

Surviving the Flood (of rosbags)

ROSCon 2023

Benji Barash, Roboto AI

whoami



Pictured: On a field collecting a rosbag for camera calibration

Benji Barash

Former simulation lead at Amazon

Now building a startup: Roboto AI

rosbagging since 2013

When people hear you
work in robotics...



When people hear you work in robotics...



Reality



Robotics engineer after using rosbag record

Surviving the Flood (of rosbags)

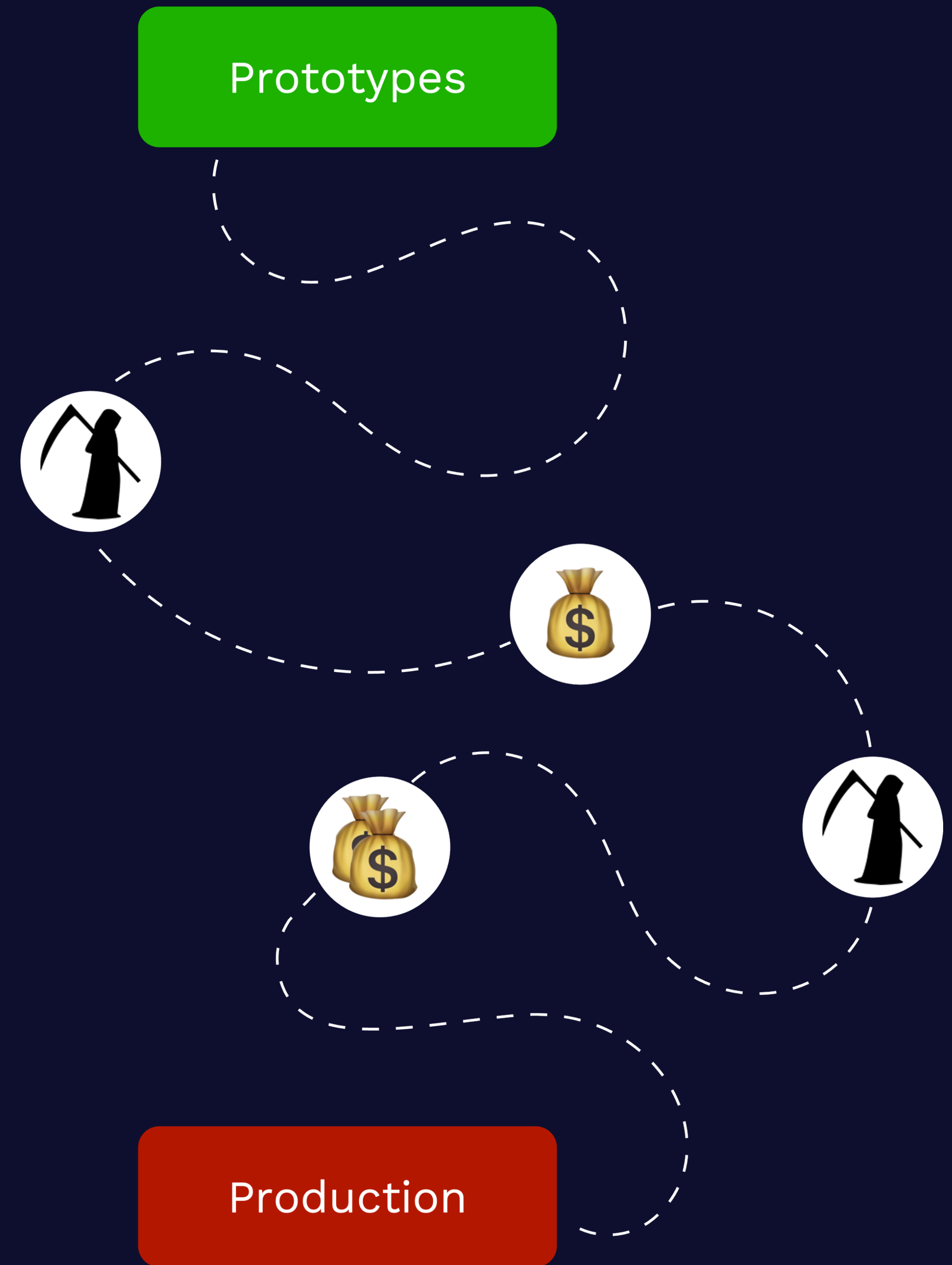


In this talk

- Overview
- Journey to Production
- Lessons Learned at Amazon
- Survive the Flood
- Roboto

