

Building GPU-Accelerated Workflows with TensorFlow and Kubernetes

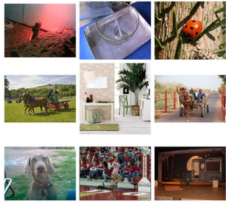
Daniel Whitenack, [@dwhitena](#)

Data Scientist and Advocate, [@pachydermIO](#)

Outline

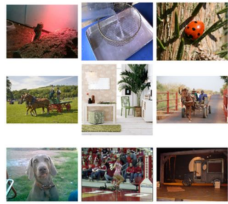
1. A typical AI workflow
2. Where GPUs come into play
3. GPU-accelerated AI on k8s
4. Live demo!
5. Q&A, resources

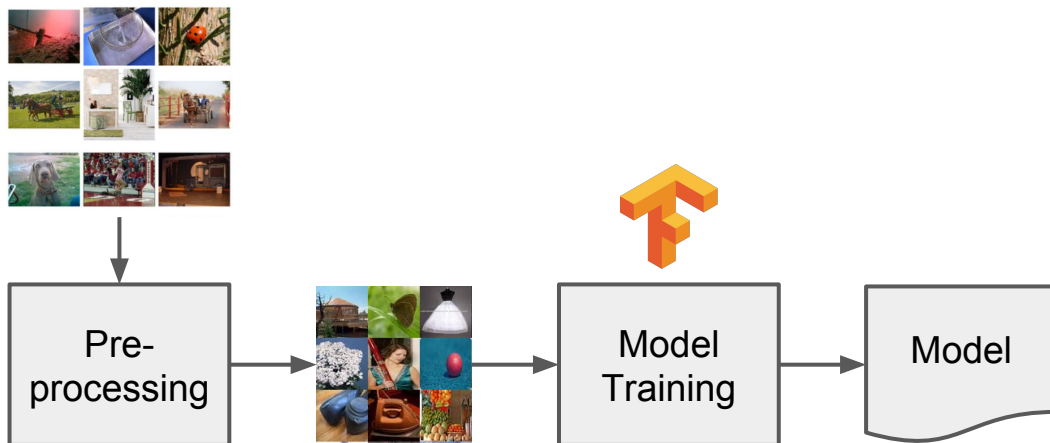
A Typical AI Workflow

