KubeCon
CloudNativeCon
North America 2019

Supercharge Kubeflow Performance on GPU Clusters

Meenakshi Kaushik (Cisco)
Neelima Mukiri (Cisco)

Agenda

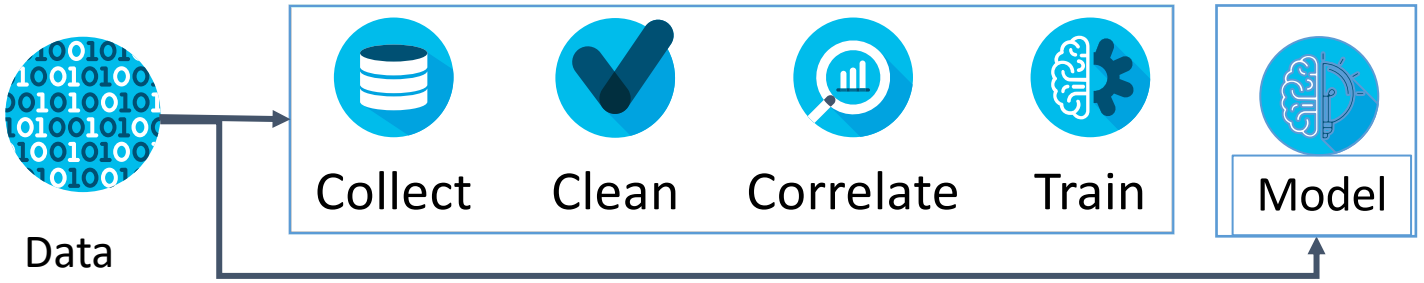
- Overview
- Introduction to KubeFlow
- Distributed Training
- Hyperparameter Tuning
- Infrastructure and OS Optimization
- Summary

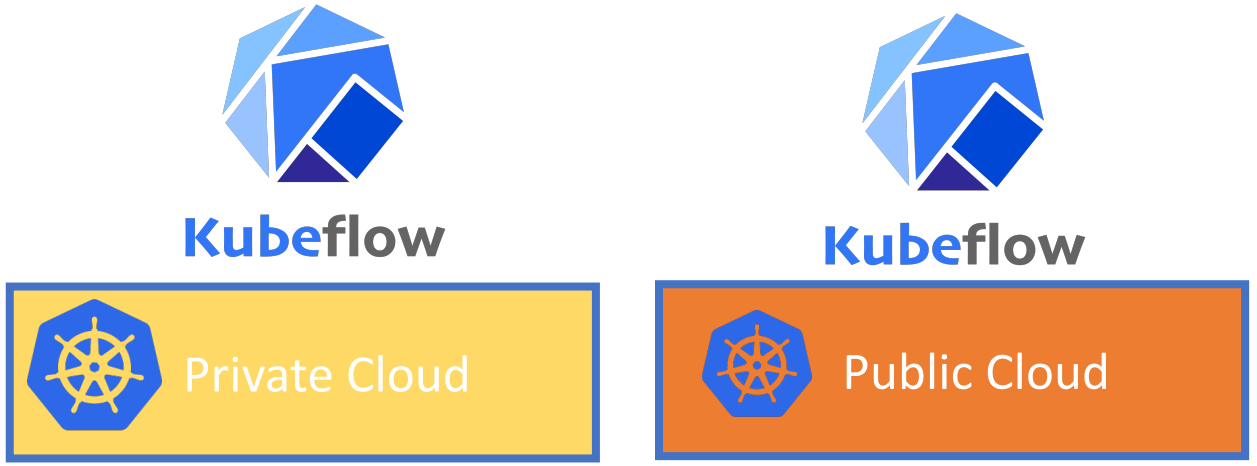
What is the Opportunity?

EASY
Distributed
ML

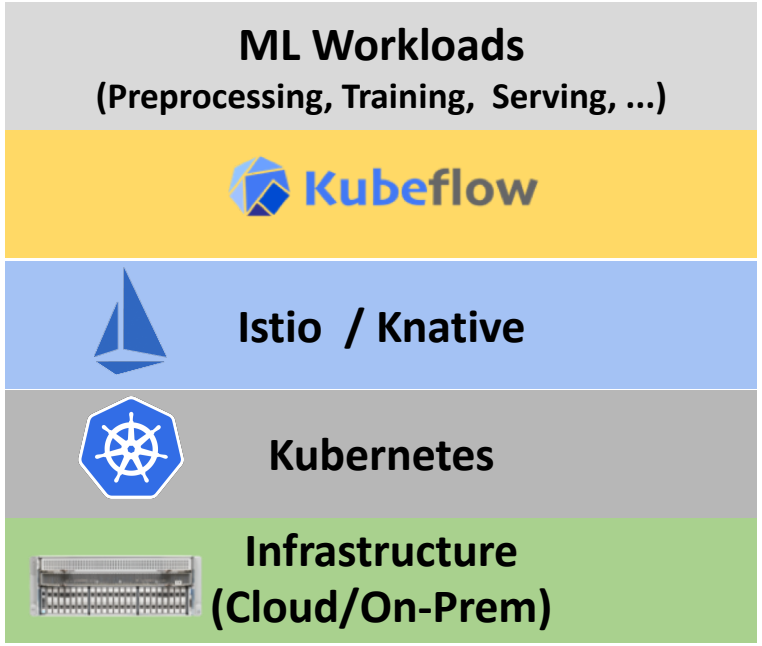
Kubeflow

Machine learning lifecycle manager that makes it easy to develop, deploy and manage portable, scalable end-to-end ML workflows everywhere





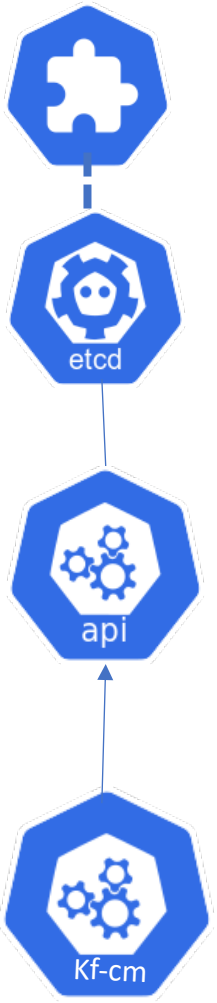
Kubeflow



Kubeflow

- **Data exploration** using Jupyter notebooks
- **Model training** using different machine learning frameworks such as TensorFlow, PyTorch, Spark, XGBoost, MPIJob, …
- **Model serving** KFServing, TFServing, Seldon, …
- **Hyperparameter tuning and NAS** via Katib
- **Pipelines**

