The Power Of Metrics
-- How to monitor and improve ML efficiency

Yan Yan, Zhilan Zweiger

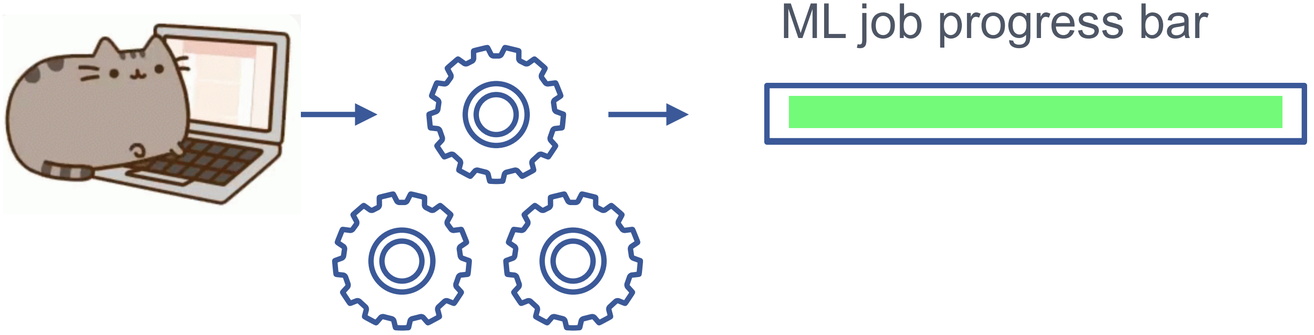
Production Engineer, Facebook

Agenda

- Background
- ML Distributed Training Architecture
- Challenges
- Tools
- Use Cases