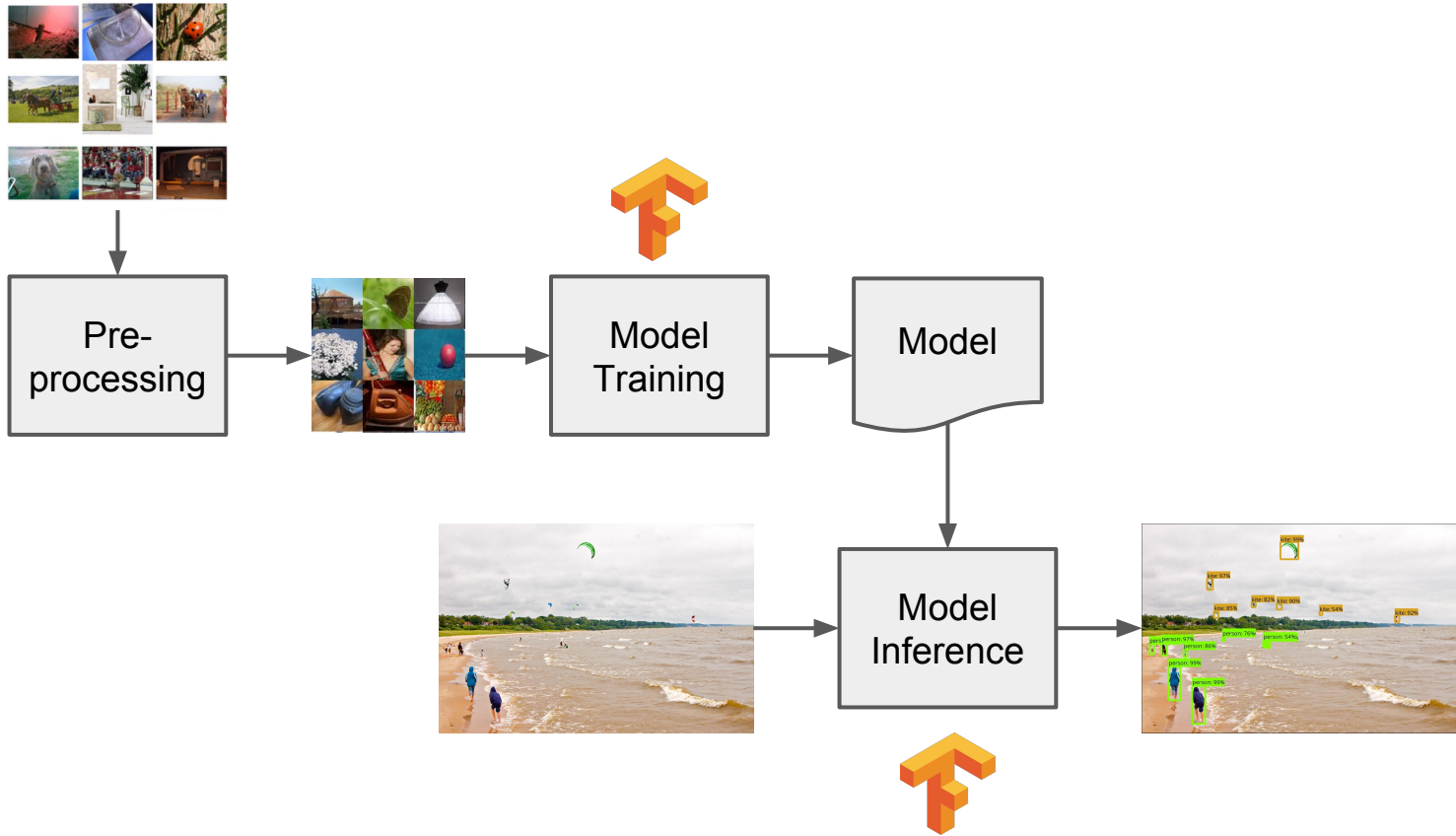


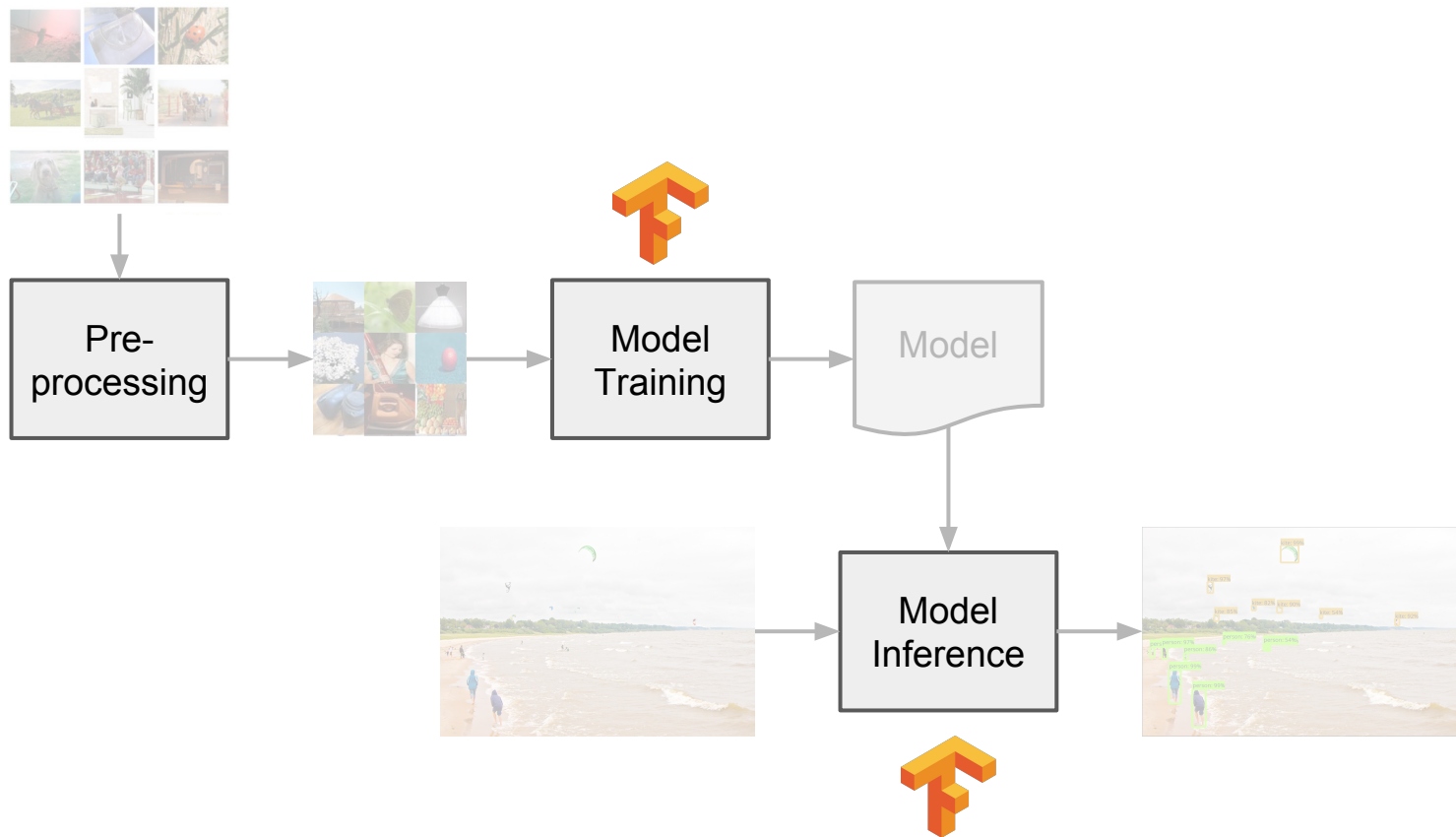
@dwhitena

#KubeCon #CloudNativeCon

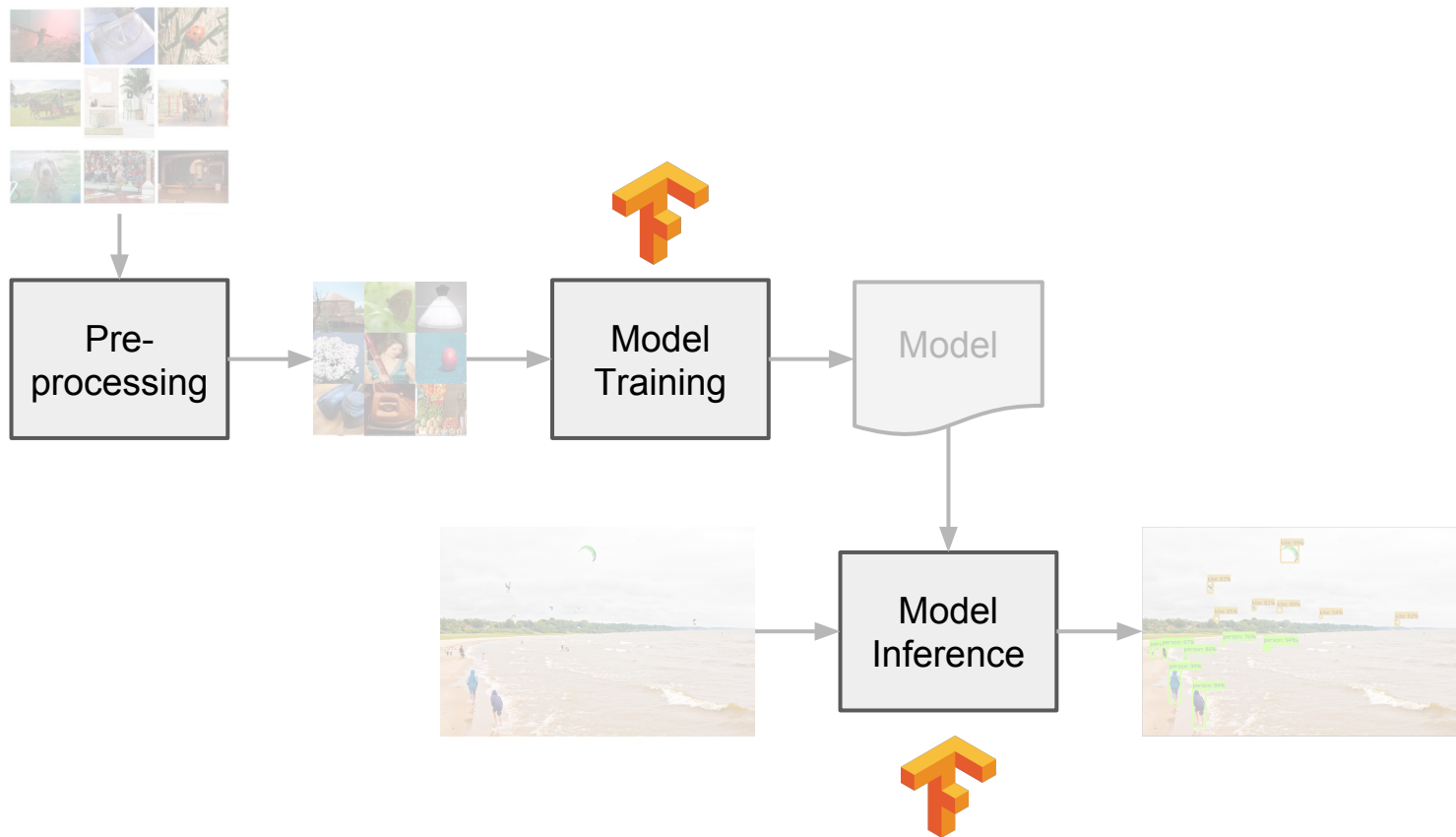


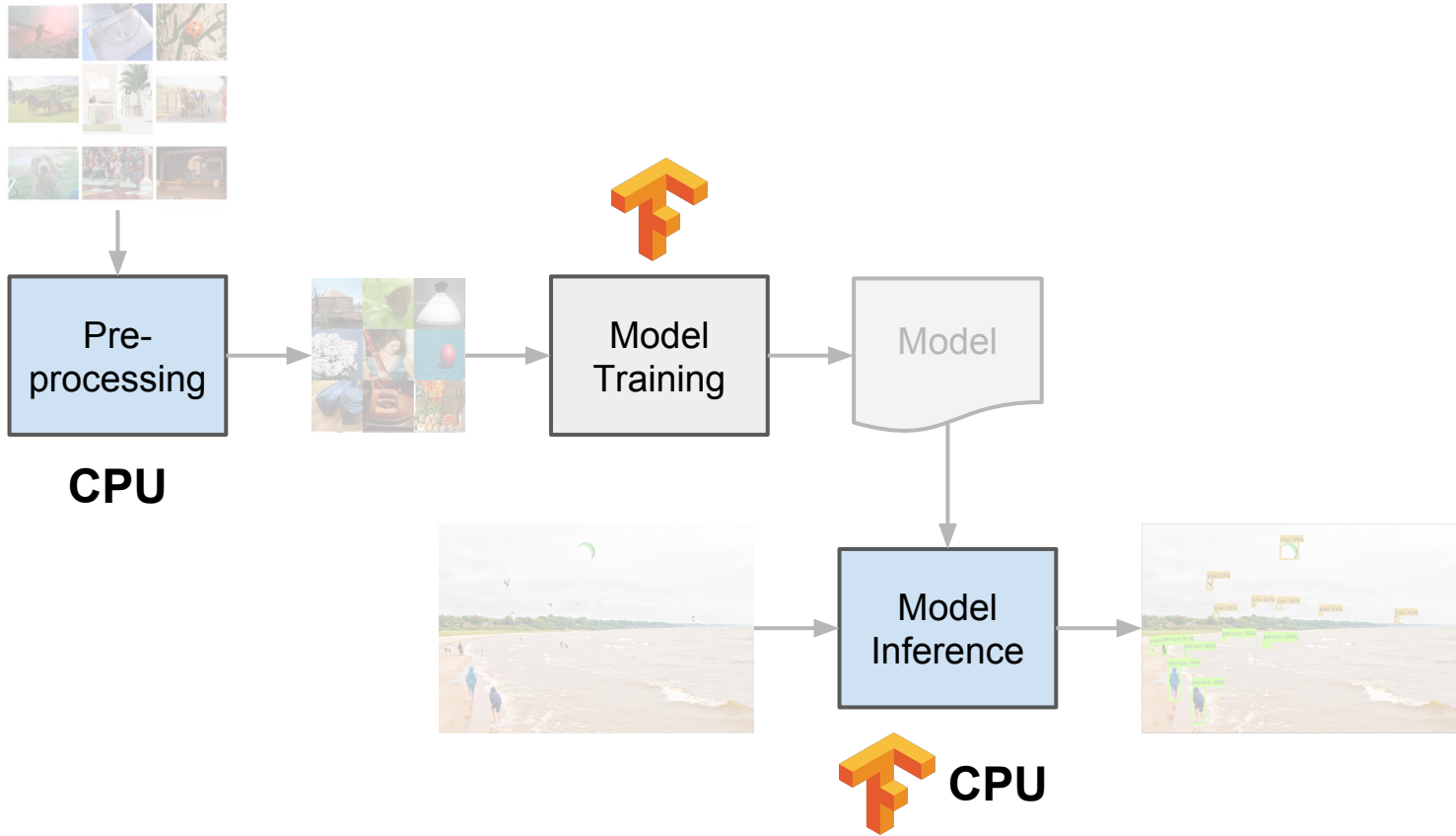


