



KubeCon



CloudNativeCon

Europe 2018

Compliant Data Management & ML on k8s

Daniel Whitenack, @dwhitena



Outline



KubeCon



CloudNativeCon

Europe 2018

1. The what and why of compliance
2. Guidelines for and properties of compliant ML
3. A k8s-based solution
4. Demo!
5. Q&A



KubeCon



CloudNativeCon

Europe 2018

The what and why of compliance



The what and why of compliance



KubeCon



CloudNativeCon

Europe 2018



daniel whitenack

daniel whitenack **github**

daniel whitenack **pachyderm**

dr. daniel whitenack

Google Search

I'm Feeling Lucky

Report inappropriate predictions

The what and why of compliance



KubeCon



CloudNativeCon

Europe 2018

We need to explain (or make transparent):

- **Which data** was processed,
- **How** it was processed,
- **Who** processed the data, and
- What **corresponding data** was generated.



KubeCon



CloudNativeCon

Europe 2018

Guidelines for and properties of compliant ML



Guidelines for and properties of compliant ML



KubeCon



CloudNativeCon

Europe 2018

Production
ML/AI Model
Training

ML/AI Model
Inference or
Prediction

Guidelines for and properties of compliant ML



KubeCon



CloudNativeCon

Europe 2018

Data
Ingestion

Data
Cleaning

Feature
Engineering

Model
Selection,
Parameter
Search

Production
ML/AI Model
Training

ML/AI Model
Inference or
Prediction

Guidelines for and properties of compliant ML



KubeCon



CloudNativeCon

