

Cloud repatriation and the death of cloud-only



Dan Swinhoe
News Editor

Companies are realizing a hybrid mix of cloud and on-prem could be a better strategy than a binary approach

"We're seeing that the application would have actually done better in a private environment from a cost and latency standpoint"

Once it began gaining traction, the cloud was billed as a panacea for companies looking to reduce their IT costs. For many, going all-in on the cloud meant the capex costs of data centers could be done away with and expensive legacy infrastructure closed and sold off.

Browse the pages of *DCD* and most of our enterprise stories are companies migrating en masse to public cloud as part of a 'cloud-first/cloud-native/cloud-only' strategy.

But recent headlines are suggesting that companies may be retreating back from the cloud and 'repatriating' data, workloads, and applications back to on-premise or colocation facilities. Are companies really growing cold on the cloud, or is this hype from hardware vendors and service providers looking to cash in?

Cloud repatriation: is it real?

Does the promise of cloud still ring true? By and large, yes. Companies of all shapes and sizes rely on public cloud infrastructure for a variety of mission-critical workloads.

However, the idea of going all-in on 'cloud-only' and abandoning on-premise and/or colocation facilities is fading away in lieu of a more hybrid 'cloud-and' approach.

Most of the service and colocation providers *DCD* spoke to for this piece said repatriation was happening, though to varying degrees. IT analyst firm IDC told us that its surveys show repatriation as a steady trend ‘essentially as soon as the public cloud became mainstream,’ with around 70 to 80 percent of companies repatriating at least some data back from public cloud each year.

“The cloud-first, cloud-only approach is still a thing, but I think it’s becoming a less prevalent approach,” says Natalya Yezhkova, research vice president within IDC’s Enterprise Infrastructure Practice. “Some organizations have this cloud-only approach, which is okay if you’re a small company. If you’re a startup and you don’t have any IT professionals on your team it can be a great solution.”

While it may be common to move some workloads back, it’s important to note a wholesale withdrawal from the cloud is incredibly rare.

“What we see now is a greater number of companies who are thinking more about a cloud-also approach,” adds Yezhkova. “They think about public cloud as an essential element of the IT strategy, but they don’t need to put all the eggs into one basket and then suffer when something happens. Instead, they have a more balanced approach; see the pros and cons of having workloads in the public cloud vs having workloads running in dedicated environments.”

“We are seeing it, it is real. But we’re really seeing a future that’s hybrid,” says DataBank CTO Vlad Friedman. “What we’re really seeing is people are more intelligently thinking about placing their workloads.”

Who is repatriating data?

While any company could reclaim data back from the cloud, most of the companies *DCD* spoke to said enterprises that made wholesale migrations to the cloud without enough preparation and cloud-native startups that had reached a certain scale were the most likely candidates for repatriation.

Dropbox notably pulled significant amounts of data back from Amazon Web Services (AWS) between 2013 and 2016. According to a 2021 interview with DCD, the decision to remain on-prem – known as ‘Magic Pocket’ – worked out significantly cheaper and gave Dropbox more control over the data the company hosted. The company still also uses AWS where required, however.

More recently, web company 37signals – which runs project management platform Basecamp and subscription-based email

service Hey – announced the two services were migrating off of AWS and Google Cloud.

“We’ve seen all the cloud has to offer, and tried most of it,” CTO and co-founder David Heinemeier Hansson said in a blog post. “It’s finally time to conclude: Renting computers is (mostly) a bad deal for medium-sized companies like ours with stable growth. The savings promised in reduced complexity never materialized. So we’re making our plans to leave.”

37 Signals didn’t reply to *DCD*’s request for an interview, but in subsequent posts the company said it was spent more than \$3 million on the cloud in 2022 and would be saving some \$7 million over five years by switching to Dell hardware – some \$600,000 worth – located in two colocation facilities in partnership with Deft who will manage the installation.

“Any mid-sized SaaS business and above with stable workloads that does not benchmark their rental bill for servers in the cloud against buying their own boxes is committing financial malpractice at this point,” Hansson said. The migration is underway and expected to be completed this year.



"If a new workload can only be run into dedicated environments because of regulatory requirements, what does it mean for other workloads?"



