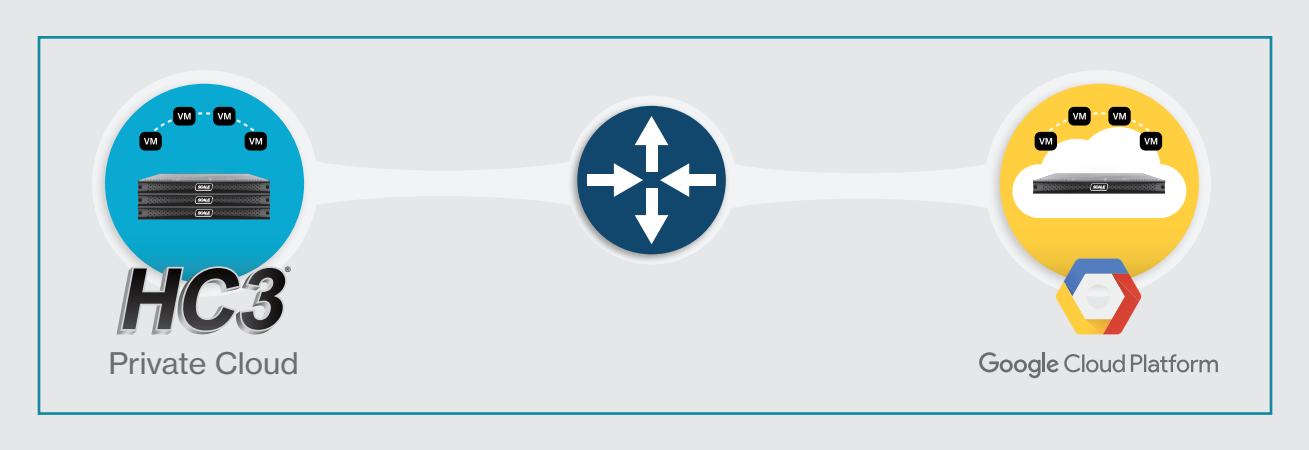
HC3 Cloud Unity℠

Theory of Operations

A uniquely easy and affordable hybrid cloud solution.

Technical White Paper

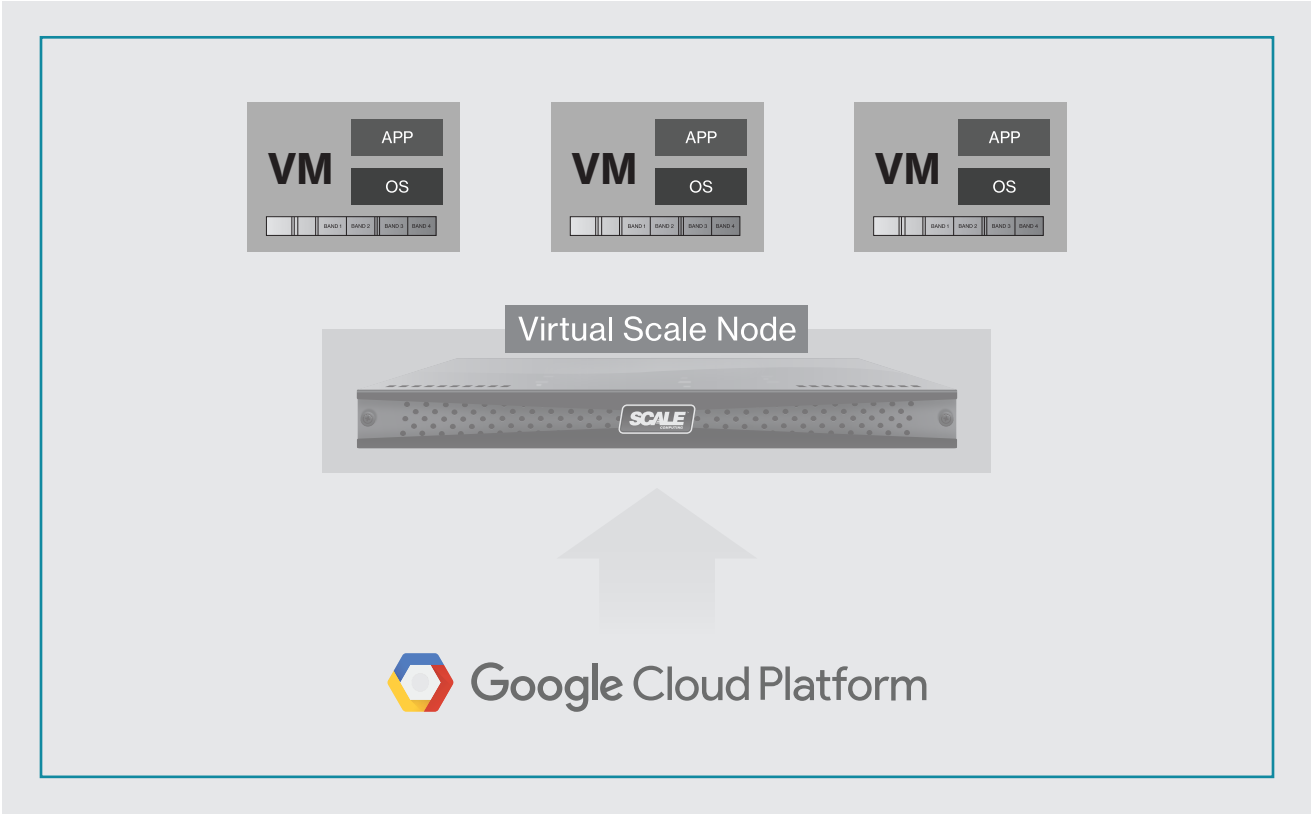
This document is intended to describe the technology, concepts, and theory behind the Scale Computing HC3 Cloud Unity platform. It is presumed that the reader is familiar with the basic theory of operations for Scale Computing HC3 and SCRIBE and high-level components and operations of an HC3 cluster.