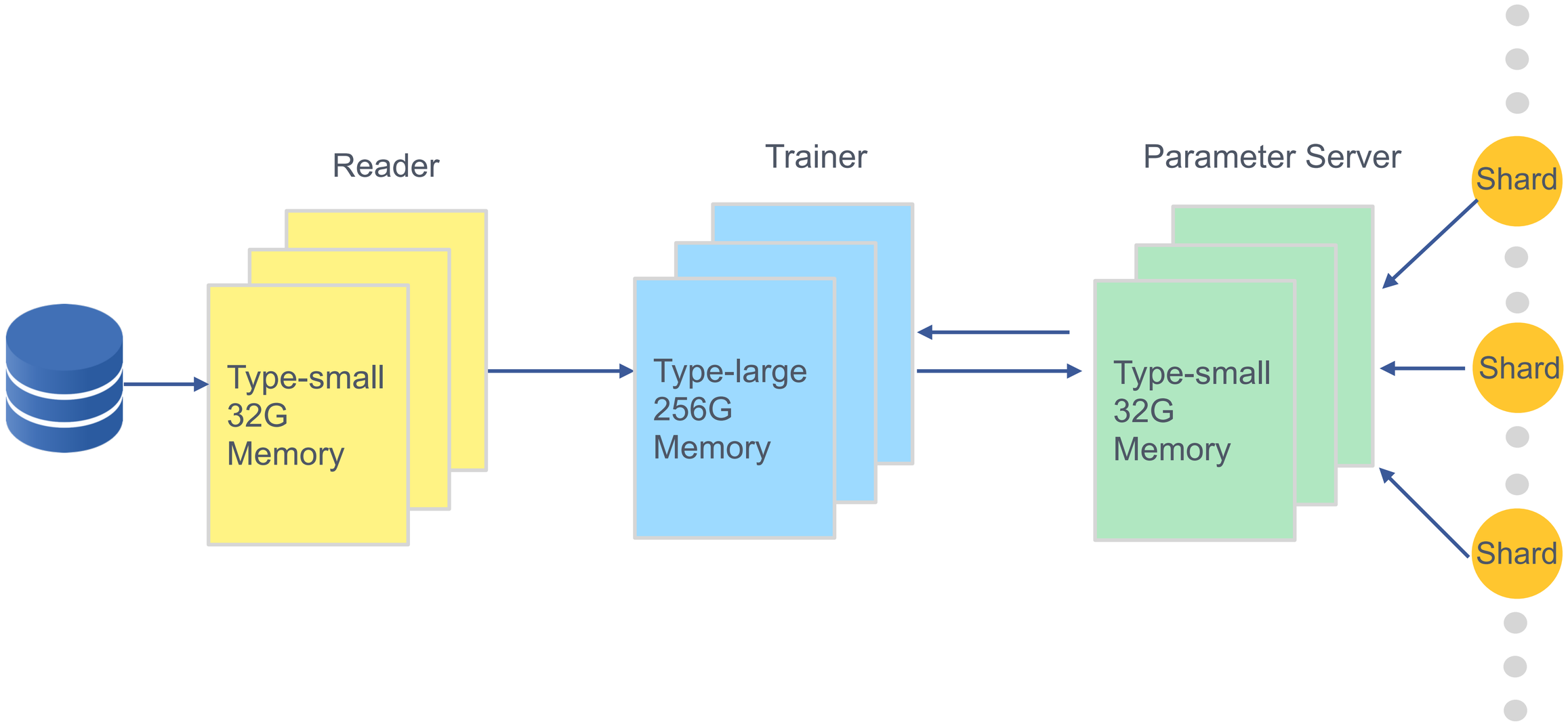
Background Story

Distributed Training

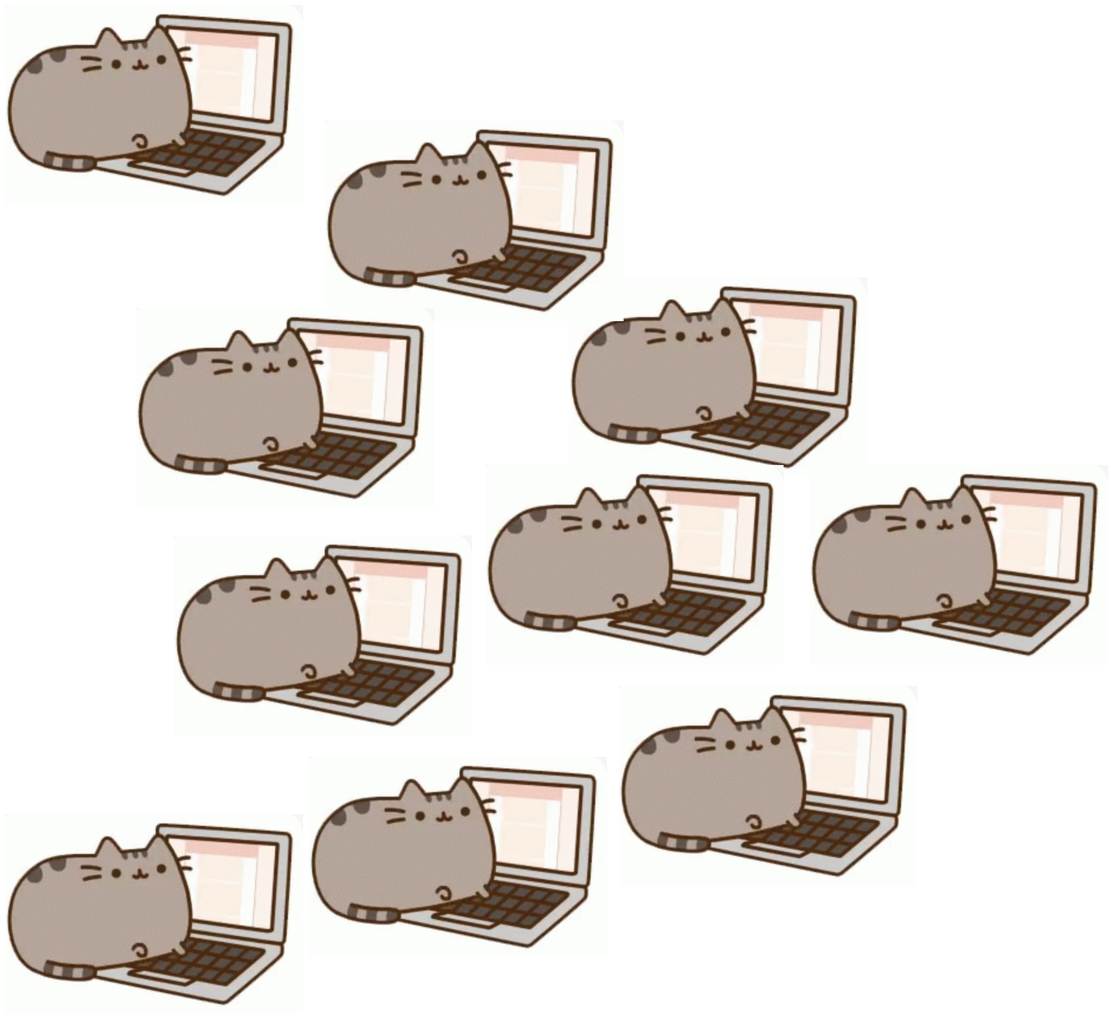


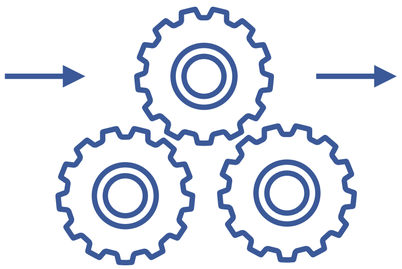
ML Distributed Training Architecture

Distributed Training Architecture

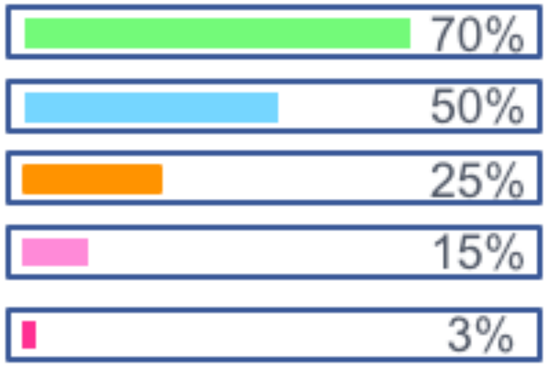


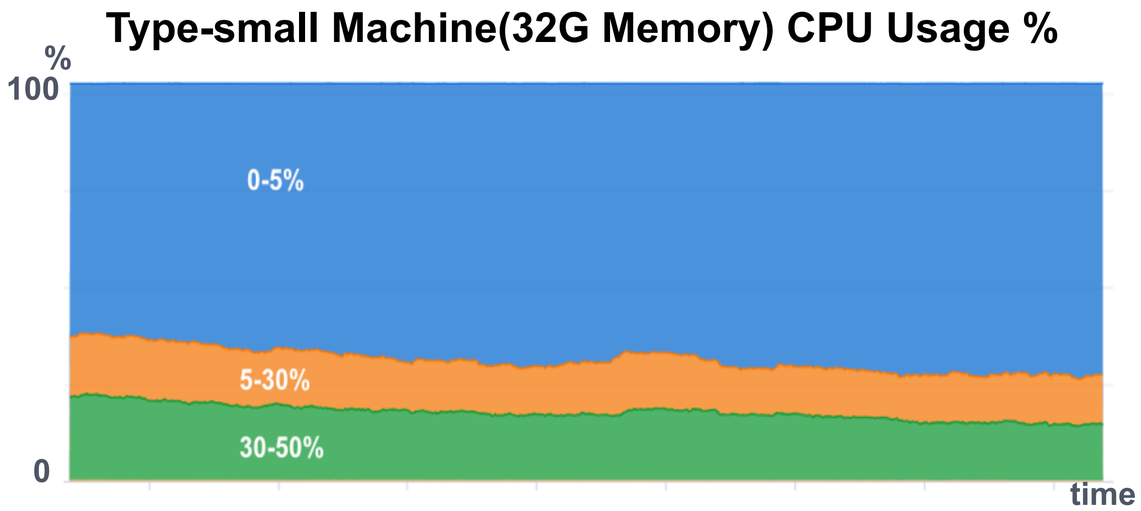
What happened to context? forever...

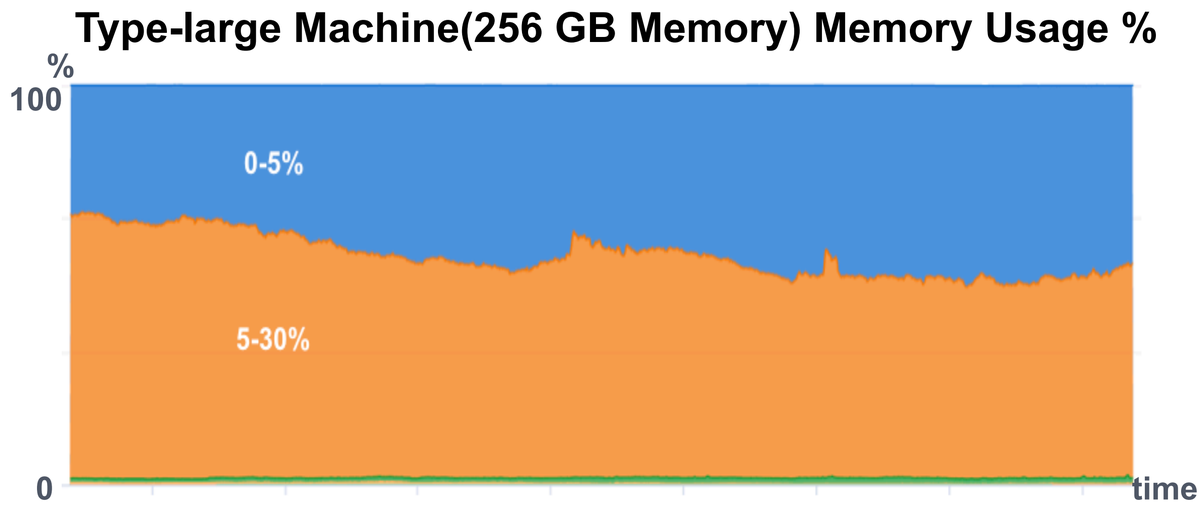




ML job progress bar




Type-small Machine(32G Memory) CPU Usage %
%
100
0
0-5%
5-30%
30-50%
time


Type-large Machine(256 GB Memory) Memory Usage %
%
100
0
0-5%
5-30%
time

How to find the best practices?

