LIQID

Case Study:

Reducing colocation OPEX and customer CAPEX with Liqid composable HPC

colovore

About Colovore

Colovore's unrivaled power and liquid cooling densities of up to 50 kilowatts (kW) per rack and a pay-by-the-kW pricing model set Colovore apart, but it's the company's fundamental knowledge of IT infrastructure and dedication to customer support and true partnership that really stand out in the colocation marketplace. The Silicon Valley-based data center operator has delivered multiple megawatts (MW) of high-density colocation services to customers ranging from Fortune 500s to startups without a single operational hiccup since inception.

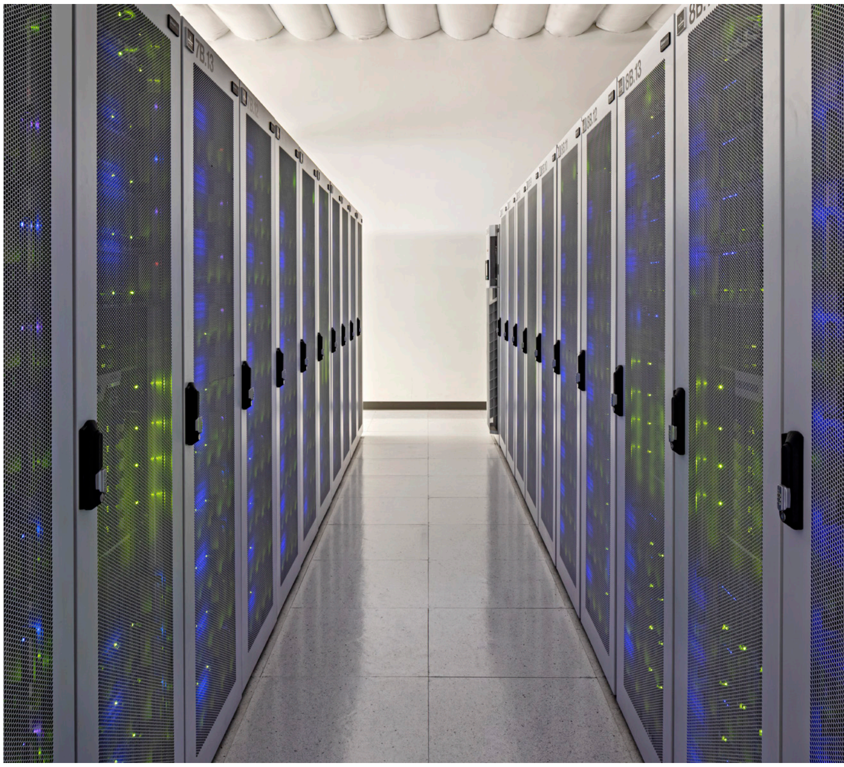
The company was founded in 2013 in Santa Clara and specifically engineered their colocation facility to power and cool the dense, multi-rack infrastructure required for next-gen artificial intelligence (AI), high-performance computing (HPC), Big Data, and other applications that are in high demand in the public and private sector.

Colovore Delivers Savings on Capital and Operational Expenditures for AI & HPC

As companies develop new AI and HPC applications, server requirements have become more sophisticated. GPUs have become table stakes, and compute architectures that support them have massive energy overhead and generate a lot of heat. Increasingly, data center environments like Colovore are designed with specialized water-based cooling to manage the thermal load generated from such large, densely-configured infrastructures. These high-value and processing-intensive servers require highly-specialized data center environments in order to optimize TCO and IT operational efficiency.

"The challenge is that most data centers are built to support only 3-5 kW of power and heat per rack. So racks can only be partially-filled with these powerful modern servers, and more and more rack and floor space is required. This increases operating costs and dramatically decreases application performance and scalability," said Ben Coughlin, Co-Founder, Colovore. "Based on what we've seen with NVIDIA and now many other GPU providers, demand for higher-power density colocation cabinets for customer deployments continues to increase. These users require the ability to cool 15-30 kW+ of heat load in a fully-packed cabinet."

