activeloop

Deep Lake

A Lakehouse for Deep Learning

Abhinav Tuli
Levon Ghukasyan
Sasun Hambardzumyan

* Image generated by AI

@activeloopai

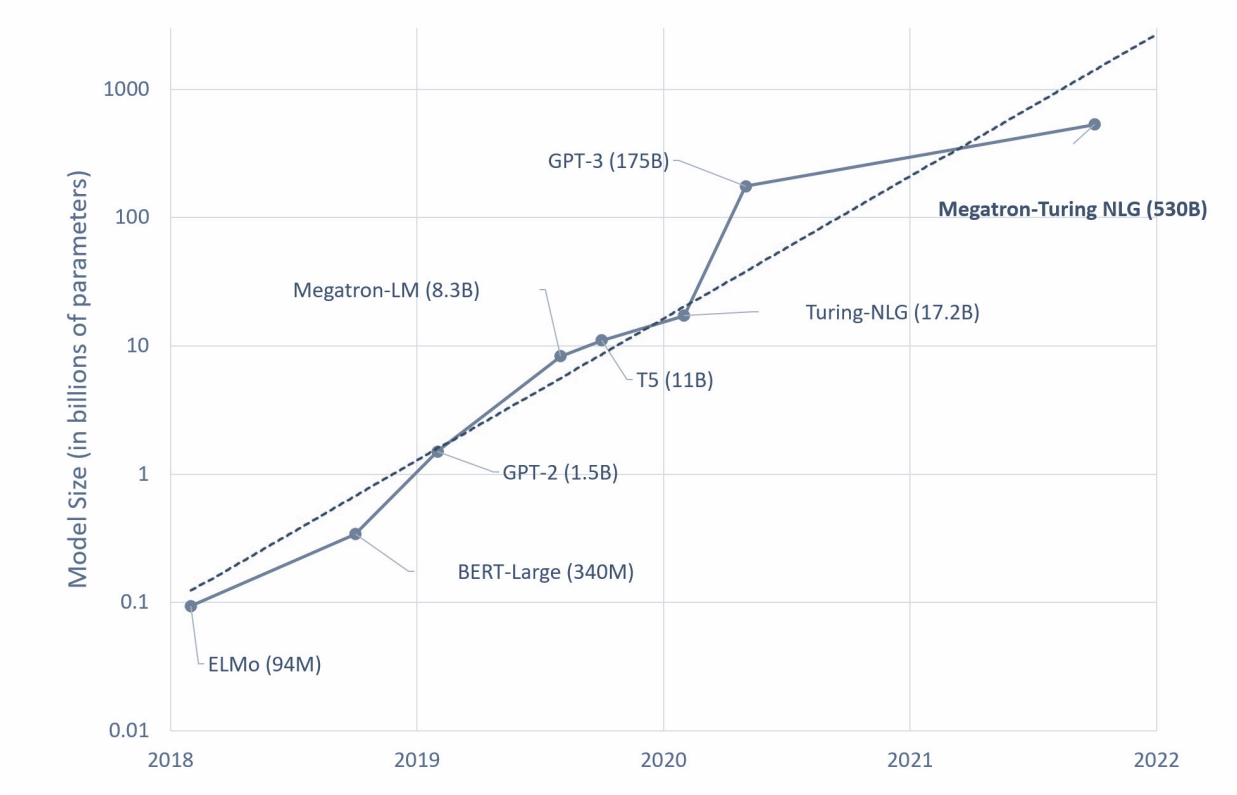
slack.activeloop.ai

activeloopai/Hub

www.activeloop.ai

Y Combinator · Stanford University · Google · TESLA · Princeton University · ETH zürich · TRENDED #1 IN PYTHON · 5.1k GITHUB STARS · 1K+ COMMUNITY MEMBERS

Deep Learning is growing at an unprecedented pace



The Data Problem

1 in

projects ends up on the ML project graveyard because of poor data development practices