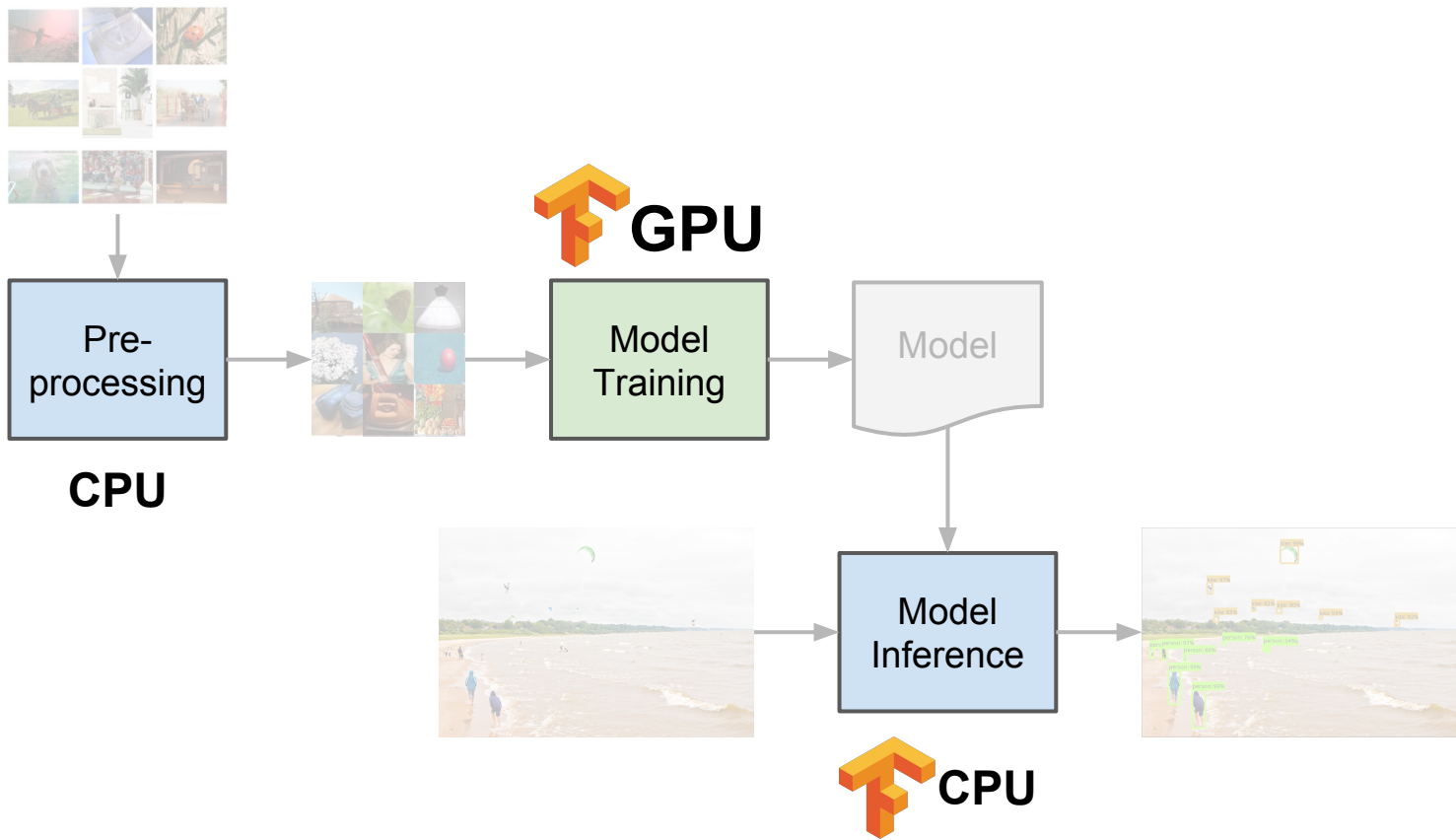




Where GPUs Come into Play





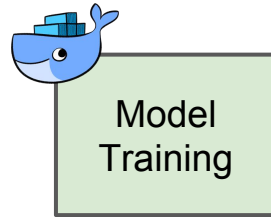
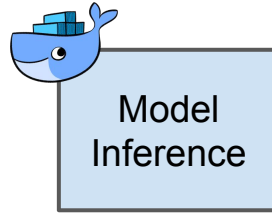
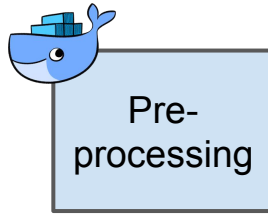


