

Deepak Marathe

Deepak Marathe is a Software Professional
Specializing as Polyglot FullStack Software
Developer, Big Data Platform Engineer, Cloud Solution
Architect, OpenSource Software Contributor,
Web3/Blockchain/AI Entrepreneur and Educator,
Sole Proprietor/Consultant, Electronic
Musician/Instrumentalist and a Vedic Astrologer.

Portfolio/Skills : <https://deepakmarathe.netlify.app>

Technical Skills

- **Programming Languages** : Solidity, SQL, Java, Python, Bash, Go-lang, NodeJS, Scala, C, C++, Ruby.
- **FullStack Software Development** : **Java**(Spring, Hibernate, Dropwizard, Spark, MySQL), **Python**(Flask, Django, Postgres), **Node**(NodeJS, Mongo, Express, ReactJS), Web3.js, Ajax, JQuery, HTML, CSS, Javascript, ReactJS
- **Big Data Platform Engineering** : Apache Kafka, Apache Flink, Apache Spark, Apache MapReduce, Pig, Hive, HBase, Cassandra, ElasticSearch/OpenSearch, Airflow, Oozie, Docker, Kubernetes, Helm, Chef, TICK (Grafana), ELK (Kibana)
- **Cloud Solution Architecture** : **AWS** : (Lambda, Amplify, Beanstalk, DynamoDB, RDS, S3, Route 53, ACM, Athena, EC2, EMR, SQS, Cognito), **GCP** (BigQuery, DataProc, BigTable, FireBase), **Azure** (Azure Data Factory and Azure DataBricks), **Databricks**, Heroku, Netlify.
- **Artificial Intelligence / Machine Learning** : Generative AI(GPT-3, 4), Natural Language Processing(spacy.io, **OpenAI**), Text-To-Speech(pytsx3), Speech-To-Text(SpeechRecognition), Time Series Forecasting (prophet, pandas, matplotlib), Classification (numpy, scipy, sklearn),
- **Web3/Blockchain/SmartContracts** : Solidity(0.8.1), Remix, Truffle, Ganache, Metamask, Web3.js, **Alchemy**, Geth (Private Ethereum Blockchain).
- **Rule Engine** : Drools
- **Android** : Optical Character Recognition Application to Extract Aadhar ID from Image.
- **Plugin Development** : EditorOpenFileCount Plugin for IntelliJ IDEA IDE
- **Speech Technologies** : Alexa Skill for Intuit QuickBooks (Alexa Voice Services)
- **Automation** : Spotify Music Streaming Platform Automation using Selenium.
- **Operating System** : Macintosh, Linux and Windows.
- **Source Version Control** : Gitlab, Github, Bitbucket
- **Continuous Integration** : Jenkins/Hudson, Sonar.
- **Unit Testing** : JUnit, TestNG, Mockito/PowerMock, PyTest
- **Database** : Postgresql, MySQL, DynamoDB, HBase/BigTable, MongoDB, Cassandra, ElasticSearch, InfluxDB
- **Electronic Music Production** : Ambisonic Music (Spatial 360), Ambient Noise Cancellation, Ableton Live, Reaper

Entrepreneurship Experience



Co-Founder/CTO, DFacto.AI (Dec 2022 - Present)

Developed a Fully Functional No-Code Analytics Fullstack Web Application for Business Owners featuring Generative AI Integration, Scalability, Security, Payments Integration, Social Login Integration : <https://dfacto.ai>

Generative AI : OpenAI/GPT3

Backend : Java, Spring, Hibernate, MySQL

Frontend : HTML5, CSS, Javascript, React

Deployment : AWS Elastic Beanstalk, AWS Amplify, AWS RDS, AWS ACM

Domain : GoDaddy, AWS Route53

Payments : Stripe

Social Login : Google Firebase

Security : ACM/HTTPS, GWT for API

Achievements : Raised infrastructure credits from AWS for \$1000.
Raised infrastructure credits from Microsoft for OpenAI/GPT3 for \$2500.