Source: [Rackspace Technologies report]

activeloop

Solving the problem using data lakes

Benefits

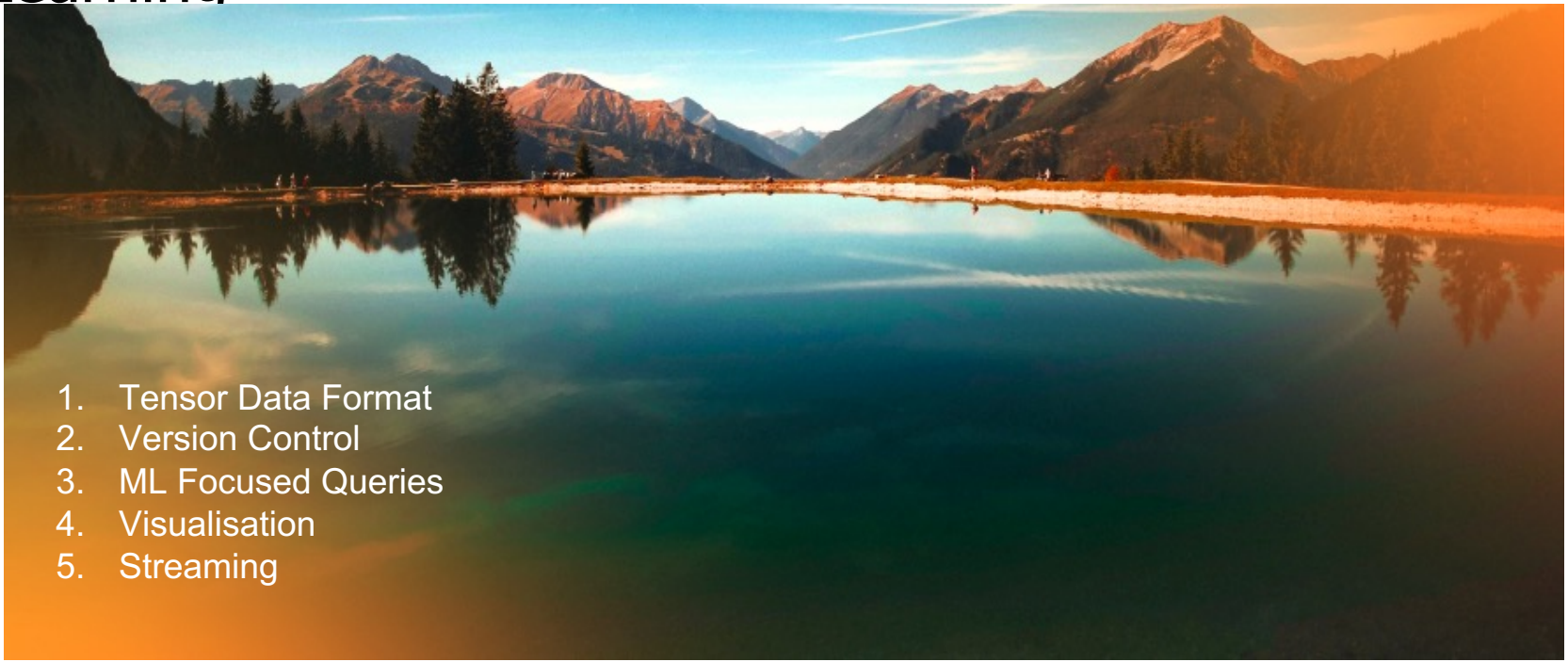
> Break down Data Silos
> Enable data-driven Decision Making
> Improve Operational Efficiency
> Reduce Costs

Limitations

> Complex data isn’t supported
> No Deep Learning integration
> Missing gap between MLOps and MDS
> Queries only for analytics

