**NAVIN CHAGANTI**

**Contact Details**: 245 E 35th Street New York Apt#11D ,NY-10016 **Cell number**: +1-(520)-440-8389 **Email**: navinchaganti@gmail.com **GitHub:** <https://github.com/navin619> **LinkedIn:** [Profile Link](http://www.linkedin.com/pub/navin-chaganti/60/5bb/283/" \o "View public profile)

**Education:**

**Master's in Computer Engineering (**University of Arizona, Tucson, AZ), **GPA:** 3.75.0/4.0  **Aug 2015**

* **Coursework:** Software Engineering Concepts, Engineering of Computer Based Systems, Operating Systems, Introduction to Machine Learning, Big Data Analytics, Distributed Computing, Statistical Natural Language Processing, Algorithms Graphs and Networks.

**Bachelor of Engineering in Instrumentation (**Mumbai University, India**) May 2011**

* **Coursework:** Data Structures and Algorithms, Introduction to Databases, Real-time Operating Systems, Device Drivers, Embedded Systems, Advanced Embedded Systems, Computer Architecture.

**Skills:**

* Four years of experience in Software Development using Python, C/C++ and Java.
* Experience in Ecommerce, Finance and healthcare IT systems and workflows.
* Experience in building ETL data pipelines using Spark SQL in Python, Scala and Java.
* Developed Datalake (Big data warehousing) on Cloudera, Hortonworks and AWS Hadoop Frameworks
* Experience in SQL or other relational databases and good understanding of data warehousing Extract, Transform and Load (ETL) principles
* Experience knowledge of Python scripts and scikit-learn Packages for implementing machine learning concepts and converting them to scalable Apache Spark Machine learning pipelines using the ML\Mlib libraries.

**Experience:**

**Data Engineer Jan 2018 – Present**

Adobe, New York, NY, USA

**Project:** Adobe Audience Manager

* Built ETL systems using Python, Apache spark SQL framework. Input sources of DCM data is pull from google Storage, AWS s3 etc. and then processed on AWS EMR cluster using Pyspark Scripts. All the final results are streamed into AWS kinesis which are later consumed by Spark streaming consumers to write into AWS s3 and Snowflake DWH.
* ETL jobs on Snowflake are pushed into the Redshift for the reporting capabilities using Tableau.
* Processing about 30 TB of incoming data per day and datawarehouse around 700TB of data on Snowflake, Cloudera and AWS as a mixed platform.

**Big Data Software Engineer Nov 2015 – Jan 2018**

KPMG LLP, New York, NY, USA

**Project:** Data Lake forReporting/ Analytics Data warehousing on Hortornworks Hadoop Platform

**Client:** Goldman Sachs

* Built ETL systems using Python, HIVE and Apache spark SQL framework. Storing all the result files in Apache parquet and mapping them to HIVE for Enterprise Data warehousing.
* Built Real-time data pipelines using Kafka and Python consumers to ingest data through Adobe Real-time Firehose API into Elastic Search and built real-time dashboards using Kibana.
* Developed DataMart’s in Sybase and connected them to Tableau for reporting dashboards.

**Client:** Nike

**Project:** Recommendation Engine for Nike

* Developed Machine learning pipelines with Apache Spark Ml\MLib libraries using Python to make predictions on 70 TB data.
* Automated the workflow using Airbnb Airflow tool, to run the machine learning scripts in a DAG manner.
* Developed data validation and test cases using Python Nose framework.
* Worked with Data scientist of Nike to understand the machine learning concepts and models and implement them on PySpark.

**Client:** Bank of America

**Project:** Common Ingestion Framework (Real-time data Ingestion)

* Implemented Spring Kafka framework to ingest data from different data source (file, RDBMS, MQ)
* Developed WEB API framework as ingestion point to Apache Kafka, routing all the HTTP calls into the Kafka topics.
* Built Flume Kafka Pipelines to consume data from Kafka and ingest into HDFS in Avro format.
* Implemented K-Streams API to transform data real-time and route it to different topics in Kafka, for seamless real-time consumption.
* Implemented Kafka Connect source and sink connectors to ingest and consume data from the Kafka pipeline.

**Software Engineer (Data) Intern Jun 2015 – Nov 2015**

Xerox (Midas+), Tucson, USA

**Project:** Live feed Collection

* Collaborated with the Advanced Analytics Team Members, and developed data integrations, developed and maintained Patient-Centric Analytics Data Warehouse including updates to data models, built and maintained services to existing NLP and Medical Coding products, supported data scientists by providing research data sets, interacted with data scientists to develop production services derived from statistical models.

**Graduate Research Assistant Jan 2014 – Aug 2015**

Autonomic Computing Lab, UA, Tucson, USA| Research Assistant

**Project:** Securing Cyber Space (Understanding the Cyber Attackers and Attacks via Social Media Analytics**)**

* Developed a NoSQL Hadoop Database which helps in recording the chats of hackers from different IRC (Internet Relay Chat) channels to detect situational awareness of the sources of attacks and provides better understanding of the community of cyber attackers.
* Developed MapReduce programs in java and implemented big data analytics of the hacker chat data.

**Associate Software Developer Feb 2012 - Dec 2013**

Rockwell Collins, Hyderabad, India

**Project:** Data Link DLCA6500(Software Developer)

**Client:** Bombardier Aerospace, Executive Embraer jet

* Software Developer for the Data-Link System (known as Controller Pilot Data Link Communications)
* Developed Software and the GUI for the cockpit display according to the client requirements and the FAA standards (DO-178B level C).

**Academic Projects:**

**Pothole Mapper Jan 2014 - May 2014**

* Developed android App to detect Potholes on the streets. This app uses accelerometer sensor to detect vibrations (potholes) while the user is driving .Demo regarding this app is given in the below link.
* Demo Link: <https://www.youtube.com/watch?v=jCwmcXcNoSw> Website: <http://navin619.github.io/Pothole_Mapper>.