Europe 2018

Data
Ingestion

Data
Cleaning

Feature
Engineering

Model
Selection,
Parameter
Search

Feature
Transforms

Production
ML/AI Model
Training

Production
Model
Testing

Model
Export &
Optimization

ML/AI Model
Inference or
Prediction

Guidelines for and properties of compliant ML



KubeCon



CloudNativeCon

Europe 2018

Data
Ingestion

Data
Cleaning

Feature
Engineering

Model
Selection,
Parameter
Search

Feature
Transforms

Production
ML/AI Model
Training

Production
Model
Testing

Model
Export &
Optimization

Data Input
Transforms

ML/AI Model
Inference or
Prediction

Post
Processing

Guidelines for and properties of compliant ML

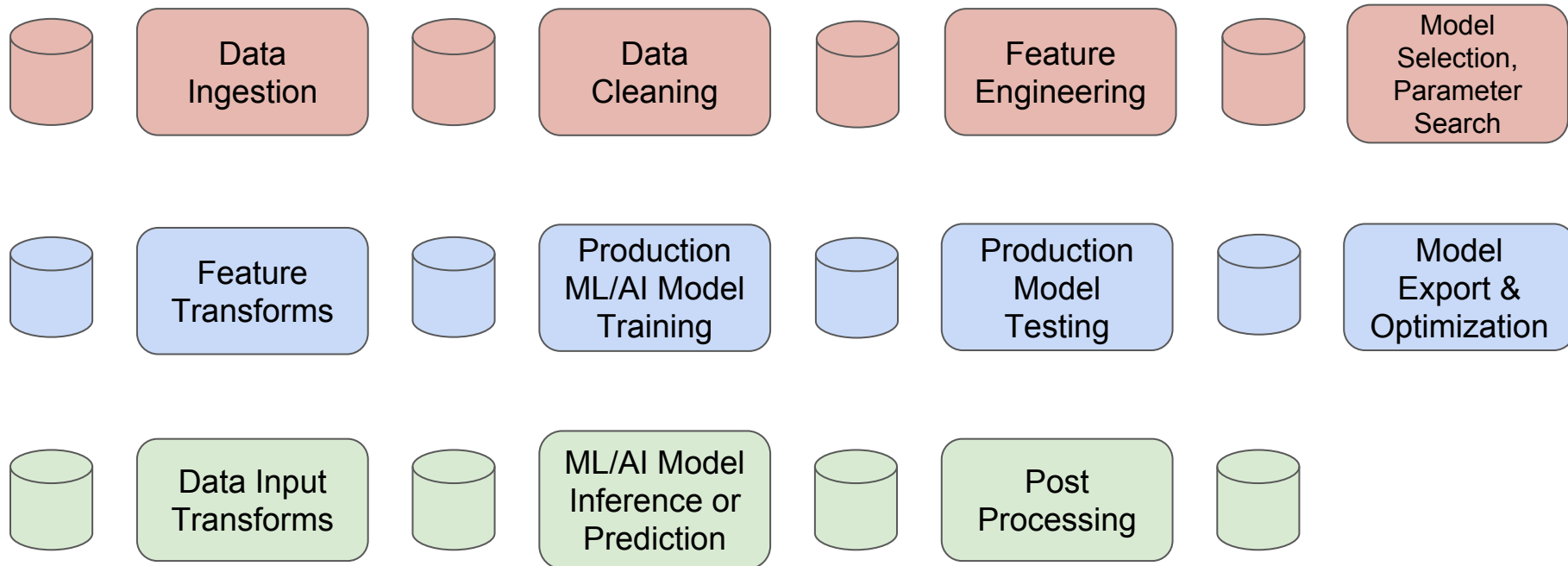


KubeCon



CloudNativeCon

Europe 2018



Guidelines for and properties of compliant ML

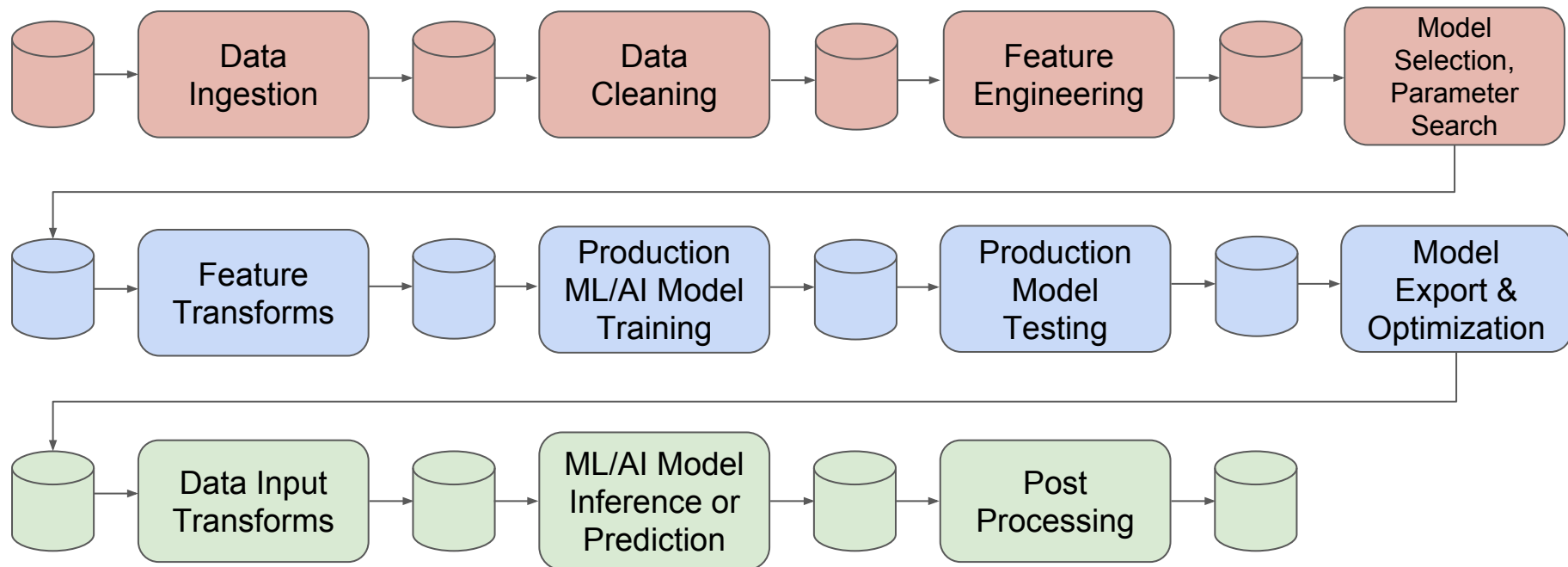


KubeCon



CloudNativeCon

Europe 2018



Guidelines for and properties of compliant ML



KubeCon



CloudNativeCon

Europe 2018

We need:

@dwhitena

#KubeCon + #CloudNativeCon

Guidelines for and properties of compliant ML



KubeCon



CloudNativeCon

Europe 2018

We need:

- Framework/infrastructure agnostic way to deploy/manage/define our workflow.

Guidelines for and properties of compliant ML



KubeCon



CloudNativeCon

Europe 2018

We need:

- Framework/infrastructure agnostic way to deploy/manage/define our workflow.
- Versioning system for all data, models, and processing in our workflow.

