KubeCon | CloudNativeCon

North America 2018

Eco-Friendly ML

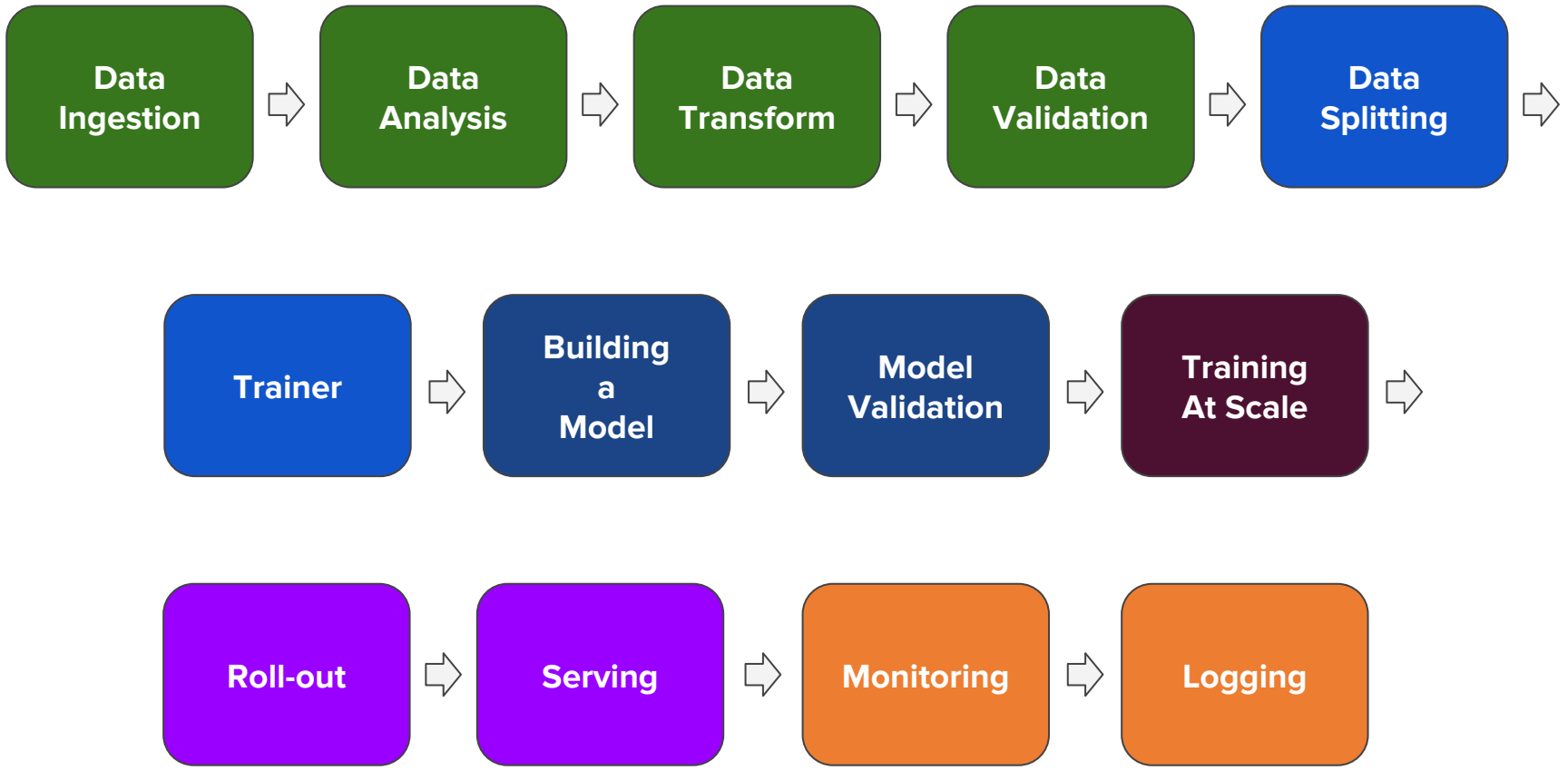
How the Kubeflow Ecosystem Bootstrapped Itself

