Springpath is latest startup to attack hyperconvergence opportunity

Analyst: Simon Robinson Tim Stammers

19 Feb, 2015

Springpath has become the latest startup to reveal itself as targeting the growing enterprise appetite for converged infrastructure offerings. It will initially aim at the emerging 'hyperconverged' opportunity, but the company – founded by two former senior VMware engineers – is ultimately focusing on a broader range of virtualized and non-virtualized applications and workloads with a pure-software approach it calls the Springpath Data Platform. Springpath has raised $34m in venture funding, and says it already has almost two dozen customers.

The 451 Take

As we recently noted, the market for hyperconverged offerings has materialized from virtually nothing into one of the hottest topics in infrastructure IT in little over a year. Against this backdrop, any new startup needs to make a splash, with an idea that moves the industry narrative forward. In this respect, Springpath does not disappoint; its Data Platform is one of the most comprehensive approaches we have seen to building out a next-generation converged platform. It nicely piggybacks on latent interest in new platforms such as Docker, and should appeal to customers looking to buy into a longer-term vision across a broader range of workloads and applications. The full complement of features and support is still a work in progress, and in the short term at least, Springpath needs to demonstrate it can compete effectively in the intensely competitive VMware space. It also faces the challenge of building a presence in what is becoming a white-hot market – but Springpath's initial proposition looks likely to generate plenty of attention.

Context

Sunnyvale, California-based Springpath was formed in 2012 (and has been operating in stealth as Storvisor) by two ex-VMware engineers. CTO Mallik Mahalingam was a principal engineer at VMware, and for a decade led both storage and networking product development. He is credited as the inventor of VxLAN – a foundation stone of SDN – and prior to VMware, he worked at HP Labs and held engineering roles at NetFrame Systems and Intel.

Meanwhile, Krishna Yadappanavar is an eight-year VMware veteran and an architect working on the VMware file clustered system (VMFS) as well as vFlash. He previously worked at CommVault, EMC and Veritas. Product management is led by Ravi Parthasarathy (Omneon Video Networks, NetApp), while Suresh Ravoor (VMware, Brocade, Sanrise) is VP of engineering. The company currently has over 60 employees, and has raised $34m in venture funding from Sequoia Capital, NEA and Redpoint Ventures.

Strategy

As with others, Springpath has a vision of a 'software defined' enterprise IT infrastructure that is capable of supporting a variety of distributed, container-based applications operating in a highly automated, self-healing and efficient manner. All of this should be capable of delivering 'enterprise grade' features, and scale in line with business requirements across a pool of storage, networking and compute resources. The question is how to move to this utopian state from the reality of today's environment – where data and applications often reside in a mix of multivendor silos, resulting in inefficient utilization, fragmented data management and severe scale limitations.

Springpath's answer is what it calls a 'single data platform.' It is designed as a 100% software offering running on commodity servers to deliver enterprise-grade features and elastic scaling. Although Springpath's product can be considered part of the broader move to hyperconverged infrastructure offerings, it also believes it has taken a more comprehensive approach to convergence. This stems from the fact that Springpath chose to build a distributed file system from scratch; the company claims this provides the flexibility to support a broader range of environments, offers greater scale and performance, and optimally implements features such as snapshot, compression and de-duplication.

Springpath is targeting midsized to large enterprises. It's building out a channel-centric model, and has signed up Tech Data Corp as a distributor, which will provide resellers with access to servers pre-loaded with Springpath software. It supports deployment on HP, Dell, Cisco and Supermicro servers, and is working on reference architecture designs that resellers and other partners can

build to. A beta version of the product supporting VMware vSphere has been running for more than a year, and became generally available this January. Springpath says it now has 'close to two dozen' deployed customers, the largest running an eight-node cluster. The product is priced on a subscription basis, starting at $4,000 per server per year, with a minimum of three servers.