CPU Usage%

Network Read Speed

Memory Usage%

Network Write Speed



Network Usage%

QPS

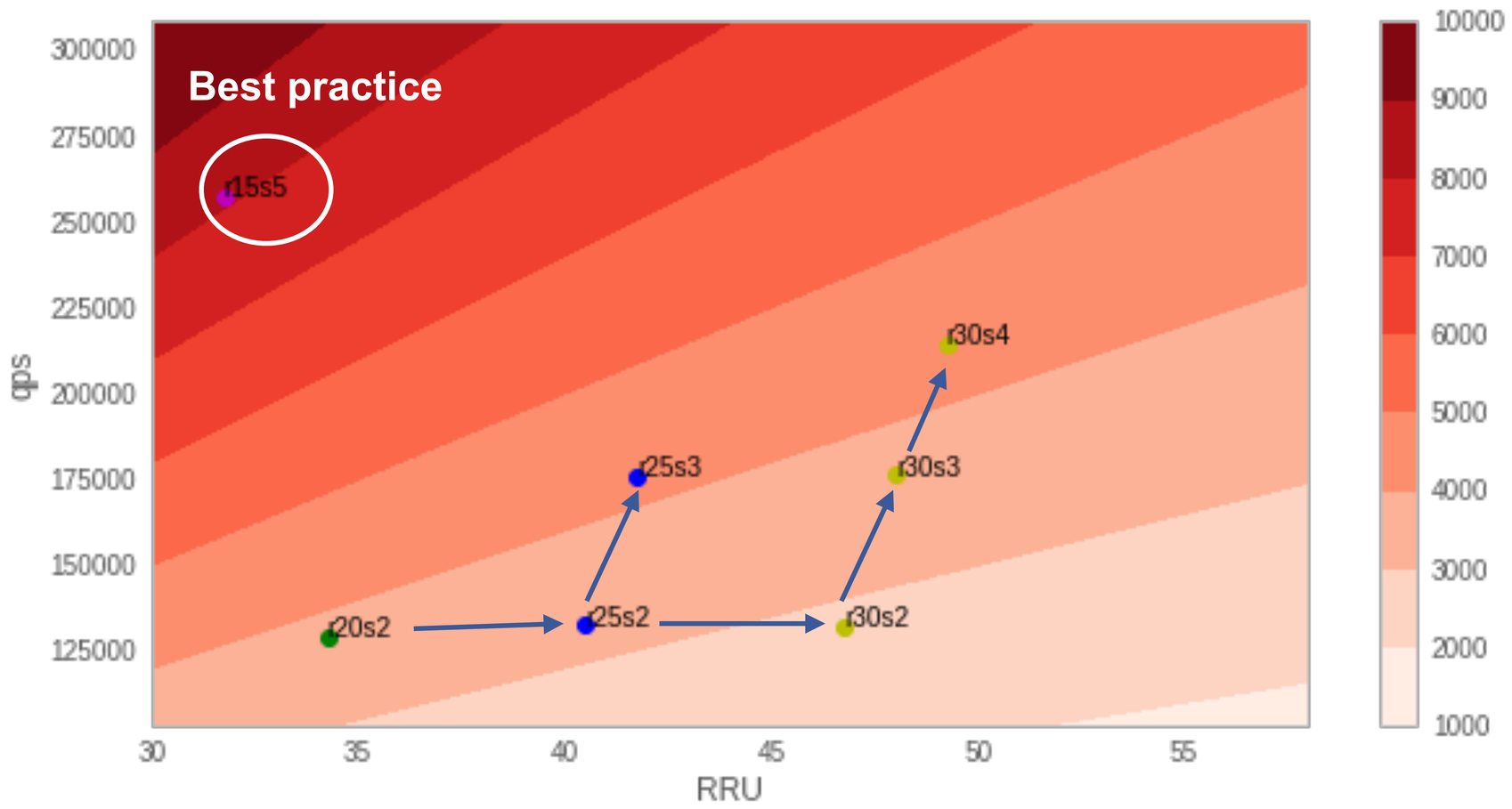
IOwait

Memory Bandwidth

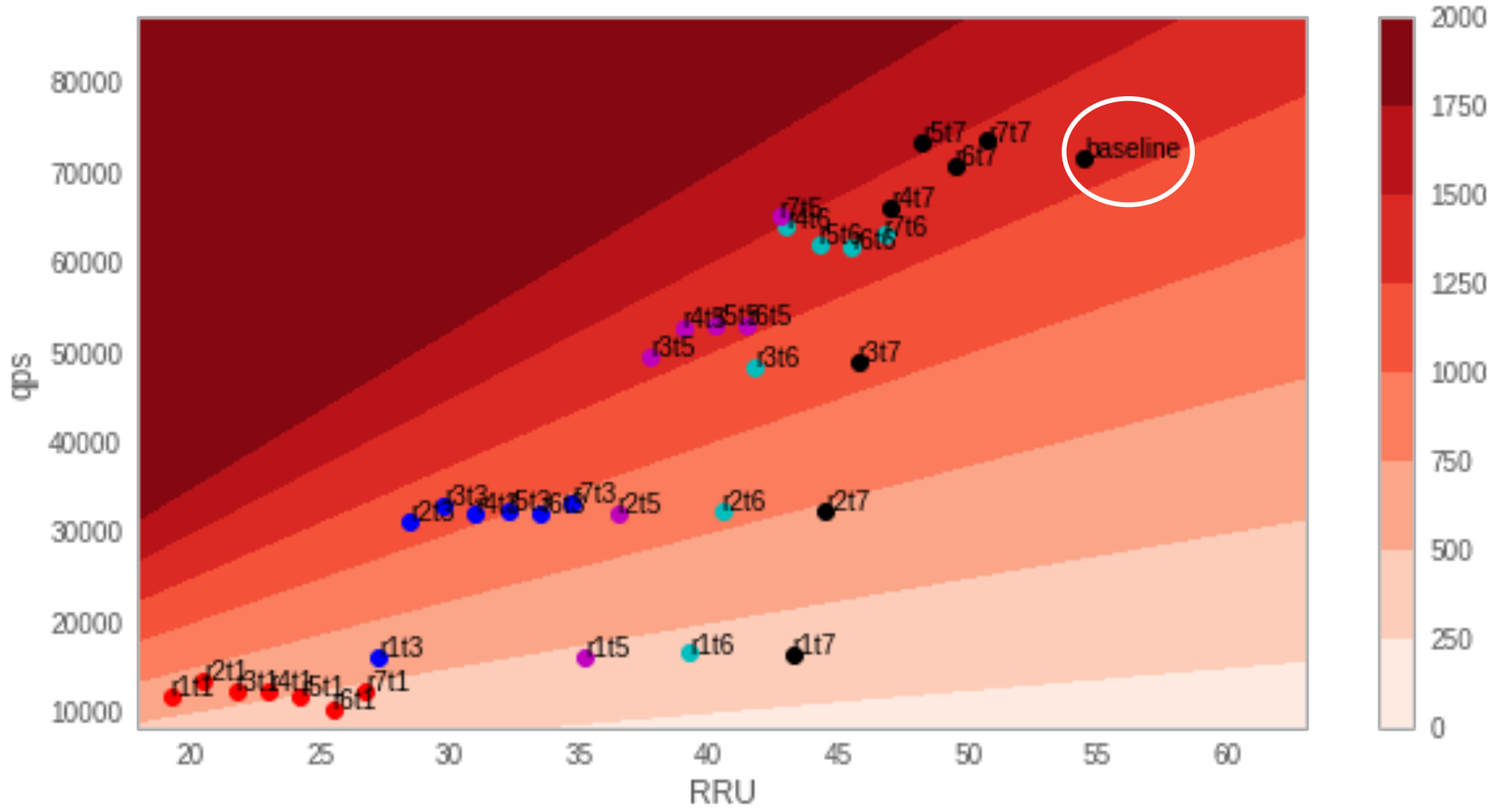
Gflops