Pete MacKinnon
Principal Software Engineer
Red Hat Inc., AI Center of Excellence
pmackinn@redhat.com

Agenda

- The ML DevOps problem space
- Project goals and aspirations
- Piecing it together
- Open Source ethos and communities
- Q&A

A ML pipeline



ML for everyone

- A lot of pieces needed to enable a comprehensive ML platform
- Difficult for one project or institution to solve all these challenges
- Start with a high-level mission statement
  - Portable: bare metal to cloud
  - Scalable: from 1 machine to 100s
  - Composable: microservice architecture
- Kubernetes is an obvious starting point
- What else is out there to support the project goals?

Deep thoughts

"If you want to go quickly, go alone. If you want to go far, go together."
– *African Proverb*

"Yeah, but what if we want to go quickly and far?"
– *Jeremy Lewi*
Kubeflow Lead Engineer

The Kubeflow mission

- Dedicated to making deployments of machine learning (ML) workflows on Kubernetes simple, portable, and scalable
- Goal is not to recreate other services, but to provide a straightforward way to deploy best-of-breed open source systems for ML to diverse infrastructures
- Anywhere you run Kubernetes, you should be able to run Kubeflow

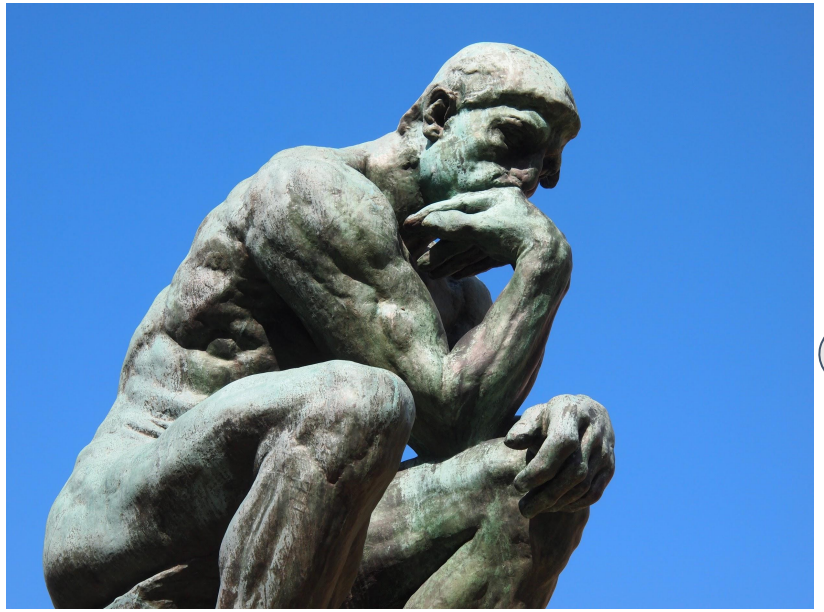
That’s all well and fine, but...

How do we run data science notebooks?

What about access to notebooks? Ingress control?

How will we do CI/CD?

Data pipelines and provenance?



Inference. That’s tricky...

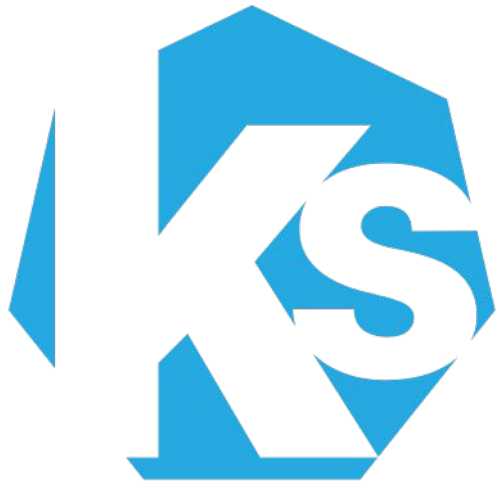
Deployment. Different environments....