Design Goals

Scale Computing's architecture has always been designed to provide simple, scalable, and highly available compute and storage services with the intelligence and automation built into HC3 and SCRIBE. As a 'private cloud' which operates on premises, HC3 provides a cloud-like experience with the security of on-premises hardware. Leveraging the ubiquity of public cloud providers was a natural extension of the capabilities offered by HC3.

Nested virtualization, supported natively by our core KVM hypervisor, has been a feature since 2012 but has only recently been exposed by select public cloud providers such as Google for production use. With nested virtualization, hardware extensions (CPU, VMX) enable the ability to create hypervisors-in-hypervisors. Combining the performance, stability, and simplicity of HC3, native nested virtualization and the elastic resources offered by cloud providers creates the foundation for an ideal platform to diversify workloads and create hybrid cloud environments.



For Scale Computing's commitment to ease of use, it was imperative to simplify the connectivity problem inherent in cloud solutions. Creating and running workloads on a public cloud platform is half of the equation. Plumbing those compute instances on a discrete, public cloud to a private enterprise network is challenging. With the maturity of software-defined networking technologies such as VXLAN and full-support for the prevailing open-source SDN technologies built into the Linux kernel, HC3 Cloud Unity provides a turn-key, zero-config network fabric to extend your ground and cloud topologies easily and without confusing (and potentially costly) VPNs.

Google Compute Engine is Google's cloud platform and shares the same hypervisor (KVM) as Scale Computing's HC3 platform. This is the basis for Cloud Unity.

HC3 Cloud Unity Overview

HC3 Cloud Unity is an Infrastructure-as-a-Service (IaaS) offering, providing the capabilities of Scale Computing's HC3 platform in a cloud-hosted environment with built-in ground-to-cloud networking. In essence, virtualized HC3 single nodes run on another KVM instance, hosted by Google Compute Engine. Hypervisor-in-hypervisor is allowed by nested virtualization, and is the key feature that unlocks the power of HC3 running on Google's Compute Engine.

Flexibility

One of the principal advantages that cloud providers have always had is the fact that you can have your resources be demand-driven and elastic. You can create virtual machines that match specific needs and requirements with virtual computing. Often, virtual configurations can be achieved which would be either impossible or exceedingly costly with physical hardware.

The same advantages are in-place with HC3 Cloud Unity. Virtual computing instances, or Virtual Single Node Appliance Configurations (vSNACs), are also able to be designed and configured with resources to meet specific needs for performance or storage which would otherwise be impossible to achieve with commodity, off-the-shelf (COTS) hardware.