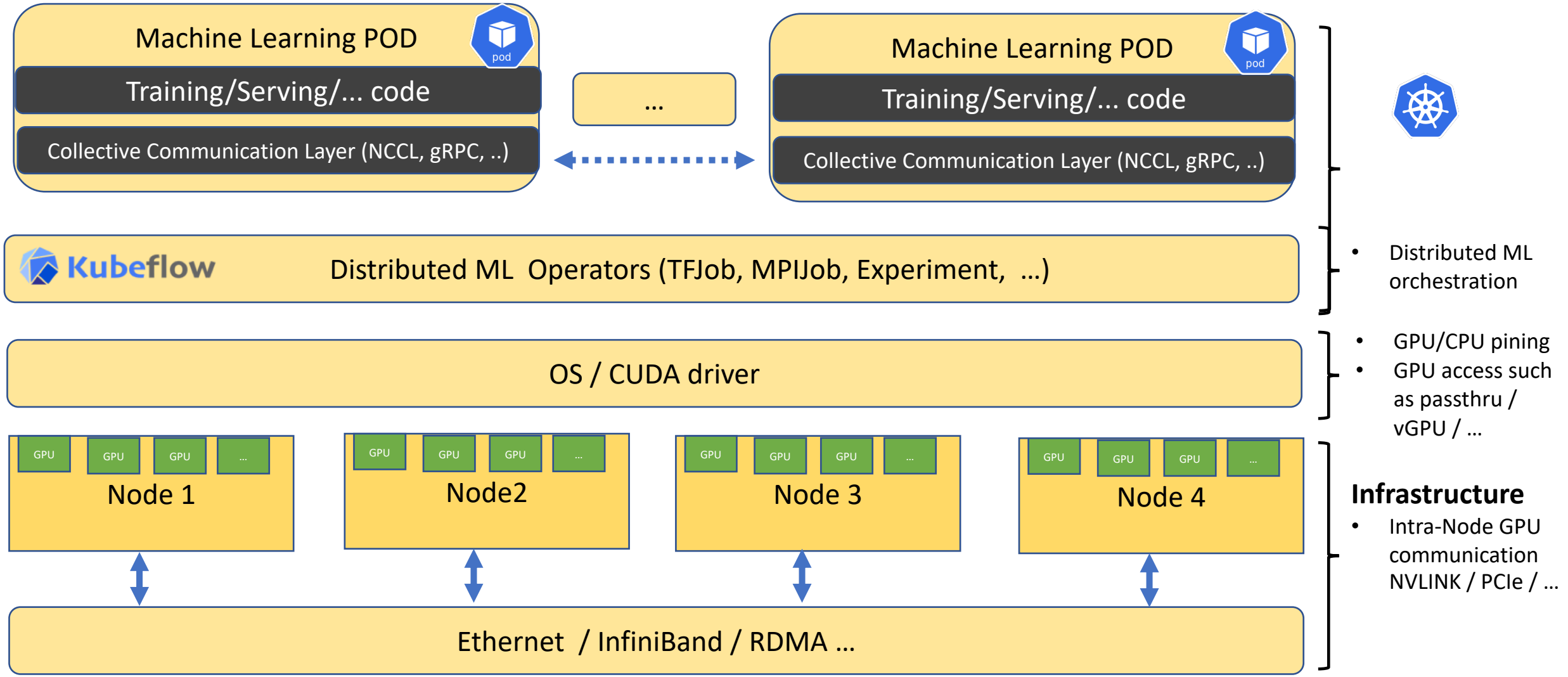
Kubeflow Operators



kubeflow.org/v1alpha2

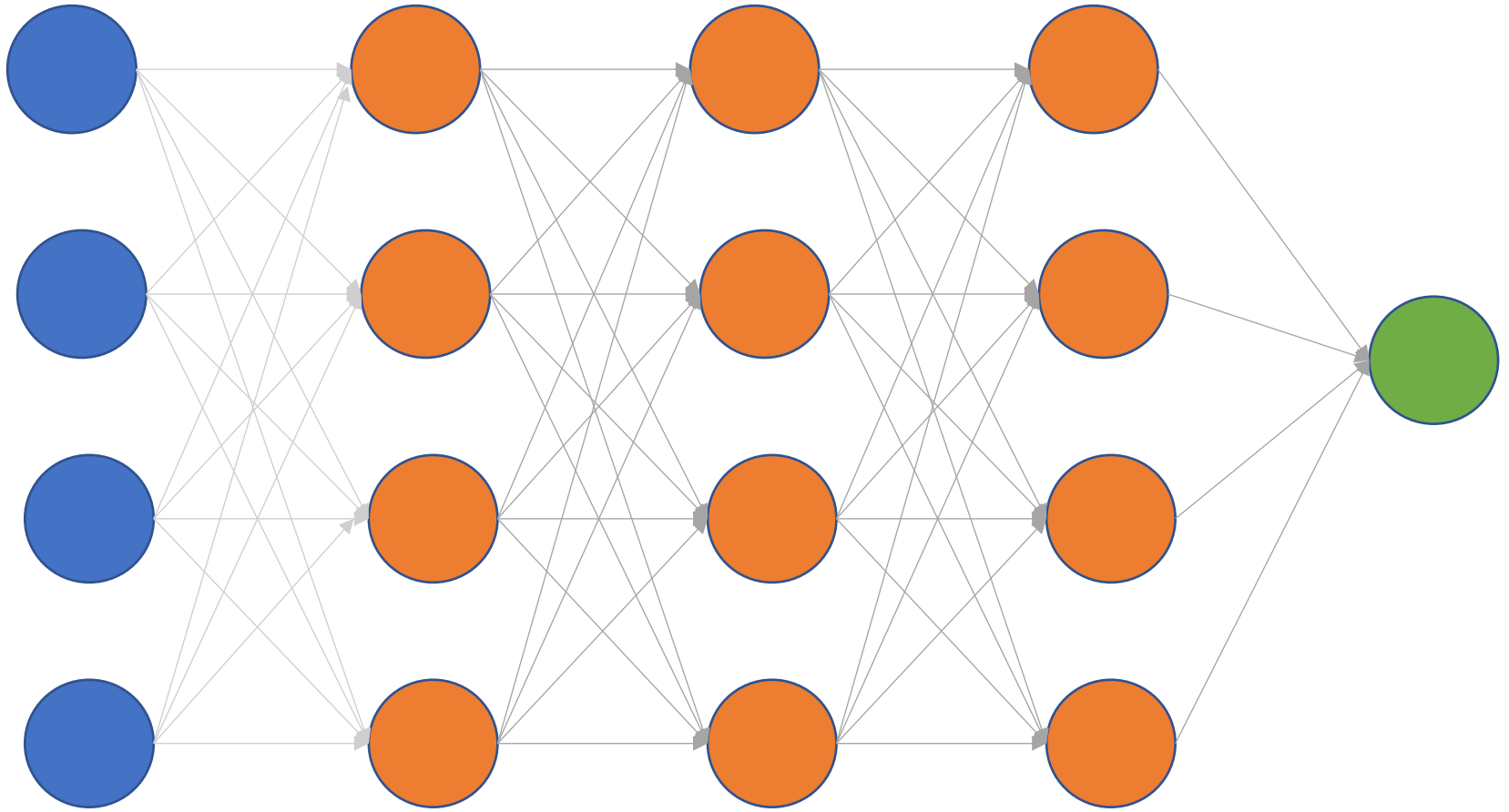
Pre-processing	Training	Hyperparameter Tuning	Model Serving
• Spark • ..	• MPIJob • TFJob • XGBoost • PyTorchJob • …	• Suggestion • Experiment • Trials • …	• Inference Services • Configurations • Revisions • Routes • ..

Distributed Machine Learning (ML)