A single rack unit of these servers can draw an entire kW of power—a massive leap from even three years ago. This doesn't include NVMe storage, GPU, FPGA, and other accelerator technologies that are needed to drive the data throughput requirements of these applications.



Previously, when customers deployed an HPC/AI stack, they could support just four or five systems in a rack. This lead to hardware sprawl, rack sprawl, and other inefficiencies, with some resources sitting idle and others overtaxed, and an IT footprint consisting of lots of empty, unusable, and wasted space.

"Customers deploying these dense servers come to us today saying 'wait, this doesn't work in my old data center environment, I'm running into power and cooling constraints. I need something different," said Coughlin. "That's what we provide. It ultimately comes down to enough power and cooling per square foot; these are the building blocks important for optimizing IT infrastructure deployments. For lower-density general-purpose compute servers, IT admins haven't had to think about this in the past, but when it comes to AI and HPC servers and their heat loads, the data center environment can make or break the deployment. If you don't have the right architecture, you're going to run into lots of problems, fast."

In addition to providing monthly operational savings for AI and HPC applications with facilities that can more efficiently handle modern data workloads, Colovore provides significant capital expenditure savings for its customers as well.

"The simple reality is that the fewer cabinets and floor space required to support your servers, the more you save. In terms of upfront capex, you save on racks, power circuits, in-cabinet PDUs, cabling, and top-of-rack switches right out of the gates. The build-out costs and deployment timeframes are significantly reduced. And the same holds true for capex savings as a customer expands in the future." Coughlin said.

Provide Data Dexterity for Dense Computing Requirements

For Colovore and its customers, massive data performance in the most efficient, cost-effective fashion is key to shared IT success, and computing driven by GPU and other accelerator technologies is essential to providing that performance. Investigating ways to offer the company's clients even further performance and efficiencies, as well as ways to more effectively pool and scale GPU, FPGA, NVMe storage, and storage class memory, Coughlin was introduced to Liqid's composable infrastructure solutions by way of a shared investor.

"Our clients come to us because they require enormous compute power that they can't effectively achieve on-premises or in legacy colocations. This is especially true of our Silicon Valley customers, where real estate is at a premium and data center demand far outpaces supply," said Coughlin. "It's very important that we provide them with the ability to pack the maximum amount of hardware and computing capacity into one cabinet before scaling out another. You can run out of space quickly."

Colovore's value proposition builds on the concept of 'scaling out' performance, which allows for IT users to share workloads across individual compute and accelerator devices in large quantities, but physically can be accomplished within an existing server cabinet footprint. "We call this 'scaling within.' Our customers want to be able to increase their workload capacities as simply as possible, and if this requires new servers to be added, they don't want to then have to pay for more new racks and floorspace due to power and cooling constraints in their existing racks. Scaling in this manner takes time, adds expense, and is far less efficient. Because we offer so much more power and cooling per cabinet, our customers can simply add those servers in the same existing racks and away they go."

The Liqid composable disaggregated infrastructure platform is perfectly suited for Colovore customers who can extend their ability to scale resources further within their high-density cabinets. Liqid Command Center software disaggregates the server components —CPU, GPU, FPGA, NVMe, Intel® Optane™ memory technology, and NICs—and places them into resource pools. Liqid's intuitive GUI, REST API, or CLI enables an end user to create servers in seconds from those resource pools to deploy applications on-demand.

For customers who choose to deploy Liqid CDI solutions, Colovore's high-density racks provide the ability to pool hundreds of GPUs in combination with other accelerators to create a system perfectly balanced to suit the needs of individual workloads, into a small, tight rack footprint. Resources can be shared across PCIe for top performance or across Ethernet or InfiniBand over longer distance. There's no need to physically reconfigure servers, manually add additional devices, or overprovision expensive hardware to support heavy workloads. These combined capabilities make the Liqid CDI platform the most comprehensive such solution in the business, and highly complementary to the core value proposition that Colovore offers its customers.

