Data Collection & Processing for AI

Avi Gopal

Co-founder & CTO @ Metabob
CTO @ Mahn Arc Ventures
Co-Founder @ Clyste

"Avi has an illustrious background in Aerospace Engineering, with expertise in every breadth of development, specifically versed within AI and machine learning"

Chapter04
Session07
Step Up. Speak Up. Go Up

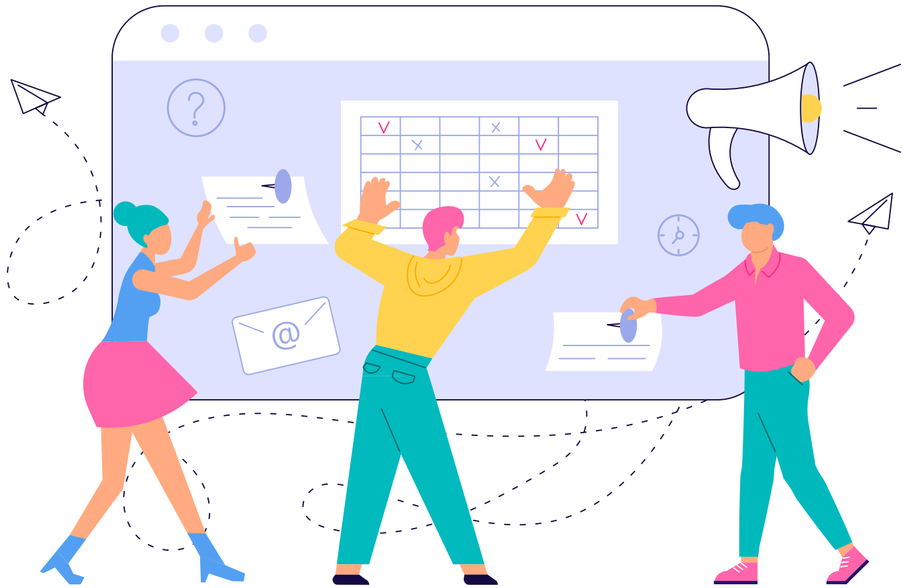
June 27th 10AM PST

zoom.goupaz.com

Hosted by

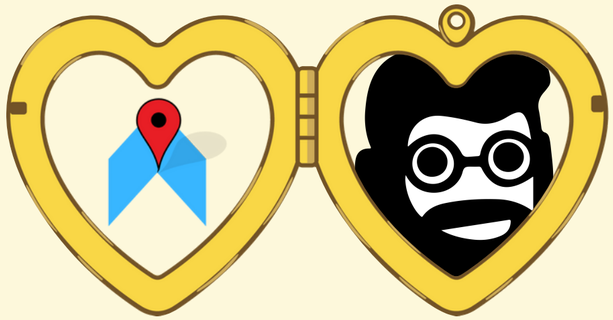
Metabob

Agenda



- **10:00–10:10 | Kickoff**
  - Introduction
  - Chit Chat
- **10:10-11:10 | Presentation**
  - Speaker
- **11:10-11:15 | Break**
  - Pre Q&A
- **11:15-11:30 | Q&A**
  - Discussion
- **11:30-11:35 | Wrap Up**
  - Contact Speaker

Chapter04
Session07



Hosted by

Axel L

Growth Marketing

Passionate about sports and technologies enhancing employee efficiency

Metabob

It's the fast, easy, and visual way of debugging code.

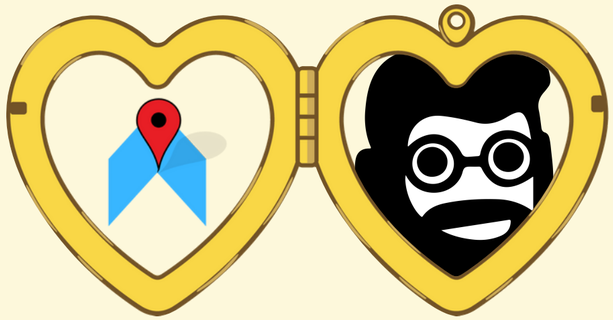
https://metabob.com

GOUP

Community driven Open source accelerator!
https://goupaz.com

Chapter04

#contributor



YOU

GOUP 02hero/Chapter04
Contributor

About YOU

Step Up. Speak Up. Go Up.