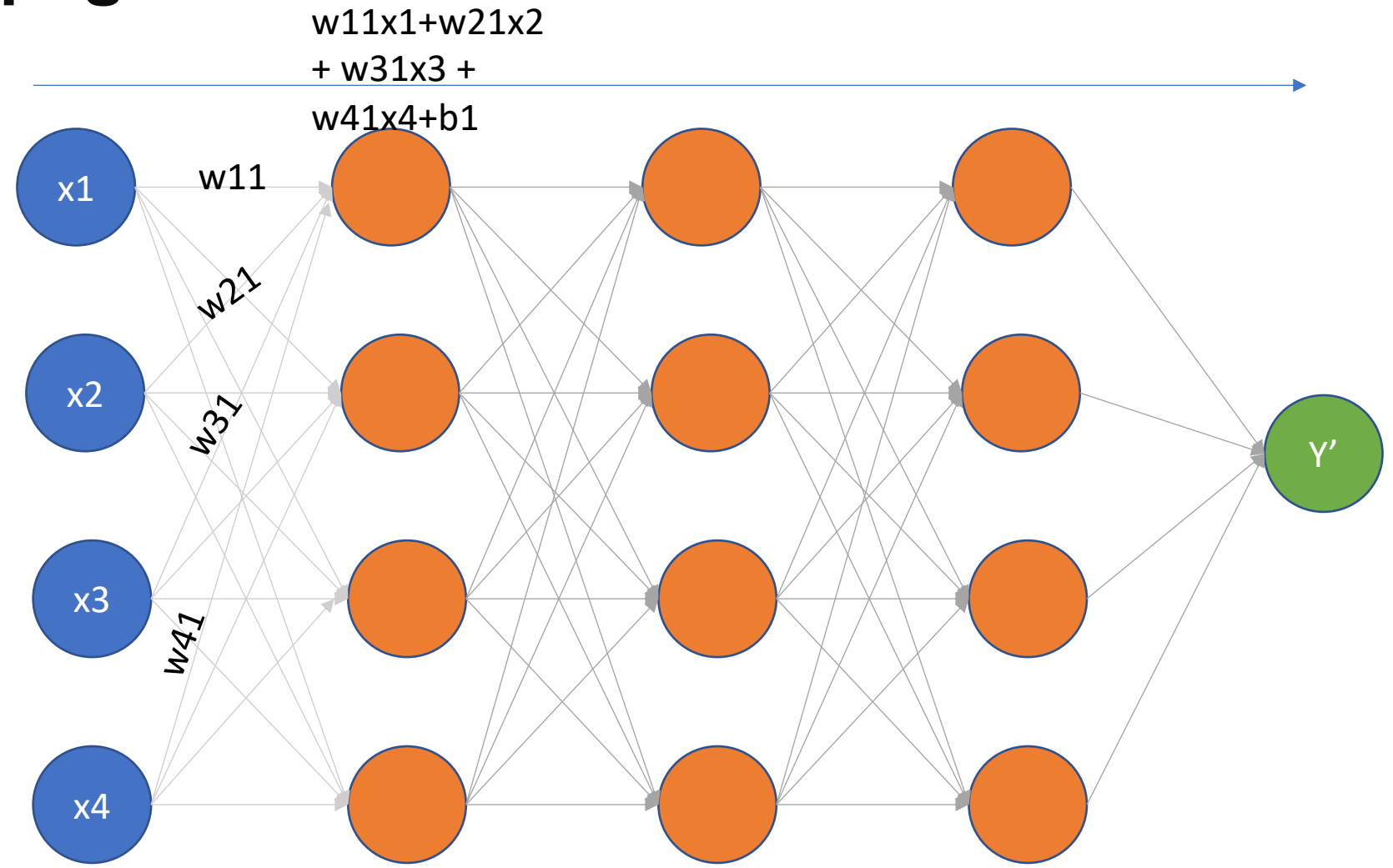


Distributed Training

Deep Neural Network

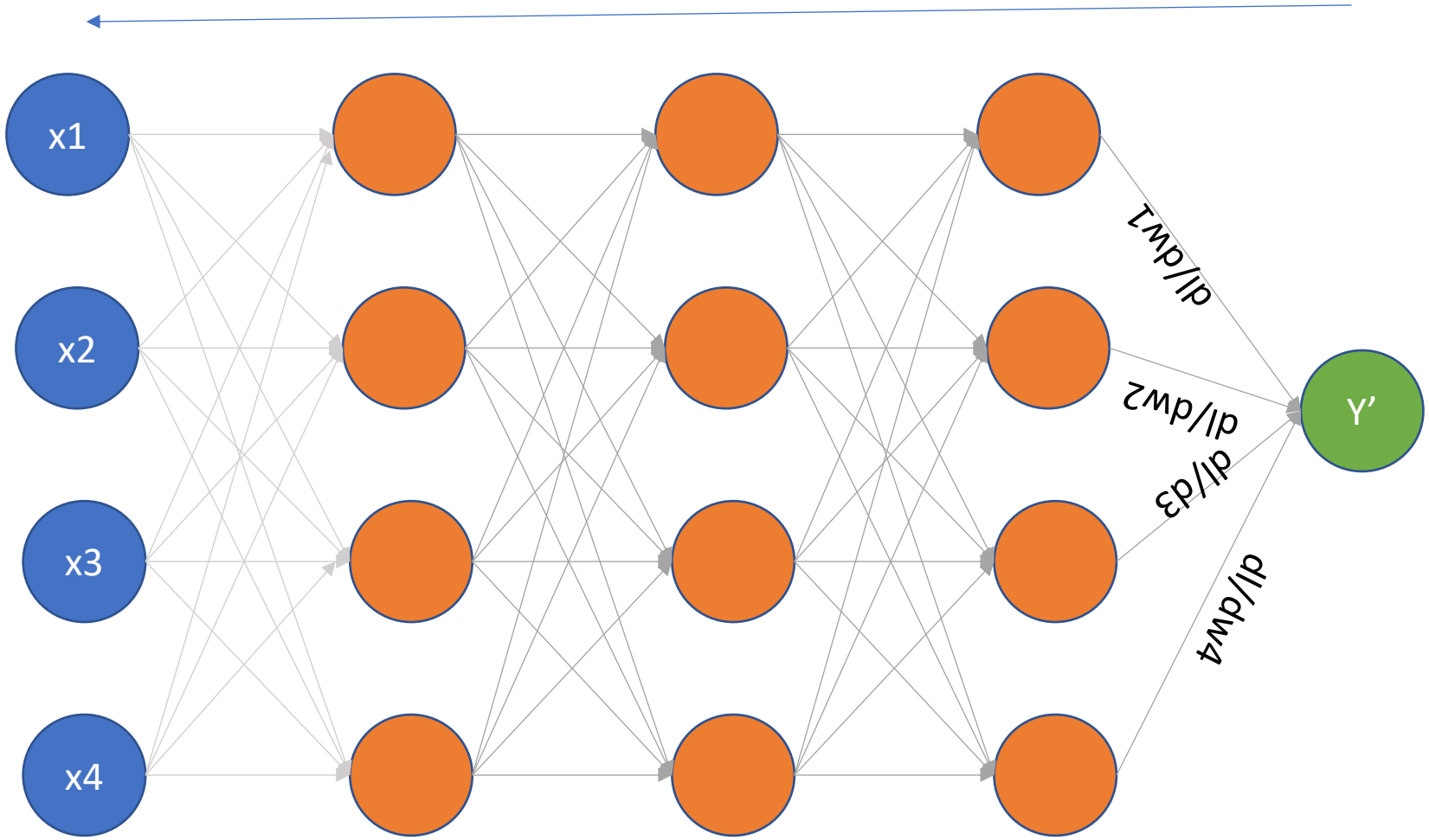


Deep Neural Network
Forward Propagation



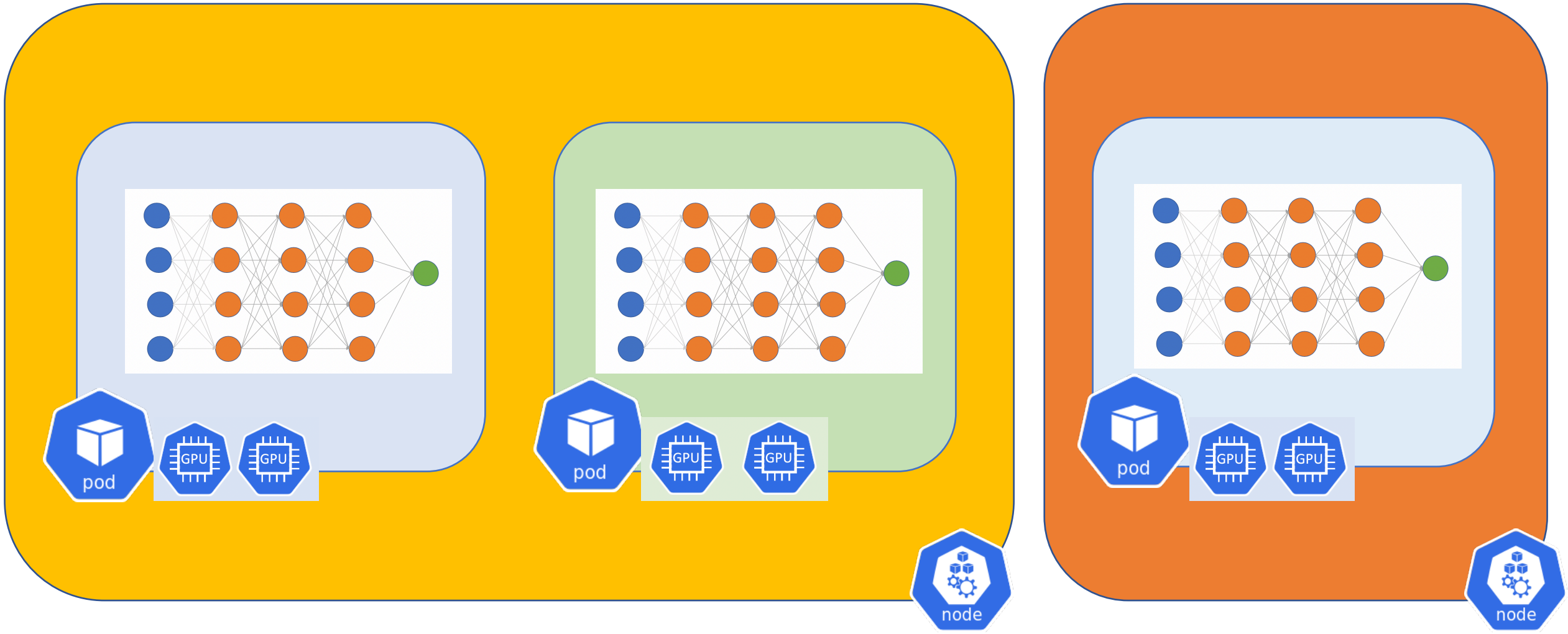
For a fun example, see https://www.youtube.com/playlist?list=PLZHQObOWTQDNU6R1_67000Dx_ZCJB-3pi

Deep Neural Network
Backward Propagation

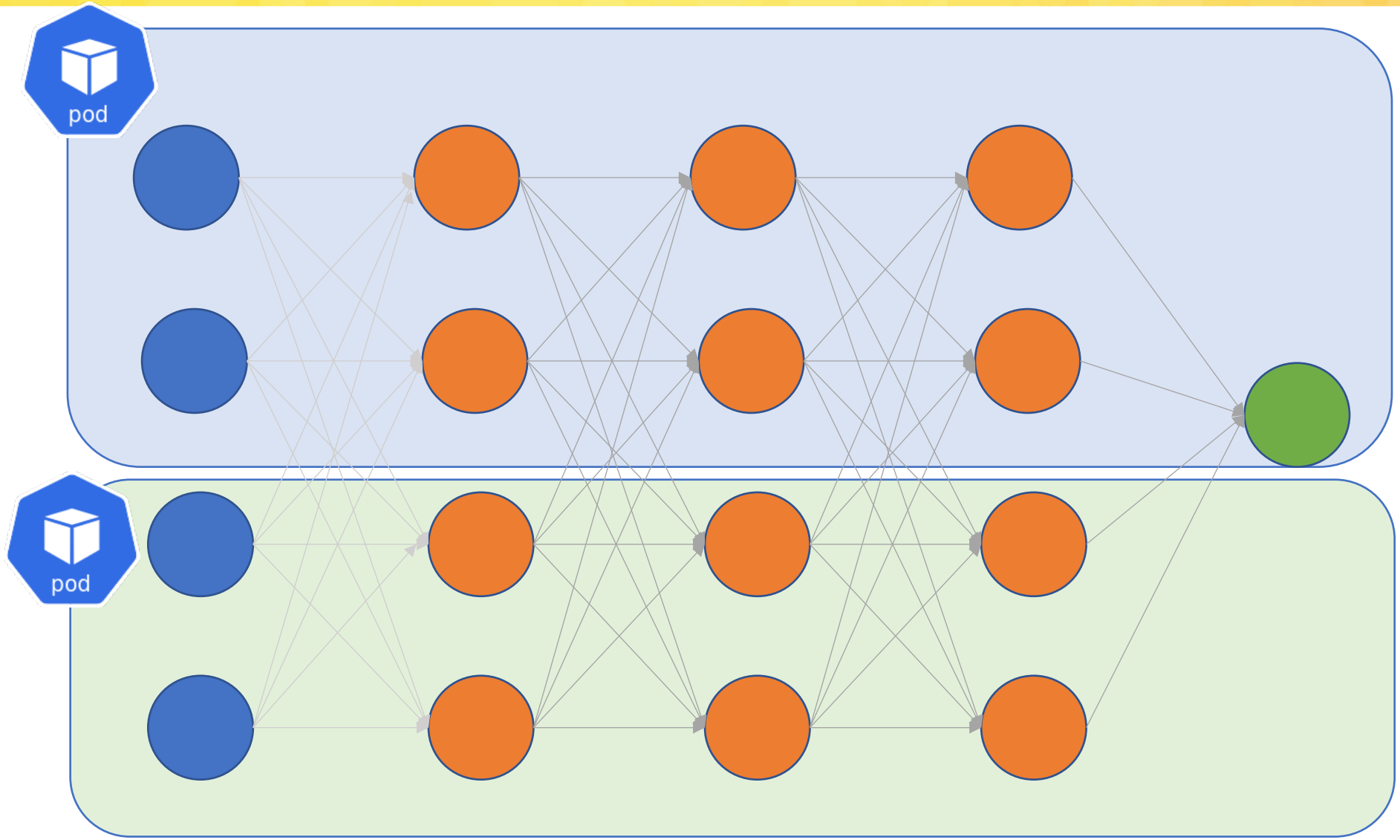


For a fun example, see https://www.youtube.com/playlist?list=PLZHQObOWTQDNU6R1_67000Dx_ZCJB-3pi