Some repatriation is like 37signals; cloud-native startups that have reached a scale where it can be more economical to switch to on-prem. Another part of the repatriation picture is companies that may have done a full 'lift and shift' of their IT estate and later realized not everything may be suited for cloud, especially if it hasn't been refactored and modernized.

For companies that have a heavy on-prem/colo footprint, many new workloads may start in the cloud during development and while they are ramping up to benefit from the speed and flexibility cloud offers. But after a certain level of maturation or compliance threshold – or as soon as IT finds out the workload exists in some cases – applications will then need to be brought home.

DataBank CTO Friedman notes that he has seen a number of service providers repatriate once they reach a certain scale, and are moving back steady-state, computationally- or I/O-intensive applications.

"They're figuring out a hybrid architecture that allows them to achieve savings. But I don't think it's a pure-play move back to colo; it's about moving the right workloads back to colo. Because 'colo' or 'cloud' is not the desired outcome; it's efficiency, it's performance, it's lower latency."

Planning is important – companies that just look to do a straight lift and shift to the cloud could see costs increase and performance suffer and lead to regret and repatriation.

"I previously worked with a large system integrator in the Nordic region that had set out to move 80 percent of its workloads into the public cloud. After three years of laborious efforts, they had moved just 10 percent before aborting the project in its entirety and deferring back to on-premise," says Tom Christensen, global technology advisor and executive analyst at Hitachi Vantara.

Why are companies bringing workloads back?

IDC's Yezhkova tells *DCD* that security remains a major driver of repatriation, though this has declined in recent years.

One of the biggest drivers is simply cost – the wrong workloads in the wrong configurations can cost more in the cloud than on-prem.

"Public cloud might be cheaper, but not always," says Yezhkova. "Ingress-egress fees, the data transfer fees, they add up. And as the workload grows, companies might realize that it's actually cheaper to run these workloads in on-premises environments."

A *2021 report* from Andreessen Horowitz noted that cloud repatriation could drive a 50 percent reduction in cloud spend, but notes it is a "major decision" to start moving workloads off of the cloud.

"If you're operating at scale, the cost of cloud can at least double your infrastructure bill," the report said. "You're crazy if you don't start in the cloud; you're crazy if you stay on it."

Likewise, cloud costs can be far more unpredictable than on-premise equivalents, especially if configured incorrectly without spending controls in place. For complex deployments, the added cost of a service provider to manage cloud environment can add up too.

Data sovereignty demands can be a driver in some markets. Countries with stricter data residency laws may force enterprises to keep data within their own borders – and in some cases out of the hands of certain companies under the purview of data-hungry governments. Many cloud providers are looking to offer 'sovereign cloud' solutions that hand over controls to a trusted domestic partner to overcome some of these issues.

Latency, performance, and management may be a driver – super latency-sensitive applications may not meet performance expectations in public cloud environments to the same degree they might on-premise or in colocation sites.

"We're seeing these companies that may have done a lift and shift and thought that this would have been better," says DataBank's Friedman, "but the application would have actually done better in that private environment from a cost and latency standpoint."

Add in the fact that companies inevitably need as many (if not more) people to manage increasingly complex and intertwined environments as they did on-premise, and some companies may prefer to just keep things in-house.

SEO software firm Ahrefs recently posted its own calculations, saying it had saved around \$400 million in two years by not having a cloud-only approach.

The company calculated the costs of having its equivalent hardware and workloads entirely within AWS' Singapore region over the last two years, and estimates the cost would be \$440m, versus the \$40m it actually paid for 850 on-premise servers during that time.

"Ahrefs wouldn't be profitable, or even exist, if our products were 100 percent on AWS," Ahref's data center operations lead Efim Mirochnik said, although critics have noted that its cloud cost estimates were severely underoptimized and that a more honest evaluation would have shown a smaller gulf in costs.

What workloads should and shouldn't be in the cloud?

What workloads make the most sense to sit in the cloud versus on-premise will vary depending on a number of factors including the application, its users, geography, and more. The same workload may have different answers at different times.

"A workload is like a living thing and the requirements might change," says IDC's Yezhkova, who says that the only definable 'tipping point' for when a workload should be repatriated is when performance could be impacted by keeping it untouched.