Partnering with Liqid for Massive GPU Composability

In collaboration, Liqid and Colovore can deliver systems that offer unparalleled GPU performance for AI workloads:

Configuration Overview	
Solution	Bespoke GPU Supercomputer
Compute	5x compute nodes (5TB DRAM)
GPU	100 NVIDIA Quadro RTX 8000 (3840GB RAM)
Storage	300TB NVMe SSD
Networking	20x 100GB/s (IB or Eth)
Video	Yes – Video Out Available
Architecture	Liqid CDI
Multi-Node	Yes

In application testing, the composable, disaggregated system achieved the following results:

Peer-to-Peer Latency Testing:

P2P Enabled (Latency)

P2P=Enabled Latency (P2P Writes) Matrix (us)

GPU	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
0	2.58	2.37	2.61	2.43	2.56	2.49	2.70	2.54	2.69	2.42	2.52	2.62	2.50	2.49	2.48	2.73	2.73	2.43	2.39	2.69
1	2.73	2.58	2.40	2.51	2.42	2.64	2.78	2.52	2.58	2.44	2.52	2.57	2.44	2.68	2.44	2.78	2.47	2.60	2.63	2.62
2	2.35	2.73	2.63	2.58	2.57	2.55	2.51	2.41	2.58	2.68	2.56	2.46	2.56	2.57	2.47	2.50	2.52	2.65	2.58	2.51
3	2.58	2.62	2.66	2.85	2.58	2.47	2.58	2.53	2.57	2.66	2.42	2.49	2.45	2.71	2.53	2.57	2.52	2.51	2.44	2.58
4	2.54	2.69	2.45	2.61	2.62	2.47	2.49	2.52	2.55	2.53	2.65	2.67	2.66	2.47	2.64	2.61	2.64	2.62	2.83	2.66
5	2.60	2.62	2.49	2.71	2.51	2.75	2.42	2.95	2.46	2.76	2.41	2.52	2.52	2.53	2.69	2.59	2.75	2.48	2.40	2.66
6	2.81	2.57	2.61	2.64	2.47	2.53	2.78	2.65	2.77	2.60	2.53	2.68	2.43	2.86	2.49	2.47	2.53	2.69	2.53	2.67
7	2.68	2.47	2.57	2.59	2.54	2.45	2.59	2.70	2.61	2.47	2.43	2.85	2.79	2.54	2.80	2.59	2.55	2.65	2.62	2.49
8	2.59	2.53	2.64	2.65	2.67	2.70	2.47	2.72	2.81	2.66	2.68	2.78	2.75	2.55	2.68	2.50	2.54	2.59	2.54	2.71
9	2.77	2.47	2.65	2.49	2.67	2.51	2.51	2.58	2.50	2.91	2.56	2.58	2.58	2.74	2.65	2.70	2.54	2.58	2.49	2.72
10	2.71	2.52	2.53	2.51	2.69	2.44	2.64	2.48	2.55	2.61	2.84	2.67	2.66	2.52	2.82	2.99	2.92	2.80	2.52	2.64
11	2.74	2.72	2.56	2.82	2.58	2.68	2.59	2.74	2.50	2.55	2.54	2.74	2.53	2.47	2.63	2.61	2.53	2.54	2.61	2.45
12	2.64	2.59	2.64	2.46	2.53	2.42	2.58	2.52	2.56	2.56	2.43	2.57	2.83	2.57	2.57	2.68	2.70	2.70	2.48	2.74
13	2.80	2.55	2.66	2.46	2.78	2.54	2.41	2.68	2.49	2.74	2.66	2.73	2.65	2.82	2.68	2.60	2.62	2.63	2.70	2.73
14	2.68	2.70	2.43	2.59	2.70	2.70	2.54	2.47	2.64	2.50	2.67	2.58	2.54	2.43	2.76	2.57	2.79	2.54	2.71	2.50
15	2.45	2.47	2.50	2.50	2.62	2.52	2.49	2.78	2.84	2.53	2.55	2.50	2.56	2.52	2.57	2.63	2.47	2.71	2.53	2.40
16	2.42	2.58	2.74	2.64	2.45	2.51	2.57	2.52	2.42	2.53	2.51	2.76	2.54	2.68	2.52	2.53	2.73	2.78	2.62	2.64
17	2.54	2.78	2.57	2.58	2.80	2.63	2.48	2.69	2.55	2.64	2.77	2.41	2.56	2.53	2.55	2.43	2.63	2.88	2.59	2.81
18	2.56	2.59	2.49	2.53	2.53	2.73	2.80	2.71	2.54	2.48	2.54	2.53	2.83	2.39	2.46	2.57	2.47	2.56	2.78	2.69
19	2.66	2.85	2.87	2.60	2.49	2.47	2.51	2.49	2.75	2.70	2.65	2.56	2.67	2.63	2.51	2.69	2.62	2.69	2.67	2.94