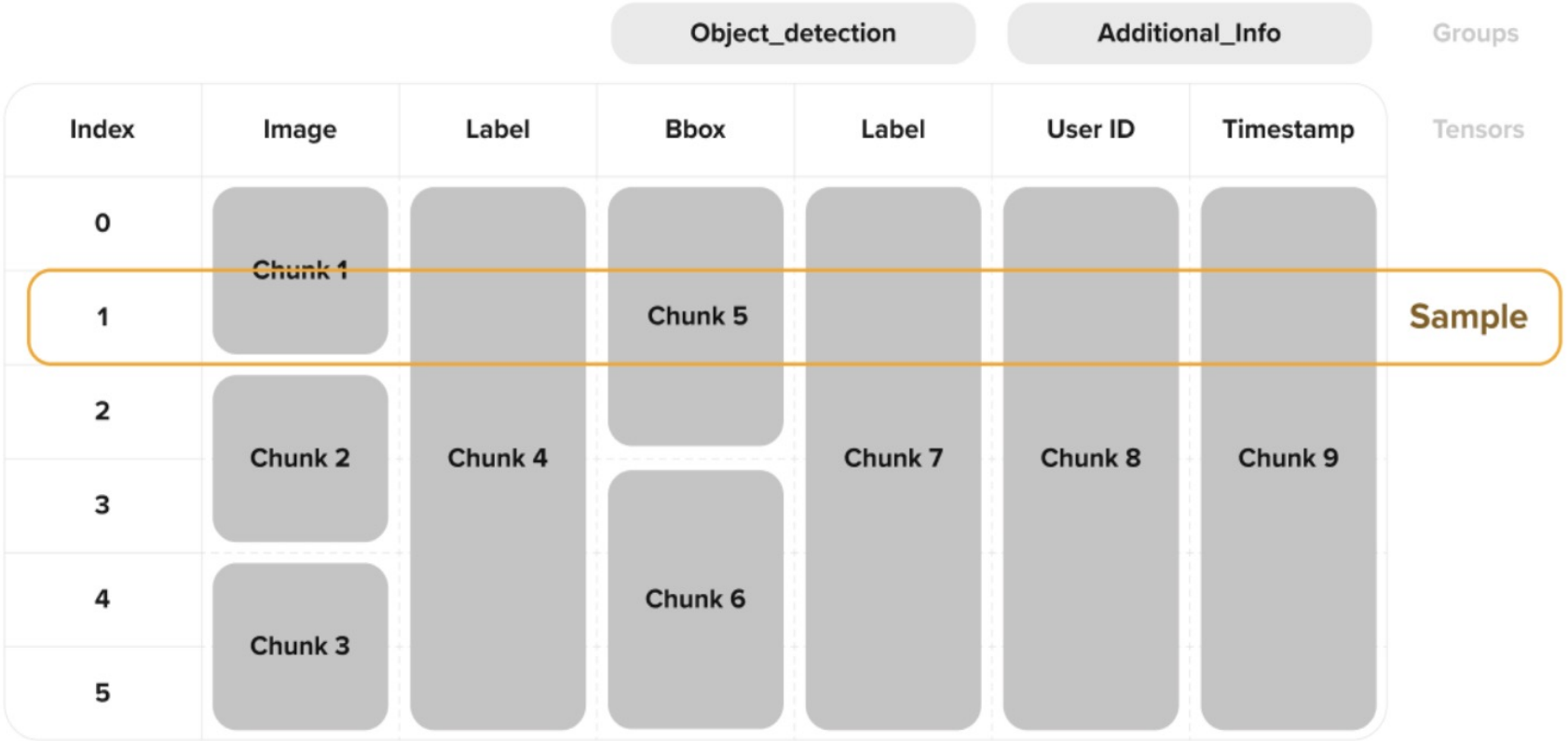
activeloop

Introducing Deep Lake: Lakehouse for Deep Learning

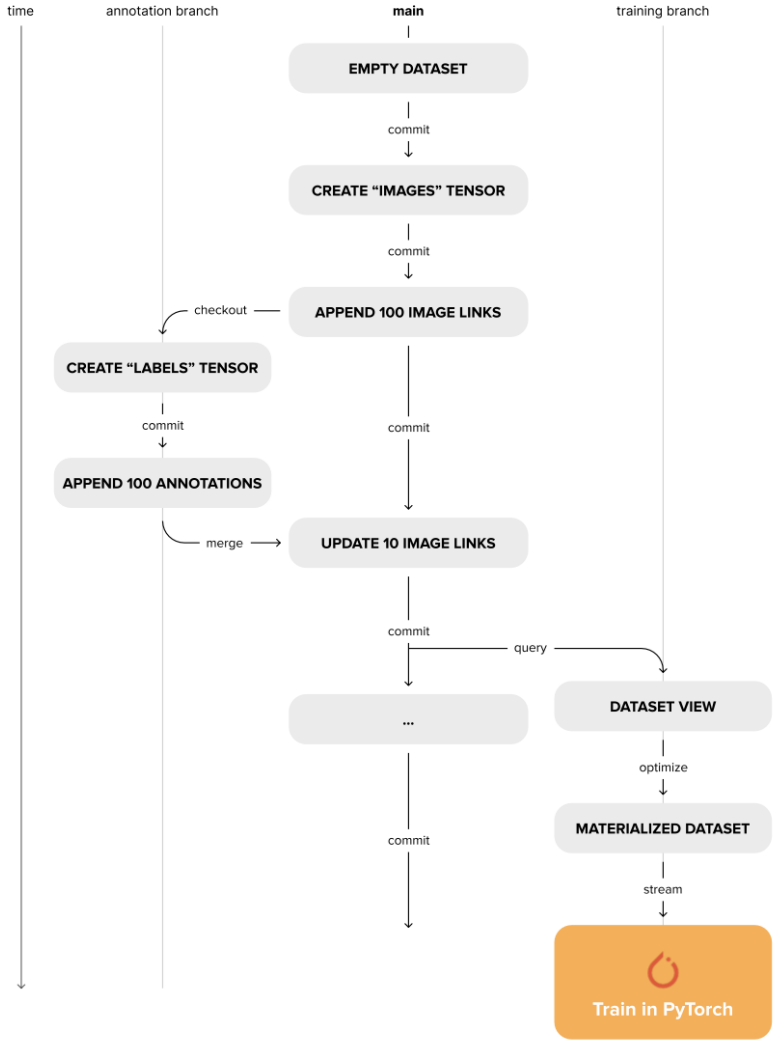


activeloop

Tensor Storage Format: Native to Deep Learning



Version Control: Track Data Lineag

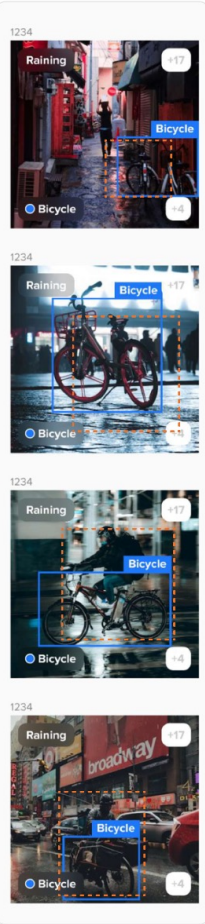


activeloop

Query: Rapid queries with Tensor Query Language (**TQL**)

