

# Sujee Maniyam

## Big Data Expert, Author and Entrepreneur

San Francisco Bay Area, USA | [resume@sujee.net](mailto:resume@sujee.net) | [LinkedIn](#) | [GitHub](#) | [portfolio](#) | resume ([html](#) , [pdf](#) )

## STRENGTHS

- Five+ years of hands-on expertise in big data, analytics and cloud computing
- Fifteen+ years in Software Development
- Strong leadership, mentoring and interpersonal skills ; believe in leading by doing
- Constantly learning and leveraging emerging technologies
- Speaker at conferences and meetups ([talks](#))

## SKILLS

- **Domains:** Big Data, Analytics, NoSQL, Cloud Computing
- **Hadoop / Big Data :**
  - Expert consulting and training services to Fortune 500 clients and startups.  
Verticals : Financial, Legal / eDiscovery, Online Advertising, Appliances
  - Implemented Hadoop based data warehouses, integrated Hadoop with Enterprise Data Warehouse systems
  - Built real-time Big Data solutions using HBase handling billions of records
  - Built scalable, cost-effective solutions using Cloud technologies
  - Implemented Big Data analytical solutions that 'close the loop' and provide actionable intelligence
  - Developed free text search solution with Hadoop and Solr. Analyzing emails for compliance and eDiscovery.
  - Taught courses and seminars on Big Data and Cloud technologies
  - Expert knowledge in Hadoop/HDFS, MapReduce, HBase, Pig, Sqoop, Amazon Elastic Map Reduce (EMR), Accumulo
  - Cloud : Amazon EC2, Rackspace, Google Cloud
  - Author of [Hadoop Illuminated](#) (open source Hadoop book), [HOWTOs & tutorials](#)
  - Founder & organizer of [Big Data Gurus meetup](#) .
- **Certifications :** Cloudera Certified Developer, Hortonworks Certified Developer (Hadoop2), HortonWorks Certified Admin
- **Languages :** Java, Python, Ruby, R, Shell (bash), PHP, C / C++ / Objective-C
- **Databases :** MySQL, MongoDB, HBase, Accumulo
- **Dev. Tools :** Linux/Mac OS environment, Eclipse, git, Xcode, Subversion
- **Stacks :** Ruby on Rails, Tomado (python), LAMP stack
- **Open Source Projects :**
  - [github.com/sujee](https://github.com/sujee) : [DNS tool for Hadoop / HBase clusters](#) , [launch scripts for Amazon EMR](#) , Hadoop, HBase tutorials
  - Hadoop project contributions : [HBASE-5577 \(improve patch submission process\)](#) , [HBASE-4440 \(Improve HBASE performance benchmark\)](#) , [HBASE-5555 \(HBase DNS verification\)](#) ,
- **Training available for :** Hadoop, Big Data for Executives, HBase, Accumulo, Cassandra, Data Science with R
- **Business Domains :** Distributed systems, online advertising (RTB, targeting), social media advertising

## WORK EXPERIENCE

- [Elephant Scale](#) ( *Expert Big Data Services* ) [2013 - Present]  
**Role : Co-Founder & Principal Consultant**  
Provide expert consulting services, solutions and training around Big Data ecosystem (Hadoop, NoSQL, Cloud) to Fortune 500 clients and startups.
- [WanDisco](#) ( *Big Data platform* ) [2012 Sept - 2013 Apr]  
**Role : Hadoop Consultant**  
Advised executives on Big Data strategy. Worked on WanDisco Non-Stop-Namenode product. Conducted internal Hadoop training for the team. Spun up a popular webinar series on Hadoop.
- [Hitachi Data Systems](#) ( *Storage, Hadoop Appliance* ) [2011 May - 2012 Aug]  
**Roles : Principal Hadoop Consultant, Analytics Lead, Mentor**

Worked with the Big Data Storage Team (acquired from Parascale) that was building Hitachi's Big Data Storage appliance. Virtualized Hadoop in Linux KVM to provide a safer, scalable analytics sandbox on the appliance. Developed a HDFS plugin to interface with proprietary file system. Advised file system engineers on Hadoop specific optimizations. Implemented component based tests / benchmarks.

Mentored an intern working on recommendation engine using Hadoop / Mahout.

- [Adpredictive](#) ( *Online Advertising / Targeting* ) [2009 Nov - 2011 Feb]

**Role : Hadoop Consultant**

Worked closely with the CTO in helping them scale to 100s millions of impressions per day. Built a scalable, cost effective, and fault tolerant data ware house system on Amazon EC2 Cloud. Developed MapReduce/EMR jobs to analyze the data and provide heuristics and reports. The heuristics were used for improving campaign targeting and efficiency.

Technologies used : Hadoop, HBase, Sqoop, Scribe, Java, Map Reduce , Amazon EC2 infrastructure, Amazon Elastic Map Reduce (EMR), MySQL, shell scripts

[my talk on EMR experience](#)

- **Past Experiences**

- Consulting Engineer, **Hitachi America Ltd**, Santa Clara. (iPhone / Ruby-on-Rails applications) [2009 Oct - 2011 Mar]
- Lead engineer @ **Uloop**. (Web development on LAMP stack) [2008 - 2009 Oct]
- Software Engineer @ **IBM**, Burlingame, CA. (Java, Enterprise work flow engine) [2002 Feb - 2008]
- Software Engineer @ **Crossworlds** , Burlingame, CA. (Java, Solaris, Enterprise work flow engine) [1999 July - 2002 Feb]

## PATENTS AND PUBLICATIONS

- Patent : [Autogenerating reusable web service modules \[# 20060271537\]](#) @ IBM
- Publication : Introduction to Hadoop Map Reduce on [Java Tech Journal February 2012 edition](#) ( [pdf](#) )

## UNIVERSITY EDUCATION

B.S. in Computer Engineering, [University of Melbourne](#) Australia , 1998