P2P Disabled (Latency)

P2P=Disabled Latency Matrix (us)

GPU	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
0	2.67	29.98	28.96	30.18	31.08	30.81	29.18	29.94	29.63	30.77	30.05	28.80	29.32	30.68	29.52	30.58	30.63	30.27	29.22	29.39
1	29.20	2.74	30.19	30.36	30.47	29.58	31.12	30.63	30.10	30.08	30.59	31.33	31.08	29.07	30.15	29.44	29.14	29.64	30.99	30.88
2	29.40	29.75	2.66	30.53	30.01	29.61	30.54	31.00	30.43	29.72	29.46	29.49	29.00	30.53	30.74	29.23	30.47	29.81	29.22	30.45
3	30.36	29.56	29.08	2.74	30.35	28.75	28.10	30.58	29.46	29.82	30.66	29.02	29.87	30.33	30.50	30.41	29.36	28.86	29.04	29.85
4	30.41	29.93	29.67	30.71	2.68	30.13	28.72	32.19	28.35	30.17	29.90	28.71	28.22	29.82	29.79	30.80	29.39	29.95	30.23	30.69
5	30.53	27.94	29.39	29.15	29.51	2.85	28.34	29.55	30.08	30.37	28.91	29.44	30.14	30.24	31.05	29.56	30.73	30.19	30.30	28.79
6	28.44	30.91	31.02	30.54	30.76	30.14	2.66	29.94	30.81	29.56	30.72	29.61	30.70	31.12	30.58	29.30	30.15	30.09	29.70	31.28
7	29.56	30.72	30.92	30.01	29.45	30.22	30.43	2.75	29.86	30.07	30.44	29.64	30.00	29.08	29.48	29.88	31.23	29.22	30.77	31.16
8	28.96	29.61	29.43	28.96	29.70	29.12	30.61	30.88	2.72	29.03	30.92	30.20	29.19	29.78	30.35	29.84	29.23	30.76	29.69	30.05
9	29.61	29.79	29.42	30.81	30.58	29.37	29.09	27.03	29.53	2.90	29.17	30.30	30.52	30.58	30.08	29.71	31.42	30.02	30.27	29.48
10	30.20	28.45	29.31	29.42	29.63	28.75	28.77	29.51	29.97	28.77	2.85	30.06	27.57	30.84	30.93	29.04	30.24	30.44	29.57	29.75
11	30.56	30.09	29.23	28.62	29.40	29.77	30.24	30.12	30.16	29.27	30.36	2.71	29.56	29.56	29.70	30.45	29.63	29.75	30.89	30.61
12	28.84	30.78	29.92	30.61	29.96	30.69	30.28	28.78	29.95	29.84	28.64	30.65	2.80	30.06	30.35	31.25	30.27	29.17	29.63	30.48
13	30.95	30.03	28.58	28.60	28.00	30.38	29.94	30.12	30.08	28.88	30.00	30.40	30.01	2.68	30.57	29.25	29.41	29.52	30.30	30.03
14	30.55	28.60	30.23	30.06	29.79	28.57	29.85	30.57	29.53	29.78	30.33	29.89	29.88	29.10	2.76	30.76	29.97	29.83	30.17	30.34
15	29.89	29.41	29.60	28.64	27.56	30.20	29.00	29.67	29.51	30.03	29.95	28.40	29.73	31.22	30.43	2.62	30.58	30.05	30.57	30.48
16	29.88	29.91	29.24	29.81	28.29	30.24	29.97	29.18	30.85	30.42	31.57	30.42	30.38	29.69	30.25	29.83	2.74	28.78	30.46	29.84
17	31.14	30.51	30.57	29.22	29.62	27.83	30.84	29.75	28.91	30.03	27.37	28.50	30.82	30.08	31.46	30.08	29.86	2.87	30.20	30.66
18	29.74	30.25	30.55	31.01	30.28	28.18	30.51	29.49	29.96	29.05	29.06	30.48	30.07	30.28	29.93	30.00	29.96	31.07	2.90	29.57
19	30.10	30.23	29.62	29.83	29.58	29.49	29.83	29.71	29.09	29.59	30.26	30.51	30.30	30.47	30.76	30.59	30.69	30.50	30.28	2.58

