Adapting Kubernetes For Machine Learning Workflows

Open FinTech Forum 2018
October 10, 2018

Ania Musial, Senior Software Engineer
Keith Laban, Senior Infrastructure Engineer

TechAtBloomberg.com
About Us

• Ania Musial
  — Senior Software Engineer
  — Machine Learning Platform, AI Group

• Keith Laban
  — Senior Infrastructure Engineer
  — Data & Analytics Infrastructure

• https://www.techatbloomberg.com/post-topic/data-science/
Outline

- Machine Learning at Bloomberg
- Toward a Model Development Life Cycle
- Machine Learning the Kubernetes Way
- Concluding Remarks
Machine Learning at Bloomberg
Scale and Performance: Market Data

100 Billion Messages per day

2008 Financial Crisis
Flash Crash
US Debt Ceiling Crisis/ S&P Downgrade
Japan Earthquake & Tsunami
Stimulus Tapering Discussion
Natural Language Processing & Events

First Bloomberg Headline

New York Times Story

SEC Announcement

12%

<20 minutes
Over $136 Billion Wiped Out in Minutes
Applications Contingent on ML

SoFi Raises $500 Million Led by Silver Lake for Global Expansion

By Selina Wang

(Bloomberg) -- Social Finance Inc. said it raised $500 million in a financing round led by private equity firm Silver Lake Partners. It plans to use the funds for international expansion of its online lending business and development of new financial products.

Other investors in the round include SoftBank Group Corp. and GPI Capital, and the funding brings SoFi's total investment to $1.9 billion, the company said. Several sovereign wealth funds from countries in Asia and Europe invested as well, said Mike Cagney, SoFi's chief executive officer and co-founder. The international group will purchase SoFi's loans in addition to taking an equity stake, he said.

“Our challenge has always been our reliance on wholesale capital markets,” Cagney said. “Silver Lake is not a balance sheet partner, but their investor base is hugely important to us, and there’s a big overlap between their limited partners and our loan buying universe.”

TechAtBloomberg.com
Toward a Model Development Life Cycle
Different Roles, Different Priorities

**Customers**
- Smarter Functionality
- Privacy
- Correctness
- Uptime/Stability

**ML Practitioners**
- Keep up with latest trends in academia and forefront of Deep Learning innovations
- Access to specialized hardware
- Rapid iteration over research ideas
- Focus on business problem, not building infrastructure

**Engineers**
- Less operational burden
- Support for multi-tenancy
- Easy data access but also secure and properly privileged
- Avoid duplicated code
- Reproducibility of results

TechAtBloomberg.com
Embarking on a New Problem: What?

• What **data** is relevant to my business problem?
• What does the data **look** like?
• What **features** should I create from my data?
• What metric should I use to **measure** my model performance?
• What **algorithms** or **parameters** should I test out?
• How is my model doing on live data in **production**?
Embarking on a New Problem: How?

• Where is my data located?
• How do I get permissioned for my data?
• How do I load this data in a native visualization tool?
• How do I get this toolkit installed and operational?
• How do I schedule my long-running job?
• How much compute power is available for me to exploit?
• How do I deploy my model?
The Infinite Model Development Loop

Train
- Data
- Task
- Evaluate
- Train

Predict
- Publish
- Release
- Monitor
Achieving Nirvana

Foundational Building Blocks
• Specialized Hardware
• Standardized Runtimes
• Batch Scheduling

Data Liquidity
• Data Access
• Security

Developer Experience
• Ease of Use
• Rapid Iteration
• Native Tooling

ML Specifics
• Experiment Management
• Hyperparameter Tuning
• Result Reproducibility
Machine Learning the Kubernetes Way
What is Kubernetes?