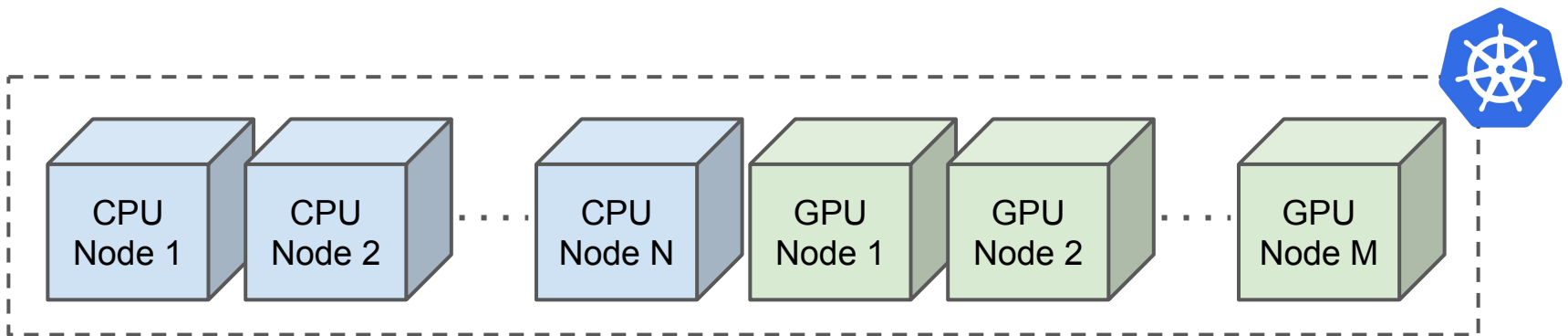
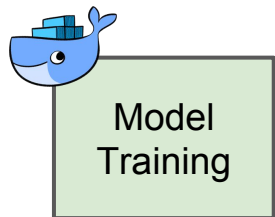
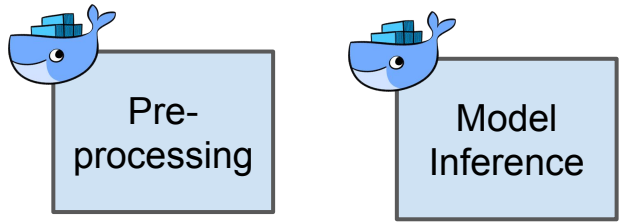
GPU-accelerated AI on Kubernetes

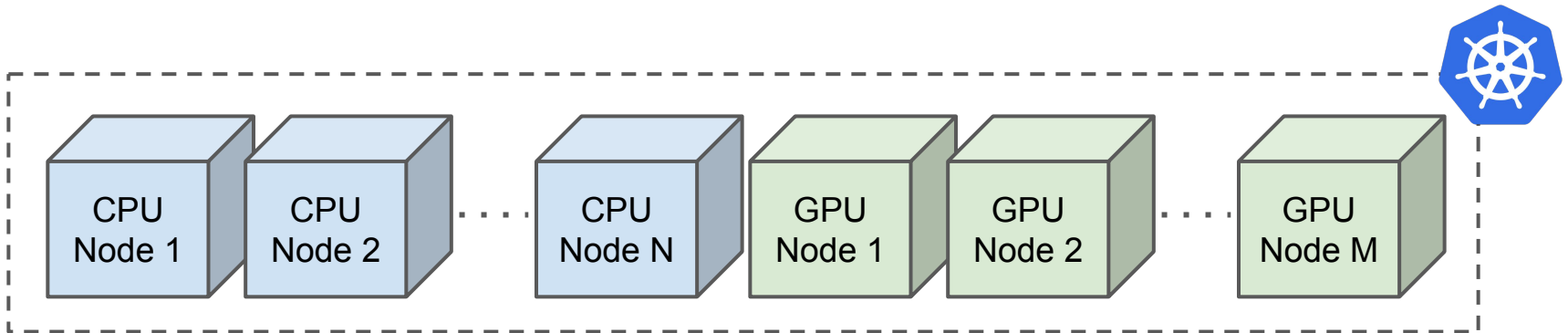
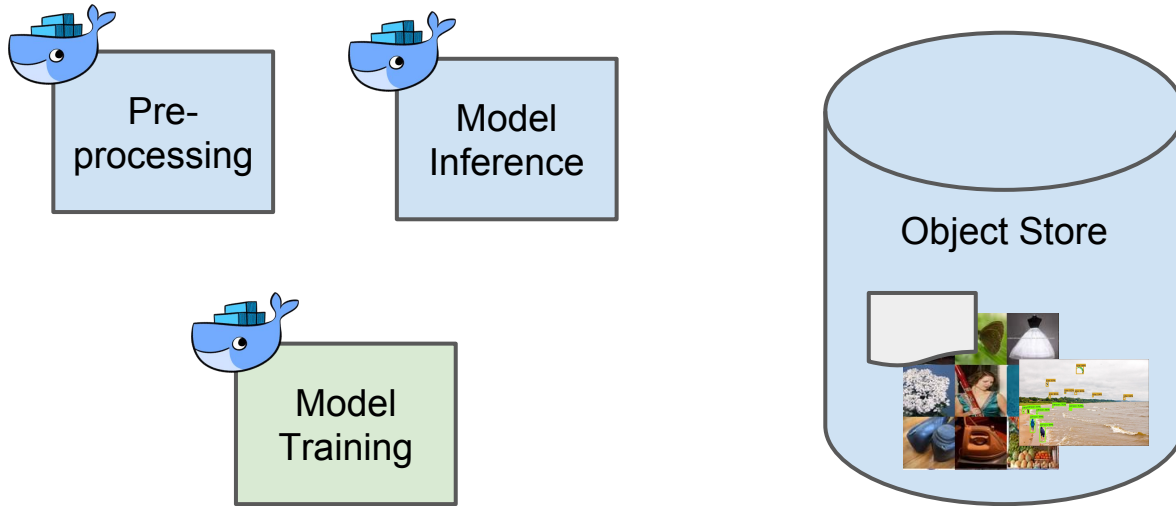
Pre-
processing