Peer-to-Peer Bandwidth Testing:

P2P Enabled (Bandwidth)

Bidirectional P2P=Enabled Bandwidth Matrix (GB/s)

D\D	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
0	540.28	25.03	25.02	25.02	25.03	24.33	24.33	24.20	24.34	24.29	24.28	24.30	24.30	24.30	24.18	24.35	24.31	24.31	24.29	24.28
1	25.02	543.05	25.02	25.02	25.03	24.33	24.33	24.35	24.34	24.21	24.29	24.29	24.31	24.30	24.18	24.34	24.35	24.34	24.32	24.27
2	25.04	25.02	541.77	25.02	25.01	24.34	24.36	24.35	24.34	24.36	24.31	24.25	24.16	24.31	24.29	24.35	24.33	24.35	24.35	24.26
3	25.03	25.02	25.02	539.61	25.02	24.33	24.35	24.32	24.34	24.21	24.31	24.30	24.27	24.31	24.32	24.35	24.35	24.34	24.34	24.32
4	25.02	25.03	25.02	25.03	543.56	24.21	24.33	24.23	24.20	24.20	24.18	24.24	24.31	24.26	24.34	24.22	24.22	24.38	24.38	24.29
5	24.34	24.34	24.31	24.34	24.28	540.42	25.03	25.03	25.02	25.03	24.34	24.35	24.31	24.35	24.22	24.30	24.29	24.31	24.31	24.22
6	24.34	24.34	24.32	24.35	24.30	25.02	532.94	25.04	25.02	25.03	24.34	24.35	24.36	24.22	24.34	24.31	24.20	24.23	24.31	24.26
7	24.36	24.35	24.24	24.34	24.32	25.03	25.03	535.84	25.02	25.02	24.35	24.36	24.36	24.33	24.31	24.31	24.32	24.30	24.29	24.29
8	24.33	24.34	24.31	24.33	24.21	25.03	25.03	25.03	540.28	25.03	24.35	24.25	24.34	24.34	24.19	24.30	24.25	24.32	24.28	24.34
9	24.28	24.23	24.35	24.27	24.25	25.02	25.03	25.01	25.02	532.50	24.36	24.27	24.37	24.36	24.27	24.24	24.23	24.25	24.18	24.33
10	24.30	24.29	24.30	24.29	24.21	24.34	24.25	24.33	24.35	24.25	533.64	25.04	25.03	25.03	25.04	24.33	24.33	24.24	24.27	24.19
11	24.29	24.30	24.29	24.31	24.17	24.35	24.34	24.27	24.33	24.26	25.04	535.10	25.03	25.01	25.03	24.34	24.34	24.31	24.26	24.24
12	24.30	24.28	24.25	24.28	24.24	24.34	24.36	24.34	24.35	24.34	25.04	25.02	531.46	25.03	25.02	24.28	24.35	24.34	24.35	24.32
13	24.21	24.26	24.31	24.30	24.32	24.25	24.35	24.36	24.34	24.32	25.02	25.03	25.03	533.53	25.03	24.35	24.35	24.35	24.35	24.36
14	24.23	24.23	24.24	24.32	24.17	24.22	24.21	24.30	24.22	24.29	25.03	25.03	25.04	25.03	529.80	24.22	24.27	24.33	24.28	24.36
15	24.35	24.35	24.35	24.32	24.29	24.31	24.28	24.29	24.30	24.16	24.33	24.33	24.31	24.33	24.27	528.59	25.04	25.03	25.03	25.03
16	24.35	24.26	24.34	24.36	24.19	24.28	24.29	24.31	24.31	24.26	24.32	24.34	24.33	24.33	24.31	25.01	530.50	25.04	25.03	25.03
17	24.35	24.35	24.37	24.36	24.36	24.32	24.31	24.29	24.29	24.33	24.27	24.33	24.35	24.34	24.34	25.03	25.02	529.96	25.02	25.04
18	24.30	24.35	24.28	24.36	24.32	24.31	24.31	24.32	24.32	24.15	24.35	24.32	24.35	24.33	24.20	25.03	25.03	25.03	532.27	25.03
19	24.22	24.36	24.22	24.35	24.39	24.24	24.18	24.32	24.25	24.34	24.20	24.26	24.29	24.21	24.26	25.03	25.01	25.04	25.03	532.91