Unique Advantages

With a cloud provider, the underlying physical hardware is obfuscated from the virtual guest computing instances that are created. As with HC3, the virtual guests are presented with virtualized hardware, VIRTIO disks and NICs, virtual display adapters, etc.; virtual nodes have virtual components presented to them from the Google Compute Engine (GCE). A given Windows guest VM running on HC3 would have no awareness of the underlying, physical hardware that the system is using to create a virtual disk. In the same fashion, a vSNAC is not aware that it is running on a cloud platform as opposed to physical x86 hardware.

HC3 Cloud Unity Gateway

The HC3 Cloud Unity Gateway is a micro-VM that provides connectivity to the virtual nodes running in Google's cloud as well as Scale Computing HC3 VMs running in the cloud. The remote virtual nodes have a built-in, software-defined networking endpoint that is linked to a static, public IP address. The HC3 Cloud Unity gateway VM is not a proprietary HC3 VM (there is no such thing), and can be exported and run on any hypervisor. The only fully-tested configuration for remotely running the HC3 Unity Gateway VM is currently on an HC3 system or SNAC.

For tests, or actual disaster recovery declarations, the HC3 Cloud Unity Gateway still remains a critical component for accessing your cloud node. The core requirements for the HC3 Cloud Unity Gateway VM are an x86-capable hypervisor and Internet access. The HC3 Cloud Unity Gateway is designed only to be run remotely. The HC3 Cloud Unity nodes which run on Google already have the components for the software defined networking modules built into the image and an HC3 Cloud Unity Gateway VM is not required to be running on a cloud instance.

Specifically, the HC3 Cloud Unity Gateway will need outbound TCP port 22 access (ssh) in order to establish the initial connection between the gateway and the cloud node.

Cloud Unity Platform Details

Cloud Overview

The Scale Computing HC3 Cloud Unity Platform currently runs on Google's Compute Engine, which is a part of their suite of cloud-based offerings.

Ultimately, these providers are offering demand-based resources, with granular billing on an infrastructure that is invisible to the end-users. These resources can be hosted databases, log-analysis buckets, or virtual machines. Cloud-computing offers the appeal of not having to manage or deal with any underlying hardware or physical infrastructure. Ideally, it would also be complemented by an easy UI/orchestration layer that enhances the ease of use.

The simplification of the experience is what HC3 delivers as well, in a private-cloud that is on-premises. You are using a simple web-based interface to create and provision virtual machines, without having to manage or coordinate the underlying compute and storage resources.

HC3 Cloud Unity Physical Layer

HC3 Cloud Unity is an extension of the HC3 platform, designed to run on cloud infrastructure, rather than physical, on-premises x86 hardware. With the nested virtualization, it became possible to allow for virtual HC3 nodes to be created on cloud infrastructure.

As cloud providers guarantee fault tolerance and data consistency on their hardware, and in what they present to computing guests running in their cloud, the management of physical hardware components and ensuring their availability to on-premises hardware is no longer required. There is no need to create a node or a cluster with multiple-disks for redundancy, no need for multiple power-supplies, or to monitor temperature values or SMART attributes for disks. The virtual nature of the hardware allows for performance to meet the performance profiles of your ground HC3 systems.

As a single-node with a large pool of resources that are managed by the cloud provider, the need to create clusters for high-availability is also deprecated. Clustering provides performance, fault tolerance and scalability. However, clustering is a very complicated endeavor, whether it is a file-system, volume, group of compute nodes, or distributed storage; clustering concepts add sophistication, but also complexity.

Storage

For HC3 Cloud Unity VM storage, there is a persistent disk that is presented to the virtual Cloud Unity nodes.The disk provides consistency and a guaranteed number of IOPS. See the following table for the IOPS that are provided on a linear, per-GB basis, with maximum sustained write IOPS of 15,000 and read IOPS of 3,000.