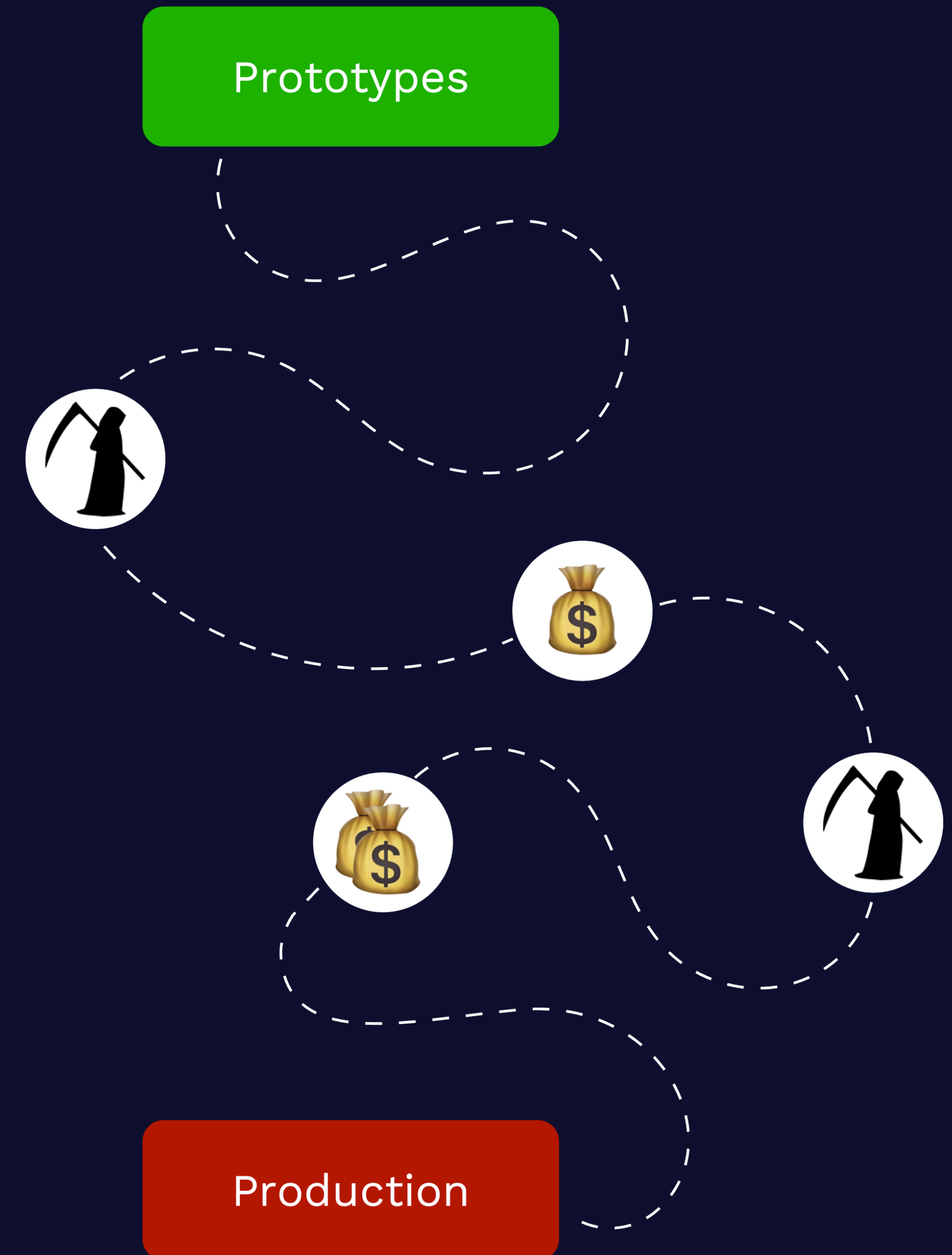
Getting to production is hard

Many robots don't make it...



Oh and also...

- 📈 Amount of data grows *very* quickly
- ⚠️ Unexpected edge cases show up
- 💣 Bugs become costly and harder to fix
- 👮 Regulators need evidence of reliability



Let's look at how things change along the way...



Prototyping: 1-5 people

 I get to work on everything from perception to path planning

 Quick iteration and limited testing; it's ok to break things






 We use mostly off-the-shelf components

 Everyone is pretty technical, we share scripts

 I store my rosbags locally and use RViz








Building a V1: 15-20 people

-  Teams start to form: hardware, perception, controls, etc.
-  Some dev ops, system failures slow us down but are expected
-  We're building some of our own hardware and algorithms
-  System complexity is higher, only a few people have full context
-  Let's put data in Google Drive / SharePoint etc.