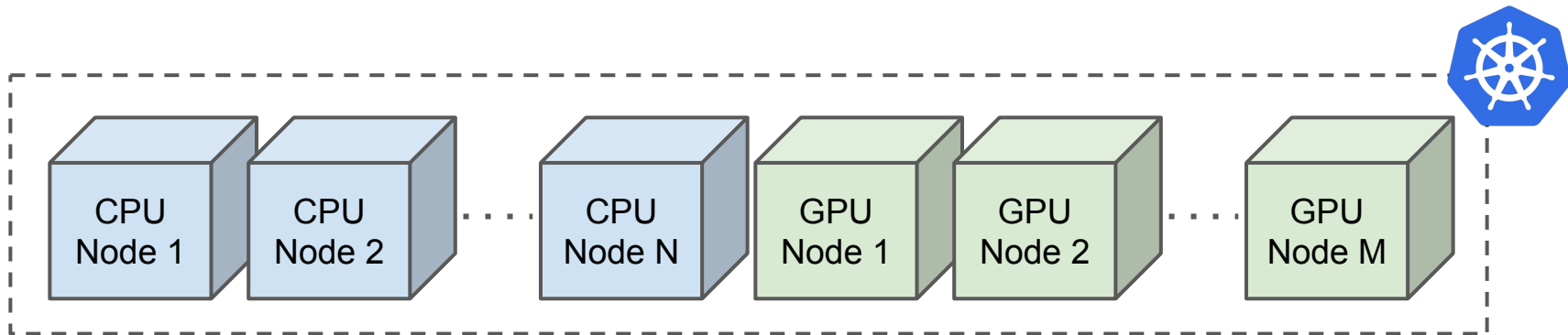
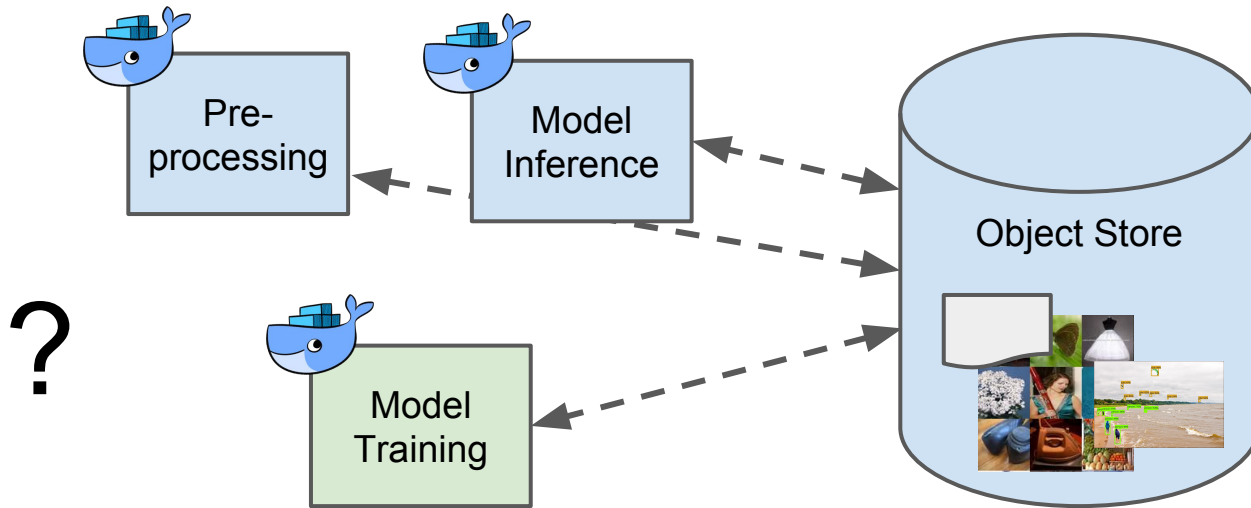
Model
Inference