Co-Founder/CTO, Web3Works.in (Sep 2022 - Present)

Developed a Fully Functional Fullstack Web3 DAO on Ethereum Blockchain, on an Enterprise grade Network from Alchemy.
<https://www.web3works.in>

Backend : Solidity(0.8.1), Remix, Truffle, Ganache, Metamask, Alchemy, Geth

Frontend : Web3.js, Node, React

Deployment : AWS S3, AWS Amplify, AWS Cognito

- Developed Contracts in Solidity 0.8.1 and Deployed to Goerli Test Network.
- Migrated the Contracts written in Solidity 0.8.1 (using Remix Web IDE) onto Solana, a Layer 1 Third Generation Blockchain network to yield Performance and Cost Efficiency for Write Operations and Lookups on Contracts between any 2 Verified Entities.
- Governance : Provided Identity management and Verification.
- Product Demo can be found here :

📺 #SmartContracts On #Ethereum #Blockchain

https://www.youtube.com/watch?v=ADIIOWXp9Pw&ab_channel=Web3Works

Industry Experience



Founder, Jnyana Consultancy [GST:29AXXP1681J1ZU]
(Feb 2020 - Present)

Independent Consultant, remote post pandemic

Roles/Designations :

Unravel Data/USA/Remote : Chief Architect, Observability.

A5 Labs/USA/Remote : Product Owner, Bonus Engine

Indeed/USA/Remote : Lead, Big Data Migration

- **Invented a theorem** in **Data Engineering** involving **Consistency, Availability and Partitioning** for a Data System that works on Polling to update the State. (**Unravel Data Systems INC, Dr Shivnath Babu**)
- **Implemented Case-Insensitive Search** on **ElasticSearch** for majority of Attributes in **Gaming Domain**.
- **Designed and Delivered a DataLake** product on top of **Google Cloud BigTable** using a Geographically Distributed, Remote team of 7 data engineers. (**Python, Flask, Google Cloud Bigtable**)
 - Migrated **Data Warehousing ETL Workflows** from **Hive-SQL** to **Spark (Python, Scala)** on **AWS EMR/Azure DataBricks** Infrastructure. Used **Airflow/Google Cloud Composer/Azure Data Factory** for Orchestration.
 - Assisted in **Data Migration** from an On-Premise Data Platform onto **AWS**. (**Spark, Pig, MapReduce**)
 - **Optimized the Computation Time** on **AWS Athena** by distributing the processing using the **ppss** tool in python.
 - Helped **Migrate the Data** from **AWS S3** Data Lake to **GCP BigQuery**.
 - **Developed Distributed Application** to perform **Window Based Aggregation** on Streaming Data using **Flink** using data from **Kafka** and Visualised using **Influx** Time Series Database and **Grafana** for Visualisation.
 - **Developed and Deployed a Web Application** to host the Blogs. (**Django, Python, Heroku**)
- **Developed NanoCube - Scalable, Distributed, Spatio Temporal DataCube**
Abstraction for Time Series Datasets Along with an Event Based Framework for Real Time Visualisation of Rolling, Additive Aggregation.
 - **RESTful Microservices APIs** for Distributed Multi-Dimensional Spatio Temporal Time Series DataCube supporting **Ingestion, Search, Scan Operations** on **GDelt** Event Dataset.
 - **Query Interface modeled on CNF** implemented using **Sparse BitSets**.
 - **Point and Range Queries** over time, all/subset of dimensions with **Pagination** for Results.
 - **Supports Interactive Aggregations** over 10 years worth of **gdelt** data at current content generation pace. (**dropwizard, java, gcp, docker,git**)
 - **MemoryMapped Files** are used to efficiently retrieve data on disk, supporting **100 ms response time** over 10 million data points, with a result page size of 100.
 - **InMemory Index Periodically Persisted** to **Disk/Cloud Storage Object**.
 - Distributed Processing for Machine Learning Production Pipelines Using **Apache Beam** on **Google Cloud DataFlow, PubSub**.

Deepak Marathe



Senior Software Engineer at Rippling, a YC Company (Dec 2021 - May 2022)

- Developed APIs for Reporting, Analytics (Python, Django, MongoDB)
- Improved Code Coverage to 95% by Systematic Refactoring to make the modules Unit Testable.
- I quit my full time employment to pursue my Entrepreneurship ventures in Generative AI.



Staff Software Engineer at Quartic.ai (Nov 2019 - March 2020)

- Optimized Distributed Rule Engine by 8X (Kafka, Drools, Scala, Spark, Postgres)
- Improved Code Coverage to 90% by Systematic Refactoring to make the modules Unit Testable.
- Implemented Data Retention Policies on Elasticsearch (Python, Django, Airflow)
- Onset of Covid affected my Full-Time Employment with Quartic. I moved on to Freelancing working Remote.



SDE3/Lead Software Engineer at Halodoc (Dec 2018 - Dec 2019)

Developed and Deployed Features in the Halodoc Backend System.
Developed RESTful Microservices (Java, Dropwizard, Hibernate), Domain modeling on MySQL, API development (Go-Lang) and deployment on Serverless Architecture (AWS Lambda).