Guidelines for and properties of compliant ML



KubeCon



CloudNativeCon

Europe 2018

We need:

- Framework/infrastructure agnostic way to deploy/manage/define our workflow.
- Versioning system for all data, models, and processing in our workflow.
- Tools for combining the two features above.



KubeCon



CloudNativeCon

Europe 2018

A k8s-based solution



A k8s-based solution



KubeCon

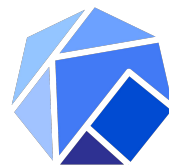


CloudNativeCon

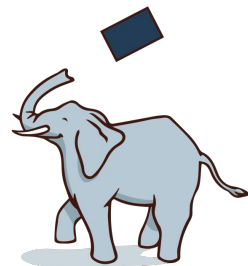
Europe 2018

We need:

- Framework/infrastructure agnostic way to deploy/manage/define our workflow.
- Versioning system for all data, models, and processing in our workflow.
- Tools for combining the two features above.



Kubeflow



Pachyderm

A k8s-based solution



KubeCon

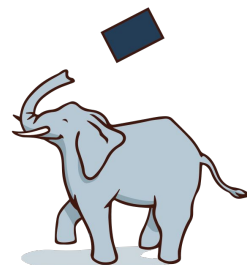


CloudNativeCon

Europe 2018

We need:

- Framework/infrastructure agnostic way to deploy/manage/define our workflow.
- Versioning system for all data, models, and processing in our workflow.
- Tools for combining the two features above.



Pachyderm



ROOK

A k8s-based solution



KubeCon

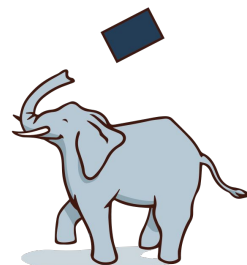


CloudNativeCon

Europe 2018

We need:

- Framework/infrastructure agnostic way to deploy/manage/define our workflow.
- Versioning system for all data, models, and processing in our workflow.
- Tools for combining the two features above.



Pachyderm

+

KVC

A k8s-based solution



KubeCon



CloudNativeCon

Europe 2018

KubeFlow (<https://github.com/kubeflow/kubeflow>):

- Standards and custom resources for deploying distributed ML tooling on k8s
- Compatibility with supplementary tools including notebooks, serving, and monitoring

A k8s-based solution



KubeCon



CloudNativeCon

Europe 2018

Pachyderm (<https://github.com/pachyderm/pachyderm>):

- Data agnostic data versioning (model versioning) built on top of object storage
- Language/framework agnostic pipelining on k8s
- Data provenance
- Access controls for data

A k8s-based solution



KubeCon



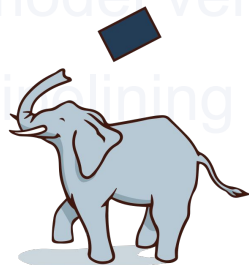
CloudNativeCon

Europe 2018

Announcing!



Kubeflow



Pachyderm

Deploy: <https://goo.gl/AiUjR8>, Proposal: <https://goo.gl/8ghTMH>

A k8s-based solution



KubeCon



CloudNativeCon

Europe 2018

KVC (<https://github.com/kubeflow/experimental-kvc>):

- Standards and tooling for exposing data to custom resources (e.g., TFJob)