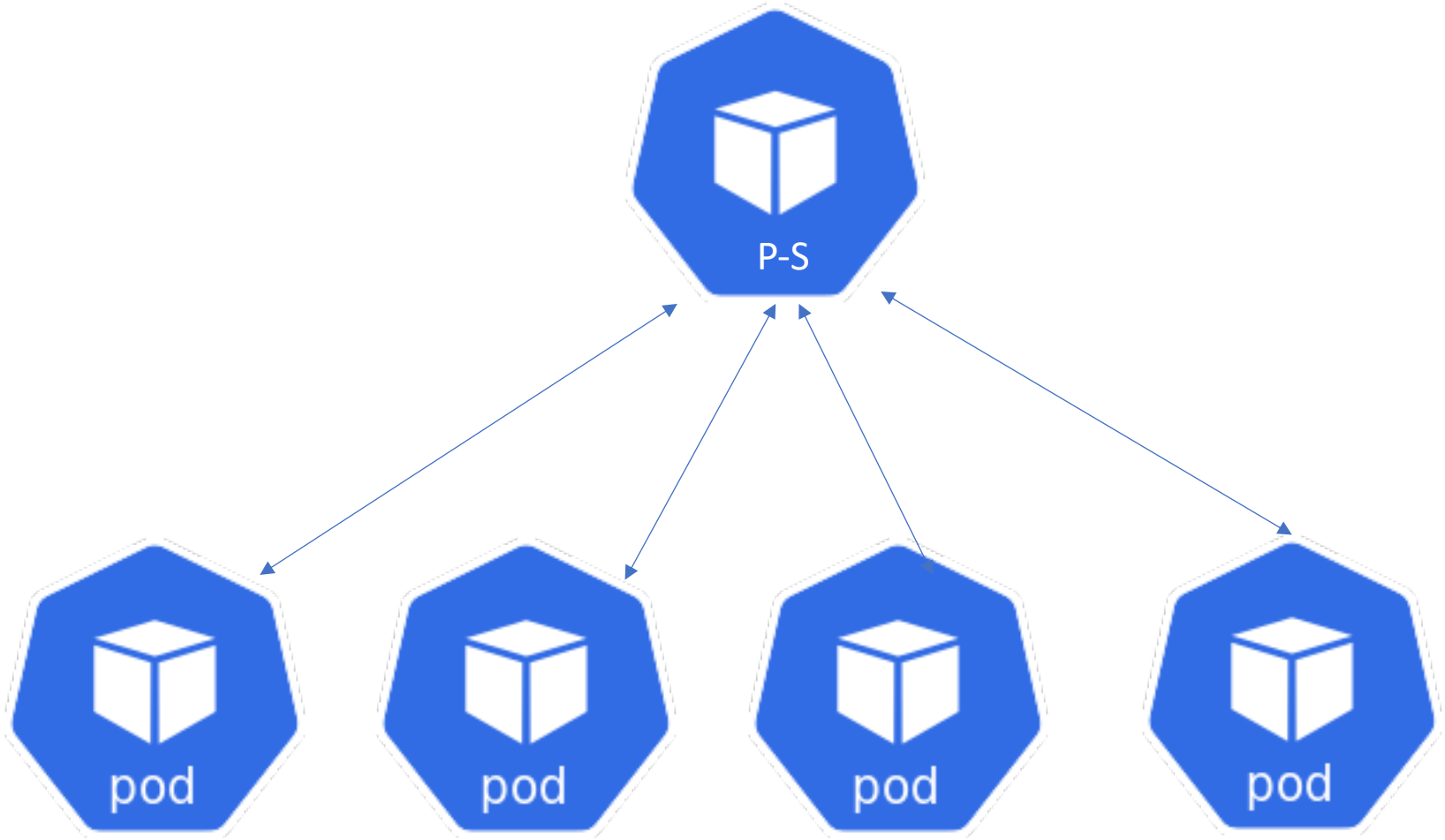
Data Parallelism



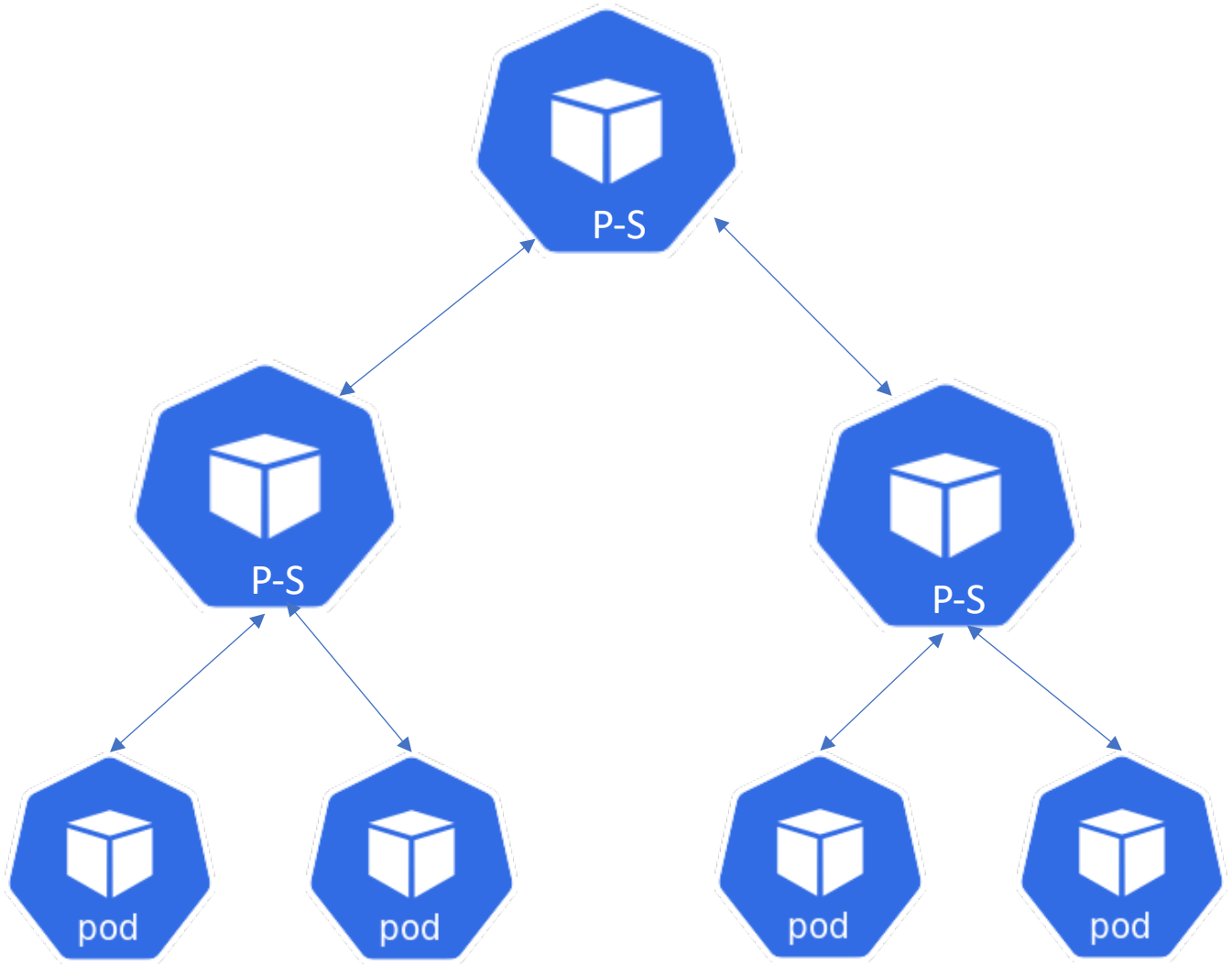
Model Parallelism



Parameter Server

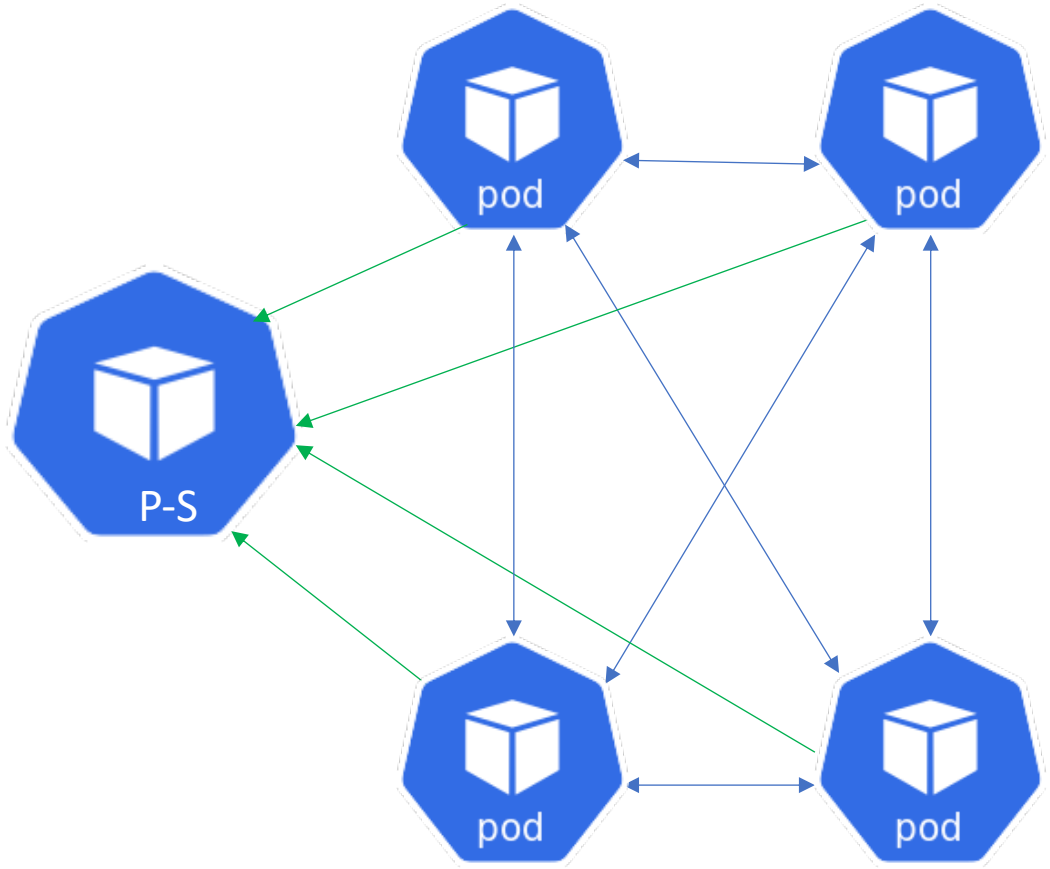


Async Parameter Server



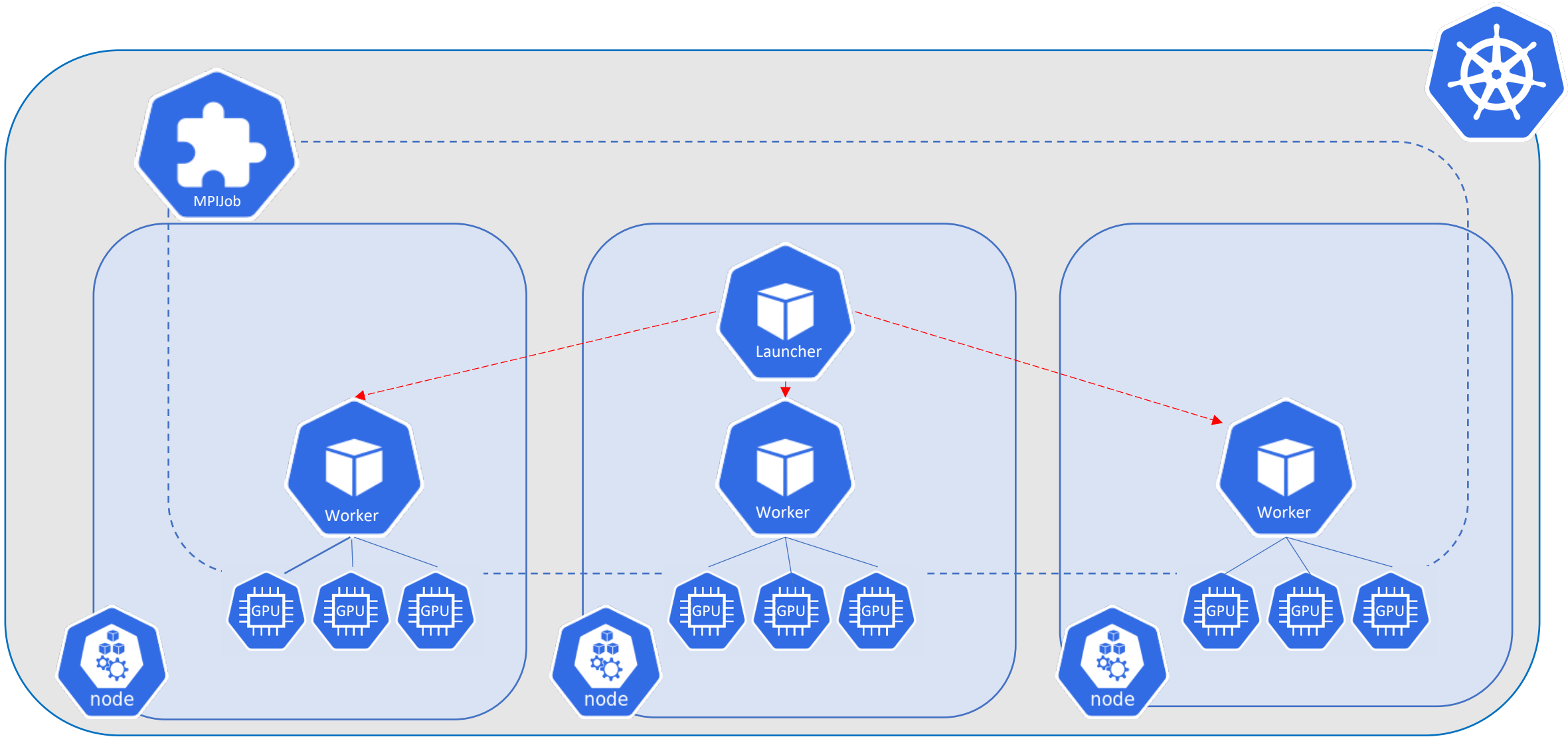

```
apiVersion: "kubeflow.org/v1"
kind: "TFJob"
spec:
  tfReplicaSpecs:
    PS:
      replicas: 2
      template:
        spec:
          containers:
            - name: tensorflow
    Worker:
      replicas: 4
      template:
        spec:
          containers:
            - name: tensorflow
```

Sync Parameter Server




```
apiVersion: "kubeflow.org/v1"
kind: "TFJob"
spec:
  tfReplicaSpecs:
    Worker:
      replicas: 3
      template:
        spec:
          containers:
            - name: tensorflow
              image: gcr.io/kubeflow-
examples/distributed_worker:v20181031-
513e107c
```

MPI Job


MPIJob
Launcher
Worker
Worker
Worker
GPU
GPU
GPU
GPU
GPU
GPU
GPU
GPU
GPU
node
node
node

MPI Job Spec


```
apiVersion: kubeflow.org/v1alpha2
kind: MPIJob
spec:
  slotsPerWorker: 1
  mpiReplicaSpecs:
    Launcher:
      replicas: 1
      template:
         spec:
           containers:
           - image: mpioperator/tensorflow-benchmarks:latest
             name: tensorflow-benchmarks
             command:
             - mpirun
             - -np