GOUP

Community driven Open source accelerator!
https://goupaz.com

Metabob

It's the fast, easy, and visual way of debugging code.

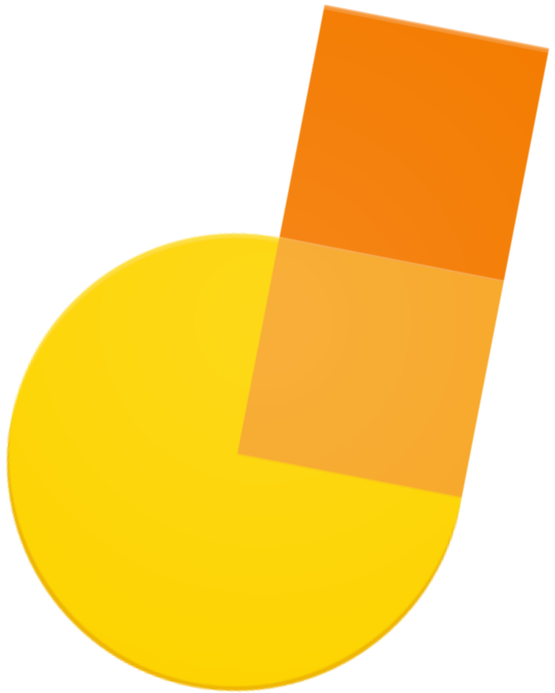
https://metabob.com

Are you ready?

Let's Begin!

CHITCHAT

JAMBOARD



What do you wish to learn from this webinar?

I want learn more about the importance of collecting data for AI