Water is wet

- Where to start?
  - blogs, forums, and conferences
  - is it possible that other communities are working on your problem?
- And then?
  - participate in those communities
  - Slack, gitter.im, IRC, etc.
  - raise issues, offer pull requests
    - **get your git technique right**
    - GitHub makes it easy to cross-reference issues
  - treat their community with the respect and importance of your own
- And then?
  - need a point of integration, can’t just throw code over the wall
  - for Kubeflow, Ksonnet is that on-ramp...

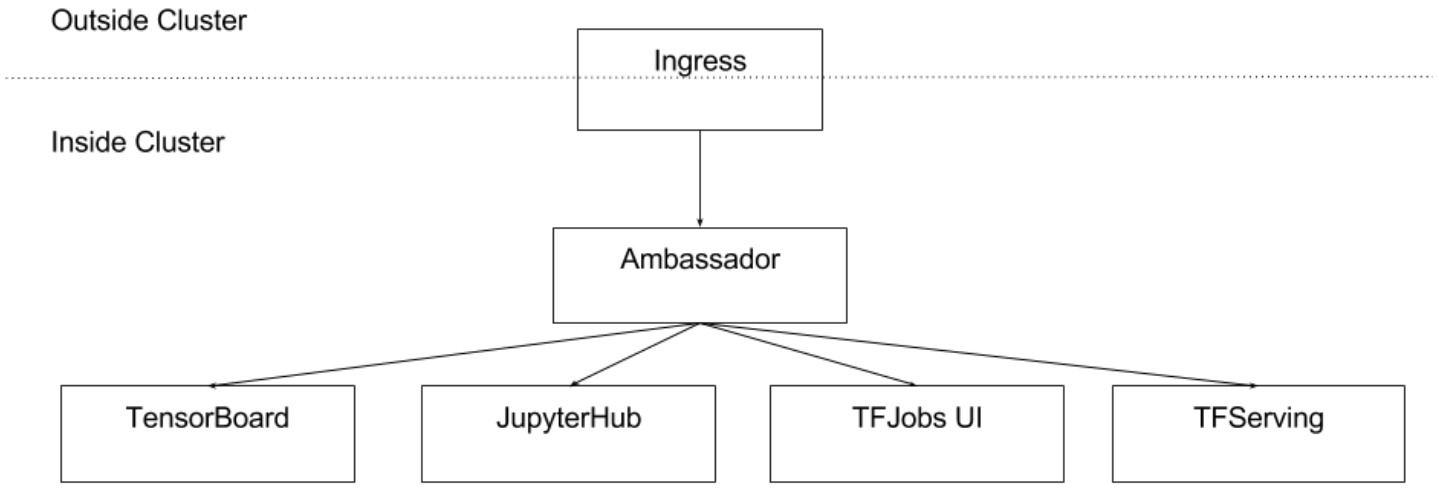
Ksonnet

- JSON-based CLI tool to generate Kubernetes deployment objects
- Based on the Jsonnet language
- More expressive than managing and patching YAML descriptors directly
- Generates resources that are entirely compatible with Kubernetes
  - you can still edit/patch the generated resources yourself as YAML/JSON
- Enables re-use of components for different envs
  - dev, test, staging, prod, etc.
- Parameter settings for different components
- Continuous deployment
  - just apply your changes and resources will be immediately updated
  - deletion of resources



Ambassador

- Cloud-native ingress controller (based on Envoy)
- Reverse proxy
  - enhancement over Node.js proxy in JupyterHub
- Annotation-based URL mapping to services
- IAP auth and can integrate with cert-manager