             - "2"
             - python
             - scripts/tf_cnn_benchmarks/tf_cnn_benchmarks.py
             - --model=resnet101
             - --batch_size=64
             - --variable_update=horovod
```

MPI Job Spec - Worker


```
Worker:
  replicas: 2
  template:
    spec:
      containers:
      - image: mpioperator/tensor    -benchmarks:latest
        name: tensorflow-benchma
        resources:
          limits:
            nvidia.com/gpu: 1
```

Matches the launcher spec

```
slotsPerWorker: 1
```

MPI Job Execution


```
+ POD_NAME=tensorflow-benchmarks-worker-1
+ /opt/kube/kubectl exec tensorflow-benchmarks-worker-1 -- /bin/sh -c  …
+ POD_NAME=tensorflow-benchmarks-worker-0
+ /opt/kube/kubectl exec tensorflow-benchmarks-worker-0 -- /bin/sh -c  …
2019-11-
11 :I tensorflow/core/common_runtime/gpu/gpu_device.cc:1326] Created TensorFlow
device (/job:localhost/replica:0/task:0/device:GPU:0 with 30568 MB memory) ->
  physical GPU (device: 0, name: Tesla V100-PCIE
32GB, pci bus id: 0000:13:00.0, compute capability: 7.0)

TensorFlow:  1.14
Model:       resnet101
Dataset:     imagenet (synthetic)
Mode:        training
SingleSess:  False
Batch size:  128 global
             64 per device
Num batches: 100
Num epochs:  0.01
Devices:     ['horovod/gpu:0', 'horovod/gpu:1']
```