Products

The Springpath Data Platform runs across three or more servers to form a cluster. A software-based controller running on each of these servers provides both compute and data services, and takes control of the internal flash SSDs and HDDs to store data. They communicate with each other over 10GbE to present a single pool of storage that spans all nodes in the cluster. Springpath says compute, storage and I/O can be scaled linearly by adding more servers to the cluster. Its data rebalancing is online and non-disruptive, and new resources can handle both new and existing data.

At the top of Springpath's software stack is a 'data access' layer. Unusually for a hyperconverged offering, this has been designed to support file, block, object and Hadoop interfaces connecting to both physical and virtual environments. As a result, Springpath says it offers a more 'horizontal' platform than many alternatives. Initially, GA support is limited to VMware vSphere, NFS for file environments and a RESTful API that is Swift and S3 compatible.

However, support for OpenStack – which will include the KVM hypervisor and the Cinder API – is now in beta, as is support for Docker. The latter is important insofar as Springpath is also targeting non-virtualized environments; support for bare metal environments and applications such as Hadoop is also planned. Additionally, Springpath plans to support Microsoft Hyper-V, along with the SMB file protocol, in the second half of the year.

At the heart of the platform is what Springpath calls the HALO architecture. HALO is a homegrown distributed file system, and stands for hardware-agnostic log-structured objects. It describes the data storage model supporting data distribution, caching, persistence and optimization – the latter includes in-line de-duplication and compression. Running across all of these aspects are data services, including thin provisioning, space-efficient snapshots, cloning, backup and DR. One notable service initially missing is native support for replication between clusters for DR; instead, Springpath says it will work with vSphere replication/SRM, although it plans to add native asynchronous replication in the second half of the year.

Springpath says there's no architectural restriction to the scale of the platform; scale is essentially limited by the scale of the hypervisor (i.e., up to 64 nodes/cluster for VMware vSphere), although

Springpath's official tested limit today is 24 nodes. It expects most VMware-based deployments to be in the 8-16 node range. Springpath says it's able to support this scale because of HALO. By writing its own distributed file system, the company says, it can offer infinite scale without compromising reliability or performance. Much of this comes down to the way it decouples the persistence and caching layers; its data-distribution layer is similar in design to pNFS, and distributes incoming data across all nodes in the cluster. This eliminates performance bottlenecks, and ensures all compute resources are not left idle, it says.

For delivering high I/O at low latencies, all reads and writes are cached in SSDs. Incoming writes are replicated to one or two other SSDs (according to user policy) in the same cluster before being acknowledge as persistent. Writes are then de-staged to HDD, while 'hot data sets' are cached in both SSDs and DRAM. Importantly, caching and persistence tiers can also be scaled independently. This might be useful if a customer wants to utilize a range of hardware elements, such as traditional server racks and storage-limited blade servers.

Springpath emphasizes that this can be done under a consistent data-services framework; for example, it notes that data de-duplicated in the persistent tier remains de-duplicated when read into the caching tier, at the SSD and memory level. Additionally, Springpath says it's using a form of inline de-duplication – based on a patent-pending Top-K Majority algorithm that enables a high rate of de-dupe with a minimal amount of memory. This approach also allows it to adjust the amount of memory for de-duplication, which it believes is important when the de-dupe rate may be uncertain. De-dupe runs at the 4k block level, and is always turned on, as is compression.

Springpath also calls out that it's using pointer-based, zero-copy snapshots, which are a space-efficient way to quickly create and store 'unlimited' backup copies online. They can also be deleted instantly. This efficiency also applies to the creation of clones, it says. The company highlights snapshot granularity – snaps can also be taken on an individual file level, which it says map to per-VM-level snaps in a VMware environment, while different policies can be applied to different VMs. The company is also planning to support the new VMware VVOLS API.

Springpath says the log-structured nature of HALO is important to overall system performance and data availability. All incoming I/O – even if it is random – is laid out sequentially on both flash and HDD and compressed into a 'self addressable object.' Springpath says this sequential layout increases flash endurance and maximizes HDD performance, while allowing rapid recovery from a media or server failure. The company says this also means objects are addressable via a key in a way similar to an object storage system; indeed, this layer can be exposed via a REST-based API to provide an object storage capability.