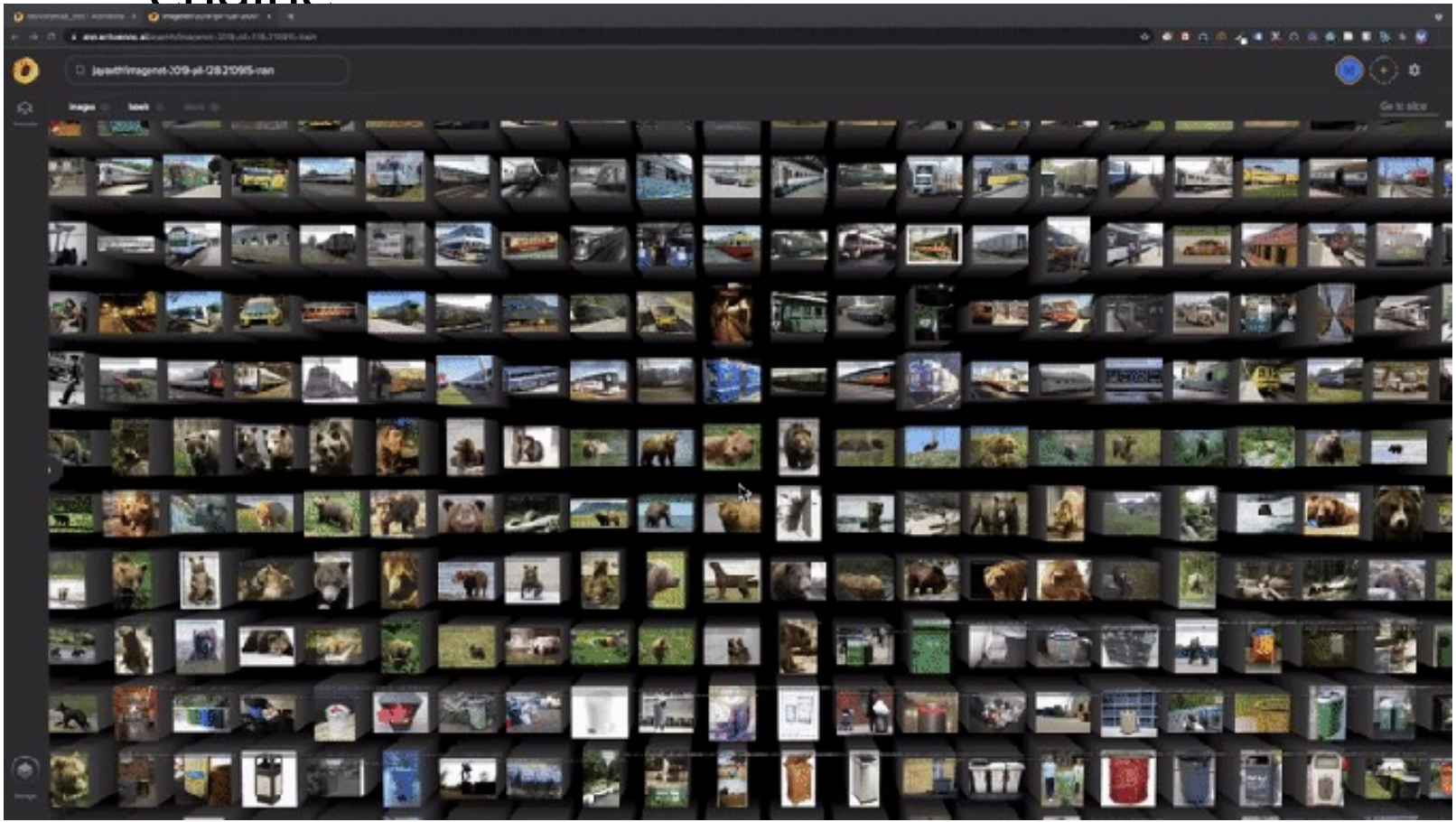
```
SELECT images [100:500, 100:500], boxes + ARRAY[-100, -100, 0, 0]
WHERE contains(categories, 'bicycle') and weather == 'raining'
ORDER BY AOI(boxes, prediction) desc
LIMIT 1000
```

Example query with indexing tensors inline with select and ordered by user-defined function computation.