Tool: flow-info CLI

Tuning without flow-info

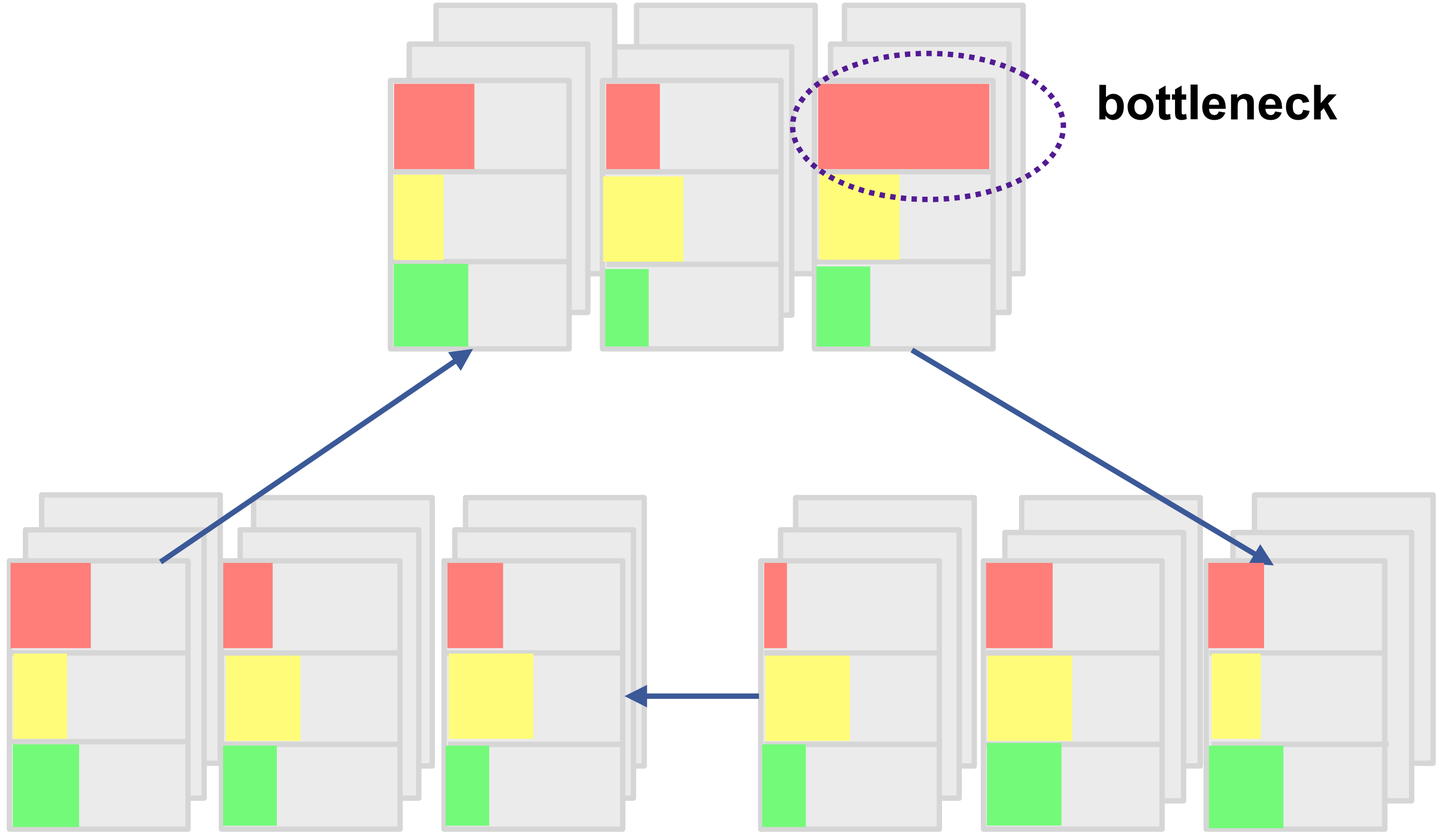


Tuning without flow-info

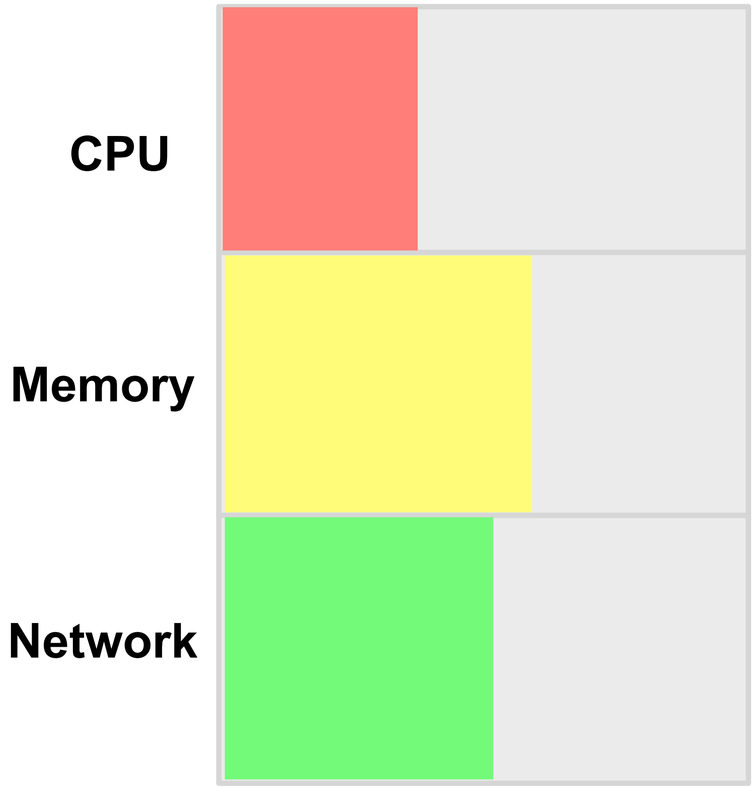


Detailed Resource Usage

Example Job:



Example Machine:



