HyperCloud Product Review "Half a rack in half a day"

A BROADBAND-TESTING REPORT



First published February 2023 (V1.3)

Published by Broadband-Testing

E-mail : info@broadband-testing.co.uk Internet: http://www.broadband-testing.co.uk

@2023 Broadband-Testing

All rights reserved. No part of this publication may be reproduced, photocopied, stored on a retrieval system, or transmitted without the express written consent of the authors.

Please note that access to or use of this Report is conditioned on the following:

- 1. The information in this Report is subject to change by Broadband-Testing without notice.
- 2. NO WARRANTIES, EXPRESS OR IMPLIED ARE GIVEN BY Broadband-Testing. ALL IMPLIED WARRANTIES, INCLUDING IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NON-INFRINGEMENT ARE DISCLAIMED AND EXCLUDED BY Broadband-Testing. IN NO EVENT SHALL Broadband-Testing BE LIABLE FOR ANY CONSEQUENTIAL, INCIDENTAL OR INDIRECT DAMAGES, OR FOR ANY LOSS OF PROFIT, REVENUE, DATA, COMPUTER PROGRAMS, OR OTHER ASSETS, EVEN IF ADVISED OF THE POSSIBILITY THEREOF.
- 3. This Report does not constitute an endorsement, recommendation or guarantee of any of the products (hardware or software) tested or the hardware and software used in testing the products. The testing does not guarantee that there are no errors or defects in the products, or that the products will meet your expectations, requirements, needs or specifications, or that they will operate without interruption.
- 4. This Report does not imply any endorsement, sponsorship, affiliation or verification by or with any companies mentioned in this report.
- 5. All trademarks, service marks, and trade names used in this Report are the trademarks, service marks, and trade names of their respective owners, and no endorsement of, sponsorship of, affiliation with, or involvement in, any of the testing, this Report or Broadband-Testing is implied, nor should it be inferred.



Broadband Testing

Broadband-Testing is an independent testing operation, based in Europe. Broadband-Testing interacts directly with the vendor, media, analyst, consultancy and investment communities equally and is therefore in a unique space within IT.

Testing covers all aspects of a product/ service from business rationalisation in the first instance to every element – from speed of deployment and ease of use/ management, through to performance and accuracy.

Testing itself takes many forms, from providing due diligence for potential investors through to public domain test reports.

Broadband-Testing is completely vendor neutral and independent. If a product does what it says on the tin, then we say so. If it doesn't, we don't tell the world it does what it cannot do... The testing is wholly complementary to analyst-related reports; think of it as analysts getting their hands dirty by actually testing what's on the latest hype curve to make sure it delivers on its claims and potential.

Broadband-Testing operates an **Approvals** scheme which prioritises products to be short-listed for purchase by end-users, based on their successful approval, and thereby short-cutting the evaluation process.

Output from the testing, including detailed research reports, articles and white papers on the latest IT technologies, are made available free of charge on our web site at http://www.broadband-testing.co.uk

Contents

Executive summary	4
Rationale behind a private cloud: The Basic Premise	5
HyperCloud Overview	7
Fancy Taking The DIY Approach?	8
HyperCloud: Deploying a private cloud	9
In Conclusion	16









Executive summary

- IT is essentially broken; the emergence of cloud computing – albeit itself a variation on similar past "outsourcing" themes – has created a dilemma for IT teams. Do they go with the "ease of use and scalability" that Cloud Service Providers (CSPs) evangelise about – and the costs associated with that route and loss of direct control over and visibility into their data and applications – or do they create their own next generation compute, storage and networking platform, AKA private and/or hybrid cloud?
- The concept of a "cloud type" experience within an OnPrem environment is appealing but one that is full of expensive and explosive potential minefields if attacked in a "DIY" manner. Put simply, the standard tools at our disposal are essentially the same ones that have been around for decades in the networking, storage, compute and security worlds. Moreover, they are still very much "islands of IT" in terms of integration. Creating your own version of an AWS or Azure, for example, is anything but trivial. At the same time, staying with a traditional, OnPrem solution is patently not the answer to scalability requirements and taking advantage of the new generation of cloud native applications and application building blocks – the DevOp era, we might call it.
- SoftIron, with its HyperCloud platform, claims to have the answer to this conundrum, providing a "private cloud in a package" style solution for creating private and hybrid clouds. Key to its approach is in providing all the requirement elements within the one platform - secure compute, storage and networking with zero association with, or requirement of, the public cloud to deliver any elements of that solution. This differentiates it significantly from the various "as a service" solutions out there that provide scalable alternatives to traditional OnPrem IT. but are based wholly or partially within the public cloud for data and application access and delivery. Equally it differentiates itself from those providing the storage element of the private cloud - there are plenty of players there, large and small - by fully integrating the compute and networking elements.
- Key to the concept of private cloud is speed of deployment. This is not simply a great benefit for new converts to any form of cloud infrastructure but equally for those finding public cloud very expensive – i.e., most – and wishing to repatriate their cloud-based data and applications as quickly (and easily) as possible.