P2P Disabled (Bandwidth)

Bidirectional P2P=Disabled Bandwidth Matrix (GB/s)

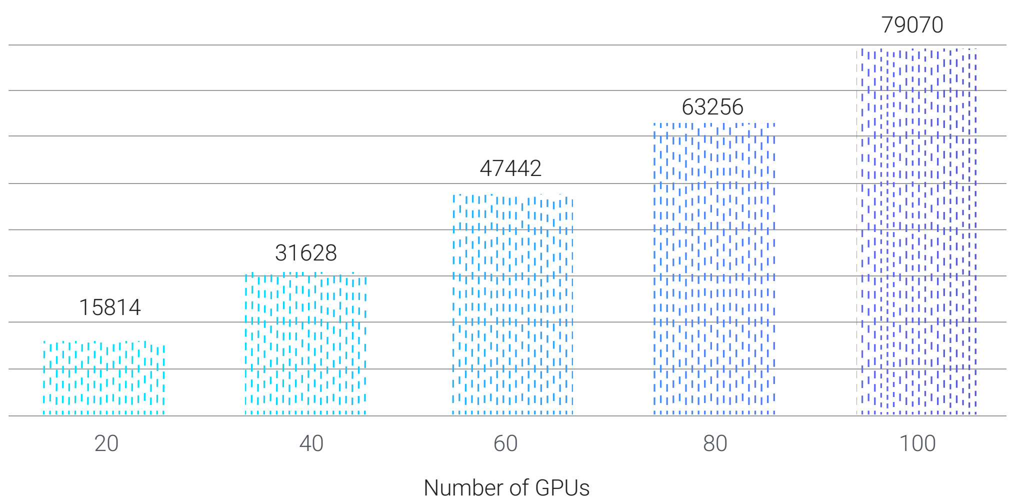
D\D	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
0	548.38	6.92	6.97	7.02	7.00	7.12	7.23	6.93	6.78	7.29	6.73	7.27	7.25	7.30	7.11	7.18	6.85	7.17	7.17	7.24
1	7.00	542.37	6.97	6.94	6.95	6.87	7.20	7.22	7.28	7.29	7.23	7.23	7.26	7.04	6.84	7.11	7.21	7.04	7.23	7.21
2	7.02	6.98	544.32	6.41	6.97	7.20	7.19	7.22	7.11	7.29	7.26	7.21	6.83	6.84	7.29	7.21	7.25	7.17	7.23	7.15
3	7.00	6.98	6.82	543.92	6.99	7.25	6.65	7.09	7.25	7.32	7.21	7.19	7.22	7.29	7.21	6.67	7.22	7.19	7.19	7.20
4	7.03	6.98	6.98	6.97	543.30	7.15	7.19	7.22	7.21	7.22	7.21	7.21	6.95	7.02	7.20	7.22	7.15	7.20	7.22	7.19
5	7.18	7.23	7.17	7.07	7.22	535.83	6.85	6.93	6.97	6.56	7.08	7.25	7.22	7.26	7.26	7.32	7.27	7.18	7.15	7.18
6	7.15	7.12	7.16	7.20	7.22	7.02	536.60	6.99	6.96	7.03	6.77	7.19	7.22	7.26	7.24	7.24	7.24	7.27	7.15	7.23
7	7.16	7.27	7.13	7.34	7.16	7.01	6.95	535.84	6.99	6.99	7.22	7.25	7.22	7.22	7.19	7.16	7.25	7.22	7.16	7.28
8	7.24	7.17	7.18	7.17	7.22	6.99	7.01	6.98	536.57	6.95	7.26	7.20	7.03	6.95	7.21	7.26	7.12	7.22	7.28	7.16
9	7.27	7.24	7.20	7.18	7.13	7.08	6.84	6.98	6.79	530.12	7.25	7.21	7.23	7.18	7.20	7.21	7.18	7.21	7.22	7.21
10	7.22	7.19	7.11	7.19	7.08	7.15	7.21	7.20	6.92	7.25	537.22	6.37	6.38	6.35	6.40	7.24	7.05	7.29	6.88	7.30
11	7.26	7.22	7.18	7.14	7.21	7.15	7.25	7.23	6.95	7.23	6.39	536.57	6.36	6.40	6.40	7.26	7.15	7.21	7.27	7.22