Towards Production: 50 to 100+ people

-  Dedicated teams, including operations
-  Procedures for development and verification
-  Lots of custom hardware and algorithms
-  I don't know everyone in the org => finger pointing
-  System failures and incidents are a BIG deal...

Data Deluge



Let's put logs into S3 and hire an intern to build a frontend

Our intern left, things are a mess again

⚡ CRASH ⚡
VP wants to know how many times we've had this issue before?

PM:
Can someone make me a video of our recent crash?



Systems Engineer:
I need state estimates in a CSV. How do I get them out of 500 x 20GB rosbags?

Leadership:
We're going to hire an infra team

Infra team:
We need to rewrite everything from scratch. See you in 6+ months

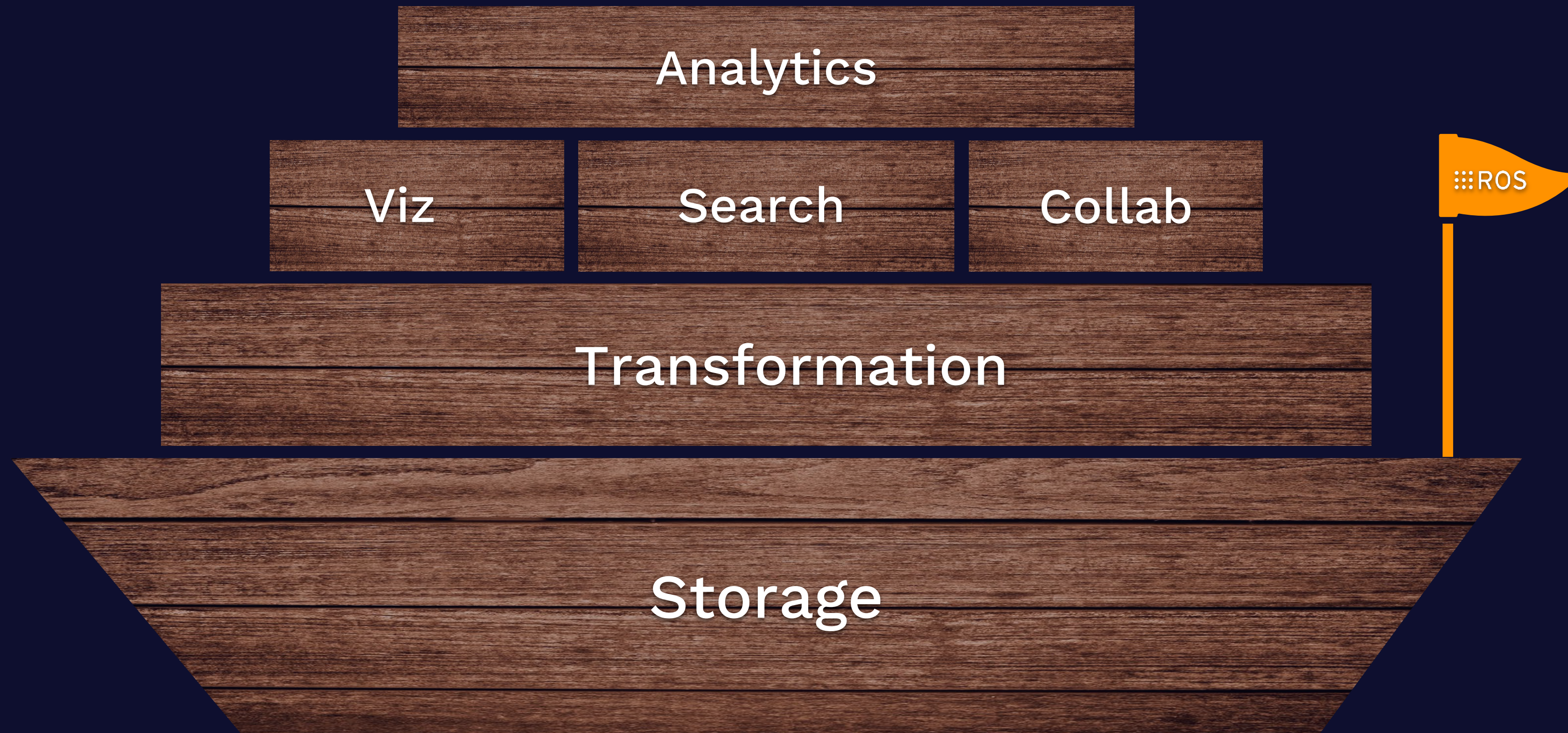
CV Scientist:
I need to extract images from bags... I'm building my own data infra instead

