

## About

Data Engineer with around 5 years of experience in Designing and Implementing ML Models, ETL pipelines, Data Mesh and Analytics.

Possess PGD in Big Data Analytics from CDAC ACTS (HQ), Pune.

Seeking to leverage my Mathematical knowledge, Analytics abilities, and ML models development competency in the role of Data Analytics.

## Contact

-  (+91) 8239-3838-47
-  rsohlot@gmail.com
-  Bangalore, IN
-  [linkedin.com/in/rsohlot](https://www.linkedin.com/in/rsohlot)
- <https://rsohlot.github.io>

## Professional experience

2020 - Present

Data Engineer-II, Infinx HealthCare, Bangalore

- **My Responsibility: Optimizing, designing and scaling the product (Data pipeline)**
- **Tools/services used: python, pyspark (apache spark), airflow, aws EMR, aws Redshift, glue crawler, glue catalog, Athena, Quicksight, glue visual ETL.**
- IRIS: An Insights dashboard to help client analyzing their revenue. Build on generalized ETL platform to make easier onboard for newer client implementation.
- IROS: A Data processing product with the capability to scale, contains pkgs for data quality check, logging as an event that then tracked with the help of kibana. Output: A/R Inventory data to maximize revenue recovery by prioritize using tree-based ML models.
- Data Processing: Transforming (containing 80 col derivation) 70 Lakh record in 5 min. (EMR + pyspark)

2019 - 2020

Associate Data Scientist, KratinLLC, Nagpur

- **Tools used prophet, optaplanner, Django, plotly, pandas, docker.**
- Designed a Time Series Model give accuracy up to 97% used to create an AI smart SAP tool to maximize the workforce utilization considering all the preferences and requirements. Results in saving the 500 miles every week of the workforce traveling.
- Estimated the future demand to get the best assignment of workforce.
- Designed and implemented the warehouse for the BI and analytics. Saved daily 1 hr. task of the team to fetch, validate and store the data. **This was done with Hadoop, hive, spark, hue, Apache Atlas, Apache ranger.**

## EDUCATION

CDAC -PGD Big Data Analytics

*Pune*

*Maharashtra*

*2018 – 2019*

B.Tech. ECE

*Rajasthan Technical University*

*Jaipur, RJ*

*2010 – 2014*

## SKILLS

Machine Learning

Reinforcement Learning, GAN

Python, Java, R

Big Data, Spark, Hive, Kafka

SQL (DBMS)

Mathematics (Statistics, Calculus)

Pytorch, TensorFlow

## CERTIFICATION

- Market Analysis Course (Jan 2017 – March 2017)
- ML, Coursera (May 2016- June 2017)
- Spark Working (Dec 2019 – Dec 2019)
- GAN with Pytorch (May 2020 – June 2020)
- AI: Feature Engineering (July 2020 – July 2020)

2016 - 2018

---

Co-Founder, WHT, Bhiwadi, RJ

- Founded WHT. It was a Logistics Firm. Learnt the Logistics business flow of Northern India. Firm manages complete workflow of goods for the various Industries.

## Personal Projects

---

Open-Source  
Contribution

---

Chaos-genius (github)

- Design and discussion in adding new KPIs related to data metrics.
- Adding new features to product.

Personal  
Projects

---

Iceberg wrapper

- Building a project that make easy to use iceberg. Like to see the data difference before rolling back to the older data version.
- Keeping the data as a branches of git to keep all the data snapshots, and many more production useful features.

Open-Source  
Contribution

Data engineering util

- A lib which will contain use ful function to maintain the data for production for pandas and pyspark.
- Some Features:
  - .1. Schema versioning capability for DF
  - .2. Data type standardization with metadata for daily pipeline run.
  - .3. Data quality check capability with parallel processing, and many more.