- Doctor Referral - Domain Modeling, implementation (java, dropwizard), stage and production deployment (AWS). Objective was to enable patients to have a second opinion with other doctors or the ability of consulting doctors to refer the patient to a specialist doctor.
- Recent Doctors - For the users that have consulted with doctors via tele-consultation, recent doctors have to be shown to the user. The objective was to increase engagement with doctors. Used Go-Lang with AWS Lambda to generate the recent doctor list and store it as users' attributes.
- Doctor Categories - Grouping the available doctors with categories - helps the user of the app search doctors in relevant categories.
- Actively involved in debugging and resolving bugs in production.

Deepak Marathe



Product Engineer / SDE3 - Data Engineering at GO-JEK/Halodoc (Dec 2016 – Dec 2018)

Worked towards building a **Self-Serve Data Platform** scaling to **Millions per day of Transactional data** in an **Agile Environment** adhering to **Pair Programming** and **Test Driven Development**.

- **Framework for Data Consumption from Kafka**, filtering based on data, and relaying it to pluggable sinks[console, database, http server] and evolved to plug dynamic sinks based on application configuration from Consul. Fault tolerance and recovery using retry with **Exponential Backoff** semantics. **Monitoring using Datadog/Statsd/JMX**.
- Data archival to cold storage(GCS/Secor) and analysis(Spark/Zeppelin).
- **Streaming Data Aggregation** using [Kafka,Flink/Yarn,Influx,Grafana]. Creation and management of **HA Flink cluster** on **Google DataProc** Service. Remote submission/launch of data aggregators on flink/yarn cluster.
- **Microservices: Customer Wallet Suspension Service for Fraud**.
Java/spark, consul/HashiCorp for configuration, gradle as build tool, gitlab as version control and CI/CD. Used Hystrix dashboard/turbine for latency and fault tolerance, guava rate limiter for rate limiting with Instrumentation(Newrelic), monitoring(datadog) and alerting(slack/pagerduty).



Senior Data Engineer at Intuit (Apr 2015 – Nov 2016)

Learned and Contributed to major projects in the IDEA BU. Ensure **Code Quality** through **Unit Tests, Functional Tests, Code Coverage**.
Developed Voice Capabilities to Intuit's flagship product - **QuickBooks Online** using **Amazon Voice Service/Alexa API** on **Amazon Echo** hardware.

- **Distributed Configuration Management Service**
 - o Client using Archaius from Netflix. Guice dependency injection, testable code- code coverage, unit test, integration test, mocking.
 - o REST server for configuration service using AWS-EC2, Dropwizard framework, AWS-DynamoDB.
- **Dragline: data copy tool from HDFS/Hive to Vertica**.
 - o Feature implementation – configuration validator, logging improvements, bug fixes, code refactor for code quality.
 - o Kerberos, HP Vertica,
 - o Unit/Functional Tests for dragline: Improved code coverage numbers from 20% to 80%.
 - o CI/CD for the project using Jenkins + Sonar.
- **Test Framework - API to generate Mock data for testing**.
 - o Data Source and Data Sink abstraction, extensible/pluggable data sources (Local FS, HDFS, S3, Vertica, Hive). Random data preparation using schema definition supplied via configuration and pluggable schema adapters. Support for Vertica data types.
- **Sangria Console: Web Application for admin to blacklist/whitelist EINs**
 - o developed using Spring java + hibernate ORM tool. App deployed on tomcat.

Deepak Marathe

- o SSO integration and login module.
- **Extra contributions:**
 - o Hackathon - develop speech capabilities to Intuit's flagship product - QuickBooks Online using amazon voice service API to run QBO skill on Amazon Echo device. Host a rest server in AWS using Elastic Beanstalk.
QBO Virtual Personal Assistant
ID: amzn1.echo-sdk-ams.app.caef4128-7866-4e4f-8f32-c7234f27e8d4
https://developer.amazon.com/edw/home.html#/skill/amzn1.echo-sdk-ams.app.caef4128-7866-4e4f-8f32-c7234f27e8d4/en_US/info
 - o Develop plugins to IntelliJ IDEA :
<https://github.com/deepakmarathe/EditorOpenFilesCountPlugin>
 - o Home Audio System using Raspberry Pi : install VLC player, use LUA HTTP interface to control the home theater connected to raspberry pi, using a handheld device which has VLC player installed in it. (VLC remote).



**Data Engineer/Senior Data Engineer
at InMobi (March 2012 - March 2015)**

Developed ETL Pipelines for Batch Systems. Helped the team scale from handling a few GB of data to a few hundred TBs contributing to a user base of hundreds of million. Built visualization tools for easy onboarding, analyzed large amounts of data to get insights on user patterns and helped business by building products atop those.