When it comes to data sovereignty, application interdependency may mean it's easier to bring all the workloads closer to the most tightly regulated workloads, rather than constantly shifting information between on-premise and the cloud.

"If a new workload can only be run into dedicated environments because of regulatory requirements, what does it mean for other workloads?" says Yezhkova. "It might have this snowball effect on other workloads."

"It's about moving the right workloads back to colo. Because 'colo' or 'cloud' is not the desired outcome; it's efficiency, it's performance, it's lower latency"

Workloads that see data moving constantly back and forth in and out of clouds between applications could be candidates for moving as a way to avoid ingress/egress fees; getting data out of the cloud is more expensive than putting it in there.

"Companies need to do cost analyses on a regular basis," Yezhkova adds. "It shouldn't be done once and everybody forgets about it, it should be a regular exercise."

Workloads with high variability are natural bedfellows to sit in the cloud. Backup data that are fairly static in nature could also be good options to keep in the cloud.

Latency-sensitive Edge use cases are another where the public cloud might not make as much sense – Ensono principal consultant Simon Ratcliffe notes manufacturing often needs on-site compute that may not be suitable for cloud – although the introduction of Edge locations such as AWS Local and Wavelength Zones may alleviate issues there.

Hardware performance may also be a driving factor. Jake Madders, director of Hyve Managed Hosting, tells *DCD* he recently deal with a finance client that needed a particularly high clockspeed that wasn't available through public cloud instances, and required a customized server distributed to around 20 locations worldwide.

And what about HPC and mainframes? Most of the cloud providers now offer mainframe modernization services aimed to get those workloads off-premise, while many are *making major head roads* into cloud-based HPC.

Most of those we spoke to for this piece agree mainframes are unlikely to move quickly in the cloud, simply due to their speed, cost, and reliability, coupled with the difficulty of migrating the most legacy of complex workloads.

Ensono's Ratcliffe even notes that companies with spare capacity on their existing mainframes may well find Linux workloads can be run cheaper and more efficiently on something like an IBM Z-System than a cloud instance.

"One of our biggest customers is Next," he says. "The whole Next directory is run on a very modern mainframe, they have no intention of changing it."

"However, everything that goes around it that connects it to the user experience, is written in public cloud."

On HPC, there are arguments for and against on both sides. "We recently built two high-performance computing scenarios for two clients: One each in the data center and the equivalent inside Azure," says Ratcliffe.

"One decided to keep it on-prem, because of how they collect and feed the data into it. For the other client Azure came out as the right answer for them, again driven about how they collect their data."

"The first organization collects all their data in one place and it's right next to their data center and that works for them. The second company is all over the world and always connecting in through the Internet anyway. The answer is both because they're driven by different use cases, and by different behaviors."

Hyve's Madders adds that his company did a similar cost analysis for a BP subsidiary looking at an HPC-type deployment looking at seismic processing data for well-digging.

The companies looked at two options; one was leasing five or so racks within a Hyve environment in an Equinix facility, or in the cloud and running. The system would generally be running one week of intensive processing calculations until the next batch a month or so later.

"We worked it out that with a public cloud environment, the cost of compute required to do one week's calculation cost the same as buying the kit and then reselling it after the month."

On-premise can offer 'as-a-Service' pricing

Colocation seems to be much more of a common destination for reclaimed workloads. Many companies simply won't be interested in the investment and effort required to stand up a whole new on-premise data center.

But even if companies are looking to colos to save on costs, they should still be cognizant of the investment required in people, hardware, networking, and any other equipment that physical infrastructure will be needed.



What companies need to know for a repatriation project

Understand your network, applications, and dependencies: Understand what workloads you have where – including VMs, number of CPUs, and RAM – and how they are interconnected. Figure out where you might be reliant on locked-in cloud services and what the alternatives might be.

Figure out why you're having issues: If you're spending more than you thought, or an application isn't performing as well as it should, understand why. Sometimes it might be easier to update the application and keep it where it is.

Identify the right workloads to bring back: Steady-state, predictable, computationally or I/O-intensive workloads, and those workloads sensitive to latency will be best positioned for repatriation. Understand application dependency and the latency between any workloads affected by a move.

