

# Karthik Balasubramanian

1063 Morse Avenue 6-303  
Sunnyvale, CA, 94089

<https://karthikbalasubramanian.github.io/>

+1(412)-326-8433

[balasubramaniankarthik102@gmail.com](mailto:balasubramaniankarthik102@gmail.com)

## Education

**Master of Information Systems Management, Heinz College, Carnegie Mellon University** May 2017

**Relevant coursework:** Machine Learning (ML), Practical Data Science, Research in Data Science, Data Structures, Math for ML, Econometrics, Distributed Systems, Statistics and Database Management.

**Bachelor of Information Technology, SRM University, Chennai**

Apr 2013

## Skills

**Languages:** Python, Java/J2ee, Ruby, Scala

**Databases:** MySQL, CloudSql, Redis.

**Frameworks:** Apache Airflow, Akka,, Apache Hadoop, Django, Docker, Tungsten

**Machine learning packages:** Numpy, Scipy, Sci-Kit learn, Pandas, Keras, Matplotlib

## Professional Experience

**Marketo, An Adobe Company, San Mateo** Associate Software Engineer October 2017- Present

- Marketo Subscription Migration to Google Cloud – Architected a distributed workflow automation using Apache Airflow from scratch to move Marketo subscriptions from on premise data center to google cloud
- Showcased the company's technology debt - “**Double encoding problem**” to database replication vendors by creating a testbed environment of 2 3 node tungsten replication clusters and replicating data originated from latin1 driver enabled utf-8 databases using google translate apis.
- Architected a solution to dump and restore very large databases from database servers which has endured database version upgrades using Xtrabackup. Improved the efficiency of database dump and restore process of 1 Terabyte database from 6-7 days (logical backup) to 6 and half hours. Resolved Database replication problems between master and slaves, long running query problems in slaves and improved overall database server health.
- Conducted performance tests on usability of open source migration tools and frameworks at Marketo. Documented the advantages and disadvantages of using them and concluded the recommendation based on my observation
- Debugged and managed a ruby-rake based ETL metrics reporting system which reports weekly customers' Marketo platform utility metrics to product managers, customer success associates and engineering managers at Marketo.
- Wrote a multi-processed query report generator tool which helps software quality and engineering teams to run a set of queries on multiple customer databases and returns the aggregated query report in emails.

**SingTel DataSpark Analytics, Singapore** Data Engineering Intern Summer 2016

- Created an interactive Apache Spark Job configurator tool from the scratch using React.Js and Spring Data framework in an agile environment to auto-write ETL spark job configuration files.
- Exported the tool as a proprietary of DataSpark to its clients with documentation and tutorials.

**Tata Consultancy Services, Chennai, India** Assistant Systems Engineer, Jun 2013 – May 2015

- Held accountable for project estimation, use-case design, development, performance testing and rollout activities of a critical data integration project for a US based retailer.
- Optimized data transformations to facilitate large scale data transfer from client systems to enterprise cloud. Data movement efficiency was 99% which reduced data reconciliation process for the client.
- Actively involved in developing and documenting proof-of-concepts with data transformation and processing frameworks like Apache Camel which was later utilized across the client projects as boilerplates.
- Received Technical excellence award for successfully completing three of the eight major releases in the project

## Academic Projects

**Udacity Machine Learning Engineer Nanodegree** October 2018- Present

- Completed the first part - Machine Learning foundations course and working on second part - Advanced machine learning course in the two-part Nanodegree program.

- Acquired hands on experience in developing supervised, unsupervised, convolutional and deep neural net machine learning models on different problem sets as part of the course.

***Demand Forecasting Singapore Bus traffic based on Taxi movements***

***Jan 2016-Present***

- Visualized, observed and suggested improvements in Singapore's bus distribution schedules based on ad hoc taxi distributions for a 30 day time window.

***Airbnb price prediction tool***

***Sep 2016- Dec 2016***

- Created a pure python tool which predicts the price of an Airbnb listing based on its features and locational significance; Also considers competitor pricing and accordingly suggests a price and provides information like other nearby Airbnb listing price stats and advertisement suggestions.
- Performed hyper-parameter tuning with selected models and achieved accuracy score of 70% with limited dataset for the City of Pittsburgh.

***Hilti Quiz-Train***

***Aug 2015- Dec 2015***

- Created Hilti Quiz-Train, a native android application for Hilti mobile challenge 2016.
- Developed the backend using Spring Data framework and Synchronous Volley Framework. Responsible for material design philosophy adherence in mobile app. Completed second in the competition which saw a participation of over 120 teams around the world.