12	7.18	7.24	7.25	7.24	7.23	7.22	7.23	7.18	7.21	7.23	6.31	6.26	534.28	6.38	6.18	6.98	7.15	7.26	7.31	7.25
13	7.29	6.82	7.23	7.26	7.25	7.27	7.27	7.28	7.20	7.25	6.38	6.40	6.41	530.62	6.40	7.29	6.64	7.30	7.21	7.20
14	7.23	6.76	7.25	7.24	7.18	7.33	7.29	7.24	7.29	7.16	6.39	6.39	6.39	6.42	535.86	7.26	7.23	7.23	7.17	7.18
15	7.21	7.20	7.19	7.14	7.27	7.22	7.21	7.14	7.23	7.21	6.86	7.27	7.15	6.85	7.27	532.72	6.56	6.96	7.02	7.07
16	6.54	7.25	7.17	7.19	7.18	7.26	7.22	7.24	6.70	6.99	7.19	7.23	6.94	7.33	7.25	6.92	537.31	6.94	6.66	7.05
17	6.63	7.18	6.84	7.23	7.27	7.24	7.18	7.07	7.25	7.12	7.15	7.14	7.19	7.12	7.22	6.95	7.01	535.55	6.92	6.61
18	7.22	7.21	6.68	7.26	7.22	7.18	7.22	7.16	7.21	7.25	6.86	7.23	7.28	7.24	7.22	7.08	6.98	7.04	537.32	6.95
19	7.22	7.24	7.19	7.28	7.29	6.74	7.17	7.30	7.20	7.18	7.20	7.14	7.15	7.16	6.61	7.06	6.95	7.00	6.92	536.57

Resnet Performance Testing Summary

The Resnet50 benchmark is used to measure performance capabilities of GPU-based systems. The below data was measured on a TensorFlow Resnet50 benchmark using up to 100x NVIDIA Quadro RTX 8000 GPUs connected to five compute nodes using Liqid Command Center and a PCIe fabric. For testing, a TensorFlow batch size of 1024 was used and 100x NVIDIA RTX 8000 GPUs were composed and the Liqid composable system was able to achieve an image training throughput of more than 79K images per second.