Argo

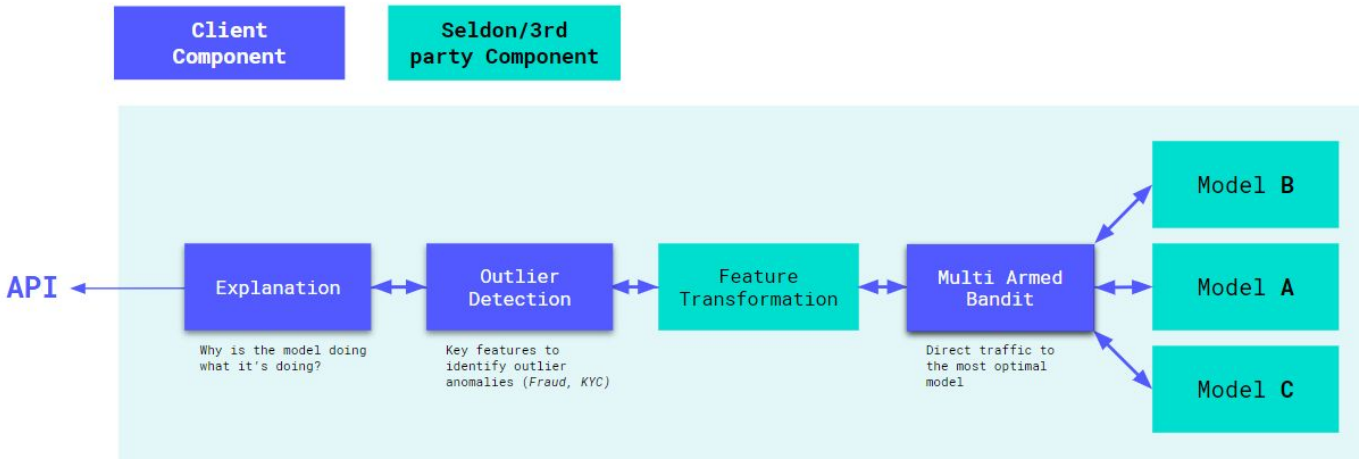
- Kubernetes native workflow manager (CRD and operator)
- Used by project in pre/post-submit testing in GCP, Kubebench, and Kubeflow Pipelines sub-project
- Define workflows where each step in the workflow is a container
- Model multi-step workflows as a sequence of tasks or capture the dependencies between tasks using a graph (DAG)
- Well suited to compute intensive jobs for machine learning or data processing on Kubernetes
- Argo is cloud agnostic and can run on any Kubernetes cluster

Argo - The Workflow Engine for Kubernetes

argo

Seldon

- Deploy and manage runtime inference graphs at scale
  - often different from training graphs
  - the graphs are deployed as microservices in Kubernetes
    - Model, Transformer, Router, Combiner, Output Transformer
    - A/B testing, multi-armed bandit
- TensorFlow, scikit-learn, Spark, H2O, R models
- Mix and match components for advanced graph deployments
- gRPC and REST interfaces
- Model wrappers for Python, R, Java and Node.js
  - uses OpenShift s2i for generating the wrappers