“Kubernetes is an open-source system for automating deployment, scaling, and management of containerized applications.”
Kubernetes Life Cycle

Watch Resource Changes

React to Updates
Kubernetes Life Cycle II

Pod → Pod Resource → Kube API Server → Controller → node

Pod Resource

Schedule Pod
Deployment Resource (YAML)

```yaml
apiVersion: apps/v1
kind: Deployment
metadata:
  name: nginx-deployment
spec:
  replicas: 3
  template:
    spec:
      containers:
        - name: nginx
          image: nginx:1.7.9
          ports:
            - containerPort: 80
```

https://kubernetes.io/docs/concepts/workloads/controllers/deployment/
Deployment Resource: Kind

```yaml
apiVersion: apps/v1
kind: Deployment
metadata:
  name: nginx-deployment
spec:
  replicas: 3
  template:
    spec:
      containers:
        - name: nginx
          image: nginx:1.7.9
          ports:
            - containerPort: 80
```

https://kubernetes.io/docs/concepts/workloads/controllers/deployment/
Deployment Resource: Spec (Replicas)

```yaml
apiVersion: apps/v1
kind: Deployment
metadata:
  name: nginx-deployment
spec:
  replicas: 3
  template:
    spec:
      containers:
        - name: nginx
          image: nginx:1.7.9
          ports:
            - containerPort: 80

https://kubernetes.io/docs/concepts/workloads/controllers/deployment/
```
Deployment Resource: Spec (Pod Template)

```yaml
apiVersion: apps/v1
kind: Deployment
metadata:
  name: nginx-deployment
spec:
  replicas: 3
  template:
    spec:
      containers:
        - name: nginx
          image: nginx:1.7.9
          ports:
            - containerPort: 80
```

[https://kubernetes.io/docs/concepts/workloads/controllers/deployment/](https://kubernetes.io/docs/concepts/workloads/controllers/deployment/)
Benefits of Kubernetes

- Declarative deployment
- Hardware indifference
- Monitoring
- Replication
- Autoscaling
- Security Extensions
  - Ingress + SSL termination
  - Istio service mesh
Custom Resources

• **Resource** - a set of API objects in Kubernetes like Deployments or Pods.

• **Custom Resource** - an extension of the Kubernetes API defined by a user that declares new types of API Objects.
Kubernetes Meets ML: TensorFlowJob

```yaml
apiVersion: ds.bloomberg.com/v1
kind: TensorFlowJob
metadata:
  name: tf-test
  annotations:
    ai.bloomberg.com/project: foo
    ai.bloomberg.com/experiment: abcde
spec:
  framework: tensorflow-1.7-python-3
  identities:
    - hadoop:
      id: keithlaban
  pipPackages: [ai.bloomberg.com.myteam.gpu_tftraining, numpy]
  module: gpu_tftraining
  size: GpuLarge
  args:
    --data-dir hdfs://CLUSTER/projects/news/news_index
    --output-dir hdfs://CLUSTER/users/klaban1/abcde/1
```
Kubernetes Meets ML: TensorFlowJob

apiVersion: ds.bloomberg.com/v1
kind: TensorFlowJob

metadata:
  name: tf-test
  annotations:
    ai.bloomberg.com/project: foo
    ai.bloomberg.com/experiment: abcde

spec:
  framework: tensorflow-1.7-python-3
  identities:
    - hadoop:
        id: keithlaban
  pipPackages: [ai.bloomberg.com.myteam.gpu_tftraining, numpy]
  module: gpu_tftraining
  size: GpuLarge
  args:
    --data-dir hdfs://CLUSTER/projects/news/news_index
    --output-dir hdfs://CLUSTER/users/klaban1/abcde/1
Kubernetes Meets ML: TensorFlowJob

```yaml
apiVersion: ds.bloomberg.com/v1
kind: TensorFlowJob
metadata:
  name: tf-test
  annotations:
    ai.bloomberg.com/project: foo
    ai.bloomberg.com/experiment: abcde
spec:
  framework: tensorflow-1.7-python-3
  identities:
    - hadoop:
        id: keithlaban
  pipPackages: [ai.bloomberg.com.myteam.gpu_tftraining, numpy]
  module: gpu_tftraining
  size: GpuLarge
  args:
    --data-dir hdfs://CLUSTER/projects/news/news_index
    --output-dir hdfs://CLUSTER/users/klaban1/abcde/1
```
Kubernetes Meets ML: TensorFlowJob

```yaml
apiVersion: ds.bloomberg.com/v1
kind: TensorFlowJob
metadata:
  name: tf-test
  annotations:
    ai.bloomberg.com/project: foo
    ai.bloomberg.com/experiment: abcde
spec:
  framework: tensorflow-1.7-python-3
  identities:
    - hadoop:
      id: keithlaban
  pipPackages: [ai.bloomberg.com.myteam.gpu_tftraining, numpy]
  module: gpu_tftraining
  size: GpuLarge
  args:
    --data-dir hdfs://CLUSTER/projects/news/news_index
    --output-dir hdfs://CLUSTER/users/klaban1/abcde/1
```
Kubernetes Meets ML: TensorFlowJob