- Fundamental to a totally in-house developed solution such as HyperCloud is that it is inherently a secure environment, not at all unlike a traditional mainframe system reinvented for contemporary requirements. Provenance for all elements and components is known – with no hidden (and nasty) secret lurking within.
- Since the advent of giant data centres (DCs) the availability of power – and sufficient provision thereof – has been a major concern. Now, rapidly escalating power costs have further highlighted the importance of minimising power consumption. With HyperCloud, SoftIron has minimised power usage in all areas.
- While still a work in progress (what isn't?)

SoftIron really does provide private cloud in a box; we created a cloud in one morning – need I say more?



Rationale behind a private cloud: The Basic Premise

The public cloud is expensive, inflexible in many ways, difficult to extract yourself from and you are not in control of your crown jewels – data is the most valuable currency in the world nowadays. There is also a related question over security, but arguably more relevant are those of data residency and sovereignty, as well as component provenance.

On the other hand, a traditional OnPrem deployment is, for most businesses, too restrictive, unscalable and high maintenance. Surely, then, there has to be some middle ground solution that provides the scalability the public cloud offers, while enabling a business to control and manage its infrastructure in a low maintenance fashion? The problem, to date with creating a private cloud, or even a hybrid environment, is that businesses are trying to build out what is a new form of IT architecture using components that were designed to create deployments from the 2000s or even the 90s.

The basics of a "cobbled together" infrastructure - the result of eras of different developments within the technology sphere - are essentially still with us, despite the origins of modern (Ethernet) networking being a creation at Xerox (Bob Metcalfe and team) aimed at simply making it easier to connect printers to multiple computers. So, we are talking endpoints of some description, from desktop PCs to smartphones and tablets; servers, from standalone, PCderived, to rackmount blades, networking, often using several different technologies and speeds, via switches and routers... Then, on top of these we have the security infrastructure, which itself might consist of 20-50 different technologies, or indeed many more. And all of these have to be managed and maintained 24x7, wherever they might be physically located. Moreover, this typically involves many different management platforms, processes and skillsets.

And, for all this, if we remove all the overused buzz-words and phrases, what we are essentially trying to do is nothing brand spanking new, but to recreate the mainframe. That is to say, an IT infrastructure that is easily managed, easy to secure and one that you are fully in control of, but a "mainframe" that is – literally – infinitely more flexible and scalable than existing generations of mainframe computers. Let's get real here; if a traditional mainframe delivered cost-effective scalability, flexibility and the inherent power to deliver contemporary applications, then everything we have seen developed over the past 40 years in IT would never have happened – or indeed been necessary.

From a consolidation perspective, we have seen breakout elements, such as attempts to combine security and networking as a cloud-based service, in the form of the Gartner-defined SASE/SSE market, for example. We have other vendors promising "network as a service", "hybrid access as a service" and similar cloud-based partial resolutions to the issue of modernising IT infrastructures; all worthy in their own way but neither the complete solution, nor putting you back in charge of your own IT deployment, as they typically run – at least partially – on public cloud infrastructures themselves.

Analogy: very few people – I suggest – ever stop to think where their invested or saved money is at any given time; if they did it would probably terrify them – locked inside some form of hedge, venture capital or private equity fund – certainly not in the vaults of the bank or other form of financial institution they have entrusted it with in the form of gold ingots with an 'IOU' letter attached.

So, why would businesses not worry about where their actual data and applications are at any given time of day? I know of companies historically who have deliberately employed the use of data centres within easy reach of their offices in order to be "close" to their data and apps. This is not as 'anal' as it might sound; in a world where compliance, governance and sovereignty rule, knowing exactly what lies where makes more sense than ever. Otherwise, the cost can be literally financial - and heavy at that. What, then, if you could take this "pseudo-paranoid but perfectly reasonable" approach and apply it to the flexibility the cloud offers - the aforementioned private cloud, in other words? In order for this to be truly achievable there has to be a new approach to building and deploying that next generation IT infrastructure. Here is where SoftIron, the focus of this report, enters the equation. Historically, IT infrastructures took weeks, months and even years to build out, depending on their initial scale, planned expansion and unanticipated adds and changes - a company merger, for example, being a worst-case scenario.