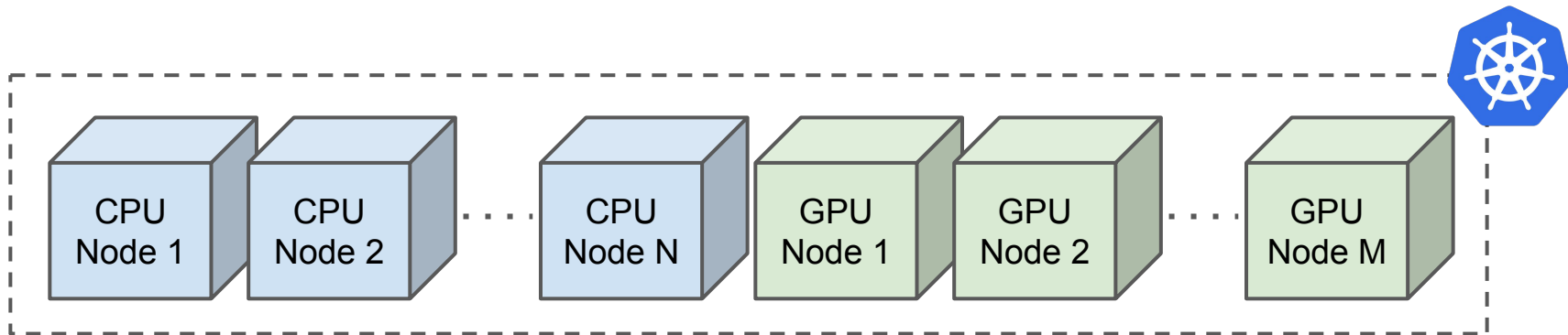
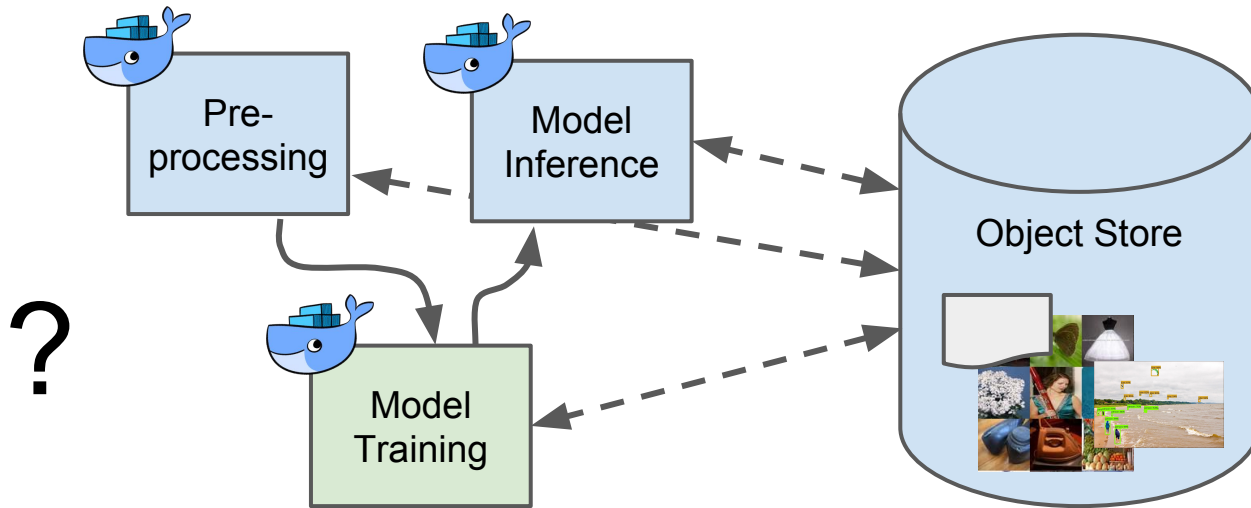
Model
Training











But how do we...

- Get the right data to the right code,

But how do we...

- Get the right data to the right code,
- Process data on the right nodes, and

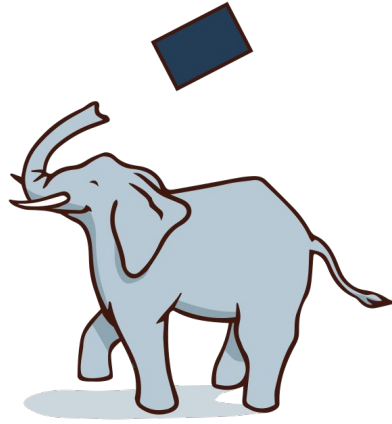
But how do we...

- Get the right data to the right code,
- Process data on the right nodes, and
- Trigger the right code at the right time.

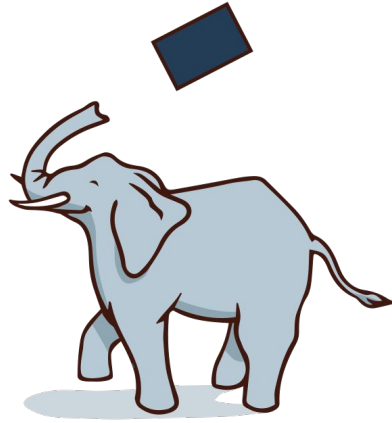
But how do we...

- Get the right data to the right code,
- Process data on the right nodes, and
- Trigger the right code at the right time.

***Bonus** - Track which versions of code and data ran to produce which results (for debugging, maintenance, and compliance)*

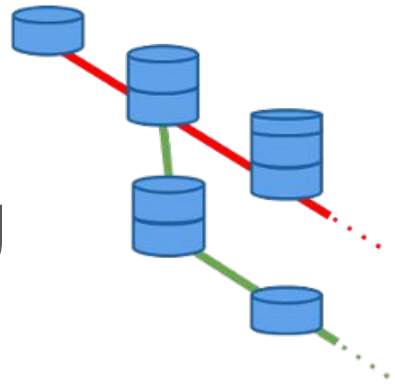


Pachyderm - The open source data
pipelining and data management layer for
Kubernetes.



Pachyderm - The open source data
pipelining and data management layer for
Kubernetes.