Use Cases

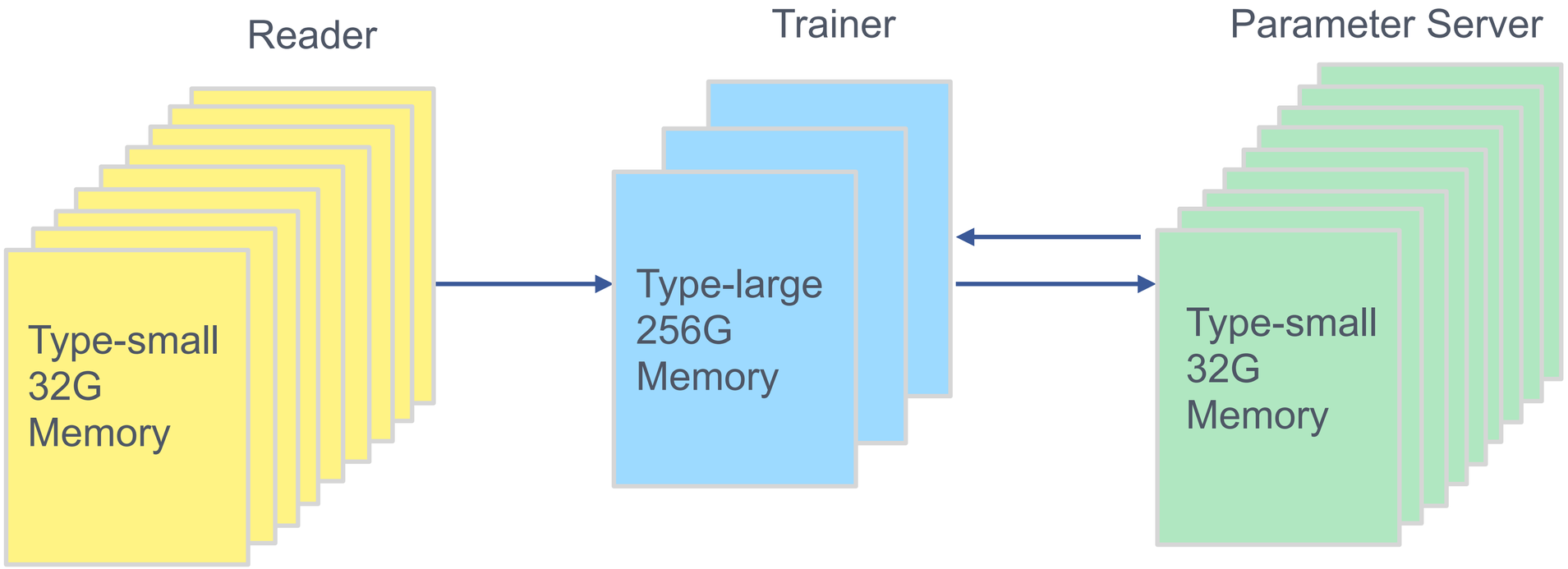
Path	Hours	Max CPU	Avg CPU	Max Mem	Avg Mem	Max IOwait	Avg IOwait	Max Eth	Avg Eth	Max Eth In	Avg Eth In	Max Eth Out	Avg Eth Out	RRU	CPU	RAM	NIC	Max Gflops	Avg Gflops
publish_op	2.83	72.15%	15.52%	51.80%	34.87%	536.32%	57.18%	10.92%	4.60%	229.41MB/s	81.03MB/s	341.38MB/s	140.17MB/s	4	40	256	25	63.00	10.02
ps_18	17.18	25.69%	16.26%	34.01%	31.93%	680.52%	12.92%	28.13%	19.96%	108.61MB/s	77.01MB/s	109.89MB/s	46.87MB/s	1	18	64	12.5	3.00	2.38
ps_19	17.18	24.32%	15.91%	34.48%	32.63%	179.56%	4.23%	35.72%	26.26%	139.53MB/s	101.64MB/s	111.43MB/s	73.66MB/s	1	18	64	12.5	3.00	2.44
ps_14	17.18	26.29%	17.41%	38.54%	36.40%	274.19%	6.27%	33.37%	24.83%	130.36MB/s	95.69MB/s	129.07MB/s	57.07MB/s	1	18	64	12.5	1.00	0.70
ps_15	17.18	30.44%	18.78%	53.29%	49.77%	588.12%	13.60%	75.52%	34.15%	174.90MB/s	130.43MB/s	295.01MB/s	102.97MB/s	1	18	64	12.5	4.00	3.29
ps_16	17.18	25.83%	16.62%	36.66%	34.35%	632.10%	12.67%	36.01%	27.23%	140.67MB/s	105.11MB/s	124.70MB/s	73.30MB/s	1	18	64	12.5	3.00	1.79
ps_17	17.18	25.42%	15.87%	31.24%	29.14%	1100.61%	11.55%	29.03%	21.60%	113.41MB/s	83.45MB/s	95.26MB/s	56.96MB/s	1	18	64	12.5	4.00	2.58
ps_10	17.18	24.14%	16.19%	40.61%	38.32%	169.38%	5.05%	33.39%	24.28%	130.42MB/s	93.78MB/s	113.65MB/s	83.06MB/s	1	18	64	12.5	6.00	5.07
ps_11	17.18	29.01%	17.98%	40.00%	37.85%	909.04%	12.31%	47.66%	36.93%	186.16MB/s	142.92MB/s	147.60MB/s	117.67MB/s	1	18	64	12.5	4.00	3.36
ps_12	17.18	28.60%	14.59%	43.03%	40.56%	1366.17%	16.04%	49.96%	23.67%	120.56MB/s	90.55MB/s	195.17MB/s	65.16MB/s	1	18	64	12.5	4.00	3.32
ps_13	17.18	24.02%	15.92%	41.44%	39.30%	279.42%	5.32%	36.43%	23.90%	125.63MB/s	91.96MB/s	142.31MB/s	74.92MB/s	1	18	64	12.5	5.00	4.16
reader_8	17.18	87.28%	16.23%	42.18%	37.91%	232.81%	5.18%	228.11%	41.88%	891.05MB/s	163.54MB/s	125.26MB/s	36.37MB/s	1	18	64	12.5	0.00	0.00
reader_9	17.18	92.69%	16.41%	42.40%	37.95%	241.15%	4.54%	222.06%	41.83%	867.42MB/s	163.34MB/s	114.15MB/s	36.22MB/s	1	18	64	12.5	0.00	0.00
reader_4	17.18	91.41%	16.27%	42.26%	37.86%	195.13%	5.74%	217.01%	41.96%	847.68MB/s	163.76MB/s	111.29MB/s	36.44MB/s	1	18	64	12.5	0.00	0.00
reader_5	17.18	92.62%	16.53%	40.60%	36.11%	782.13%	13.99%	215.16%	41.90%	840.48MB/s	163.59MB/s	108.43MB/s	36.37MB/s	1	18	64	12.5	0.00	0.00
reader_6	17.18	91.90%	16.44%	39.90%	35.47%	632.85%	14.72%	209.32%	41.93%	817.66MB/s	163.68MB/s	102.55MB/s	36.36MB/s	1	18	64	12.5	0.00	0.00
reader_7	17.18	94.55%	16.62%	38.38%	33.76%	611.94%	12.82%	222.71%	41.86%	869.97MB/s	163.48MB/s	122.66MB/s	36.29MB/s	1	18	64	12.5	0.00	0.00
reader_0	17.18	87.08%	16.31%	43.06%	38.68%	292.84%	5.22%	209.12%	41.88%	816.87MB/s	163.56MB/s	122.05MB/s	36.31MB/s	1	18	64	12.5	0.00	0.00

Example

| Path | Hours | CPU | Memory | IOwait | Network In & Out | RRU | CPU | RAM | NIC | Gflops |

ps_22	17.18	28.54%	18.77%	39.69%	37.62%	611.41%	12.38%	41.05%	27.43%	142.94MB/s	105.51MB/s	160.37MB/s	64.64MB/s	1	18	64	12.5	2.00	1.75
reader_13	17.18	93.26%	16.19%	42.43%	38.09%	179.83%	4.78%	230.77%	41.95%	901.46MB/s	163.75MB/s	115.56MB/s	36.59MB/s	1	18	64	12.5	0.00	0.00
ps_6	17.18	23.09%	14.48%	31.93%	29.81%	166.48%	4.29%	29.36%	21.18%	114.70MB/s	82.00MB/s	91.80MB/s	63.16MB/s	1	18	64	12.5	5.00	3.76
ps_7	17.18	26.48%	14.12%	32.33%	30.29%	1001.19%	12.62%	32.04%	23.22%	125.16MB/s	90.03MB/s	95.72MB/s	73.13MB/s	1	18	64	12.5	5.00	4.06
ps_5	17.18	21.44%	13.99%	35.80%	33.39%	133.38%	4.10%	32.64%	16.87%	92.92MB/s	64.61MB/s	127.50MB/s	47.08MB/s	1	18	64	12.5	5.00	3.69
ps_8	17.18	29.32%	16.14%	45.05%	41.69%	1207.63%	15.75%	52.32%	34.15%	173.80MB/s	131.29MB/s	204.38MB/s	107.10MB/s	1	18	64	12.5	4.00	2.80
ps_9	17.18	27.06%	15.87%	44.00%	41.81%	765.00%	10.90%	55.07%	23.93%	126.44MB/s	91.42MB/s	215.12MB/s	65.75MB/s	1	18	64	12.5	5.00	3.55
master	17.18	52.41%	38.06%	20.08%	19.30%	131.42%	3.54%	20.05%	14.94%	626.70MB/s	465.01MB/s	442.46MB/s	308.11MB/s	4	40	256	25	18.00	14.44
trainer_7	17.18	77.42%	44.19%	20.19%	15.18%	164.06%	4.57%	40.02%	28.70%	1250.64MB/s	889.62MB/s	1191.09MB/s	882.56MB/s	4	40	256	25	694.00	581.32
trainer_6	17.18	65.74%	50.22%	18.94%	15.80%	126.05%	5.02%	34.62%	27.86%	1081.81MB/s	870.43MB/s	808.98MB/s	629.88MB/s	4	40	256	25	532.00	361.07
trainer_5	17.18	53.87%	35.06%	17.72%	14.54%	109.88%	3.61%	37.83%	26.04%	1182.17MB/s	808.42MB/s	983.10MB/s	800.59MB/s	4	40	256	25	667.00	513.53
trainer_4	17.18	49.68%	35.43%	17.53%	14.55%	183.94%	3.63%	36.85%	28.38%	1151.59MB/s	876.01MB/s	1017.11MB/s	859.42MB/s	4	40	256	25	682.00	577.82
trainer_3	17.18	72.86%	47.97%	20.33%	17.72%	130.46%	3.74%	39.52%	30.50%	1235.08MB/s	952.67MB/s	980.70MB/s	769.01MB/s	4	40	256	25	665.00	474.88
trainer_2	17.18	70.00%	51.12%	23.14%	20.19%	121.48%	4.14%	40.38%	32.92%	1261.77MB/s	1028.13MB/s	996.26MB/s	748.26MB/s	4	40	256	25	674.00	447.59
trainer_1	17.18	81.31%	50.80%	18.42%	15.02%	76.34%	3.32%	54.65%	33.99%	1707.77MB/s	1056.90MB/s	997.32MB/s	701.01MB/s	4	40	256	25	672.00	461.61
trainer_0	17.18	62.96%	53.46%	83.63%	17.18%	1109.92%	12.66%	63.97%	24.12%	1999.07MB/s	749.76MB/s	849.39MB/s	694.90MB/s	4	28	256	25	559.00	450.15
ps_4	17.18	7.98%	3.29%	6.34%	5.94%	339.82%	10.01%	30.63%	25.15%	955.05MB/s	783.93MB/s	957.13MB/s	785.48MB/s	4	28	256	25	1.00	0.89
ps_2	17.18	9.97%	4.63%	6.35%	5.90%	463.93%	10.58%	30.23%	24.89%	942.62MB/s	775.96MB/s	944.55MB/s	777.54MB/s	4	28	256	25	1.00	0.89
ps_3	17.18	7.19%	3.41%	6.11%	5.74%	300.28%	6.32%	30.72%	25.16%	960.13MB/s	784.43MB/s	956.19MB/s	786.06MB/s	4	28	256	25	1.00	0.89
ps_0	17.18	9.38%	4.10%	5.52%	5.08%	289.92%	5.23%	30.38%	24.92%	946.92MB/s	776.64MB/s	949.37MB/s	778.64MB/s	4	28	256	25	2.00	0.89
ps_1	17.18	7.95%	3.98%	5.81%	5.35%	325.86%	8.48%	30.35%	24.92%	947.13MB/s	777.20MB/s	948.36MB/s	778.49MB/s	4	28	256	25	2.00	0.89

