

ABOUT ME

Hello, I have around 12 years of professional experience across a startup and a public company. I take pragmatic decisions and good engineering practices that pays compounding returns as we progress along rather than cutting corners to meet deadlines for today. I'm a result of an excellent set of mentors, so I take mentoring - giving and taking feedback very religiously. Finally, I also have picked up varied interests outside of Software Engineering like Welding, Stock Options Trading, Building and assembling an electric car in my parking lot, hence *The Generalist* tag.

SKILLS

Big Data	Hadoop (YARN and HDFS), HBase, Spark, Zookeeper, Kafka
Distributed Systems	Operations of above Big Data Tools · CAP Theorem, Large Scale Data Management
Programming	Scala, Java, Go, Typescript/Javascript
Operations	AWS, Cost Optimization, Kubernetes, Docker, Terraform, Basic Linux Administration
Vocational	Stock Trading, Learning to Weld, Paint and Assemble an Electric Car

EXPERIENCE

- **AK Labs** Chennai, India
Director & Fractional Software Architect *May. 2022 - Present*
 - **AK Labs:** AK Labs is my personal company under which I provide consulting services to my clients. I wear multiple hats which includes, building a pipeline of work, doing sales call with leads, managing finances for contractors when I hire them and deliver value to my clients as a Software Architect on their team(s).
 - **Hella Infra Market Pvt. Ltd:** Manage and support their internal Platform Engineering team. Also playing the role of part-time architect to their Data Platform Team that supports internal data pipelines for ETL, analytics and reporting.
 - **Astranova Labs:** Built their end to end infrastructure automation for software delivery on GCP, Kubernetes and Cloudflare.
- **Avalara Inc.** Chennai, India
Principal Engineer *Feb. 2019 - Apr. 2022*
 - **Content Sourcing Platform:** Automated detection of *Sales and Use Tax* changes across various states within US, reducing time from 3+ weeks of manual effort to a few minutes across many jurisdictions. Architected a No-Code Self-Serve Platform for Tax Researchers within Avalara to automate their workflows to identify important content changes.
- **Indix** Chennai, India
Principal Engineer & Mentor *Oct. 2015 - Jan. 2019*
 - **Data Services Platform:** Architected a Data Services Platform (*Carol*) to run our internal Big Data Workloads. This tool helped orchestrate our internal Spark based platform that ran on EMR clusters created on the fly.
 - **Product Matching:** Product Matching is the process of identifying an "iPhone 13" sold at Amazon and "iPhone 13 256GB Space Grey" sold at Walmart are the same product but sold at 2 different marketplaces. Consulted for Product Matching team to reduce their workload runtime across 2+ Billion product dataset from multiple days to about 7 - 8 hours. This saved us a lot in Infrastructure costs and also allowed us to run various experiments faster to improve our model metrics: precision and recall faster.
 - **Data Ingestion:** I was primarily part of the Data Ingestion Team which is responsible for building the necessary tools that ingests all the required data for building our catalogue. To better explain it in numbers: Crawler crawls >20 million webpages a day, Parser process >30 millions webpages a day all in realtime. This team was also responsible for managing all the *Master (Raw) Data* that flows with in our Data pipeline.
 - **Finder:** HTML Archive Storage system for storing over 3 billion HTML documents. It was built using **Suuchi**, a library that provides set of abstraction for building distributed data systems that I helped develop.
 - **Infrastructure:** During this phase we had to revamp our Devops team internally and as part of the effort, I joined as the first member to help build the team, identify tools, setup automation and processes which after the first year saved over million dollars a year on AWS by intelligently using Spot instances and Auto Scaling for 70% of Production workloads. You can find more about the experience at the talk titled "**Lessons scaling operations everyone @indix**".

- **Data Pipeline:** I was part of the team that built the first generation of Indix Data Pipeline which was completely based out of the principles of λ - architecture. I contributed in writing Map-Reduce jobs, setting up S3 file layouts for faster MR job startup times and the was monitoring the overall health of the data pipeline for a while.
- **Data Ingestion & Data Infrastructure:** Data ingestion team was responsible for a set of components like:
 - * **Crawler** - I helped build two generations of our Crawler scaling it to crawl a million pages a day across 1000s of websites.
 - * **Scheduler** - I helped build two generations of our Scheduler service that is responsible for feeding the crawler with right urls to crawl at any given point in time by consuming more than 20+ parameters.
 - * **Parser** - I helped build three generations of our Parser Service that is responsible for processing all the raw web pages that are crawled and extract semi-structured information from those pages into a JSON like structure which are feed as input to our Data Pipeline. My proudest contribution was building a DSL for our parser which expresses various transformations in a language agnostic way that helped us scale the number of parsers to 1000s of websites quite easily.
 - * **Master Data Management on S3** - Our crawlers and parsers wrote all the data that it processed onto flat files on S3 using intelligent partitions and batching so we could process them faster in our Data pipeline which was built as a bunch of Map-Reduce and Spark jobs. One of the biggest challenge in storing data in flat files was "querying" it for any analysis. This is was pre-Impala, Druid, etc. era where we didn't have much tools available in the market for these type of use-cases. I built a query tool with a custom DSL (inspired from SQL but on top of scalding) which enabled Engineers and Product Managers to access the data much faster for various analysis / debugging purposes.

- **Mu Sigma**

Bangalore, India

*Big Data Intern**Jan. 2012 - May 2012*

- Evaluating AWS for migrating to cloud from custom data centers
- Worked on running R code on Storm.

EVANGELISM ·  | 

Outside of the above outlined work, I also like to interact with developer community in general. I help host various meetups in our Chennai office and I am also an active participant on various hackathons conducted within the country. I have also given talks at national and international conferences around the world.

April 2018 **Why we built a distributed system**, DSConf 2018 Pune, India



Nov. 2017 **Lessons scaling operations to everyone @indix**, Mini Cloud Conf Chennai, India



Nov. 2017 **Using Monoids for Large Scale Aggregates**, Scala.io Lyon, France



April 2017 **Suuchi - Distributed System Primitives**, Devday, Chennai



Nov. 2016 **Lessons from managing Hadoop clusters on AWS @indix**, DevopsDays, India




Nov. 2016 **Lessons from Building Distributed RocksDB**, Geeknight Chennai



April 2016 **Lessons from running production infra on AWS Spot Reliably**, Geeknight Chennai



OPEN SOURCE

I was fortunate enough as part of my work at Indix involved writing and contributing a lot of Open source work. Take a look at my /ashwanthkumar for all of my contributions over the years.

- **Indix OSS:** **Indix OSS** is an internal community that I helped create at Indix to encourage more engineers to build and contribute to open source tools that we use on a daily basis. OSS tools that I built as part of my time at Indix are listed on the page.
- **GoCD:** I've been using **GoCD** for close to 12 years now and I've contributed a few plugins (**Github PR Plugin**, **Github Slack Notifier** and **GoCD Janitor**) which are used by quite a large number of folks in the community and also an active member on their **mailing list**.
- **Suuchi:** **Suuchi** is toolkit to build distributed data systems, that uses gRPC under the hood as the communication medium. The overall goal of this project is to build pluggable components that can be easily composed by the developer to build a data system of desired characteristics. It is written in Scala and **this page** shows how easy is to build a simple distributed in-memory database.

EDUCATION

- **SASTRA University**

Tamil Nadu, India

*Bachelor of Technology in Computer Science; GPA: (8.12/10.0)**July. 2008 - May. 2012*