With its 'private cloud as a package' approach, SoftIron is looking to absolutely minimise this build and deployment time, including ongoing scaling out, as well as providing lifecycle management, predictability and security.

There is also - very much so - the cost factor to consider. In a world where the "DX" factor looms large over the heads of IT directors, speed of transformation is already definitive. It is easy, therefore, to understand the value proposition of a CSP. The problem - often not initially perceived - is: that "value" comes indeed at a price, typically far higher than originally budgeted for. And there is the unknown factor - AKA, what will it cost going forward? Industry analysts Canalys forecast at the end of 2022 that it expects CSP costs in Europe to rise around 30%, partly to cover rising energy costs. This is anything but trivial. And it simply accelerates a trend witnessed over the past few years - cloud spending is becoming huge. An IDC Worldwide Cloud Spending Guide showed that total public cloud spending rose to \$383.6 billion in 2021, up from \$307.7 billion in 2020 - totalling almost \$700 billion spent by enterprises globally on the cloud during the pandemic. According to IDC, this translates into cloud spending representing around 30% of overall IT spend, with those costs exceeding \$1.3 trillion globally by 2025.

Moreover, how visible is that spend? A report from Anodot noted that the majority (53%) of survey respondents said their biggest challenge to controlling costs is in gaining true visibility into their cloud usage and associated costs. This was closely followed by the complexity of cloud pricing (50%) and the use of complex, multi-cloud environments (49%). The reality is that all three issues are closely related, in the sense that it is difficult to gain visibility into every aspect of cloud usage with its associated cost when that environment is excessively complex and/or extends across multiple CSP platforms.

Analysts cite the importance of having a multi-cloud strategy, but managing and optimising costs across providers is all but impossible due to the proprietary nature of those platforms and markedly dissimilar pricing models.

It has even spawned a sub-product in its own right to manage and control those multi-cloud deployments. And these products are also expensive! As a result, companies are readily looking to unload some of their cloud-based resource, aware of increasing – and uncontrollable – costs, but the backlash here is that, again, that migration comes at a significant cost and, more importantly, migrating to where exactly?

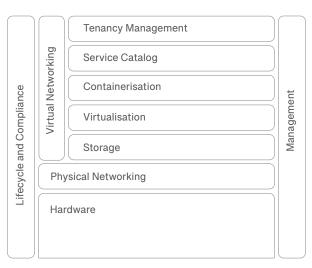
SoftIron, the focus of this report, claims it has the answer in the form of its newly released HyperCloud platform. Let us, then, take a look at exactly what HyperCloud consists of, and then see it being deployed.



HyperCloud Overview

With IT vendors it is all too easy to read a few marketing messages and promotional hyperbole that promise the earth and beyond; then when you dig deeper it isn't remotely clear what the vendor is actually offering, let alone whether it works or not.

Relevant aside: one of my roles is as a judge on IT industry awards events, such as Techtrailblazers – lots of hungry, young (2-5 years old) vendors trying to impress you with their technology in various categories. Sadly, most common is the situation where you read their entry resumé, find few clues as to why exactly their product is "the only one on the market that achieves 'xyz'" or indeed what it actually does. So, you go onto their websites – still no idea. Download an overview





SoftIron defines HyperCloud as basically consisting of the following:

Hardware: This is described as a flexible 'lego-style' construction platform, incorporating networking, storage, and compute appliances – in other words, every element of a cloud solution, that you simply plug together. On top of this is a hardware management platform, which builds, deploys and manages a standardised firmware and OS delivery across the platform.

Networking: A key aspect of the SoftIron approach is to take away the pain of integrating and supporting any number of 3rd party Ethernet switches, by incorporating its own switching, with 10/25/100G options currently, that are fully integrated within the platform. paper; no, still clueless. The result is that you have a huge number of product vendors promising a utopian solution for a given market, but don't actually appear to have the underlying technology to deliver on that promise. So, when a company such as SoftIron – albeit, in this case, a well-established vendor, significantly expanding its solutions capabilities comes along and promises to revolutionise private cloud, the first step towards validating this claim is to work out exactly what is being offered. In this instance, fortunately, there is substance behind the messaging - in other words, an actual, deliverable product solution. In essence, HyperCloud is a fully integrated 'private cloud in a box' platform which aims to massively simplify the delivery of on-premises or private cloud by taking complete ownership of the solution, not simply a cameo role. Probably the best way to describe it is as a self-assembling, self-managing, and self-healing, integrated, turn-key hardware and software solution for building private clouds. Quite a mouthful (of words), so what are the actual 'building blocks' within the platform?