A k8s-based solution



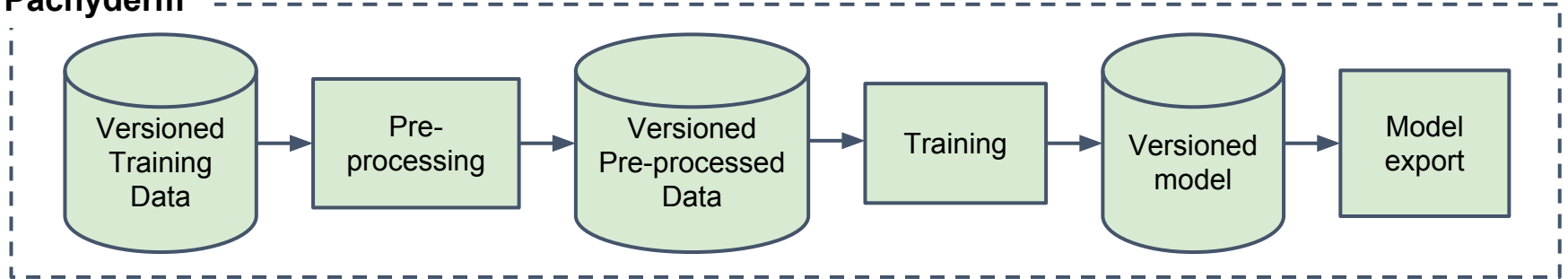
KubeCon



CloudNativeCon

Europe 2018

Pachyderm



@dwhitena

#KubeCon + #CloudNativeCon

A k8s-based solution

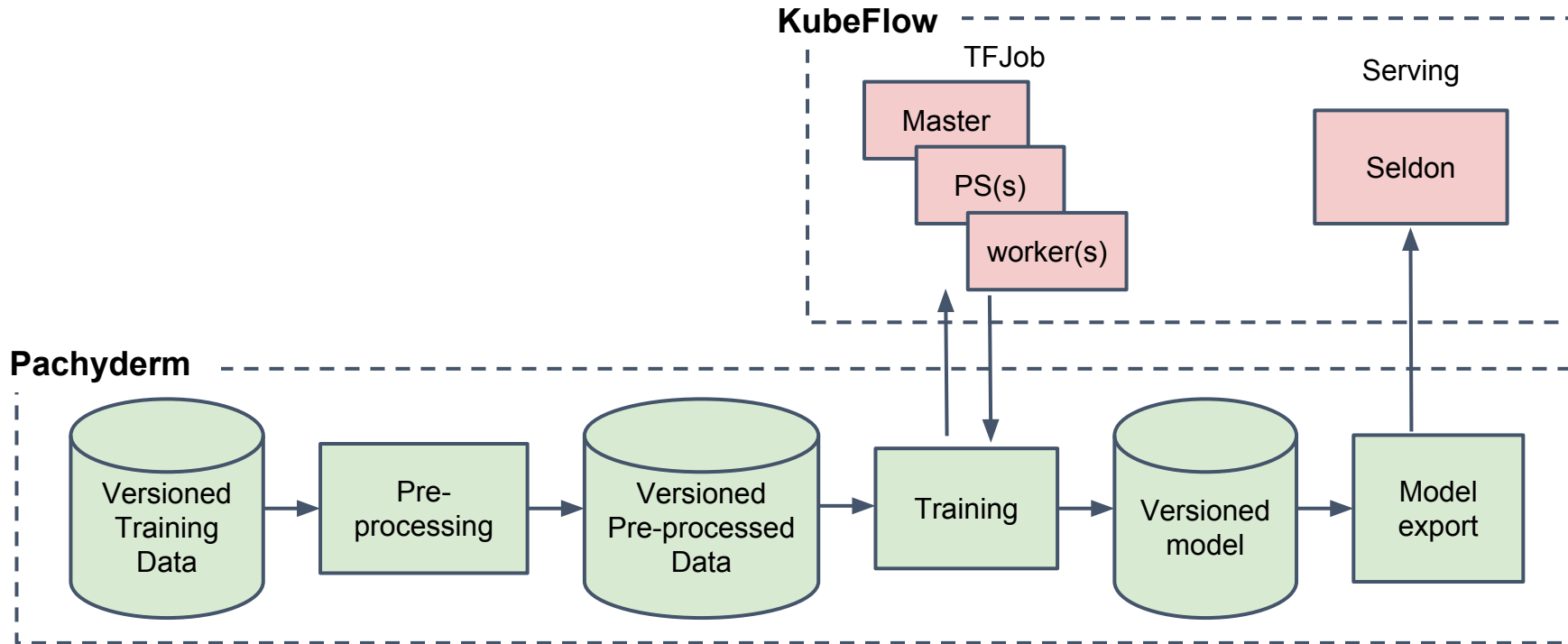


KubeCon



CloudNativeCon

Europe 2018



A k8s-based solution

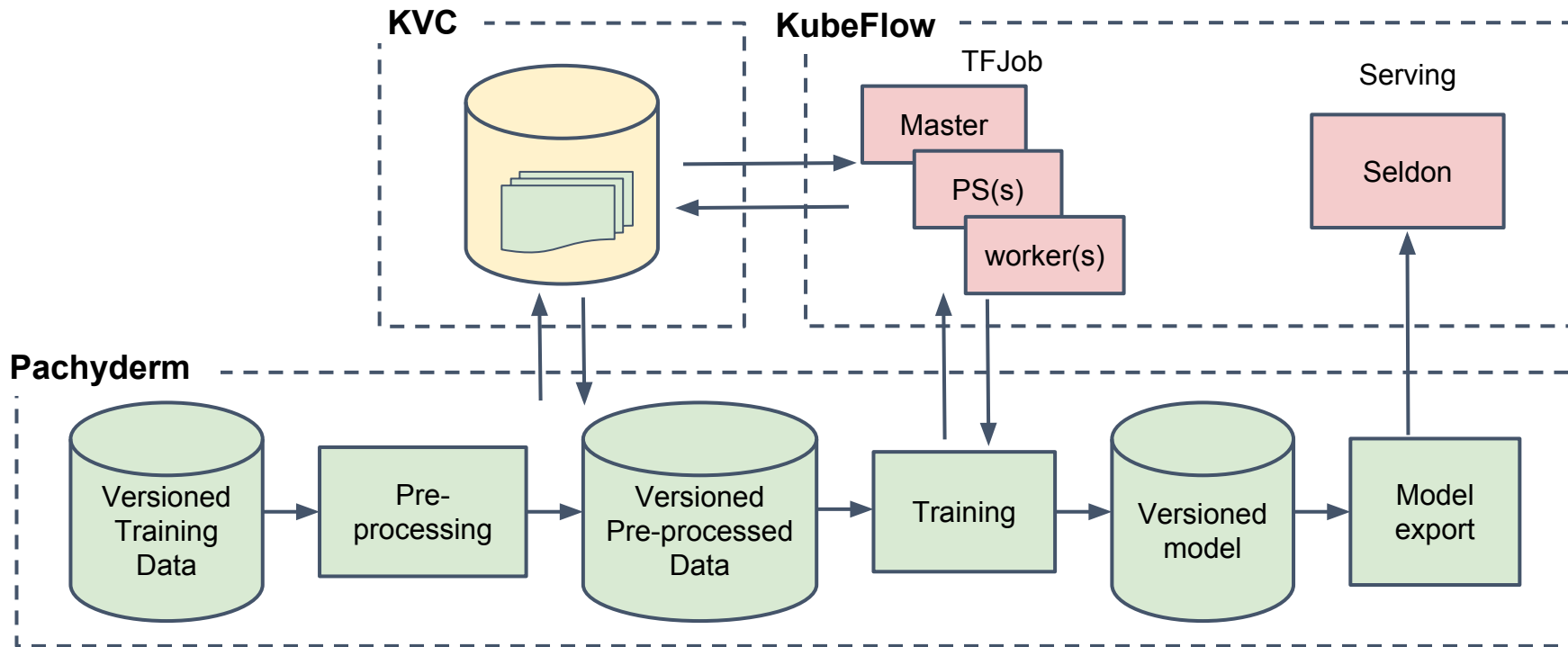


KubeCon



CloudNativeCon

Europe 2018



A k8s-based solution

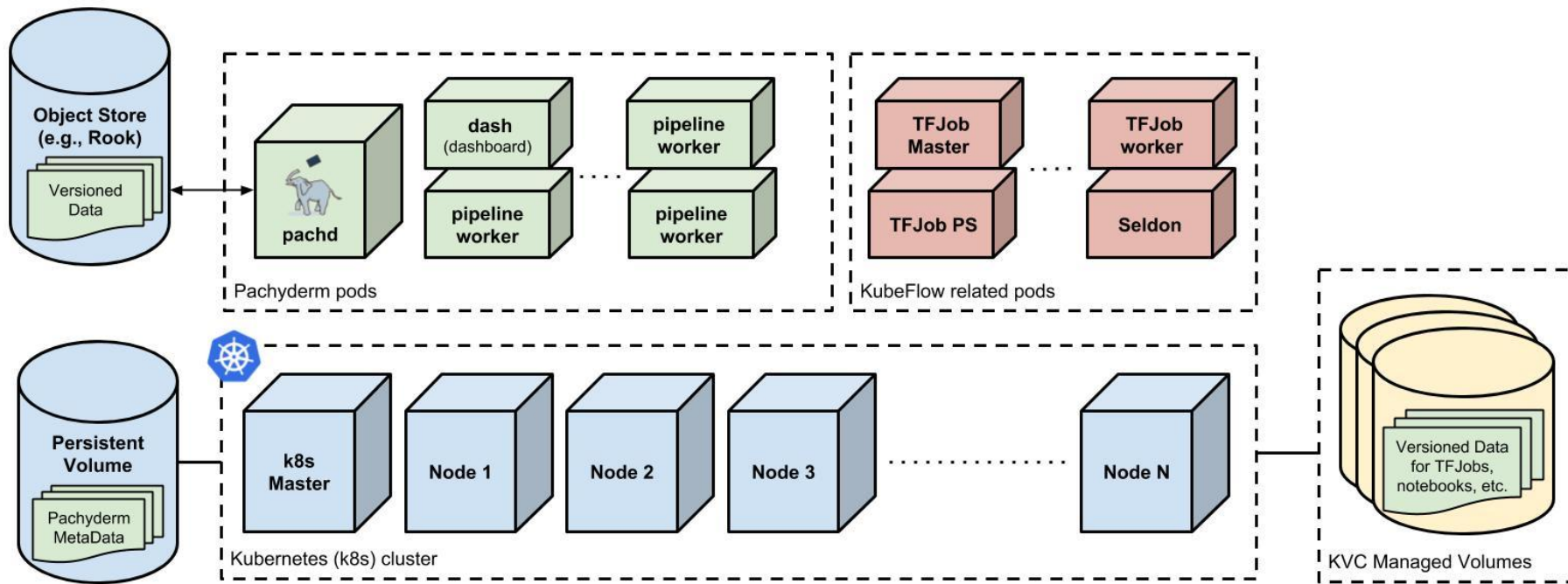


KubeCon



CloudNativeCon

Europe 2018





KubeCon



CloudNativeCon

Europe 2018

Demo!





KubeCon



CloudNativeCon

Europe 2018

Aside, Acknowledgements



Acknowledgements, this work



KubeCon

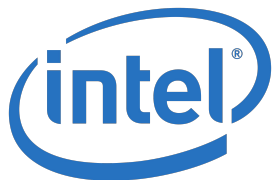


CloudNativeCon

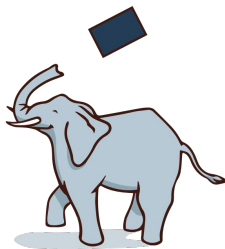
Europe 2018