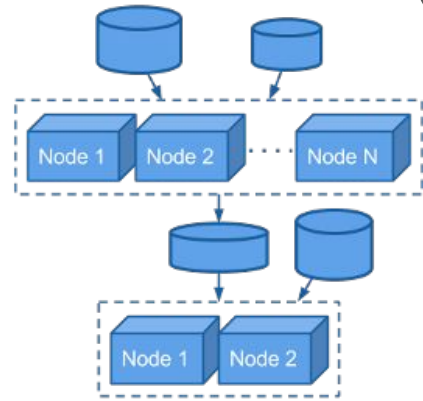
Data Versioning



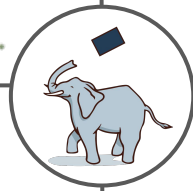
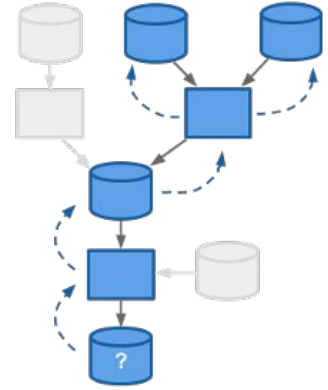
Containers for Analysis

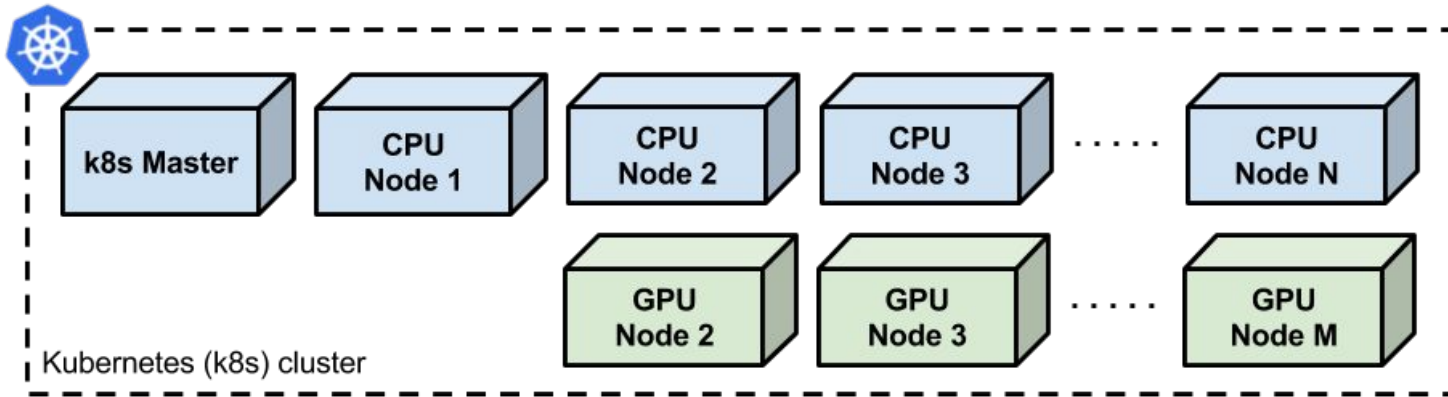
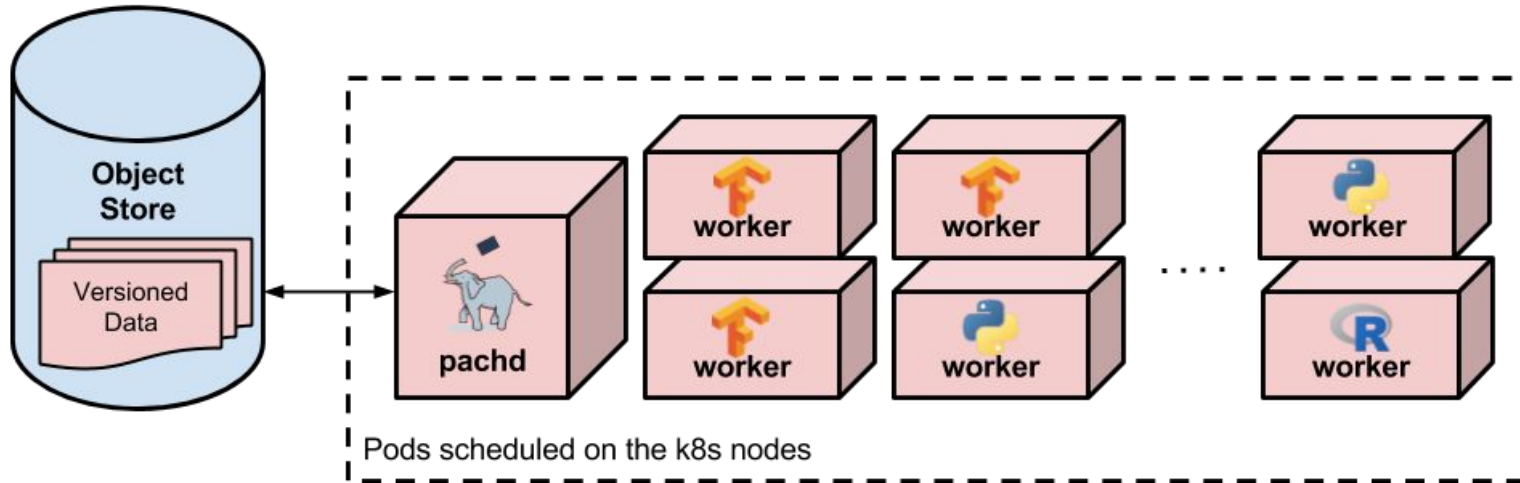


Distributed Pipelines



Data Provenance





Live Demo!

Questions/Resources

- Run [this and other ML examples](#)
- Check out the [k8s GPU docs](#)
- Join the [Pachyderm Slack channel](#)
- Check out the [Pachyderm docs](#)
- Slack/tweet [@dwhitena](#)