Storage: The storage element of SoftIron is based on the Ceph open-source, software defined platform that implements object storage on a single distributed computer cluster. It is designed to be flexible, scalable and resilient – three of the cornerstones of the SoftIron approach.

Virtualisation: This layer enables the deployment of VMs, LXC Containers and multi-VM services, all fundamental to modern scale-out approaches, but especially so with respect to a multi-tenanted scenario; for example, you can create virtualised storage templates that can be shared – and customised - across multiple tenants.

Service Catalog: HyperCloud has incorporated the idea of a Marketplace – SoftIron's, public or even build your own similar to an app store, for commonly (re)used elements such as guest images. It means you get a library of "here's one we made earlier" building blocks to further speed up the process of deployment and scale-out.

Tenancy Management: HyperCloud has been designed as an unpack and rack, internal or commercial service-provider ready platform, with multi-tenancy features inbuilt. Think, user and group ACLs, billing and accounting, resource quotes, usage limits etc.

Management: The platform includes APIs to manage every layer of the stack, which are accessible directly or through command-line or graphical management interfaces. Importantly – given the environment – this includes 'lights-out' management to remotely manage out-of-band.



Fancy Taking The DIY Approach?

Many times, over the years, as part of a Broadband-Testing analysis, vendors have identified a DIY (self-build) alternative to consider, as part of the evaluation.

The reason for this is simple: companies do often attempt, or at least seriously consider, the self-build alternative. The seeming attractions are to be independent of a proprietary lock-in, and lower costs. In practice, however, neither generally turn out to be true. Instead of being locked into a single vendor, you end up being locked into a crazily complex scenario involving multiple vendors, a huge number of interdependencies and- trust me – an impossible to manage, never-ending integration and update world.

Worse still, if any one element of that DIY solution has to be removed from the equation (discontinued, no longer supported, supplier disappears etc) then there is always the likelihood that you have to start again from scratch – or at least spend costly time rebuilding the "solution".

Being realistic for a moment with respect to the CSPs – they are expensive for a reason; clouds are very complex things

to build (more so than the fluffy ones). So, if considering a self-build version, there is as good as a never-ending list of elements to consider. Looking at a site offering graphical diagram software for illustrating a cloud, I noted that its diagram solution consisted of no less than 492 "ready-to-use" vector shapes which will "help facilitate the process of design and development of cloud computing". Mmm – that's almost 500 elements identified as being fundamental in the process of building your own private cloud – good luck with that one!

It's not as if managing a traditional IT deployment isn't difficult enough. For anyone who has had the misfortune to manage a complex, hybrid network architecture consisting of products, software and services from tens or hundreds of different suppliers - aka a "normal" network - and manage the continuing integration and updates therein, having an option to combine multiple elements into a single vendor offering, without compromise, is a good thing. This is the direction SoftIron is proposing, but not in terms of a vendor "solution" consisting of various acquired elements that are partially integrated, but a "we made it all in-house" solution. So, a tied in proprietary solution once again then? No, just the core elements all integrated out of the box for a private cloud solution. Like a flat-pack house, how you decorate internally is up to you, but all the important elements are delivered straight out of the box.

That, at least, is the premise upon which Softlron's approach is based. So how does it deliver on that promise?







HyperCloud: Deploying a private cloud

While an actual SoftIron customer might expect to see an element of "here's one we made earlier" about its deployment, because we could, we started at the very beginning.

It's a very good place to start. For the test, we used the SoftIron Berlin labs; an environment complete with rack space, cables and all the required goodies needed to help in our aim to create a private cloud consisting of a nine-device cluster (half a rack): 3x switch, 3x storage (disk arrays of 12x24TB) and 3x compute devices – a configuration itself capable of supporting a large community of users, let alone the nearinfinite scaling possibilities it offers.