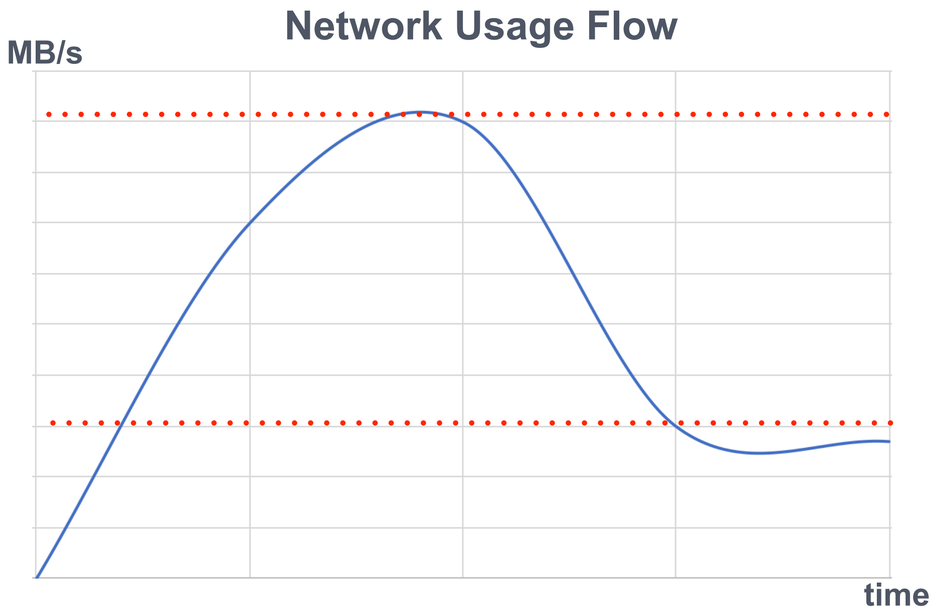
Tuning Process

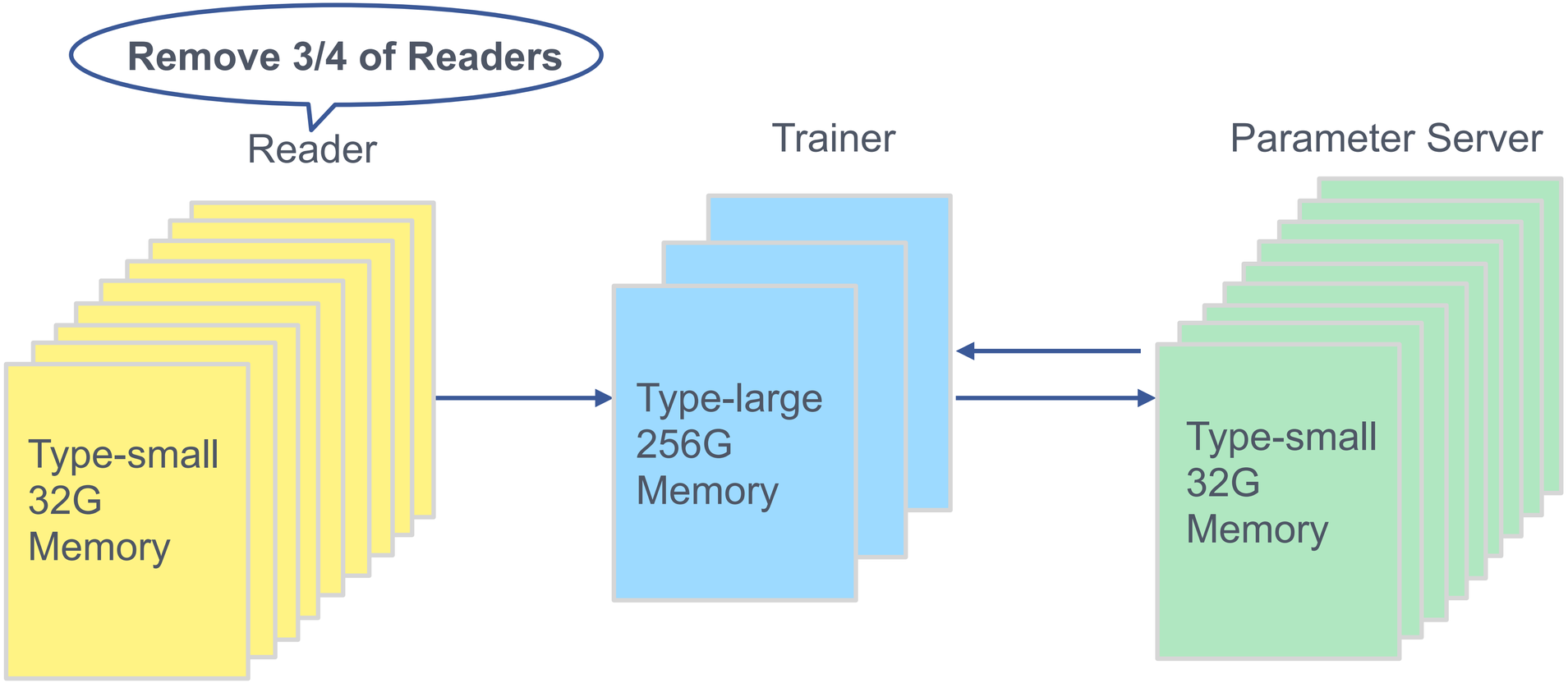


Network Usage Analysis

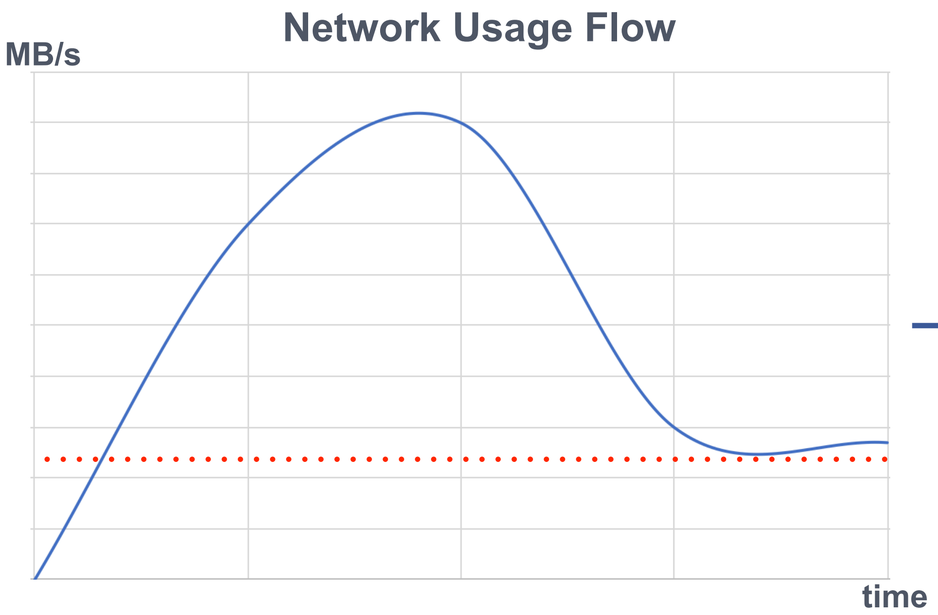
1/4

Path	Max Eth	Avg Eth	Max Eth In	Avg Eth In	Max Eth Out	Avg Eth Out
reader_8	228.11%	41.88%	891.05MB/s	163.54MB/s	125.26MB/s	36.37MB/s
reader_9	222.06%	41.83%	867.42MB/s	163.34MB/s	114.15MB/s	36.22MB/s
reader_4	217.01%	41.96%	847.68MB/s	163.76MB/s	111.29MB/s	36.44MB/s
reader_5	215.16%	41.90%	840.48MB/s	163.59MB/s	108.43MB/s	36.37MB/s
reader_6	209.32%	41.93%	817.66MB/s	163.68MB/s	102.55MB/s	36.36MB/s
reader_7	222.71%	41.86%	869.97MB/s	163.48MB/s	122.66MB/s	36.29MB/s
reader_0	209.12%	41.88%	816.87MB/s	163.56MB/s	122.05MB/s	36.31MB/s