As A Beginner How Can I Learn AI - Akhil

I want to learn how are data and AI (which I believe are mostly algorithms) are related.

Fiza: I want to learn how can we collect effective data for AI.

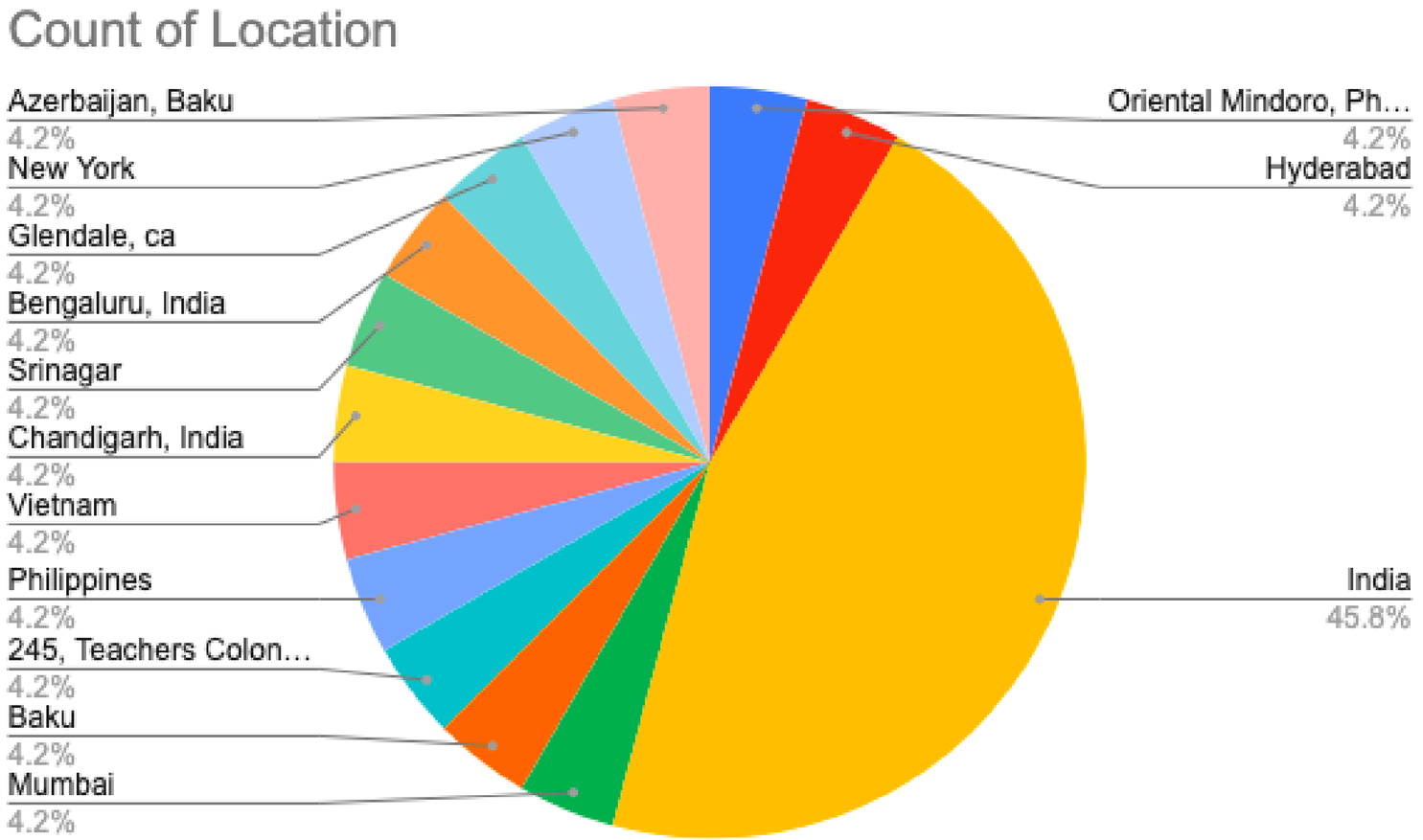
I am Chandan Kumar, a student. How to get started with AI