Figure 3 – Our rack space awaits

In truth it would be very easy to summarise the deployment in a couple of paragraphs: there's no rocket science involved – more like an Ikea assembly project, except there weren't any pieces missing! Instead, I'm detailing the processes we went through in a single morning to show what "building a private cloud" actually entails, following the zillions of hours of coding and product dev time that created our flat pack universe consists of. So, our process ran as follows:

- Unboxing Nodes: Installing media.
- Racking Nodes:
 - This is all about simply being organised; for example, using a spreadsheet to note what goes where.
 - Assembling rails.
- Cabling:
 - Identifying the number of cables and what they are.
 - Enabling remote power i.e., connecting to the rack PDU (power distribution unit).
- Networking: Creating individual VLANs for each element of the network.
- First time Connectivity: Actually connecting to the rack.
- Boot Static Nodes: Bringing up the control plane.
- Boot Storage Nodes: Checking that the storage has booted correctly.
- Boot Compute Notes: Checking that the dashboard has come up correctly.
- Deploy Infrastructure: Now we have a private cloud we can configure (for this process we have several options, including console, management UI and open-source tools such as Terraform).

Unboxing Nodes

Our first process is simply unboxing the SoftIron nodes we are using in the rack. These are typically shipped on a pallet. To give an idea of the amount of boxes/hardware involved, the "half a rack" installation we chose for the testing consists of two-three pallets.

In our case, as we're in the SoftIron labs, the boxes of nodes were in place to get us started on our "unboxing experience".



Figure 4 – Our unboxing begins



There are several components to install here: rails (inner and outer), the server itself, power and interconnect cables, media (if the node is a storage server), complete with a user guide and tool kit, though

installation is an entirely toolless affair – not even an Allen key required, let alone some proprietary tool that breaks 60 seconds into the process.

It's important to note that all SoftIron modules are 1U high only, regardless of their purpose. As well as making them extremely practical for minimising rack space in a DC or machine room, power consumption is extremely low too – for many years now, a shortage of available power to DCs has been a major problem. Factor in the recent hikes in the cost of energy and power consumption becomes a significant element of any IT solution, not least a private cloud.



Figure 5 - Snapping disks into place

Racking The Nodes

Now begins the process of racking each node, starting with the interconnect nodes. This is where good housekeeping works wonders: having a process for racking with details of what goes where. Bear in mind, in many cases, there could be a remote engineer involved, so process is everything. We used a good old-fashioned spreadsheet in our case.

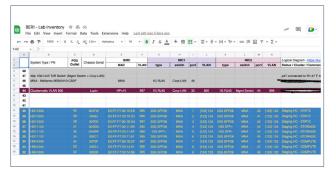


Figure 6 – Using a spreadsheet for our "housekeeping"

Common sense prevails here – noting basics such as which nodes are in which rack unit, which switch port each network adapter (NIC) is connected to and which PDU outlet they are going to draw power from. It might sound a million miles away from "cutting edge tech" but these are the fundamentals of building a high-tech solution. As we noted, the installation is toolless: for example, installing each unit consists simply of sliding back the yellow tab in the outer rails, pushing them into the rack holes, and locking with the yellow tab – on each side. The servers are shipped with the inner rails attached for ease of installation. The process is the same for each and every node.



Figure 7 - Locking units into rack

Cabling the Nodes

With units in place in the rack, our next task was to cable up: this is not entirely trivial, as anyone who has run a test lab will know; separate cables for power, networking, OOB (out of band) management and a serial console connection were the order of the day here. We used dual power supplies for redundancy, as would be standard in any similar deployment.



Figure 8 - Commencing the cabling process



From a network cable perspective, the particular nodes we were using come with 2x25GbE SFP28 NICs, so we could either use Fibre or SFP+ Direct Attach (Twinax) options. Fibre is far more flexible (no pun intended) these days than it used to be, and can readily be cut down to exact length for tidiness and ease of maintenance, but it is still relatively fragile and I'm a Yorkshireman, in a live lab in Berlin... So, we chose the latter option as a safe bet; each was connected to the two main interconnect nodes, with each NIC going to a different interconnect node, for added resilience. At the end of the day, uptime is everything.

It's worth reiterating that SoftIron also provides 10, 25, 50 and 100 GbE NICs, so basically all options are covered. For OOB (aka "lights-out") management, there is a standard RJ-45 ethernet port to allow for remote management. This simply connects to the OOB interconnect node. Finally, there is a USB-B serial console port that can be plugged into the server for direct access to a serial console. Additional serial console ports are also available, but were not required for our deployment. With every cable in place, it looks rather neat, even if I say so myself – modern IT art...