reader_1	214.98%	41.91%	839.76MB/s	163.55MB/s	319.61MB/s	36.75MB/s
reader_2	231.18%	41.89%	903.04MB/s	163.57MB/s	117.33MB/s	36.35MB/s
reader_3	227.13%	41.86%	887.22MB/s	163.46MB/s	120.73MB/s	36.18MB/s
reader_11	229.03%	41.84%	894.66MB/s	163.37MB/s	115.93MB/s	36.03MB/s
reader_12	225.51%	41.86%	880.91MB/s	163.45MB/s	117.66MB/s	35.98MB/s
reader_14	225.77%	41.92%	881.90MB/s	163.66MB/s	118.96MB/s	36.25MB/s
reader_15	216.89%	41.95%	847.24MB/s	163.80MB/s	111.61MB/s	36.23MB/s

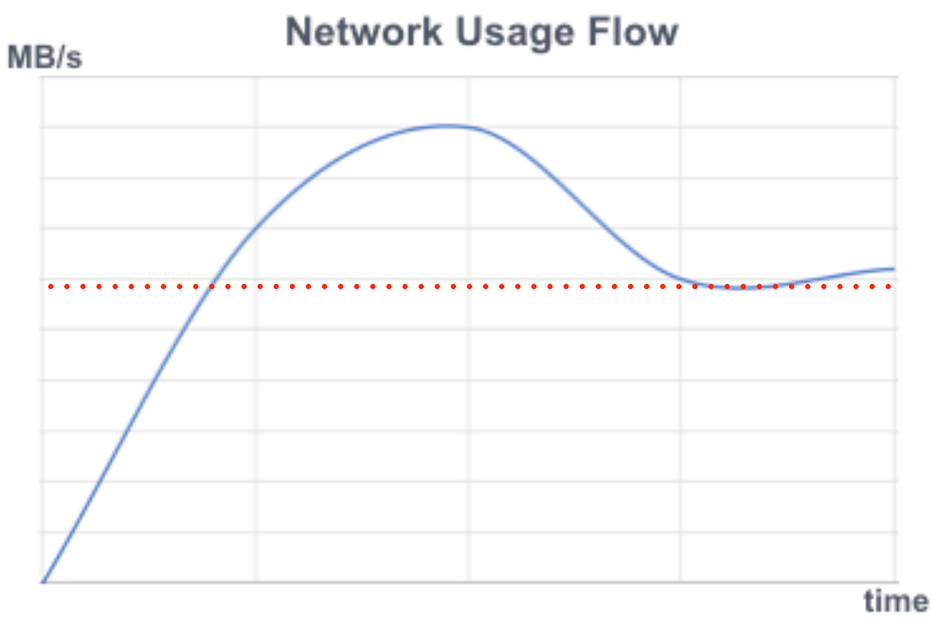


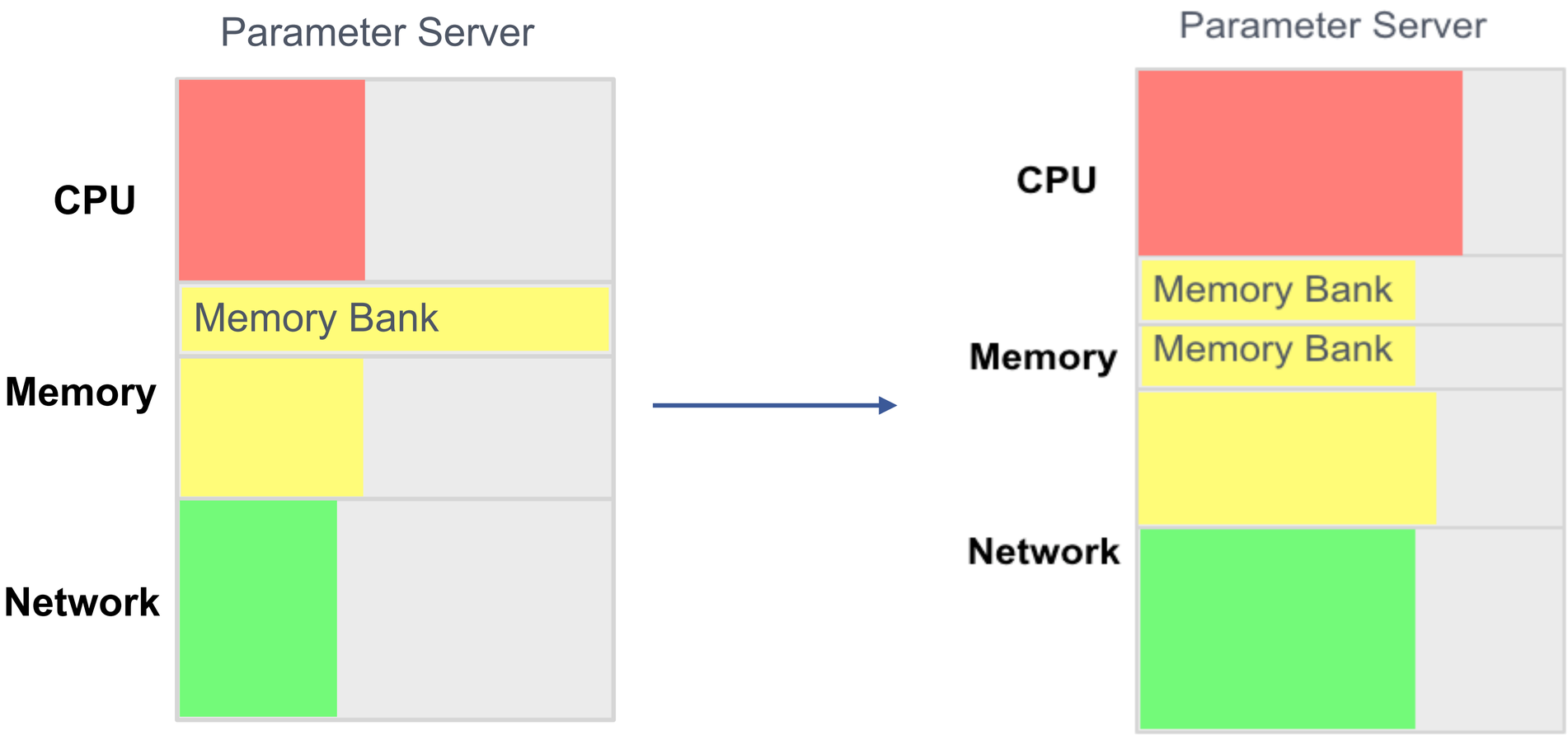

Remove 3/4 of Readers
Reader
Trainer
Parameter Server
Type-small
32G
Memory
Type-large
256G
Memory
Type-small
32G
Memory