POLL

HOW EXPERIENCED ARE YOU WITH AI DEVELOPMENT?

1	2	3	4	5
Not experienced	Somewhat experienced	Semi-experienced	Experienced	Very experienced

Audience

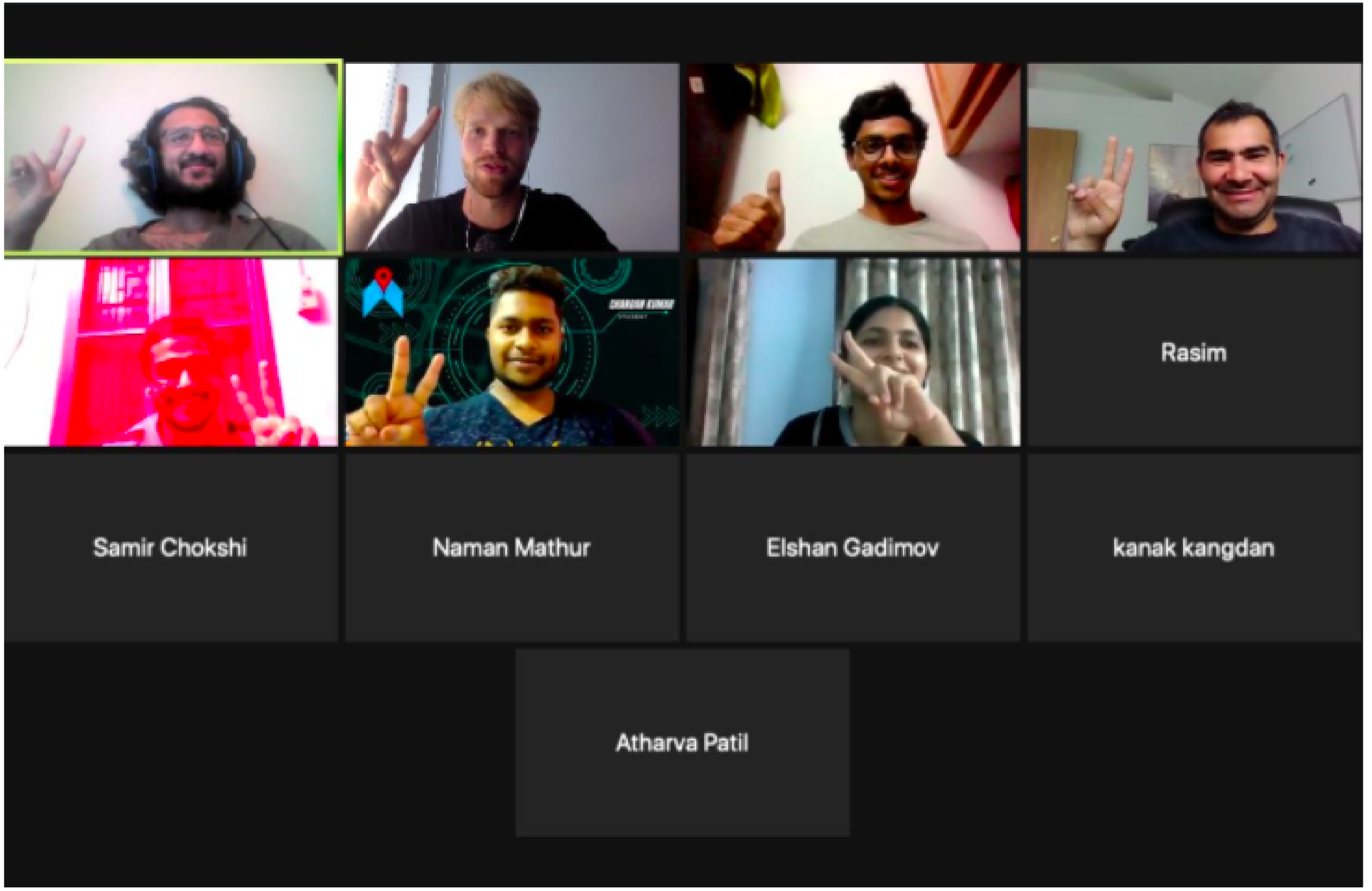

Count of Location
Azerbaijan, Baku
4.2%
New York
4.2%
Glendale, ca
4.2%
Bengaluru, India
4.2%
Srinagar
4.2%
Chandigarh, India
4.2%
Vietnam
4.2%
Philippines
4.2%
245, Teachers Colon…
4.2%
Baku
4.2%
Mumbai
4.2%
Oriental Mindoro, Ph…
4.2%
Hyderabad
4.2%
India
45.8%

Code of Conduct

1. Learn, benefit, contribute
2. No marketing, selling, competing
3. Equality despite roles & bg

Photo Shoot Time

Please turn on your camera :D


Rasim
Samir Chokshi
Naman Mathur
Elshan Gadimov
kanak kangdan
Atharva Patil

Presentation

Data Collection and Processing

Avinash G.

Avi@metabob.com

- Worked on various Aerodynamics Research Projects, specializing in novel airfoil design, during his time at CSULB as an aerospace engineer
- Created a variety of autonomously controlled craft for personal transport and area mapping
- Created new organization structures and community building applications for Clyste as a CTO
- Developed the entirely of the backend and system architecture for the code analysis platform Metabob owned by NEC