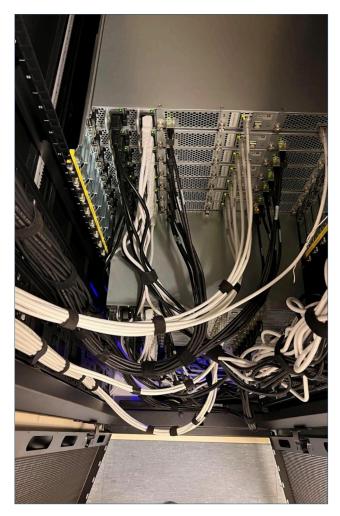


Figure 9 – Commencing the cabling process

The servers power on as soon as a power cable is plugged in, toolfree and button-free deployment. Again, it is worth noting that, from a remote management perspective, this is a great feature, especially when combined with a remote-capable PDU, so that power can be managed remotely, and servers cold-powered, even from South Devon...



Figure 10 – Our "half a rack" up and running

Networking

From a networking perspective, HyperCloud itself takes care of most tasks. The switches automatically detect nodes and will also automatically handle configuration specifics such as LACP (Link Aggregation Control Protocol) bonding and resilient path creation for yet more failsafe deployment. Our focus therefore was on creating necessary VLANs to create independent networks, but our cluster only required three, which we defined as follows:

- Internal Storage Network VLAN123
- Internal Compute Network VLAN124
- Public Network VLAN999 (pre-defined)

The internal networks exist purely for communications and management between the servers within the cluster, and the VLANs defined are only required for the HyperCloud nodes. The Public network is where the API endpoint and dashboard



are made visible to the outside world (network). This might equally be within a business or organisation, or the public internet. Importantly, the entire cloud created presents itself as just a single address on this network, but again with added resilience; it is implemented as a floating service on the cloud itself, so any one of many nodes could be offering access to the API at any point in time.

Checking Connectivity

Our next task was to confirm connectivity, by attaching a console to one of the static servers. This could have been carried out remotely via the OOB management connection, but as we were by the server rack, we used the integrated USB serial console connection. For the sake of completeness, we loaded a static virtual console to each of the nine nodes, in order that we could monitor the boot process for each device.



Figure 11 - Monitoring boot state of each node

Booting Static Nodes

Booting the static nodes – from which our cloud builds out for the first time, we are provided with a prompt asking which node in the cluster each will be, which we naturally configure as one, two and three. Who says building clouds is complex!



Figure 12 - Defining our static nodes

With the static nodes up and running, all the remaining nodes then boot using the static nodes as their information source.

With the static nodes up, we can now boot the storage nodes. As noted, all the information they need is provided by the static control plane, so it's a case of sitting and watching. As noted earlier in the report, the HyperCloud storage layer tested here is based on Ceph, so another monitoring option is to use the Ceph status command:

steve@jumphost:~\$ ceph-s
cluster:
id: 1a399cd3-79db-46be-9133-86ab2d7794cb
health: HEALTH_OK
services:
mon: 3 daemons, quorum hypercloud-storage-1,hypercloud-storage-2,hypercloud-storage-3 (age 8w)
mgr: hypercloud-storage-3(active, since 7M), standbys: hypercloud-storage-1, hypercloud-storage-2
mds: 3 up:standby
osd: 36 osds: 36 up (since 8w), 36 in (since 8w)
rgw: 3 daemons active (hypercloud-storage-1, hypercloud-storage-2, hypercloud-storage-3)
task status:
data:
pools: 11 pools, 352 pgs
objects: 38.13k objects, 176 GiB
usage: 2.8 TiB used, 327 TiB / 330 TiB avail
pas: 352 active+clean
pgs. 552 detremented
io:
client: 36 KiB/s rd, 227 KiB/s wr, 1 op/s rd, 8 op/s wr
steve@jumphost:~\$

Figure 13 – Using Ceph to monitor storage nodes booting

Booting Compute Nodes

Our final stage of the boot process involves bringing up the compute nodes. These nodes provide the resource (memory and compute) for the virtualisation layer and require no manual intervention after the initial compute node setup. This first compute node prompts for information regarding the network, in order to determine which VLANs to use and which IP to present the API endpoint to the outside world on; thereafter all is automated.

