform. The data would be returned in JSON format.

Archiving Old Data

Technology Used: Java, HDFS, Hive, Hadoop, Solr, Cassandra

Used map reduce jobs to read data from Cassandra and write it to HDFS for archival. Developed proof of concepts to analyze the data using Hive and index it using Solr.

Abstraction Layer

Technology Used: Python, Django, Postgres

Designed and developed an abstraction layer that handles all the calls made to the customer service systems. It is designed to be fault tolerant and also tolerant with respect to changes in contract between systems. This simplified the maintenance and development of the customer service systems.