Lessons Learned

- Data infrastructure gets deprioritized until it's too late
- Data and scripts get siloed across teams and cause friction
- Complacency with “nominal” data hides underlying issues
- Complex issues span subsystems and require collaboration

Survive the Flood

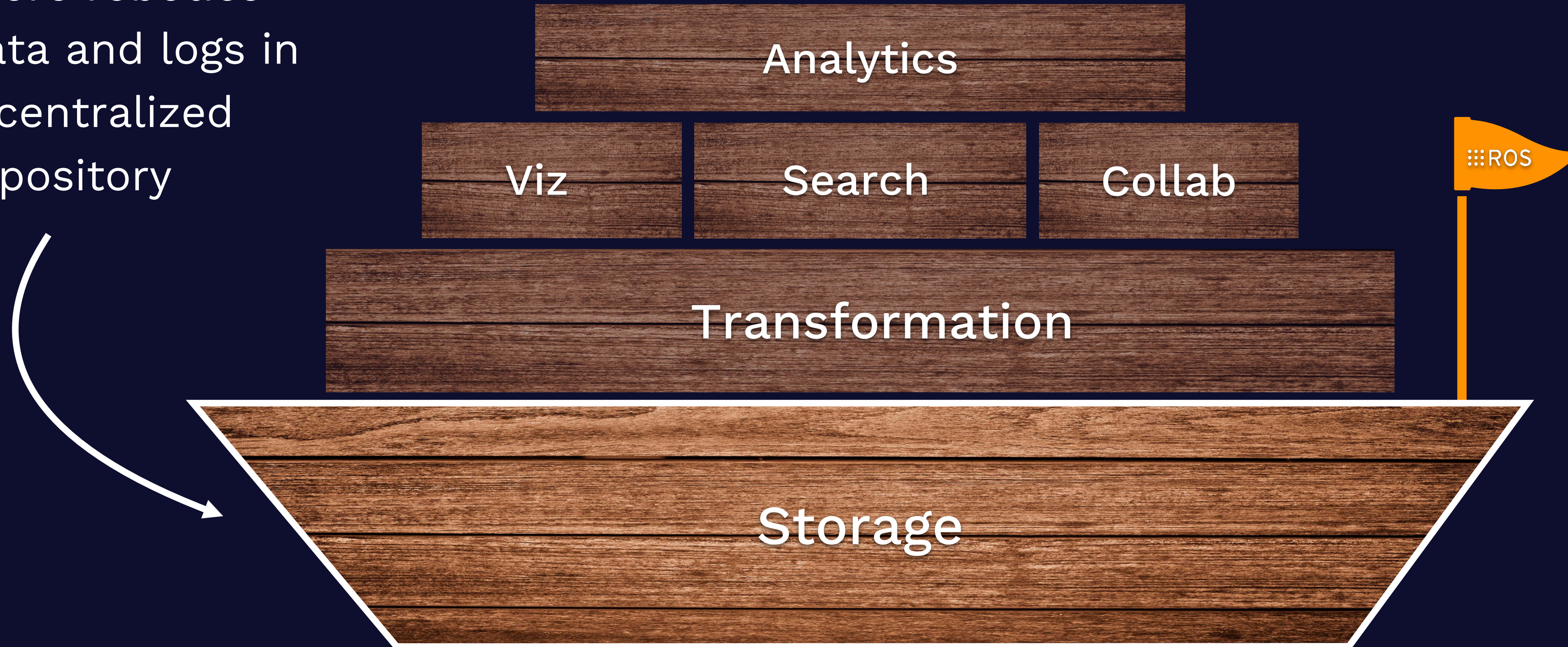
Build an ark and get everyone onboard beforehand