apiVersion: ds.bloomberg.com/v1
kind: TensorFlowJob
metadata:
  name: tf-test
  annotations:
    ai.bloomberg.com/project: foo
    ai.bloomberg.com/experiment: abcde
spec:
  framework: tensorflow-1.7-python-3
  identities:
    - hadoop:
        id: keithlaban
  pipPackages: [ai.bloomberg.com.myteam.gpu_tftrining, numpy]
  module: gpu_tftraining
  size: GpuLarge
  args:
    --data-dir hdfs://CLUSTER/projects/news/news_index
    --output-dir hdfs://CLUSTER/users/klaban1/abcde/1
Kubernetes Meets ML: TensorFlowJob

apiVersion: ds.bloomberg.com/v1
kind: TensorFlowJob
metadata:
  name: tf-test
  annotations:
    ai.bloomberg.com/project: foo
    ai.bloomberg.com/experiment: abcde
spec:
  framework: tensorflow-1.7-python-3
  identities:
    - hadoop:
      id: keithlaban
  pipPackages: [ai.bloomberg.com.myteam.gpu_tftraining, numpy]
  module: gpu_tftraining
  size: GpuLarge
  args:
    --data-dir hdfs://CLUSTER/projects/news/news_index
    --output-dir hdfs://CLUSTER/users/klaban1/abcde/1
Kubernetes Meets ML: TensorFlowJob

```yaml
apiVersion: ds.bloomberg.com/v1
kind: TensorFlowJob
metadata:
  name: tf-test
  annotations:
    ai.bloomberg.com/project: foo
    ai.bloomberg.com/experiment: abcde
spec:
  framework: tensorflow-1.7-python-3
  identities:
    - hadoop:
        id: keithlaban
  pipPackages: [ai.bloomberg.com.myteam.gpu_tftraining, numpy]
  module: gpu_tftraining
  size: GpuLarge
  args:
    --data-dir hdfs://CLUSTER/projects/news/news_index
    --output-dir hdfs://CLUSTER/users/klaban1/abcde/1
```
Kubernetes Meets ML: TensorFlowJob

apiVersion: ds.bloomberg.com/v1
kind: TensorFlowJob
metadata:
  name: tf-test
  annotations:
    ai.bloomberg.com/project: foo
    ai.bloomberg.com/experiment: abcde
spec:
  framework: tensorflow-1.7-python-3
  identities:
    - hadoop:
        id: keithlaban
  pipPackages: [ai.bloomberg.com.myteam.gpu_tftraining, numpy]
  module: gpu_tftraining
  size: GpuLarge
  args:
    --data-dir hdfs://CLUSTER/projects/news/news_index
    --output-dir hdfs://CLUSTER/users/klaban1/abcde/1
Achieving Nirvana (Reprise)

Foundational Building Blocks
- Specialized Hardware
- Standardized Runtimes
- Batch Scheduling

Developer Experience
- Ease of Use
- Rapid Iteration
- Native Tooling

Data Liquidity
- Data Access
- Security

ML Specifics
- Experiment Management
- Hyperparameter Tuning
- Result Reproducibility
A Model for the ML Ecosystem

• Other Types of Custom Resources
  — Spark
  — Python
  — JVM
  — Jupyter
  — Hyperparameter Tuning

• Foundation for reproducibility and automation
Concluding Remarks
In Conclusion

• Opportunity for Machine Learning
• Allow ML Practitioners to focus on ML
• Kubernetes is strong in the community!
Thanks!

Panel Discussion: Real-World Kubernetes Use Cases in Financial Services: Lessons Learned from Capital One, BlackRock and Bloomberg - Moderated by Ron Miller, TechCrunch

Thursday @ 4:25 PM

http://sched.co/G4Mx