Multinode Reduction Operations

Multinode Reduction Operations

The core of distributed SGD:

- start off with the same weight vector
- forward pass on all nodes (using different batches)
- backward pass to compute all gradients
- *sum all gradients across all nodes*
- add total gradient to all weights

How does the summation work on Kubernetes?

In [1]:

```
cat > kubetpl.yaml <<'EOF'
image: gcr.io/research-191823/bigdata19
memory: 4G
cpu: 1
app: bigdata19
subdomain: bigdata19
port:
  - 7880
config_map: files
env:
  - MASTER_ADDR=master.bigdata19
  - MASTER_PORT=7880
EOF
```

In [2]:

```
kubectl delete service/bigdata19 || true
kubetpl service | kubectl apply -f -
```

```
service "bigdata19" deleted
service/bigdata19 created
```

Python Script

- sum reduction across a number of machines
- core operation of synchronous distributed SGD

In [12]:

```
cat > reduce.py <<'EOF'

import os, torch
import torch.distributed as dist

rank, world = int(os.environ["rank"]), int(os.environ["world"])

dist.init_process_group("gloo", rank=rank, world_size=world)

deltas = torch.ones(1)*(rank+1)

print(f"before {rank} data {deltas[0]}")

all_nodes = dist.new_group(list(range(world)))
dist.all_reduce(deltas, op=dist.ReduceOp.SUM, group=all_nodes)

print(f"after {rank} data {deltas[0]}")

EOF
```

Upload to Shared File System

In [13]:

```
kubefcm files reduce.py
```

```
-- --from-file=reduce.py=reduce.py
configmap "files" deleted
configmap/files created
```

Create Master Node

In [14]:

```
kubectl delete pods --all || true

kubetpl pod -n master -c 'rank=0 world=4 python3 /files/reduce.py' | kubectl apply -f -
```

```
pod "master" deleted
pod "node1" deleted
pod "node2" deleted
pod "node3" deleted
pod/master created
```

In [15]:

```
sleep 15
```

In [16]:

```
kubectl get pods
```

```
NAME     READY   STATUS    RESTARTS   AGE
master   1/1     Running   0          16s
```

Create Additional Compute Nodes

In [17]:

```
for i in {1..3}; do
    kubetpl pod -n node$i -c "rank=$i world=4 python3 /files/reduce.py" | kubectl apply -f -
done
```

```
pod/node1 created
pod/node2 created
pod/node3 created
```

In [18]:

```
sleep 15
```

Result of Computation

- all values added and distributed to all nodes

In [19]:

```
kubectl logs master
```

```
before 0 data 1.0
after 0 data 10.0
```

In [20]:

```
kubectl logs node3
```

```
before 3 data 4.0
after 3 data 10.0
```

Global Reduction Operations

That's the core of synchronous distributed SGD.