Survive the Flood

Build an ark and get everyone onboard beforehand

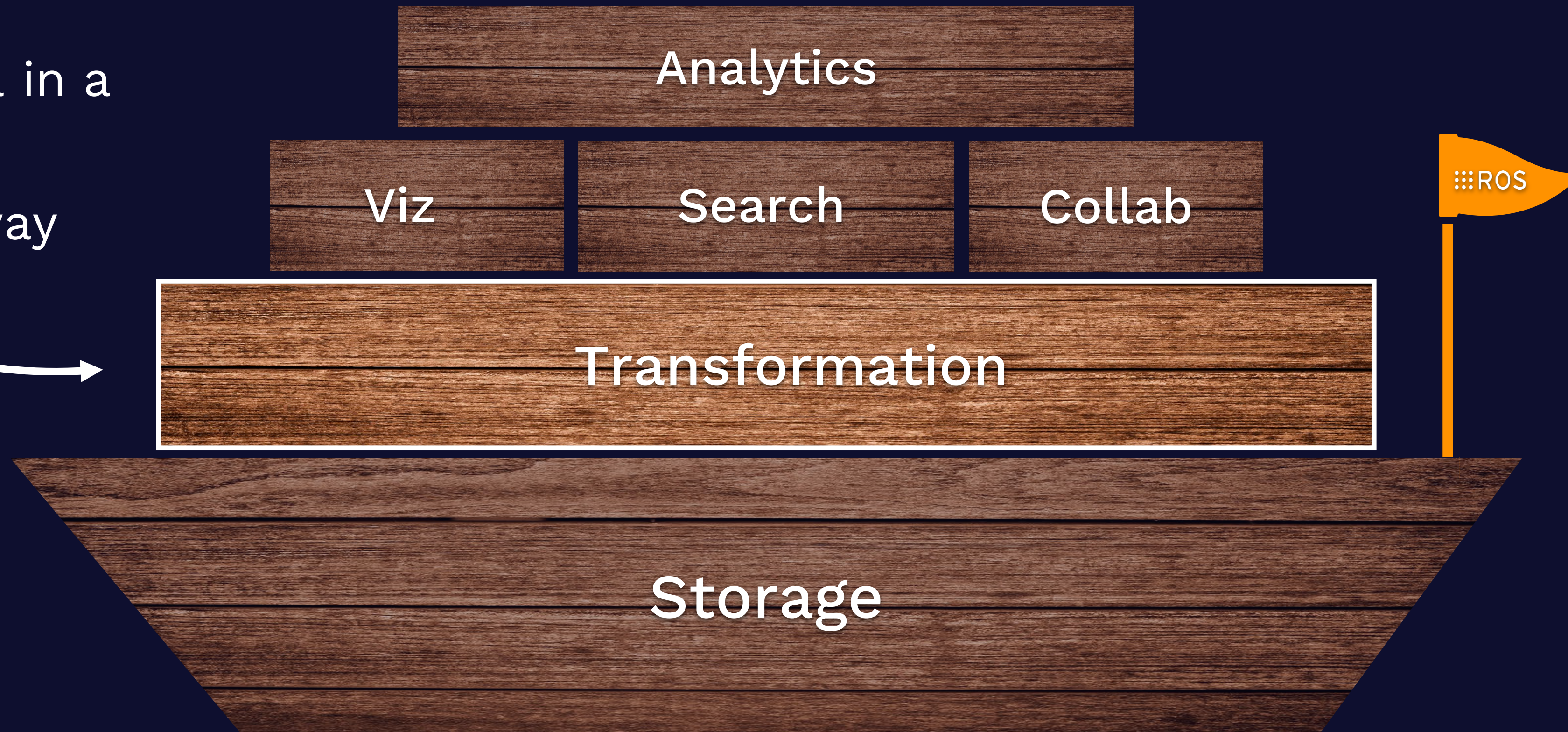
Store robotics
data and logs in
a centralized
repository