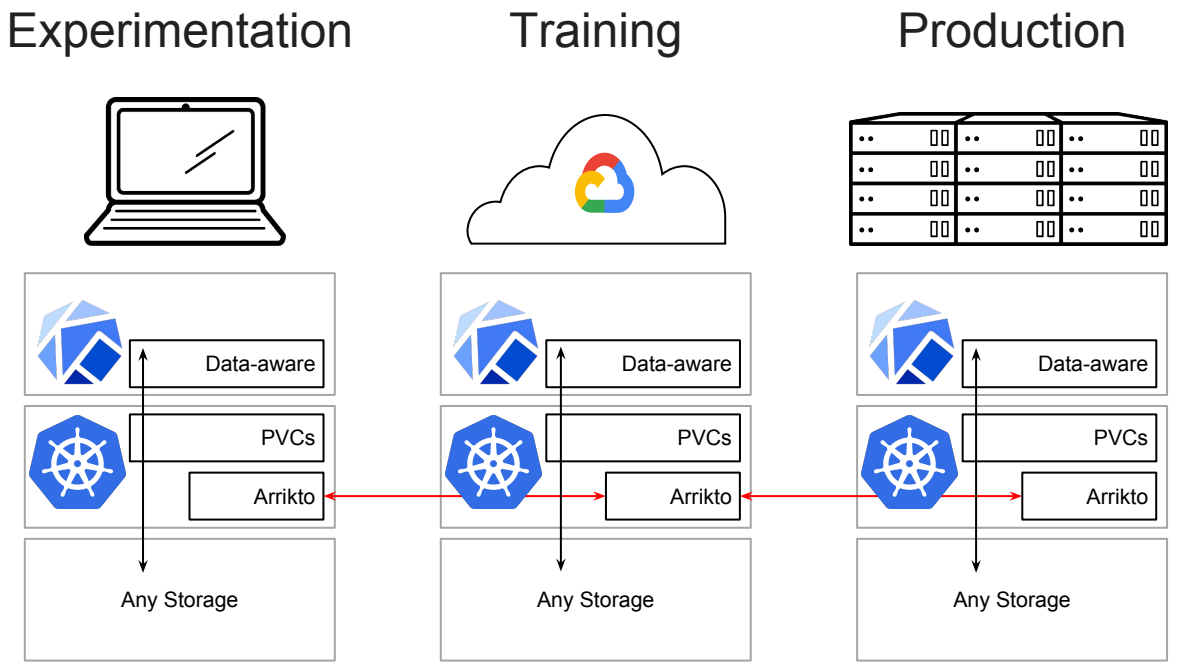
Pachyderm

- “git for data science”
- Reproducible experiments
- Data science pipelines
  - JSON pipeline specifications
    - transformation, parallelism
    - inputs: atom, cross, union, git
- CLI
  - create and manipulate pipelines
  - push and get your data in persistent storage
    - S3, GCS, or Azure
- Data provenance
  - SHA for inputs and outputs
  - deterministic results

Arrikto

- Data management integrated with Kubeflow
- Builds upon existing K8s mechanisms
- Uses PVC, PV, StorageClass and volumeClaimTemplates
- 3rd-party vendor integration
- Exposes data management in UIs
- Contributing back to community
  a new, enhanced Jupyter UI
  to expose PVC in 0.4

Arrikto



Notebooks

- JupyterHub project
  - provides the pod spawner for Kubeflow
  - ability to customize; convos around features v. customization
- Kubeflow curates TensorFlow CPU and GPU notebooks
  - based on the default builds from the TF project
- Kaggle project
  - participated in the review of an adaption of the Kaggle Python notebook for Kubeflow
- RAPIDS AI project at NVIDIA
  - guiding development of a RAPIDS AI notebook for Kubeflow
  - engineers from NVIDIA and Kubeflow collaboration on debugging in Kubernetes cloud environments

Kubeflow project governance

- Code of conduct
  - contributors and maintainers pledge to make community participation a harassment-free experience for everyone
- Standards create a healthy and positive community
  - welcoming and inclusive language
  - respectful of differing viewpoints and experiences
  - gracefully give and accept constructive criticism
  - focus on what is best for the community
  - empathy towards other community members
- Conflict resolution protocols
- We expect and encourage the same of our ecosystem communities

The community today(-ish)

- GitHub
- Apache 2.0 license
- 23 repositories
- 135 contributors
- 909 commits in core repo
- 0.3.4 release
  (heading for 0.4.0)
- New integrations
  happening continuously
- Institutional participation

And many, many more….

The power of Open Source

- Kubeflow is the sum of its parts (and people)
  - it is young but so is its ecosystem
  - much more work to be done
- Everyone wins when communities collaborate
  - free exchange of concepts and code
  - help to promote and recognize each other's achievements
  - multiplier effect for each community's user base
- Contributors feel a sense of empowerment
  - institutions and individuals alike reap the benefits in work satisfaction

Q&A

Thank you

Kubeflow BoF today!

https://github.com/kubeflow/kubeflow
kubeflow-discuss@googlegroups.com
Weekly meeting: https://zoom.us/j/799749911


KubeCon
CloudNativeCon
North America 2018