Maximum Sustained IOPS	Standard Persistent Disks
Read IOPS per GB	0.75
Write IOPS per GB	1.5
Read IOPS per instance	3,000
Write IOPS per instance	15,000
Maximum Sustained IOPS	**Standard Persistent Disks**
Read throughput per GB	0.12
Write throughput per GB	0.12
Read throughput per instance	180
Write throughput per instance	120

Compute

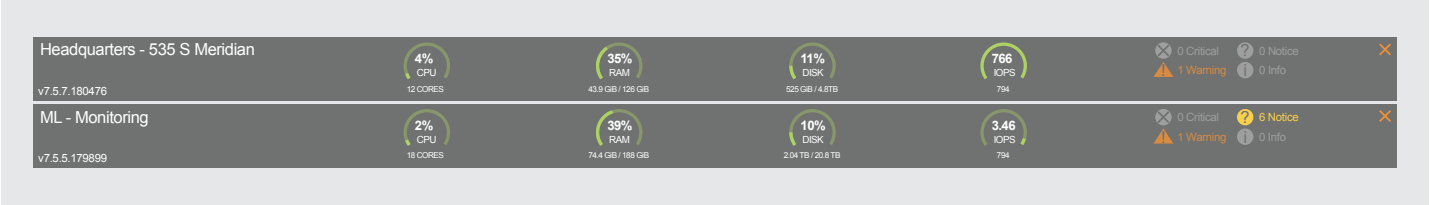
For compute resources, the physical hosts providing the compute layer have a variety of CPU platforms available. In the case of HC3 Cloud Unity nodes, Scale Computing is choosing to leverage Intel Haswell and newer platforms to take advantage of the VT-X extensions from Intel, which enable nested-virtualization.

Google presents threads to virtual machine guests as cores, and at a clock speed which matches the platform.

Management

As HC3 Cloud Unity nodes are simply virtualized Scale Computing HC3 nodes, the management of the remote node is as if there were another physical node on premises or at another datacenter. The HC3 Cloud Unity node is accessible via the configured IP address and managed with the same HC3 web interface as physical systems.

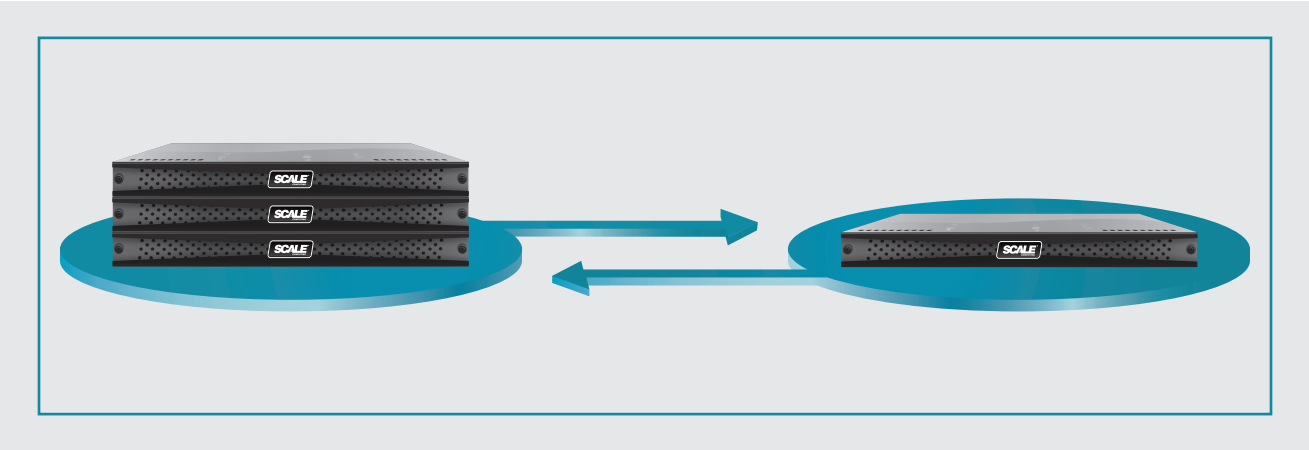
With the enhancements to remote-cluster monitoring and visibility, an at-a-glance view of the HC3 Cloud Unity instance is available within the ground cluster console[1].



[1] *Requires HCOS version 7.4 or later*

HC3 Cloud Unity DR

HC3 Cloud Unity DR is a fully-managed Disaster Recovery-as-a-Service (DRaaS) offering that is entirely built on the HC3 Cloud Unity platform.