Survive the Flood

Build an ark and get everyone onboard beforehand

Process data in a flexible but consistent way

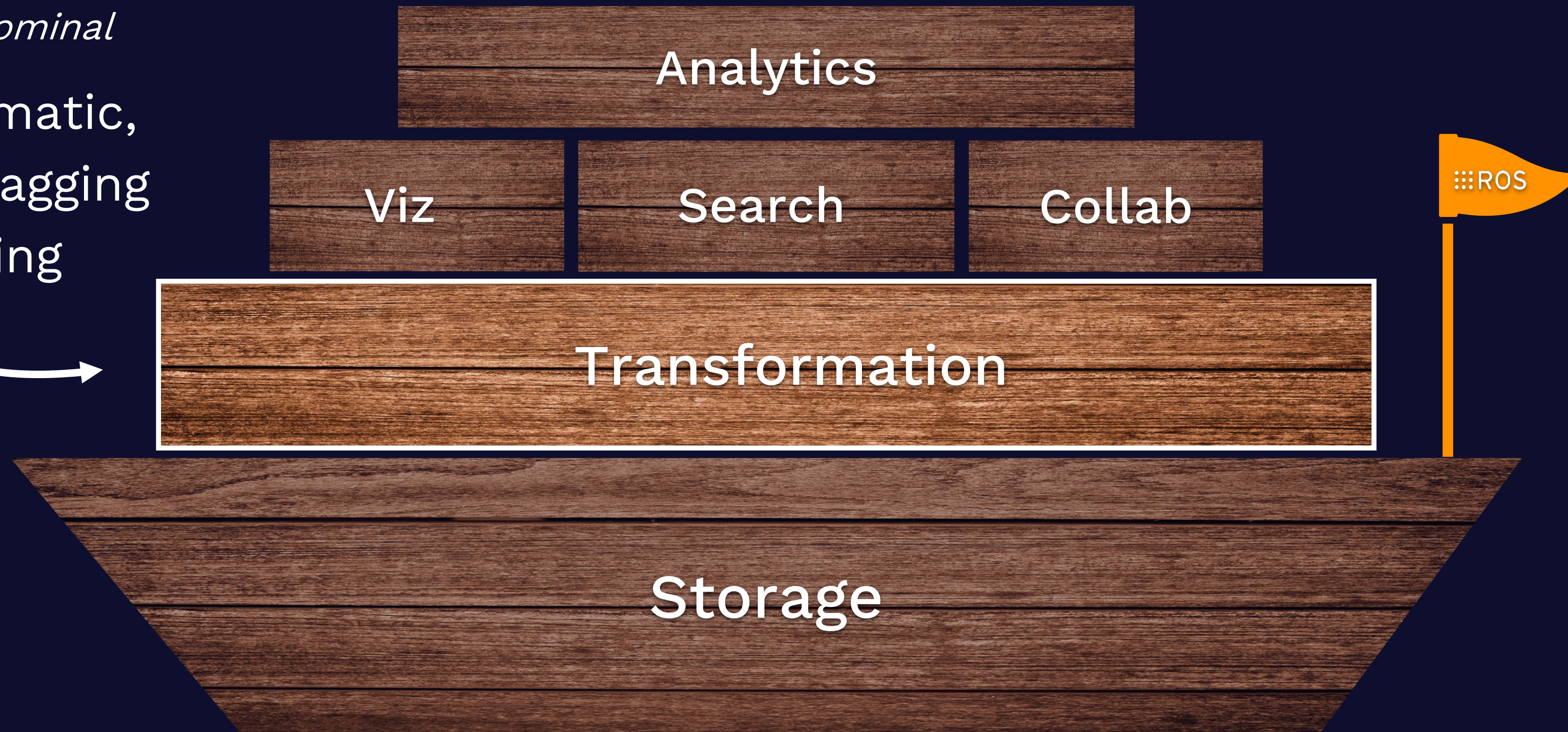


Survive the Flood