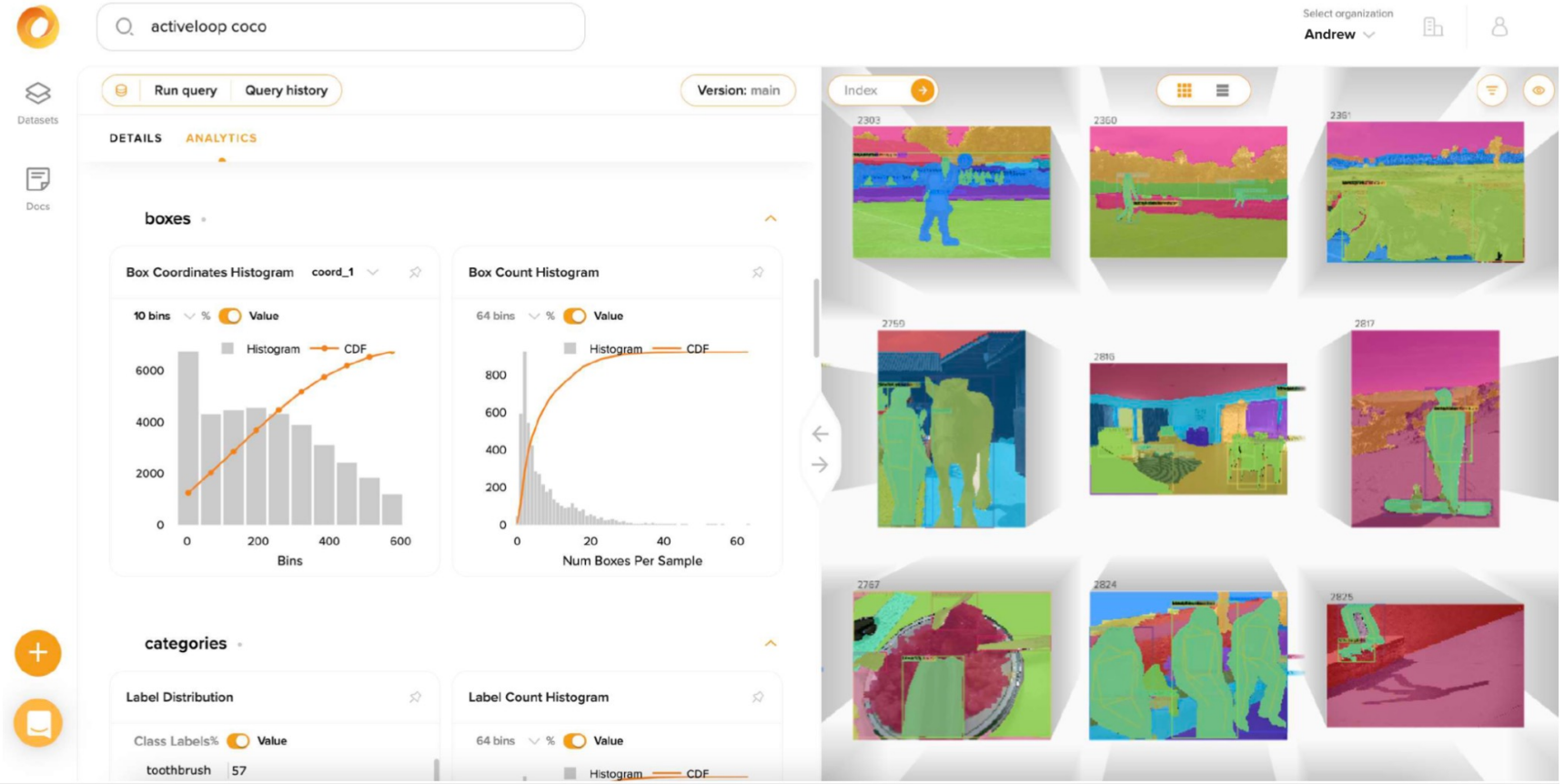
activeloop

Visualize: In-browser visualization engine



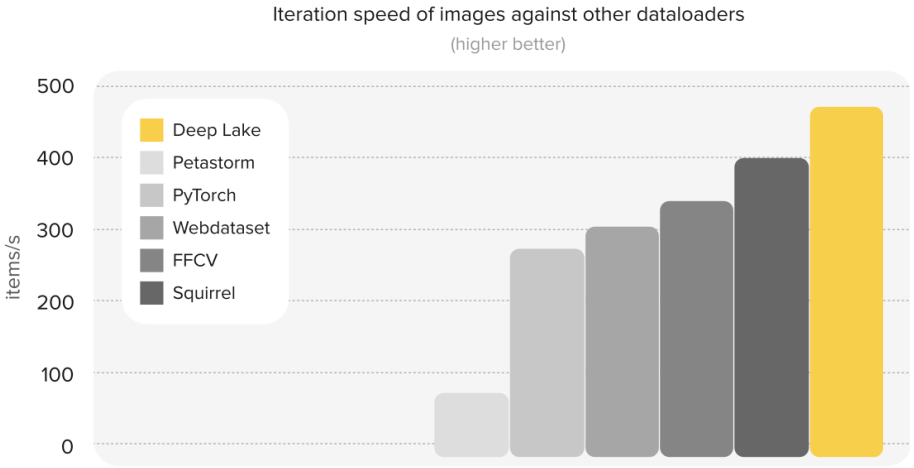
activeloop

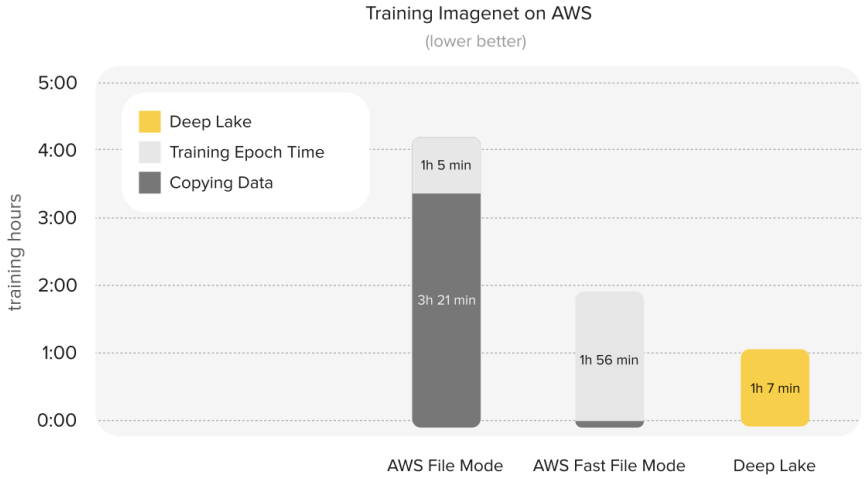
Visualize: In-browser visualization engine