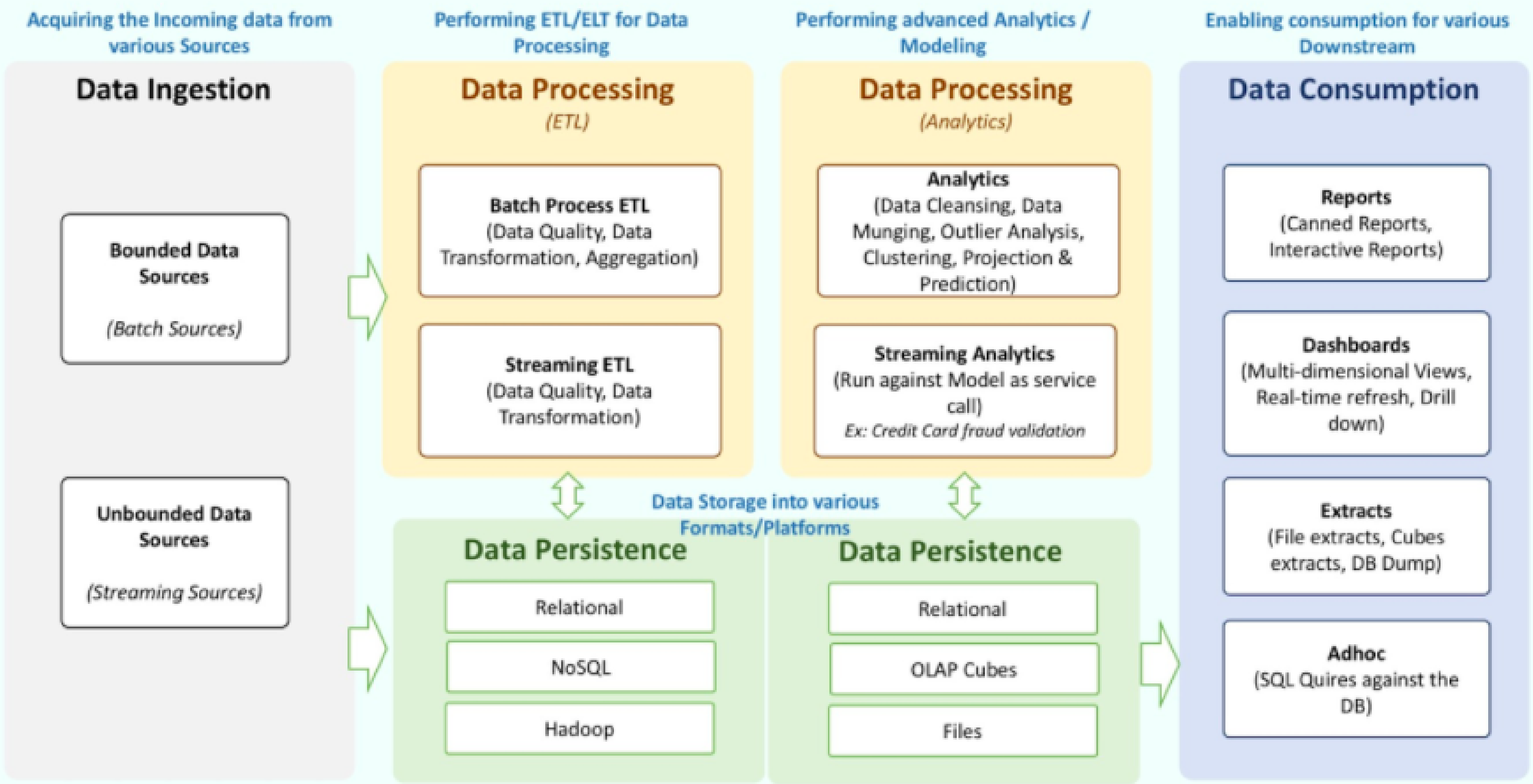
Overview

What is it?

Why do it?

Defining the Problem

Data Processing, Analytics & BI | Functional Viewpoint



Data Ingestion

- Do
  - Have a clear goal in mind
  - Focus on identifying potential data sources
  - Build the necessary integrations for each source
  - Keep track of the metadata as well
- Don’t
  - Worry about scaling when just starting out
  - Apply any processing or filters
  - Delete anything