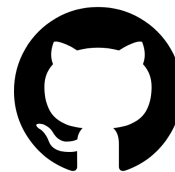
@jlewi
@aronchick



@Ajay191191
@balajismaniam



@dwhitena
@jdoliner



@hamelsmu



redhat

@pdmack

seldon[™]

@cliveseldon

@dwhitena

#KubeCon + #CloudNativeCon

Acknowledgements, Pachyderm ecosystem/integrations

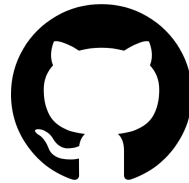
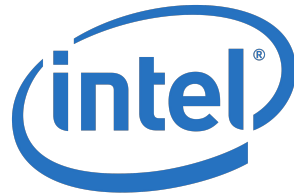


KubeCon



CloudNativeCon

Europe 2018



MINIO



Microsoft

Honeywell



KubeCon



CloudNativeCon

Europe 2018

Resources, Q&A



Resources



KubeCon



CloudNativeCon

Europe 2018

- Run this example: <https://goo.gl/vFp9JE>
- Pachyderm docs: <https://goo.gl/ezRzTd>
- Pachyderm Slack channel: <https://goo.gl/MBLUKf>
- KVC repo: <https://goo.gl/b4VnFL>
- KubeFlow docs/repo: <https://goo.gl/JMigjf>
- Slack/tweet @dwhitena



