

# Sujith Nair

---

## CONTACT

**E-mail:** [suj1th@hotmail.com](mailto:suj1th@hotmail.com)  
**Website:** [github.com/sujithjay](https://github.com/sujithjay)

## WORK EXPERIENCE

### Logistimo India

March 2016 — Present

Data Engineer II

I work on the Data Platform for Logistimo SCM product. This processes volumes of data from India's immunisation supply-chain, along with public health supply-chains in Zambia, Myanmar and others, to generate Business Intelligence reports and forecasts.

I was responsible for building the Data Platform from scratch as part of a small, competent team.

### WeAreHolidays

July 2014 — February 2016

Backend Developer

was part of a tiny team which created an entire marketplace for travel, from Zero to One.

## PROJECTS

#### ■ Data Platform at Logistimo

Transformed the data architecture at Logistimo from a Google Analytics + SQL-based one to an ETL platform based on Apache Spark, YARN and Cassandra. Accomplished as part of a team of 3 Engineers. This data platform is responsible for immunisation and public-health supply-chain data in India, Zambia, Myanmar and other emerging markets.

#### ■ Spark Job-server

was the architect for the data job-server (think Apache Beam + Oozie) which orchestrates, schedules and monitors our data pipelines. This will be open-sourced soon, along with our other data platform offerings.

#### ■ Open-source Contributions

Active contributor to the Apache Spark community; In addition, I have contributed in small capacities to Apache Impala, Scala Language documentations, and to ScalaNLP/Breeze

#### ■ A CrudRepository implementation for a NoSQL database

Conceptualised and coded an abstraction layer over the Couchbase + Elasticsearch/database stack, similar to the CrudRepository provided by Spring Data Commons.

## EDUCATION

### Masters in Technology [M.Tech] in Systems Engineering

July 2012 — June 2014

Indian Institute of Technology (BHU), Varanasi

### Bachelors in Engineering [B.E] in Computer Engineering

July 2007 — June 2011

University of Mumbai

## TECHNICAL SKILLS

### Languages:

*Proficient:* Scala, Java

*Familiar with:* Python

**Frameworks:** Spark Core, Spark SQL, YARN, HDFS

**Databases :** Apache Cassandra, MariaDB, MongoDB, Couchbase, Elasticsearch

## ACADEMIC RESEARCH

### **Gossip-Based Aggregation for Large Peer-to-Peer Networks**

*Masters' Thesis. Supervisor: Professor S.C.Gupta, EEE, IIT BHU.*

A novel gossip-based scheme for aggregation of information in a Peer-to-Peer network. The best known gossip-based algorithm for computing aggregates has a message complexity of  $O(n \log n)$ . The primary concern of my work is in reducing the message complexity of the algorithm.

### **Sampling of Large Network Topology with MapReduce**

*Independent Research. July 2013- October 2013.*

Extracting a representative sample from a very-large graph using MapReduce Programming Model. The sample graph must be 'similar' to the very-large graph.