Before Tuning:

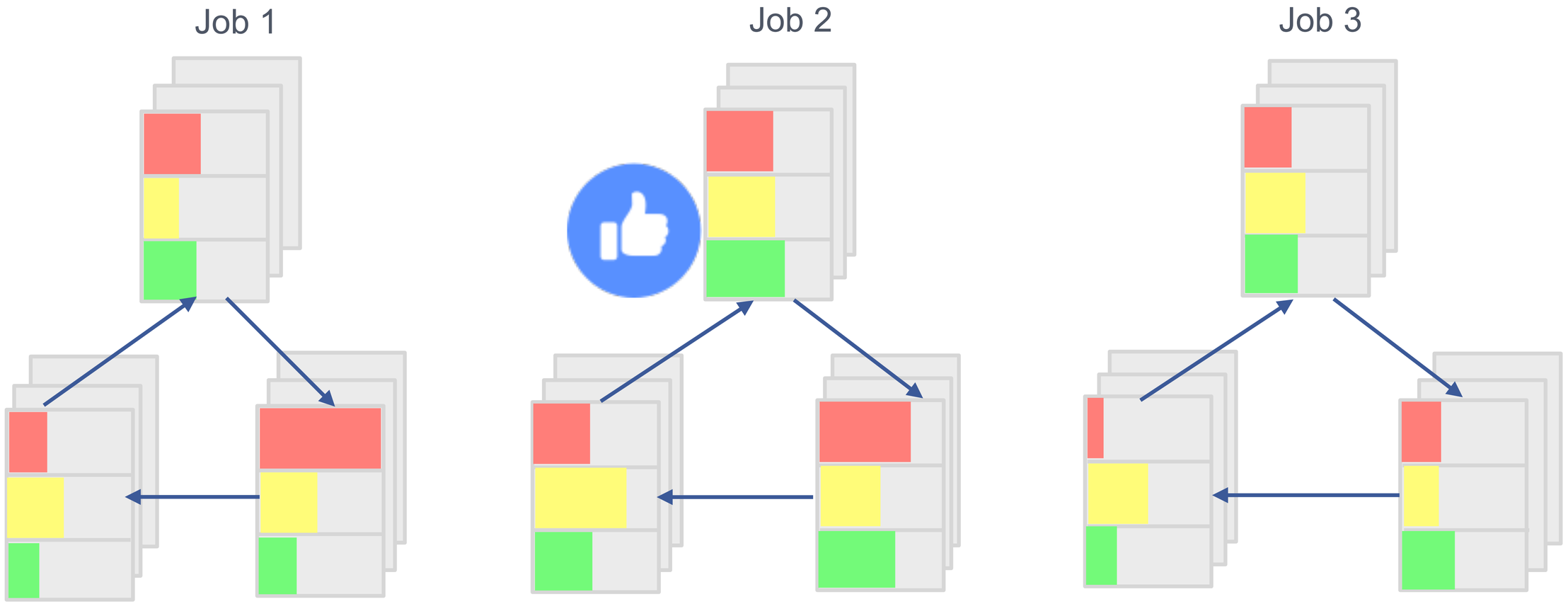


After Tuning:

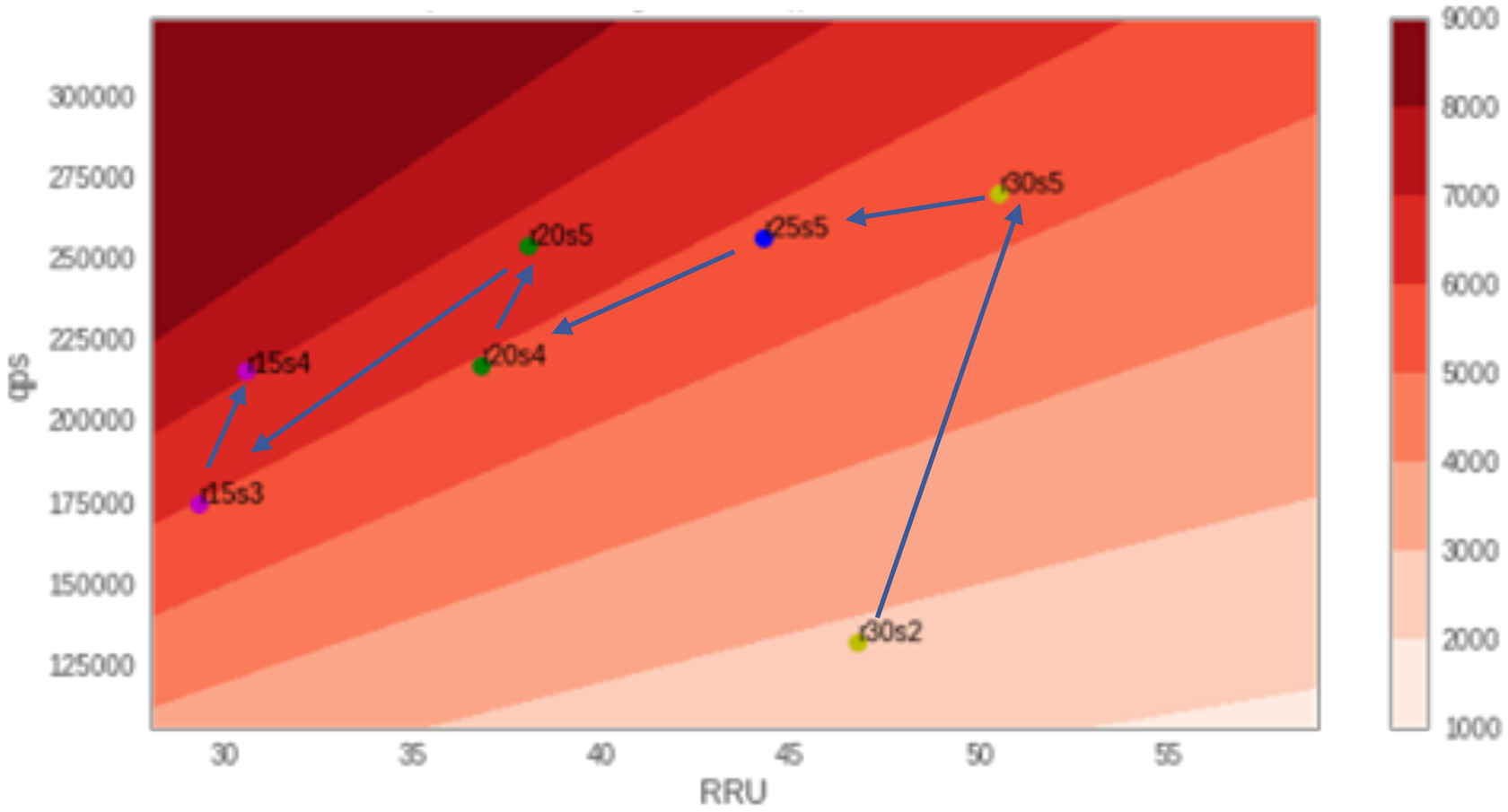



Parameter Server
CPU
Memory Bank
Memory
Network
Parameter Server
CPU
Memory Bank
Memory
Memory Bank
Network

Compare Multiple Jobs



Tuning with flow-info



facebook

Thank you

facebook

Q & A