Build an ark and get everyone onboard beforehand

No incident != Nominal

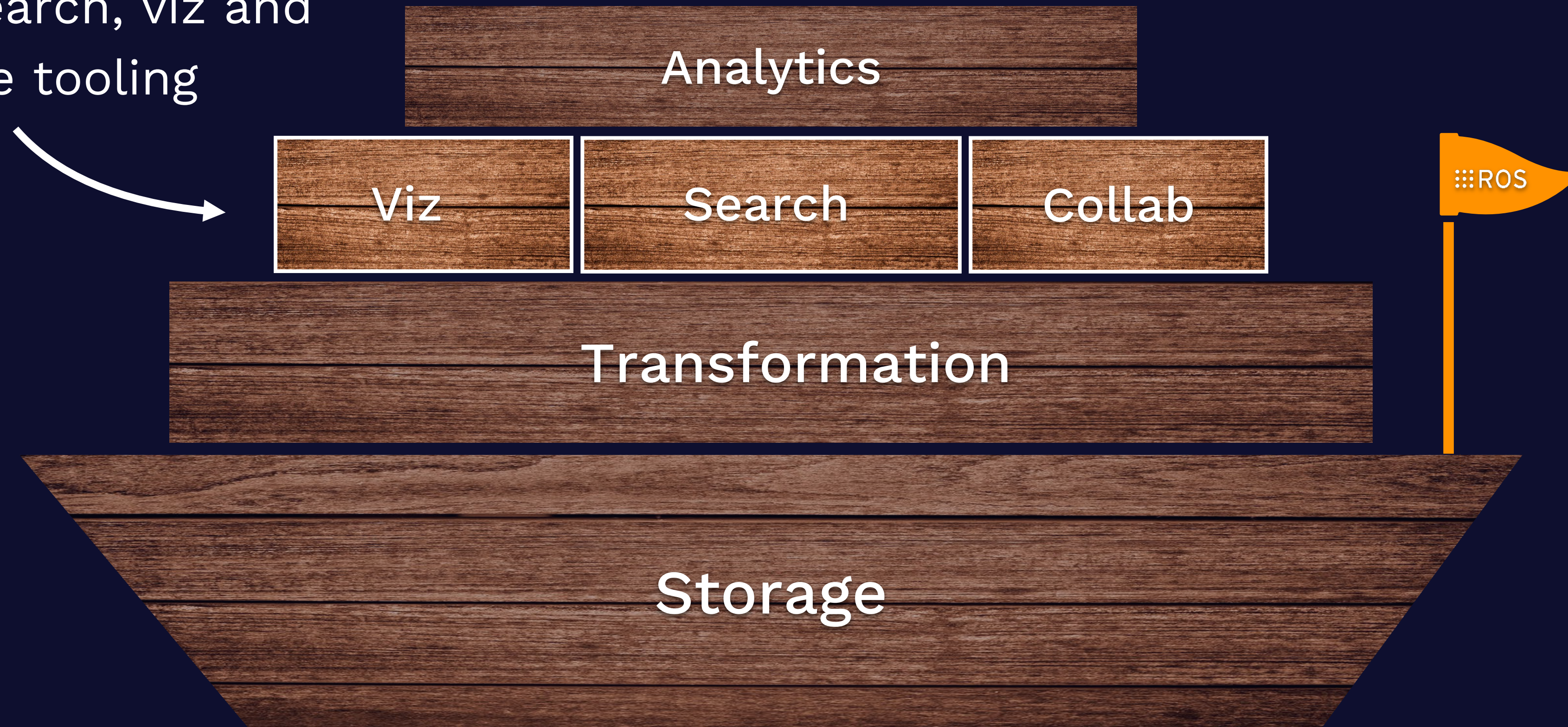
Enable automatic,
rule-based tagging
and processing