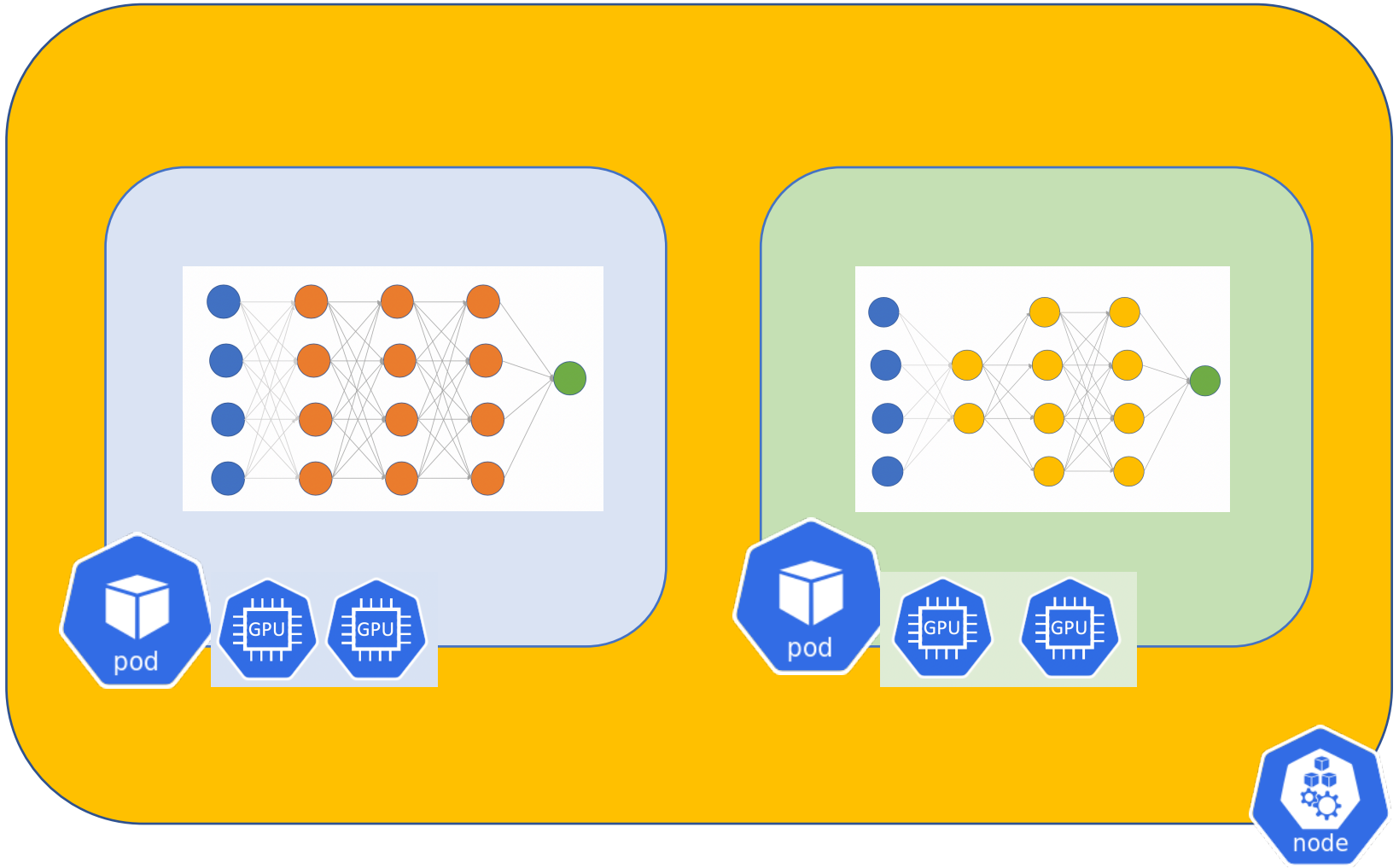
MPI Job Execution

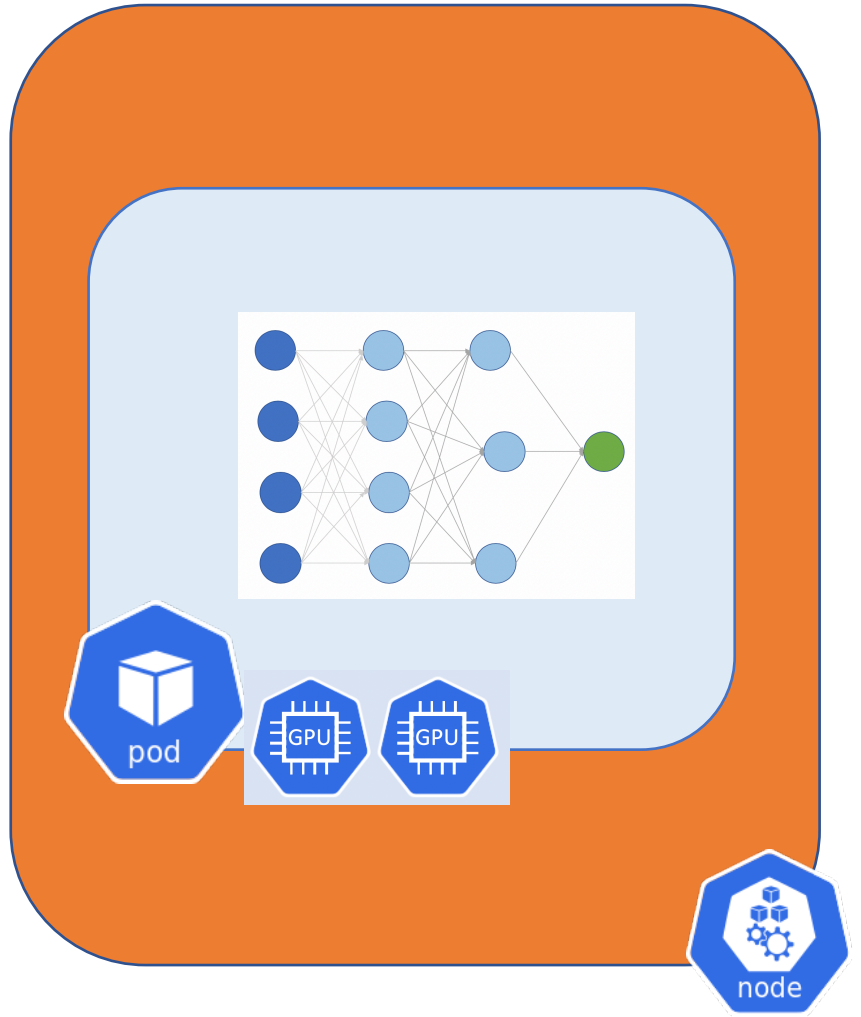

```
tensorflow-benchmarks-worker-
0:20:38 [0] NCCL INFO NET/Socket : Using [0]eth0:192.168.4.34<0>
tensorflow-benchmarks-worker-
0:20:38 [0] NCCL INFO NET/Plugin : No plugin found (libnccl-net.so).
tensorflow-benchmarks-worker-0:20:38 [0] NCCL INFO NET/IB : No device found.
NCCL version 2.4.7+cuda10.0
tensorflow-benchmarks-worker-
1:21:39 [0] NCCL INFO NET/Socket : Using [0]eth0:192.168.3.32<0>

1 images/sec: 81.5 +/- 0.0 (jitter = 0.0) 8.299
1 images/sec: 81.4 +/- 0.0 (jitter = 0.0) 8.361
10  images/sec: 93.8 +/- 3.2 (jitter = 12.7)  8.507
...
100 images/sec: 94.5 +/- 1.1 (jitter = 12.4)  8.388
----------------------------------------------------------------
total images/sec: 189.07
----------------------------------------------------------------
100 images/sec: 94.5 +/- 1.1 (jitter = 11.2)  8.524
----------------------------------------------------------------
total images/sec: 189.06
```