activeloop

Stream: Streaming Data Loaders up to 4x faster*





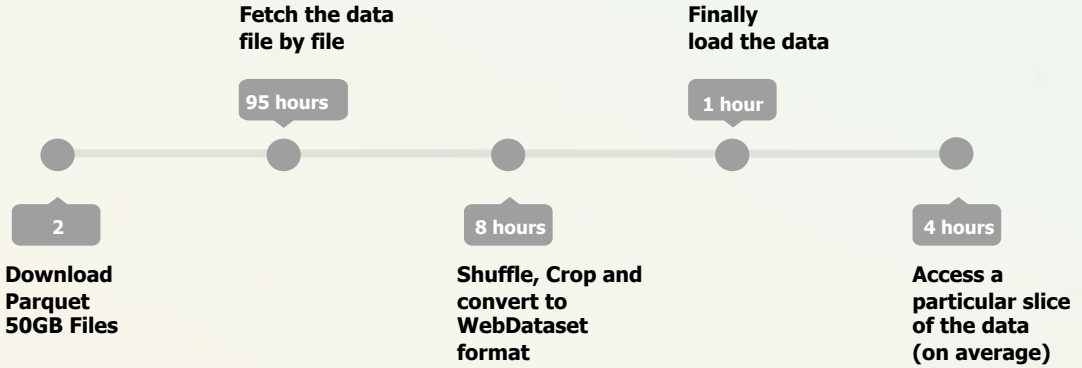
*Ofeidis et. al; ***An Overview of the Dataloader Landscape. Challenges and Promises***