Survive the Flood

Build an ark and get everyone onboard beforehand

Democratize access to data with search, viz and collaborative tooling

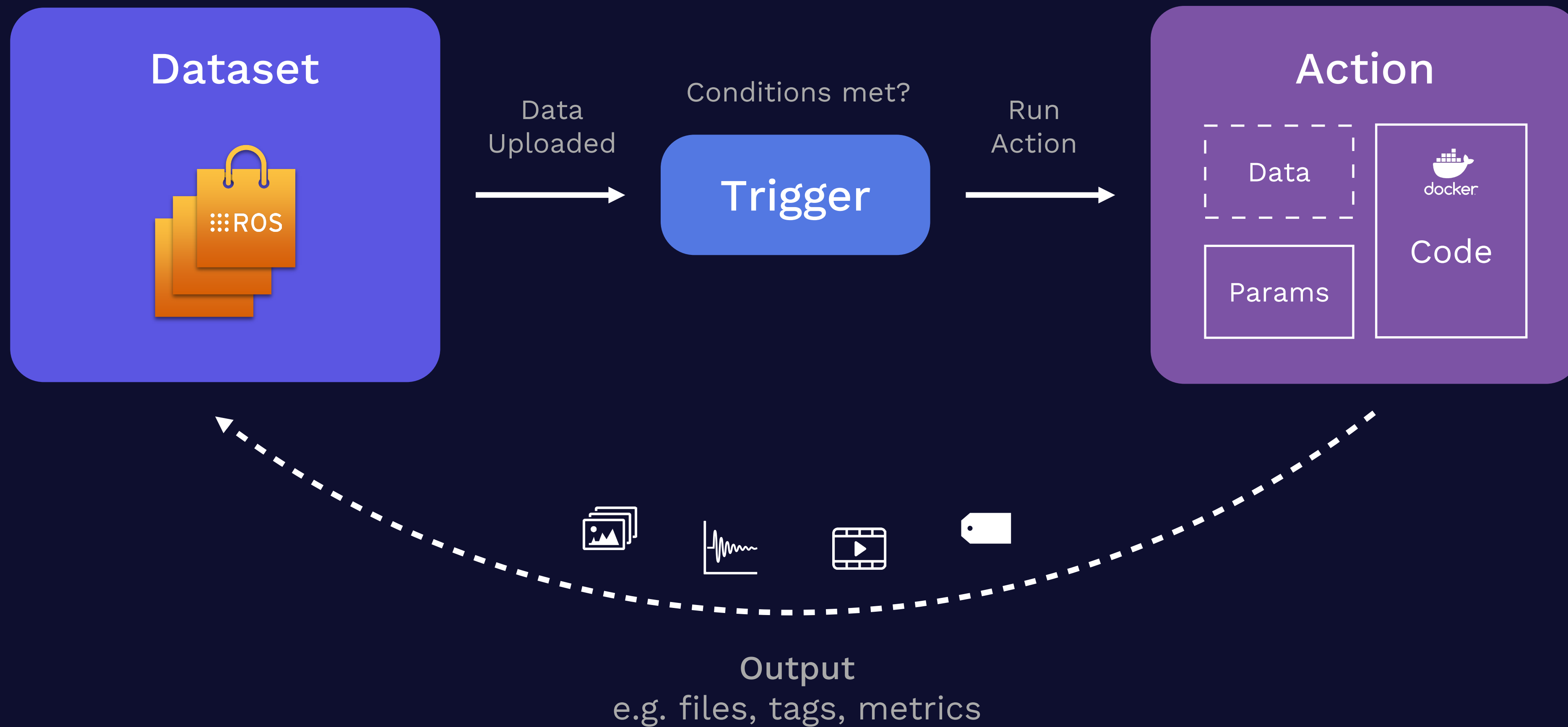


How can we build the ark?

We need new components

esp. for data processing and search

Actions Framework



action.json

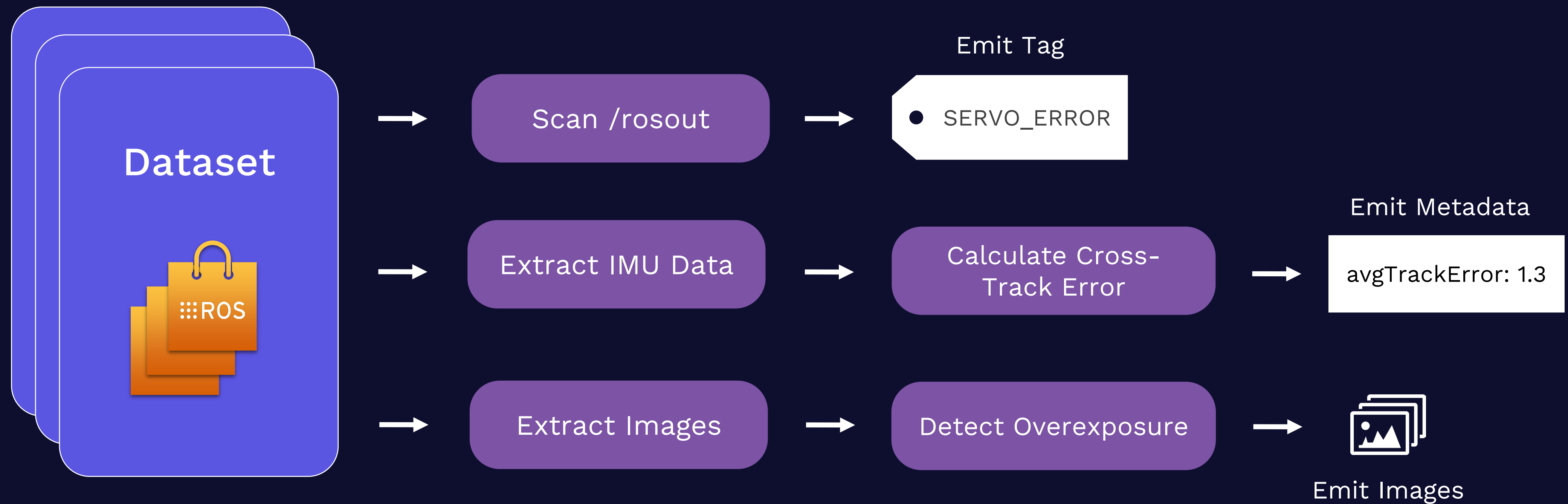
```
{
  "name": "yolov8",
  "image": "<registry/repo/name>",
  "description": "...",
  "parameters": [
    {
      "name": "topics",
      "required": false,
      "description": "..."
    },
    {
      "name": "model_name",
      "required": false,
      "description": "...",
      "default": "yolov8n"
    },
    ...
  ],
  "compute_requirements": {
    "memory": 8192,
    "vCPU": 2048
  },
  ...
}
```

Containerized source

Explicit parameters

Invokable locally and in cloud

Chain Actions to Build Workflows



Build across teams with a common foundation

Maybe we can all share useful actions...

robologs

Open source actions for robotics data



e.g.

- bag-to-mcap
- downsample-topic
- topic-to-video
- detect-overexposure
- yolov8-object-detection

Contribute: <https://github.com/roboto-ai/robologs>

Where can I run actions?

iRoboto [Preview](#)

A flexible platform to process and search your data



Sneak Preview...



Try it out!

```
$ pip install roboto
```

www.roboto.ai