ANUBHAV GUPTA

650-784-2080 | https://www.linkedin.com/in/anubhavg1| anubhav.rg@berkeley.edu

EDUCATION

University of California, Berkeley
Masters of Information Management and Systems **2016(expected)**

Managing Analytics Projects, Applied Natural Language Processing, Scalable Spatial Analytics, Agile.js, Data Mining and Analytics, Information Visualization, Scaling UP! Really Big Data on the Cloud, Distributed Computing Applications and Architecture, Web Architecture, and Quantitative Research Methods for Information Systems using R.

Motilal Nehru National Institute of Technology
Bachelor of Technology in Computer Science and Engineering **2011**

Operating Systems, Data Structures, Analysis and design of algorithms, Computer Networks, Computer Architecture.

PROFESSIONAL EXPERIENCE

IBM Research
Summer Research Intern **2015**

I worked at the Accelerated Discovery Lab with the Collaborative Research Discovery Team on Integrating Jupyter/I Python with their application called LabBook for seamless Data Curation into their Data Lake.

• Built a pip installable python package that contained wrapper functions to communicate with the LabBook using REST API calls that had been exposed.
• Built an I Python GUI extension to save and schedule the curation script to be executed as and when needed.
• Built a Flask Web server to handle authentication and communication between the I Python server and the LabBook server using REST calls.
• Implemented the task to schedule and execute the curation script as per the details provided in the LabBook.

University of California, Berkeley
Graduate Student Instructor for MIDS W251 – Scaling UP! Really Big Data **2015**

Relevant Technologies: Softlayer Cloud, Openstack, DevStack, GPFS/HDFS, Hadoop, Spark, Cloudant, web search.

McAfee India Software Pvt Ltd
Software Engineer **2011 - 2014**

• McAfee Application/Change Control Version 6.0.1: Developed five features for the current Whitelisting product namely File Diff, Salting, Strong Hashing, z switch and checksum in MAPL. This resulted in a 20% increase in performance while applying filters.
• Loadable Kernel Module (MAC/MCC Version 6.1.0) - Worked on shifting the IP of the product from kernel to user space to resolve the GPL licensing issues which was essential to roll out the new version. As a part of LKM, since the processing shifted to the user level, in order to keep the performance constant we implemented a in memory cache. Extensively worked on system calls (open, close, read, write, mmap, ioctl etc). Also worked with filesystems and kernel APIs
• Worked on run time deployment of the product on unsupported kernels and support for 3.0 kernel with this release by replacing deprecated kernel APIs. This included run time compilation of the kernel module and then packaging it with the pre shipped binary. This resulted in adoption of the new product by 20% more customers.
• Developed a Sanity suite module which tested the deployment on an unsupported kernels by checking sum basic file ops like open/close, read/write/ which reduced the run time deployment failure by 30%.
• Worked on the customer issues and bugs that have come up since the product has been released in to the market thereby improving the product and making it bug free.

SKILLS

Languages : SQL, C, C++, HTML, CSS, Python (Flask), Scala, Java, shell scripting(bash).
Front End : HTML, CSS, Java script, Angular, React, Dojo
Backend : Flask, mongoDB, Sqlite, Django, Node
Big data tools : Hadoop, HDFS/GPFS, Spark, Open stack, Salt, Solr, Cassandra, Elastic search, Brooklyn.

ACADEMIC PROJECTS

- **This is Jeopardy!** - Analyzed the Jeopardy data from J-Archive using various data visualization techniques to explore patterns, discover trends and allow a user to get an idea about the game is, what are its facts and figures in a more intuitive and user

friendly way. For this purpose we used Tableau, High charts and D3 to come up with various visualizations to achieve the same (link: http://people.ischool.berkeley.edu/~anand_rajagopal/Jeopardy/)

- **NYCe Taxi** – This overall objective of this project was to develop an algorithm that can be used to analyze the pattern in which cabs operate in a city to try and predict the tip can be expected per fare. We used both classification models like SVM, Decision Trees, Random forests and Ada-boost classifiers as well as regression models like linear regression, SVM regression and Lasso regression.
- **Mumbler Application** – Using natural language processing (NLP), I created an application that takes any corpus and produces a phrase that chains together the most frequent pairs of words (2-grams) in the text (i.e. similar to a Markov text Generator). The data set was about 60GB (around 6 billion lines of data across 100 files) of compressed data on a 3 node cluster running on GPFS (General Parallel File System).
- **TravelBites** – I worked on a web application that helps travelers (foodies) build a food itinerary based on their interests and food suggestions from local experts. This project is focused on user experience research and we used contextual inquiry, affinity diagramming, personas and scenarios to get user inputs (link: http://invis.io/V61L7CBF7).
- **Memeoke**– I worked on a Karaoke web application that displays the words of a song on its respective image or video. I developed a URL shortener for easy sharing, and implemented the backend using Flask/Sqlite and APIs from Giphy, musixmatch, and Soundcloud (www.memeoke.com).
- **Mail Finder** – I worked on implementing a web application to visually search and display the current data in my gmail inbox. The features implemented were search by name and email ID, as well as analytics to show the top ten email senders and a timeline for the top sender.