What we did at Metabob

high-quality open source repos

company SOPs

company design standard

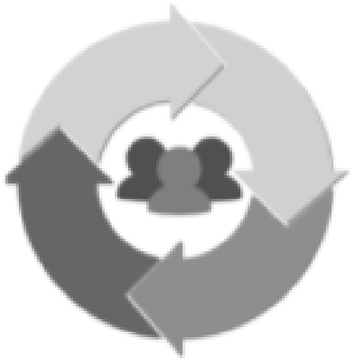
program analysis

manual review historical results

AI Agent learns the best programming practices

A set of recommended code changes

AI agent continues updating and learning through interaction



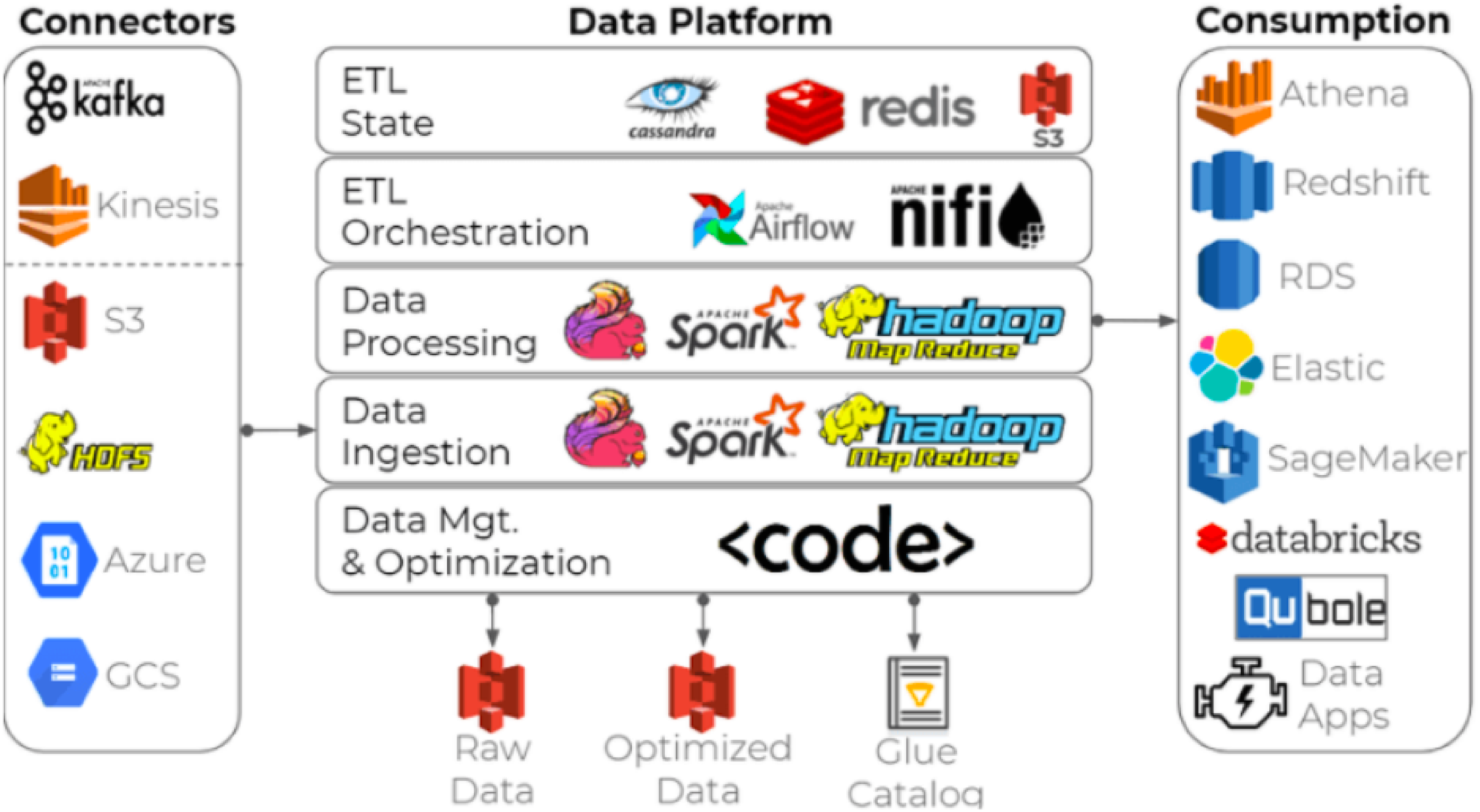
Data Processing

Only build data processing as needed

Focus initially on regularizing input and/or aggregating items from different sources

Store intermediary steps to save on processing time down the road

As capabilities develop move onto enriching sources as well


Connectors
kafka
Kinesis
S3
HDFS
Azure
GCS
Data Platform
ETL State
cassandra
redis
S3
ETL Orchestration
Airflow
nifi
Data Processing
Spark
hadoop Map Reduce
Data Ingestion
Spark
hadoop Map Reduce
Data Mgt. & Optimization
<code>
Raw Data
Optimized Data
Glue Catalog
Consumption
Athena
Redshift
RDS
Elastic
SageMaker
databricks
Qubole
Data Apps

Data Gathering
Aggregated over 50 thousand human verified bug fixes corresponding to over 15 million code changes

Data Labeling
Generated a baseline understanding of the reasons behind code changes via human labelling