activeloop

Save Time: Access LAION Dataset in <5 Seconds

100 hrs+

with in-house solutions

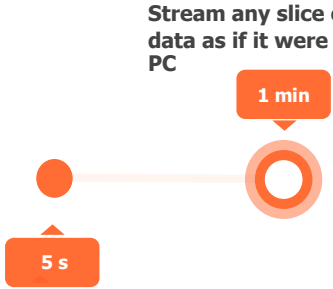


Before: no indexing and shuffling with WebDataset

After with DeepLake: + upload in ~6 hr

5 seconds

with activeloop

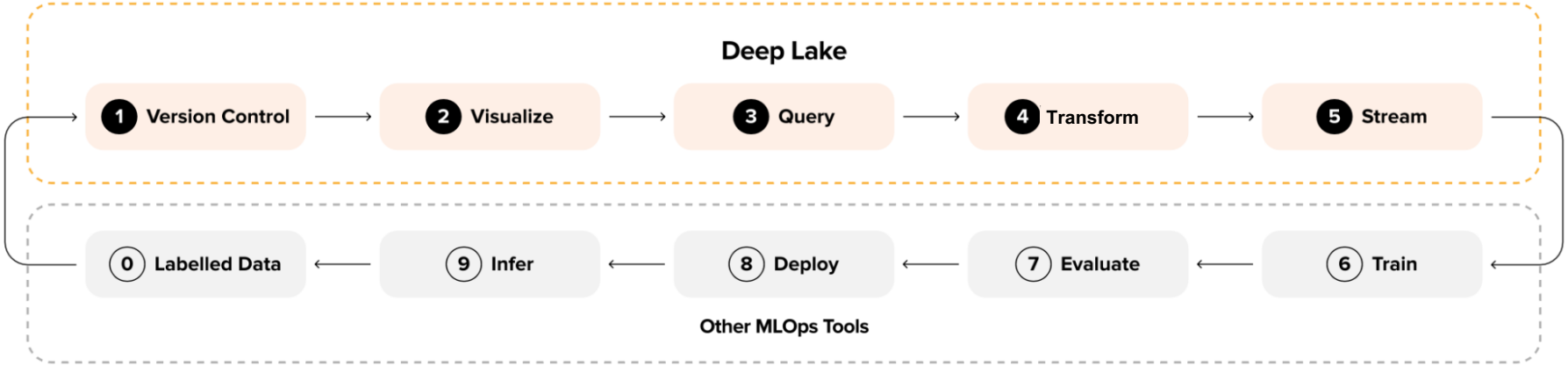



```
import deeplake
ds = deeplake.load('hub://laion/laion-400M')
dsv = ds.query('select * where …')
dl = dsv.pytorch(num_workers=16)
```

intel

activeloop

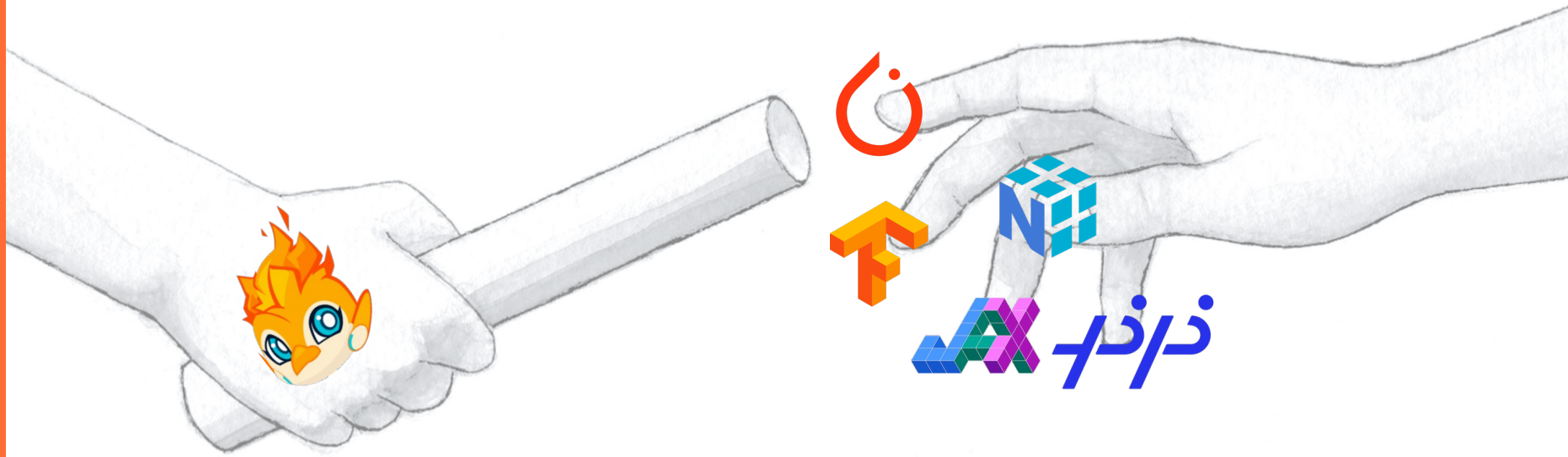
Deep Lake: Lakehouse for Deep Learning



Machine Learning Loop with Deep Lake

activeloop

Unlock compute from data bottleneck



activeloop

activeloop

Dive into Deeplake

https://github.com/activeloopai/deeplake

@activeloopai

slack.activeloop.ai

activeloopai/Hub

www.activeloop.ai

* Image generated by AI

Y Combinator · Stanford University · Google · TESLA · Princeton University · ETH zürich · TRENDED #1 IN PYTHON · 5.1k GITHUB STARS · 1K+ COMMUNITY MEMBERS