Chapin E 2076,
700 Health Sciences
Drive Stony Brook NY
11790

SANKET DIGE

(631) 652-5022
sanketdige268@gmail.com

EDUCATION

Stony Brook, NY **Stony Brook University** **Aug 2016 - Dec 2017**

- Masters in Computer and Information Sciences – Fall'16.
- Courses: Operating Systems, Analysis of algorithms, Artificial Intelligence, Machine Learning, Comp-Networks.

Pune, INDIA **College of Engineering, Pune.** **July 2011 – May 2015**

- Bachelor of Technology in Information Technology, May 2015. GPA: 8.05/10
- Undergraduate Coursework: Cloud Computing & Big Data, Algorithms & Complexity, Data Structures and Algorithms, Systems Programming and OS, DBMS, Advance DBMS.

Addition Online Courses **Coursera** **Aug 2015 – Jan 2016**

- Machine Learning by Andrew Ng - Stanford University, Introduction to Big Data Analytics – UC San Diego

EMPLOYMENT

Software Engineer **Twilio** **Feb 2018 - Present**

- Designed and implemented a multi nodes dual DC based cassandra metastore and a Data discovery platform with streaming spark applications gathering metadata and live healthcheck statistics of datasets residing on S3, Redshift, Presto and Snowflake. It supports robust backend with APIs build using Akka HTTP framework in Scala and its frontend components in react-JS being rendered by a python flask based server.
- Organize, and partition massive data sets of both structured and unstructured data by implementing Python – Celery and batched Spark applications based infrastructure tool and system to house data at data-lake on Amazon S3 and load it in different warehouses like Redshift, Snowflake or Presto for reporting and analytics purpose.
- Redesigned and implemented integration service to integrate with internal third party systems like Salesforce, Zendesk, JIRA, and others to bring data at data-lake and then use it with our warehouse data loaders in warehouses like Redshift, Snowflake or Presto for analytics.
- Build and launch real-time or batch ETL data processing pipelines supporting data warehouse models for analytics with Batch and Steaming Spark. All the batched applications were scheduled and managed with our workflow engine build using Apache Airflow.
- **Technologies Used: Airflow, Celery, Redis, Python, Spark, Scala, Chef, Ruby, Amazon EC2, RDS-postgres, Cassandra, Presto, Snowflake, S3.**

Research Assistant **Stony Brook University** **Nov 2016 – May 2017**

- Working as a part-time research programming assistant with a team of pediatric health sciences researchers in building an asthma web application.
- **Technologies Used: HTML, JavaScript, JQuery, MongoDB, NodeJS, O-Auth2.0, REST APIs, S3 buckets storage.**

Software Engineer **Varian Medical Systems** **July 2015 – July 2016**

- Contributed in project "*Remote Software Deployment v1.2*" (RSD) by identifying and resolving a security bug between the service-to-service communication.
- Created scheduler service, monitoring service and integration of other project components in "*RSD v1.3*", for gathering of client data and brought back to Varian server for further analytics perspective.
- Changed the architecture of Varian's standalone *Varian-Deployment-Tool* by providing similar working like the RSD server web app. Involved replicating and utilizing same service architecture like in RSD server architecture.
- **Technologies Used: C#, WCF, WPF, Sql server, MS test, TFS, Visual Studio, Entity framework.**

LANGUAGES AND TECHNOLOGIES

- **PROG LANG - Java; Scala; Python; SQL; NodeJS, JavaScript; J-Query, MongoDb, Cassandra, AJAX, XML, NoSQL.**
- **NON-PROG – REST APIs, Akka HTTP, Presto, Airflow, Hadoop; MapReduce; Apache Spark; Agile (Scrum).**