Below are the test results of the benchmark Resnet50 test as measured on the Liqid ThinkTank composable system with 100x NVIDIA Quadro RTX 8000 GPU.

Resnet50 - Images/Sec
Liqid ThinkTank With Up to 100x RTX8000



Resnet50/ImageNet

# of GPUs	Images/second
20	15814
40	31628
60	47442
80	63256
100	79070

Next Generation Colocation

Working together, Liqid and Colovore are designing composable systems leveraging Colovore's unique power and cooling capacities as a colocation provider for next-gen AI deployments and other high-value applications. With the ability to pool GPU resources at unprecedented volume across off-the-shelf server hardware, Liqid Command Center software delivers comprehensive, industry-leading GPU performance across all industry-standard fabrics, and maximum datacenter efficiency for Colovore's customers.

Liqid software also enables GPUs to be deployed in quantities that match the performance requirements of any given application. Any number of GPUs can be shared to any number of servers via software to ensure maximum utilization. GPUs can be moved between servers in real-time without ever touching the box. With this ability, Colovore's customers minimize waste by eliminating traditional hardware silos that limit hardware resource agility at the point of installation.

As solutions and services based on NVIDIA's powerful new A100 line of GPUs, become more common, data center power and cooling requirements will continue to increase, driving new clients to high-density colocations like Colovore.

"IT organizations are very excited about what they can accomplish with GPU performance, but they have now realized that they don't have the capabilities on-prem to power and cool them efficiently, so they come to us before any final contracts get signed," Coughlin said. "The ability to further optimize resource usage with composable infrastructure provides another compelling reason for these companies to choose Colovore to accommodate the architectural requirements of such powerful systems."

Further, borrowing from the cloud model, high performance colocation facilities like Colovore, in cooperation with partner organizations such as Liqid, can build performance-oriented, composable, disaggregated systems densely populated with data accelerators at previously impossible scale, and sell computing time-as-a-service. IT users will be able to share and scale GPU, FPGA, NVMe and more at unprecedented levels, in any amount they require, and pay for the compute time they need without the overhead of maintaining an on-prem high-density rack or paying a colocation for devices that sit idle.

"Each of our cabinets provides a high-density power circuit, for example a 415 volt/ 60amp/ 3 phase circuit delivering 35 usable kW of power. But we only charge by the kW consumed. So as workloads and power draws fluctuate based on the pooling and disaggregation of resources enabled by Liqid, customers can scale up their power and pay for it, 1 kW at a time, as they increase those workloads," said Coughlin. "They don't get stuck having to pay for the entire 35 kW, when it may only be needed from time to time. This complements the Liqid value proposition nicely."

"We look forward to working closely with the Liqid team to solve problems and create opportunities for organizations to take full advantage of the application-level advances taking place in artificial intelligence and HPC."

About Liqid

Liqid provides the world's most-comprehensive software-defined composable infrastructure platform. The Liqid platform empowers users to manage, scale, and configure physical, bare-metal server systems in seconds and then reallocate core data center devices on-demand as workflows and business needs evolve. Liqid Command Center software enables users to dynamically right size their IT resources on the fly. For more information, contact info@liqid.com or visit www.liqid.com. Follow Liqid on Twitter and LinkedIn.

About Colovore

Colovore is the Bay Area's leading provider of high-performance colocation services. Our 9MW state-of-the-art data center in Santa Clara features power densities of 35 kW per rack and a pay-by-the-kW pricing model. We offer colocation the way you want it—cost-efficient, scalable, and robust. Colovore is profitable and backed by industry leaders including Digital Realty Trust. For more information please visit www.colovore.com.