Figure out the costs: Reclaiming data could result in a change between capex and opex and may require a longer view of things. Understand what money you'll need upfront and how that affects your recurring bills down the line. Bandwidth fees might briefly increase during migration.

Have the right skills: Whether in-house or through a partner, make sure you have the right skills on hand to manage the IT and data center infrastructure.

Test: Test everything works before you begin to exit the cloud in earnest. As downtime window acceptability narrows, the margins for error narrows. Test the latency, test the hardware, run simulation workloads, do failover exercises.

Include room to grow: On-premise workloads, even those in private clouds, don't have the same near-infinite capacity to scale as public cloud. Be sure you plan ahead and have room to grow without issue.

However, the upfront costs don't have to be as high as they once were. Many companies are now offering IT hardware in an 'as-a-Service' model, reducing capex costs and turning physical IT hardware into an opex cost. Companies include HPE with its Greenlake offering, Dell with Apex, Pure Storage with Evergreen, and NetApp with Keystone.

And for those that demand on-premise enterprise data centers, companies can likewise start to expect similar pricing models. IDC research director Sean Graham recently told us that some vendors are making their data center modules available in an as-a-Service model, essentially making the containers available as a recurring opex cost.

Bringing the cloud on-premise

Many companies are seeking hybrid cloud deployments that combine some on-premise infrastructure with certain workloads in the cloud. Other are pursuing a private cloud approach, which offers some of the benefits of virtualized infrastructure in a single-tenant environment, either on-premise or in a colo.

But what about the companies wanting the actual cloud on-premise? While the cloud providers would be loath to admit it, they know they will never have all the world's workloads sitting in their environments. Proof of this is in their increasing on-premise offerings.

The likes of AWS Outpost, Azure Stack, Google Anthos, and Oracle's Cloud@Customer & Dedicated Regions offer on-premise hardware linked to their respective companies' cloud services and allow environments to be managed through one console. DataBank's Friedman adds that his company is seeing a number of these 'on-premise cloud' redeployments.

"They are for when the customer has bought into the ecosystem and the APIs, and they don't want to change their code, but they want efficiency or to place workloads tangential to their database and AI analytics. It's really a latency elimination play," he says.

"From the end-user perspective, they're getting the same experience," adds IDC's Yezhkova, "but the workloads are running in on-premises environments. Cloud providers realize that it's inevitable and if they want to be part of this game they need to introduce these types of offerings.

"Previously bringing IT back on premises would be a big deal, it would be

a new capital investment. Now with these types of services and the cloud platforms moving into on-premises, that transition is becoming easier."

Repatriation is as big of a project as the original migration

Clawing back data from the cloud is no easy or quick fix. Everyone *DCD* spoke to for this piece acknowledged it should be viewed as a standalone project with the same kind of oversight and gravitas as the migration projects that might have put data and workloads into the cloud in the first place. While the actual data transfer over the wire may only take hours if everything is set up and ready, the build-up to prepare can take months.

Prevention can be better than the cure, and where possible, companies should first look at why an application is having issues and if it can be solved without being moved. Luc van Donkersgoed, lead engineer for Dutch postal service PostNL, recently posted [on Twitter](#) how a single-line bug cost the company \$2,000 in AWS spending due to the fact it made more API calls than expected.

While some applications may be slightly easier to bring back if they've been refactored for cloud and containerized, companies still need to be aware of interdependency with other applications and databases, especially if applications being reclaimed are tied to platform-specific workloads.

A set of isolated virtual machines in a containerized system will be much easier to bring home than a large application tied to several AWS or Azure-only services and connected to latency-sensitive serverless applications, for example.

"If there really is no other alternative [than to repatriate], then look at it as if from a point of view you're planning to start an application from scratch in terms of infrastructure," says Adrian Moir, senior product management consultant & technology strategist at Quest Software.

"Make sure you have everything in place in terms of compute, storage, networking, data, protection, security, everything ready to go before you move the data set.

"It's almost like starting from scratch, but with a whole bunch of data that you've got straight away to go and deal with. It's going to be an intensive process, and it's not going to happen overnight. It's going to be one of those things that need to be thought through carefully." ■