KubeCon



CloudNativeCon

Europe 2018

Appendix



```
pachrat@kubecon:~$
```

```
pachrat@kubecore:~$
```

PACH DASH



Home



Recent Changes



Repos



Pipelines

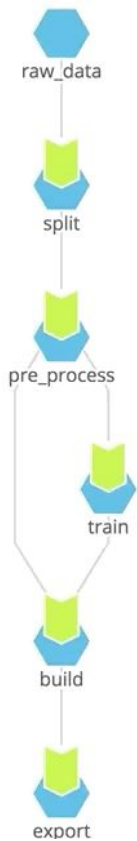


Jobs



Settings

Search Pachyderm



```
pachrat@kubecorn:~$ kubectl describe po/issue-summarization-issue-summarization-68f87d5dc9-hrfl4 --namespace=kubeflow |
grep "Image"
Image:          pachyderm/issuesummarization:f62fab38204f4767843694e944223ced
Image ID:       docker://sha256:953e85130ebc438ad1648617b76d554c7e34584fc1373fb960b0c4487388152d
Image:          seldonio/engine:0.1.5
Image ID:       docker://sha256:1730a73157a13a93dca9491a8e697cf551c34539b564d02067a9b8f07fa21a41
pachrat@kubecorn:~$
```

```
pachrat@kubecon:~$
```