Using Kubeflow to Make the Electricity Smarter in China

Julia Han & Xin Zhang, Caicloud
The Velocity of Digital Transformation

Market Competition

UnionPay  Starbucks  Suning.com  Yum
From Digital to Intelligent Transformation

AI and ML spending: from $12 billion in 2017 to $57.6 billion in 2021

Near 0.8 million GPUs in datacenters; # ML activities is doubling each year

61% interviewees plan to use ML in 2019; 58% already have ML in use

70% will adopt AI in 2030

GPU market in China is booming: 230% increase, to 3.5 billion RMB in 2017
Smart Grid for the State Grid
Scenarios – Intelligent Patrol System
Scenarios – Intelligent Usage Prediction
Scenarios – Various Predictions

Exhaust gas temperature of thermal power station

Water level of hydropower station
Scenarios – Various Object Detection

Equipment model

Equipment maintenance operation
Scenarios – Financial Operations
Intelligent Transformation is Hard

“Can I tweak the bought models and APIs?”

“I don’t want to give out my sensitive data and business ideas”
Intelligent Transformation is Hard

“Which framework, models, hyperparameters to use?”
Intelligent Transformation is Hard

“How to speed up my training using really deep network against really huge amount of data?”
### Intelligent Transformation is Hard

**“How to allocate our 400 GPUs to 20 model development?”**

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.30.1.0</td>
<td>1080Ti</td>
<td>高精模块</td>
<td>任务开始时间 2018/04/16</td>
<td>预计完成时间 2018/04/18</td>
<td>否</td>
<td></td>
</tr>
<tr>
<td>10.30.1.0</td>
<td>1080Ti</td>
<td>高精模块</td>
<td>任务开始时间 2018/04/18</td>
<td>预计完成时间 2018/04/18</td>
<td>否</td>
<td></td>
</tr>
<tr>
<td>10.30.1.1</td>
<td>1080Ti</td>
<td>高精模块</td>
<td>任务开始时间 2018/04/18</td>
<td>预计完成时间 2018/04/18</td>
<td>否</td>
<td></td>
</tr>
<tr>
<td>10.30.1.2</td>
<td>1080Ti</td>
<td>高精模块</td>
<td>任务开始时间 2018/04/18</td>
<td>预计完成时间 2018/04/18</td>
<td>否</td>
<td></td>
</tr>
<tr>
<td>10.30.1.3</td>
<td>1080Ti</td>
<td>高精模块</td>
<td>任务开始时间 2018/04/18</td>
<td>预计完成时间 2018/04/18</td>
<td>否</td>
<td></td>
</tr>
</tbody>
</table>
It is Hard by Nature

From Enterprise Almanac 2018 by Work-Bench
Kubeflow as a Great Start

Data Collection
Data Verification
Machine Resource Management
Feature Extraction
Analysis Tools
Process Management Tools
Configuration
Monitoring
Serving Infrastructure
ML Code
What Else Do ML Engineers Need?

- Data! x N
  - Data collection and storage
  - Data analysis and visualization
  - Data transformation
  - Data cleaning
  - Data validation
  - Data management

Sad example
What Else Do ML Engineers Need?

- Model Selection and Evaluation
  - Model selection
    - from linear to non-linear to deep learning… A LOT
  - Choose appropriate hyper-parameters
    - Learning rate, network layers, normalization, … A LOT
  - Model evaluation
    - First look at the curve
    - Tune the hyper parameters?
    - Need more data?
    - Chose the wrong model?
What Else Do ML Engineers Need?

• Collaborative development

I am stuck! Want a second opinion and help!

Alice: Can you help me on this problem?
Bob (delightedly): Where is your training code?
Alice: *In my laptop*
Bob: Where are your model results?
Alice: *In my laptop*
Bob: Where are your pre-processing scripts?
Alice: *In my laptop*
Bob: Where is your data so I can test-run?
Alice: *In my laptop*
Bob: Copy everything to me … doesn’t work at all
What Else Do ML Engineers Need?

- Go-to-Production
  - Model to API
  - Requires collaboration: API design, deployment testing, engineering

Tony

I know Algorithm1, Algorithm2, Algorithm3...

Dereck

I know docker, Kubernetes... Wait a minute, which service broke again?
What Else Do ML Engineers Need?

• Reproducibility

Tony

My model is 99% accuracy; sending to you via WeChat and you take care of the rest!

Dereck

WeChat transfer is slow today; will run it once received.

Why only 97% in my workstation?

Only 93% at the customer’s!
What Else Do ML Engineers Need?

• Devops for ML (CI/CD)
  • Even worse, frequent re-train is needed when we have more new data.
What Else Do ML Engineers Need?

- **System and Model OPs**
  - Monitor model effectiveness
  - Monitor data anomaly
  - Monitor system resource usage and anomaly

![We Need This and This](image)
End-to-End AI OS for the Grid Industry

Admin
Multi-tenancy, Role-based access control, resource allocation and auditing

Data Scientists
Data pre-processing, feature engineering, monitoring, model evaluation

Data Engineers
Model packaging and versioning, model deployment, model as a service
End-to-End AI OS for the Grid Industry

1. Data and resource management
2. Streamlined and interactive devops pipeline
3. Distributed and orchestrated training
4. Integrated “model to business” loop
End-to-End AI OS for the Grid Industry
End-to-End AI OS for the Grid Industry

Leveraging the unique data in the Grid, achieve AI-facilitated business innovation

<table>
<thead>
<tr>
<th>AI App</th>
<th>AI OS</th>
<th>Data lake</th>
</tr>
</thead>
<tbody>
<tr>
<td>Financial budget assisted decision</td>
<td>inferences</td>
<td>Financial data</td>
</tr>
<tr>
<td>Automated reimbursement</td>
<td>Batch training</td>
<td>Equipment operation and maintenance data</td>
</tr>
<tr>
<td>Substation security</td>
<td>Model repo</td>
<td>Substation inspection video</td>
</tr>
<tr>
<td>Abnormal electricity detection</td>
<td>AI Devops Pipeline</td>
<td>Sensor data</td>
</tr>
<tr>
<td>Customer service robot</td>
<td>Power station power generation analysis</td>
<td>Service data</td>
</tr>
<tr>
<td>Intelligent inspection</td>
<td>Hyper-parameter tuning</td>
<td>Usage data</td>
</tr>
</tbody>
</table>
End-to-End AI OS for the Grid Industry

Demo Time
# Cloud Native Intelligence

## App Store
- Face Recognition
- Object Detection
- Smart Bank
- Emotion Detection
- OCR

## Framework
- Hadoop
- Spark
- Jupyter
- TensorFlow
- Caffe
- MxNet

## Backend
- Kubeflow Admin
- Storage Manager
- Data Manager
- Clever SDK

## Resource Management
- CPU
- Memory
- GPU
- Storage

## Cross-cluster Authentication
- Application Orchestration
- Intelligent Operation
- Cyclone DevOps

## Load Balancer
- Automatically Scaling

## Infra
- Bare Metal
- Microsoft Azure
- Huawei
- Alibaba Cloud
- VMware