For the record, each boot process took around 10-15 minutes to complete. With the first compute node up, we can now see the dashboard on the endpoint specified, enabling us to pull the login information from the cluster:

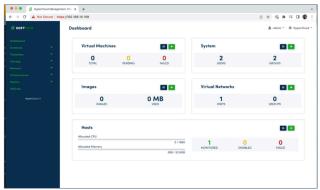


Figure 14 - Our Management UI dashboard post-boot sequence



Booting Storage Nodes

Deploying The Infrastructure

With our HyperCloud up and running, we can now spin up virtual machines (VMs) in several different ways; from the management UI it's simply a point and click exercise. As noted earlier, SoftIron also provides a marketplace for easy access to configuration options to speed deployment. Equally, for those who want to take a more hands-on approach to deployment, such as using an Infrastructure as a Code (IaaC) automated approach, it is possible to use tools such as the aforementioned Terraform. SoftIron actually provides some public examples of how this can be done:

https://github.com/SoftIron/hypercloud-examples.

It is worth reiterating here that the whole basic cloud deployment process was achieved comfortably in one morning (and no early start!). Contrast this

with, for example, trying to work out a configuration on a cloud service such as AWS EC2, in terms of exactly which options you need – let alone work out the costs involved – and, in that time, you have actually deployed HyperCloud!

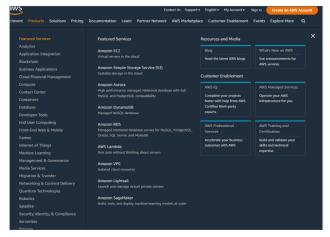


Figure 15 – Just a subset of AWS EC2 options

The Importance Of Provenance

It is understandable for potential customers of SoftIron to look at the company and ask: "why do you make everything yourselves"?

We are talking here about a vendor that even designed its own BMC (a Baseboard Management Controller is a specialised service processor that monitors the physical state of its surrounding hardware using sensors, and communicates with the system administrator through an independent connection). Normally this would just be a 3rd party component – for example, contained on the motherboard of a 3rd party device. There are two aspects of this seemingly OTT approach to manufacture; the first is in looking to provide the most robust platform possible – SoftIron's history is based around continuous availability and as close to zero downtime as possible.

The second aspect relates to inherent security: discrete security threats pre-implanted into hardware are an incredibly common form of cyber-attack. Anyone who has made the mistake of buying a cheap Chinese smartphone (surely not this author?) and seeing what was pre-loaded onto that phone (or indeed a laptop from an unqualified supplier) waiting to be triggered will understand the rationale here. Since the vast majority of vendors use some - often multiple - 3rd parties to manufacture (or even design) their devices and appliances, there are no guarantees offered from a security perspective. Just one appliance from any major IT brand can have numerous sub-manufacturers, each using components from a plethora of suppliers across the globe. Then consider that every stage; from design through component supply, manufacturing, coding and assembly, presents opportunities for tampering. As we say in the UK "you pays your money and you takes your choice".

It is also reasonable to suggest that the inbuilt security threat

in 3rd party hardware and firmware is actually understated. For this reason, SoftIron is completely obsessive about its quest for 100% ownership of design and manufacturing, including surface mount assembly –

as I witnessed as part of the test process within the SoftIron Berlin office (where even the radiator covers were designed and constructed inhouse – I kid you not!). Secure provenance does not come without costs – to verify appliance and software integrity requires a completely transparent audit of the entire design, supply chain, manufacture and delivery path. Moreover, the same approach applies to all the code that sits on any SoftIron appliance: it's what is known as full stack control: the vendor takes full authorship, ownership and total responsibility for all code that goes on its devices.

Of course, there are a very few elements – some of the silicon, such as primary processors for example – that have to be 3rd party sourced, but these are from (very) major suppliers and are still fully vetted and tested before installation. This attention to detail is very much something that is hard coded into the SoftIron philosophy. That much was suitably obvious through this test project. From the HyperCloud management UI, beyond VM definitions, you have access to all elements of the deployed cloud. In classic UI style, there is a main dashboard showing current status on all key deployment aspects - VMs, System, Images and Virtual Networks, with a Hosts resource utilisation below. On the left are the primary commands, including Storage and Network, as well as Instances, Infrastructure (consisting of Clusters, Hosts and Zones), System (Users, Groups, VDCs, ACLs) and Templates. Key to ease and speed of deployment within SoftIron is the ability to create templates for VMs, Services, Virtual Routers and VM Groups, meaning further deployments can be effectively automated. This is especially valuable in a multi-tenanted environment, such as that a service provider would host. This "create once, spawn infinitely" approach is fundamental to the SoftIron private cloud methodology. Speed of deployment is critical in many cases.