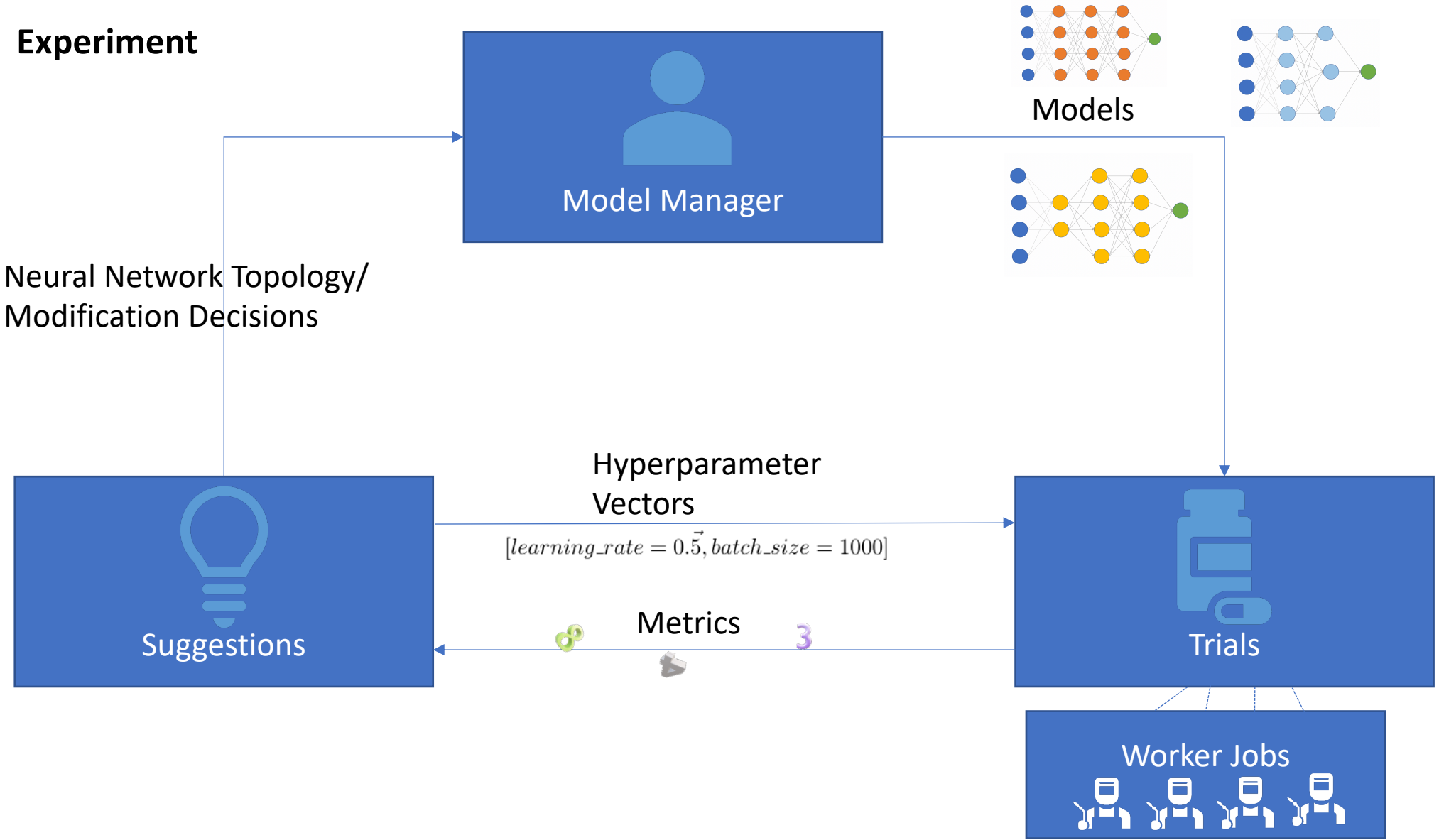
Hyperparameter tuning

Multi-Model Parallelism



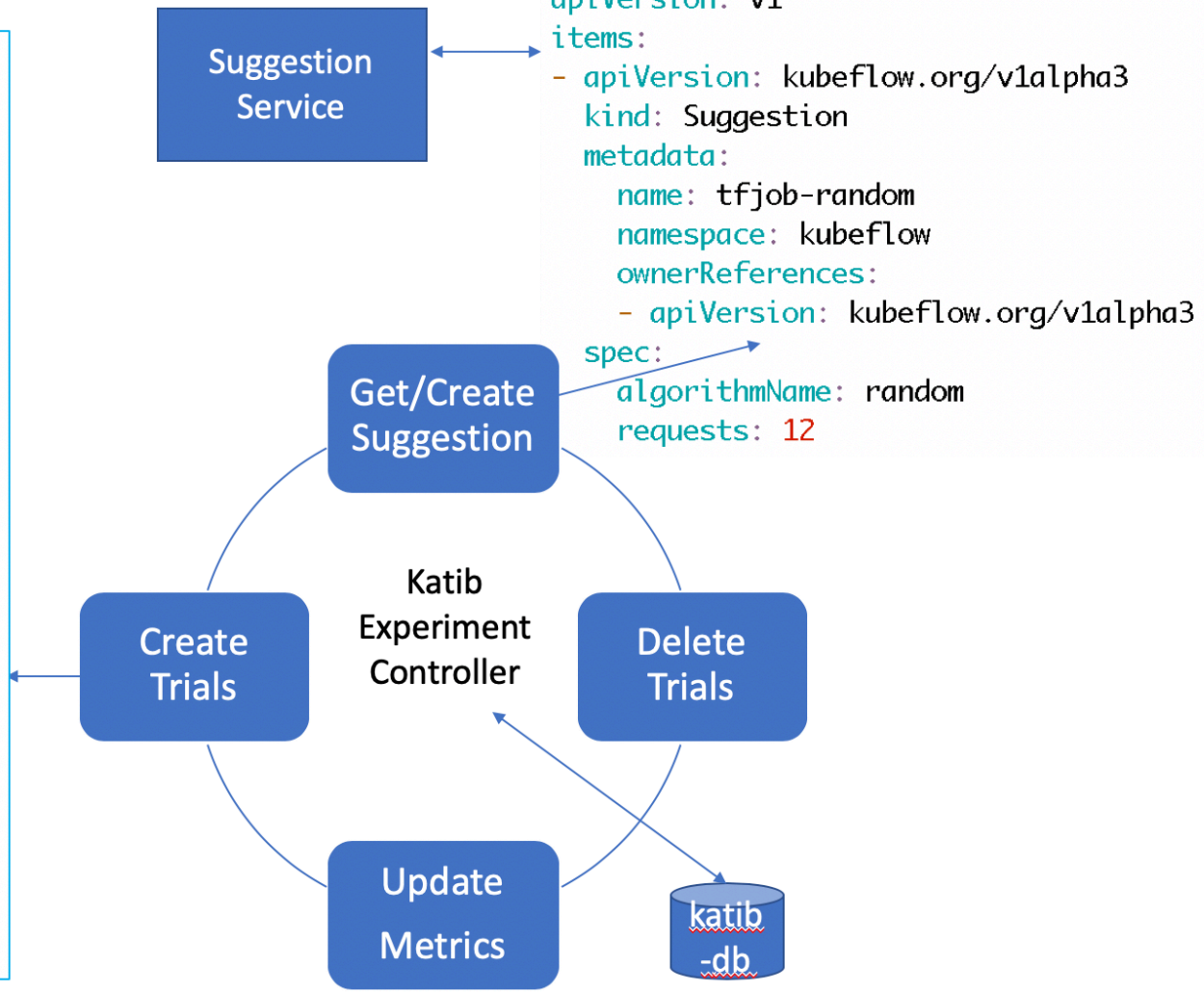


Hyperparameter Tuning



Katib Workflow


```
apiVersion: v1
items:
- apiVersion: kubeflow.org/v1alpha3
  kind: Trial
  metadata:
    name: tfjob-random-484zhpzq
    namespace: kubeflow
  spec:
    metricsCollector:
      collector:
        kind: TensorFlowEvent
      source:
        fileSystemPath:
          kind: Directory
          path: /train
    objective:
      goal: 0.99
      objectiveMetricName: accuracy_1
      type: maximize
    parameterAssignments:
    - name: --learning_rate
      value: "0.029127223437729596"
    - name: --batch_size
      value: "133"
```



Katib Experiment Spec


```yaml
apiVersion: "kubeflow.org/v1alpha3"
kind: Experiment
spec:
  parallelTrialCount: 3
  maxTrialCount: 12
  maxFailedTrialCount: 3
  objective:
    type: maximize
    goal: 0.99
    objectiveMetricName: Validation-Accuracy
  algorithm:
    algorithmName: nasrl
    algorithmSettings:
      - name: "lstm_num_cells"
        value: "64"
      - name: "lstm_num_layers"
        value: "1"
```

Katib Experiment Spec


```
trialTemplate:
  goTemplate:
      rawTemplate: |-
        apiVersion: batch/v1
        kind: Job
        metadata:
          name: {{.Trial}}
          namespace: {{.NameSpace}}
        spec:
          template:
            spec:
              containers:
              - name: {{.Trial}}
```

```
command:
- "python3.5"
- "-u"
- "RunTrial.py"
{{- with .HyperParameters}}
{{- range .}}
- "--{{.Name}}={{.Value}}"
{{- end}}
{{- end}}
resources:
  limits:
    nvidia.com/gpu: 1
```

Kubeflow Experiment Execution


```
apiVersion: v1
items:
- apiVersion: kubeflow.org/v1alpha3
  kind: Suggestion
  spec:
    algorithmName: nasrl
    requests: 3
  status:
    suggestionCount: 3
    suggestions:
    - name: nas-rl-example-gmgpfrqs
      parameterAssignments:
      - name: architecture
        - name: nn_config
        value: '{''num_layers'': 8, ''input_sizes'': [32, 32, 3], ''outp
ut_sizes'':
          [10], ''embedding'': {''19'': {''opt_id'': 19, ''opt_type'': '
'convolution'',
          ''opt_params'': {''filter_size'': ''5'', ''num_filter'': ''128
'', ''stride'':
```

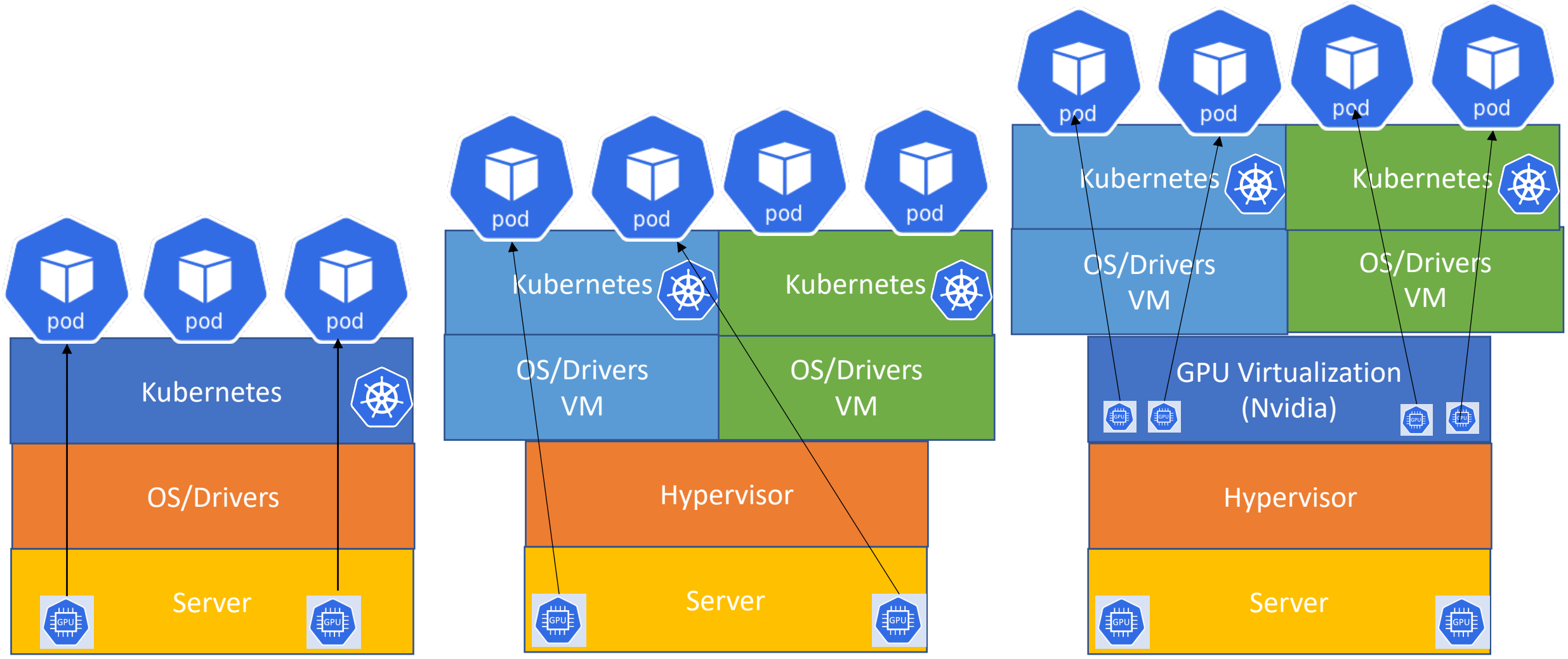
Kubeflow Experiment Execution

```
nas-rl-example-nasrl-59dd58f67d-np586  1/1 Running   192.168.4.60   tc-hx3-2-gpu-v10-1

nas-rl-example-9gnzfqlf-6ksvq  0/2 Completed 192.168.3.63   tc-hx3-2-gpu-v10-0
nas-rl-example-9gnzfqlf-72nlj  0/2 Completed 192.168.3.66   tc-hx3-2-gpu-v10-0
nas-rl-example-9gnzfqlf-sdsxt  0/2 Completed 192.168.3.64   tc-hx3-2-gpu-v10-0
nas-rl-example-skbcfxq9-62ld9  0/2 Completed 192.168.3.65   tc-hx3-2-gpu-v10-0
nas-rl-example-skbcfxq9-vqp78  2/2 Running   192.168.4.65   tc-hx3-2-gpu-v10-1
nas-rl-example-skbcfxq9-wt5d6  0/2 Completed 192.168.4.63   tc-hx3-2-gpu-v10-1
nas-rl-example-tp28clqz-bmjvs  0/2 Completed 192.168.4.61   tc-hx3-2-gpu-v10-1
nas-rl-example-tp28clqz-pq9q8  0/2 Completed 192.168.4.64   tc-hx3-2-gpu-v10-1
nas-rl-example-tp28clqz-wd48f  0/2 Completed 192.168.4.62   tc-hx3-2-gpu-v10-1
```