Besides replicating writes at the SSD/caching level, HALO also replicates data at the persistent/HDD level. Two replicas are created, making for a total of three copies, meaning that the cluster can survive two SSD/HDD/server failures without risking data loss. Springpath does not currently use erasure coding for this process – it believes the performance penalties outweigh the space savings – but may add this in the future. Additionally, Springpath says the controller software can be patched and upgraded without taking the cluster offline.

Competition

Any hyperconverged offering such as Springpath is an alternative to a conventional stand-alone storage array. The most obvious examples of that are EMC's VNX and NetApp's FAS arrays, but other devices include Dell Compellent, HDS HUS VM, HP 3PAR, and IBM's Storewize and XIV arrays. Successful startups selling midrange stand-alone storage include Nimble Storage, Tegile and Tintri.

Within hyperconvergence, HP arguably invented the architecture eight years ago when it began selling a software-only 'virtual storage appliance' based on the LeftHand Networks storage system. That software is now called HP StoreVirtual. HP has also joined VMware's EVO:RAIL program to become a channel to market for VMware's Virtual SAN; that product just updated, and the EVO:RAIL program has arguably done the most to increase interest in hyperconvergence. Other EVO partners include Dell, Fujitsu, NEC, EMC, NetApp and HDS. As things currently stand, Lenovo and Cisco are the only major server vendors not signed up to EVO:RAIL.

EMC and Dell are the only other major suppliers offering hyperconverged products that are not EVO:RAIL systems based on Virtual SAN (although they both are also EVO partners). EMC's ScaleIO is pitched further upmarket than Virtual SAN, and unlike Virtual SAN, can be installed on clusters of non-virtualized or non-VMware virtualized servers. Dell's XC Series Web-scale Appliance is a hyperconverged offering comprising software OEMed from startup Nutanix, loaded onto Dell servers.

Nutanix is currently the leader among the hyperconverged suppliers that will compete directly with Springpath. The heavily funded and IPO-bound vendor currently claims more than 1,200 customers, including over 30 that have spent $1m+ on Nutanix systems. Another fast-growing startup is SimpliVity, which by October 2014, was claiming more than 200 customers, after only 18 months on the market. Other hyperconvergence startups include suppliers Maxta, Pivot3, Scale Computing and open source specialist NIMBOXX (which just acquired Virtual Bridges' VERVE VDI business). Meanwhile, erstwhile VDI specialist Atlantis Computing has made a pivot into hyperconvergence with its USX play. DataCore and StorMagic are also in the mix.

Versus these competitors, Springpath believes it is the first and only hyperconverged player to offer best-of-breed features across the board with 'no compromises.' This, it says, comes down to its decision to build its distributed file system from the ground up. Although its main competitors claim to offer many of the same features and capabilities, Springpath believes they are compromised architecturally because they are built on existing file systems (such as VMFS for VMware vSAN). This limits their ability to scale, or affects their ability to offer features such as de-duplication and compression without adversely impacting performance, it says. Additionally, it believes its software-only approach offers differentiation against Nutanix and SimpliVity, while its multi-hypervisor/container support distinguishes it from VMware vSAN.

SWOT Analysis

Strengths

Springpath boasts an engineering-heavy management team that has worked on some of the thorniest problems in virtualization; the result is a product that, on paper, appears very comprehensive.

Weaknesses

The company is only just making its first products generally available, which means its customer count and pool of reference customers are fairly small. It also means that support for all platforms is not quite GA yet.

Opportunities

Interest in convergence as a way of addressing longstanding infrastructure challenges – and hyperconvergence in particular – is booming, especially for vendors offering a 'software defined' twist on this.

Threats

The market in its constituent components – especially in storage – was already pretty crowded before the latest wave of hyperconverged players emerged. Springpath has some differentiation from the pack, but it needs help from partners to get its voice heard.