Figure 16 – SoftIron management GUI populating the cloud

For every aspect of management, multiple permissions and ownership levels can be created, so different levels of administrator can be tasked with whatever level is suitable for their (lack of) expertise. The dashboard can be refreshed at any time to show current status. Post Berlin deployment, we did this from my offices in South Devon in the UK, with the host cloud environment in SoftIron's San Diego offices, including adds and changes via a London office and updates were instant. It's definitely a global solution....

The combination of speed of deployment and flexibility of configuration thereafter really lends itself to a new generation of service providers looking to take ownership of their own cloud service delivery, when the likes of AWS, Azure etc might have put them off from ever imagining such a business opportunity existed previously. From an enterprise perspective, looking to rationalise, modernise and yet maintain control over its IT infrastructure, again SoftIron provides a left-field solution they have probably not considered before.



The Ever-Growing Importance Of Total Cost of Ownership (TCO)

Within IT, the decision-making process for procurement has changed markedly in recent years. For decades it was primarily about the technology itself; budgets were often insanely large and IT became a "must have" marketplace as vendors brought out newer versions of what essentially was the "same old same old", regardless of the bottom-line impact it had on their business – and the associated ongoing costs. But IT teams bought into it anyway.

Now that process has massively refocused on costs; not simply upfront acquisition costs (Capex – Capital Expenditure) but more so with respect to ongoing costs - operating expenses (Opex) and what the true TCO actually is. Here, for example, is where many CSP customers have been bitten; the attraction of an easy entry point is soon overtaken by ever-increasing – and unanticipated – costs and, worse still, no actual way to estimate future costs. At the same time, TCO cannot be reduced by simply lowering IT delivery and capabilities. Valuing TCO without compromising pure performance and calculating the resulting ROI spawns a new approach to a decision-making process long overdue in IT. Companies have changed the way in which they evaluate technology and are moving towards solutions-based purchases, not bit-part procurement and the old "bolt-on" approach to moving their IT infrastructure (allegedly) forward.

What SoftIron is looking to offer is technically a no compromise solution from a private cloud perspective, both providing the customer with a complete solution and guaranteed costs and predictable TCO as a result. In other words, it rips up the accepted rule book and starts anew.

About time too!



In Conclusion

Within IT circles there is much talk about "disruptive technologies" – these don't always mean good disruption, but in the case of SoftIron with HyperCloud, it is fair to say that its directive is both disruptive and good.

Cloud as a concept is fine – a kind of mainframe for the new age; far more scalable, flexible and all-encompassing. But is it affordable? And, does it mean a business can still be in control of its prize assets – its data, applications, user base and customers? If we look at cloud from an existing CSP offering angle, then the answer to both is no. Customers of the major CSPs have been shocked by the unexpected costs – that are still rising - the difficulties in migrating away from those major providers and the realisation that they don't actually know where their data and applications reside at any given time.

What SoftIron is offering is that complete cloud experience, but "from your own home" – in other words a private cloud. So, you – whether a service provider, or a business looking to revamp its IT – are in control of everything, but the incredible complexity of a modern IT infrastructure is dramatically slimmed down. It's kind of IT reinvented – and it needed doing. Is SoftIron alone in offering a private cloud solution? Of course not – there are shedloads (technical term) of vendors offering such; however, in almost every case, what they are offering are a few of the building blocks required, and principally just the storage element. And even then, there's often a strong DIY element to it.

In providing the complete networking, storage and compute infrastructure "straight out of the box" SoftIron is daring to be different. Moreover, it is all its own offering – not some collection of 3rd party bolt-ons, but designed and manufactured almost entirely in-house. If someone had said to me, there's a vendor out there – and not one of the Silicon Valley giants – that is creating a turnkey private cloud solution, equally adapted to single or multi-tenant consumption – and it is making it all itself... Well, the word "insane" comes to mind or "pull the other one, it has bells on it" – but that is precisely what SoftIron has done – and it has pulled it off.

Is this, then, the whole story? Well, of course not; all IT is a work in progress and so is HyperCloud – the already rapid deployment will get more streamlined and more automated. That template approach to scaling out the private clouds will get ever more flexible and all-encompassing. But what we already have does what it says on the tin; it has fully automated life cycle management, huge scalability and enormous flexibility. And it works – straight out of the box. The challenge we set was "half a rack in half a day" – installing, deploying and managing a private cloud within a single morning... And we achieved that goal.