Infrastructure and OS Optimization

Kubernetes GPU Configuration



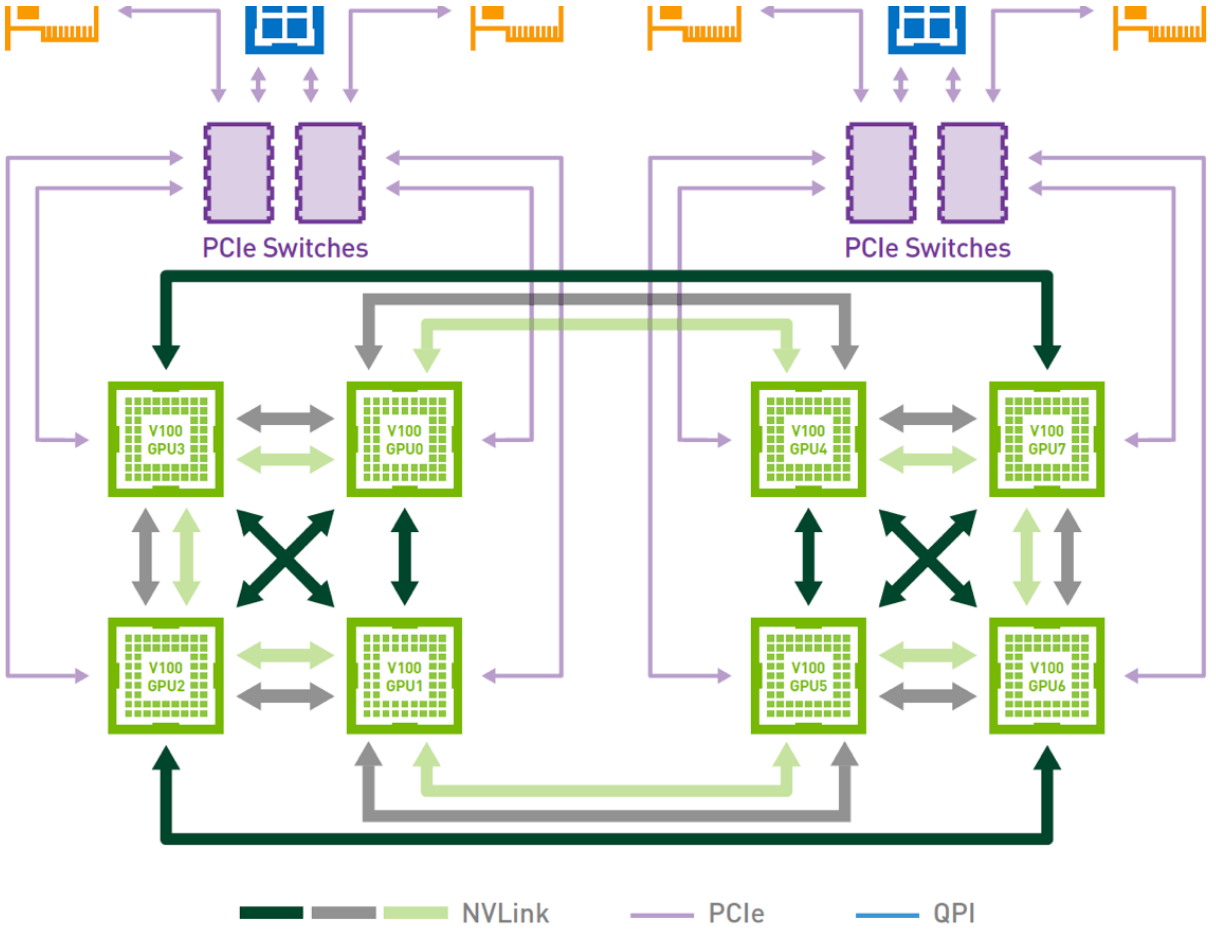
Bare metal

VMs with PCI Passthru

VMs with vGPUs

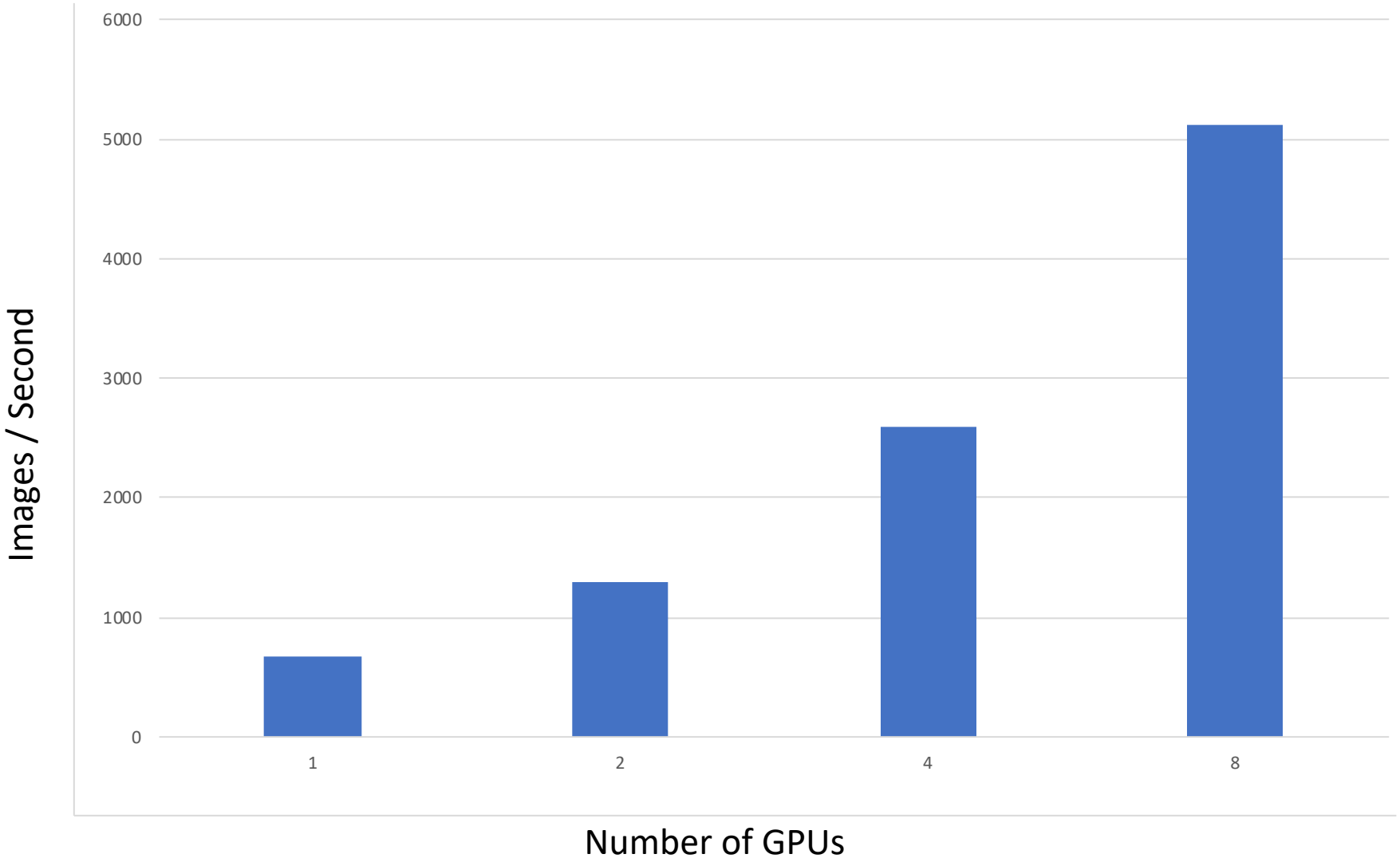
Intra-Node GPU Topology

- **NVLink**- High speed GPU interconnect, each lane is 25GB/s
- NVLink 2.0 has 6 lanes with overall 300GB/s bandwidth between 8x SXM2 V100-32GB Nvidia GPU
- 4x PCI switch for connectivity between CPU/GPU and network card
- 4 x16 PCI slot for VIC or 3rd party NICs
- Based on standard Nvidia NVLink 2.0 spec

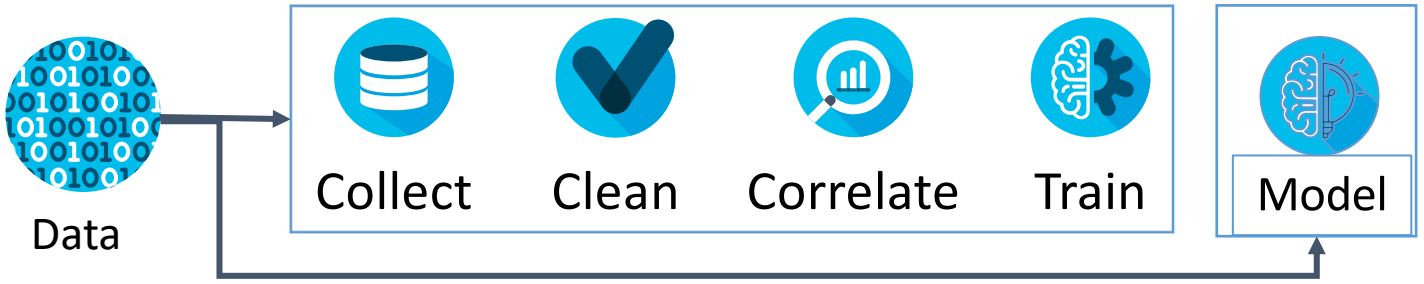


Source - https://www.cisco.com/c/dam/en/us/products/collateral/servers-unified-computing/ucs-c-series-rack-servers/whitepaper-c11-741689.pdf

Scaling with NVLink



Summary



Kubeflow

EASY
Distributed
ML