- **Data Ingestion and Curation:**
 - o Built, deployed and managed data pipelines to ingest data from multiple sources. Used HBase/HDFS for storage, Thrift-LZO for data modeling, and MR/Pig for ETL. Also did some ad hoc analysis using Pig and Hive.
 - o Augmented and automated data pipelines using in-house workflow scheduler (now Apache Falcon) built on top of Oozie. Configured a dashboard to collect and monitor metrics using Graphana.
 - o Was on pager duty to maintain SLA of service.
- **Analytics and Products:**
 - o Understanding user interactions using data analysis:
 - Improved ability to target ads by extracting user visit patterns and segregating users by devising scoring mechanisms to infer key metrics like Daily Active Users, Monthly Active Users over large data sets. Evolved it to a framework for custom Syntax Validations and Rule Based Classification.
 - o User Targeting suite: Interest targeting, Retargeting, Segment/Persona targeting).
 - Implemented user classification into various Personas - depending on business decisions. Addressed challenges involved in User Profile Creation, Shipping to Local Data Centers, and Online-Cache population for ad serving.
- **Extra:**
 - o Participated in Hackathon conducted by InMobi: Developed a Data Pipeline Designer to address the needs of product analysts. Self-Serve is a Web based visual data flow designer capable of performing data discovery and code generation for apache pig.
 - o Participated in InMobi sports day and won Runner up trophy in Table Tennis men's doubles category.

Deepak Marathe




System Software Engineer at Hewlett Packard, STSD Bangalore (Aug 2010 – March 2012)

- o **Application for Manageability of servers, named Replication Manager**
 - Active/Passive Management of HP storage servers/ control the replication.
 - Web Application, Java.
- o **Part of the Open Storage (OST) plug-in team.**
 - OST plug-in implementation for HP D2D storage servers, to work with Symantec 'NetBackup' solution, C/C++
 - Integration of client side deduplication.
 - Setup and Maintained build system and Automation for HP OST.
 - o Hudson as the CI tool for build automation.
 - o Builds automation on windows using visual studio command line utilities, ant, with code coverage tools integrated.
 - o Linux – rhel, suse build automation-using make.
 - In charge of maintaining thread related features and common code base.

ACADEMIC PROJECTS

- Designed an Animation Package (in Java language) as part of a Computer Graphics project in the 4th semester.
- Designed a project on database technologies (in java language) as a part of Database Systems, comprising basic functionalities of a typical database system.
- Worked on a mobile application "m-Bank" to mobilize the banking process with the help of j2ME.
- Worked on the Web Application 'Online Chess Club' as a part of my curriculum project.
- Designed ERP software for "Adlabs Foto Fast".

INTERNSHIPS

-  Successfully completed a project on web-technologies involving Struts framework (MVC) at Tavant Technologies, Bangalore.
- Worked on HTML, AJAX, Reporting Tools at iNAT Technologies located at NITK Campus, Surathkal.

EDUCATION

COURSE	INSTITUTION	Year	PERFORMANCE
B. Tech Information Technology	NITK, Surathkal	2006-2010	CGPA - 8.00/10
12 th Standard (C.B.S.E.)	Jawahar Navodaya Vidyalaya Chara, Udupi	2005-2006	88.00%
10 th Standard (C.B.S.E.)	Jawahar Navodaya Vidyalaya Chara, Udupi	2003-2004	86.00%

Deepak Marathe

TECHNICAL INTERESTS

- Generative AI, GPT-3, GPT-4
- Web3 : Blockchain, Hyperledger, Tokens, SmartContracts
- Distributed Computing : Data Processing at Scale, Streaming Data Processing, Complex Event Processing, Real Time Analytics
- Design and Analysis of Data Structures and Algorithms.
- Object Oriented Programming.
- Spatial Sound Processing (VST, Spatial Workstation/Facebook Audio 360)

HOBBIES

- Software Development : <https://github.com/deepakmarathe>
- Entrepreneurship, Angel Investor
- Vedic Astrologer
- Musician
 - Instrumental : Rumba Style Flamenco Music on a Nylon String Guitar : <https://www.dailymotion.com/video/x4qmuqe>
 - Electronic Music Production : Artist on Apple Music and Spotify, 2K Shazams : <https://open.spotify.com/artist/7u4oOi5lhlW4qHpsJghKih>

PERSONAL INFORMATION

- | | |
|-----------------|---|
| ▪ Name | Deepak Marathe |
| ▪ Phone Number | +918088936848 |
| ▪ E-mail | dpkmarathe@gmail.com |
| ▪ Date of birth | December 08, 1988 |
| ▪ Linkedin | https://www.linkedin.com/in/deepak-marathe/ |