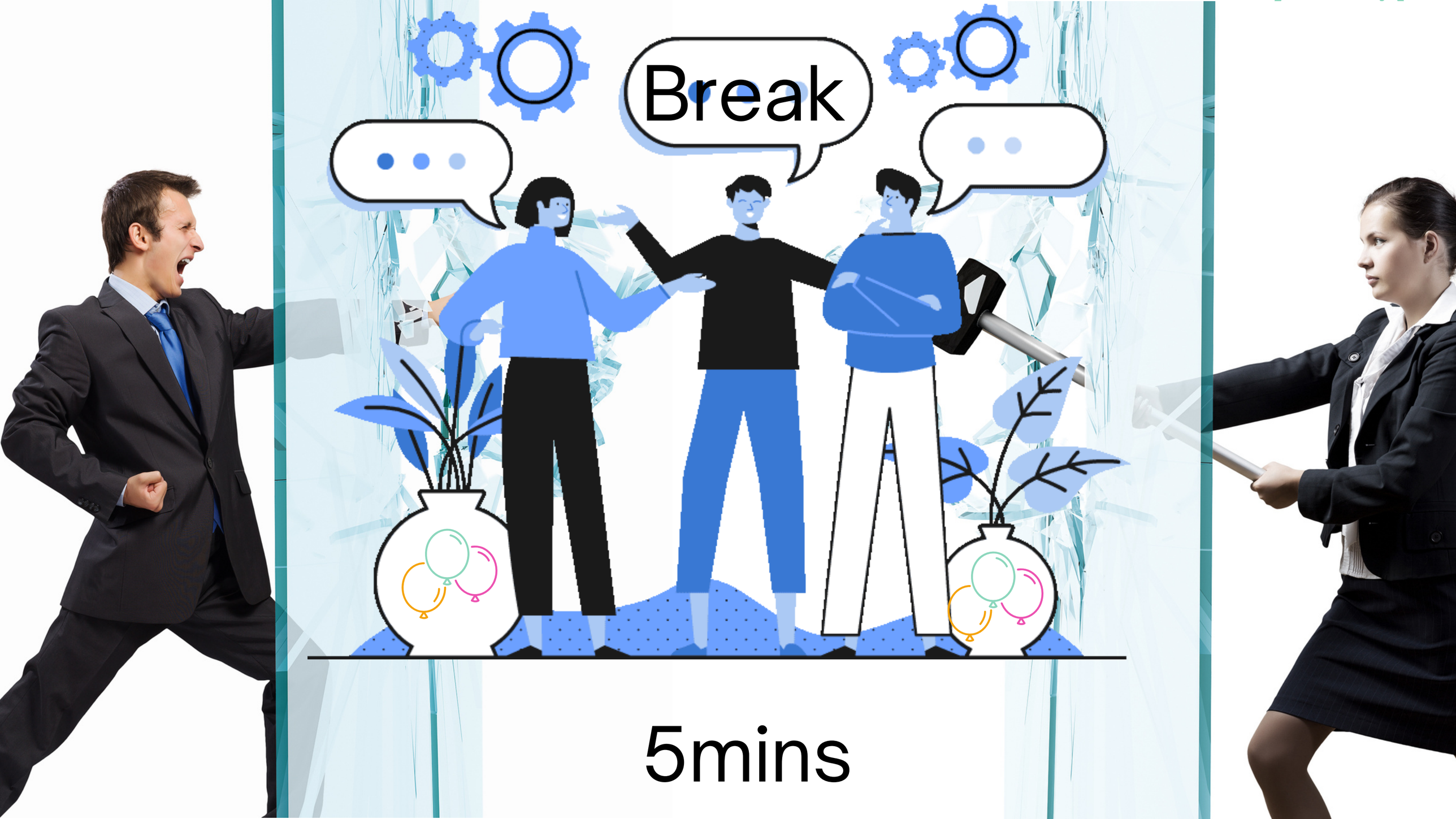
Topic Generation
Sorted code changes into a set of topics relating why and how a change was made

Q&A | Discussion

Type your thought here

Type your thought here

Type your thought here


Break
5mins

Wrap Up

Remember!

- Re-Search
- Be short and clear
- Re-mind
- Q&A over Slack

Linkedin: @avinash-gopal-440669140
GOUP Slack: @Avi CL

Avi Gopal

Join Us!

1. Community Managers
2. Tech Writers
3. In/Out Ambassadors
4. Marketing Creators & Editors
5. Course Creators
6. Project Creators

HTTPS://GOUPAZ.COM

HTTPS://METABOB.COM

Thank You

Culture
#egoless #collaborative #competent #decentralized #scalable #fun

Open source
#creator #contributor

How can we do better?

Diversity
#age #gender #location #economics #religion #politicalview