Turn-key Disaster Recover

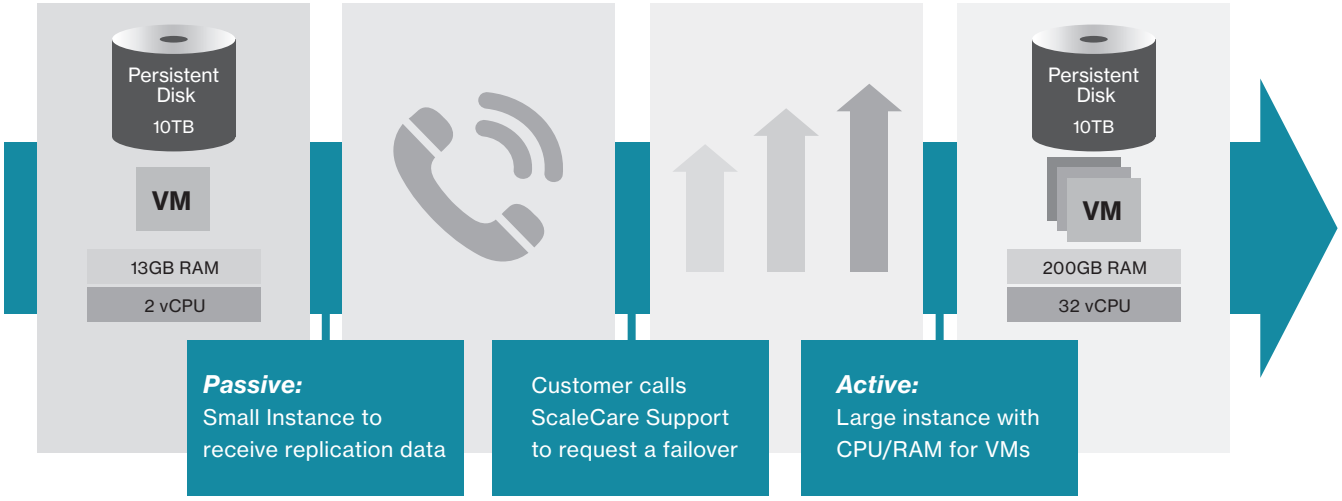
HC3 Cloud Unity DR extends the functionality of HC3 Cloud Unity to provide an end-to-end disaster recovery and workload protection platform. Combining the flexibility of HC3 Cloud Unity, the power of the HC3 Cloud Unity Gateway to provide seamless ground-cloud-ground networking, and native HC3 replication yields a holistic data protection offering.

HC3 Cloud Resource Efficiency

HC3 Cloud Unity DR leverages the built in replication of HC3 and the HC3 Cloud Unity Gateway network simplicity to provide a cloud-based platform to both protect and run your workloads. While there is a wide spectrum of HC3 Cloud Unity DR offerings, some of which have resources which equal physical systems, running resource-intensive virtual nodes constantly in order to replicate data on Cloud Unity can be costly.

However, HC3 has a unique advantage: replicating data is a low-overhead process, and you will not need to run an instance that is equivalent to your physical system in order to replicate. You can replicate your desired set of VMs actively from your physical system, but the replication target runs in a passive mode. The ***passive*** mode VM is a fixed-size instance that is small enough to run the required replication services without unnecessary overhead. A passive-mode VM consumes a fraction of the resources required of an active node, and thus offers a considerable savings over the cost of running an active-mode full-time for replication purposes.

The ***active*** mode HC3 Cloud Unity VM instance has enough RAM and CPU allocated in order to run the workloads required to properly declare and institute a full disaster recovery scenario. Active mode is only required during a disaster declaration or test.



As the diagram illustrates, the passive instance requires a few cores and a small amount of memory. This is the standard mode for replicating data. The active instance has more cores and memory and is designed to run workloads.

When a DR declaration is required or a test is underway the instances can be resized from a passive to an active mode within a few minutes.

Summary

The hybrid cloud functionality of HC3 Cloud Unity is the next evolution in extending hyperconverged infrastructure into the cloud while still preserving the simplicity at the heart of the HC3 product. HC3 Cloud Unity uniquely combines on-premise and cloud infrastructure concepts by leveraging the efficiency, resiliency, and scalability offered by Google's Compute Engine in concert with the simplicity, scalability, reliability, and affordability of the HC3 system.

With the foundation in place for a true hybrid cloud that is affordable and accessible with the HC3 product, the path for further innovation has been paved. Scale Computing and Google will continue to develop and enhance HC3 Cloud Unity and the Google Compute Engine respectively and with the opportunity for additional functionality coming in the future.

Below are links to additional resources you might find useful. If you are interested in learning more about HC3 Cloud Unity, please contact Scale Computing at 877.SCALE.59.

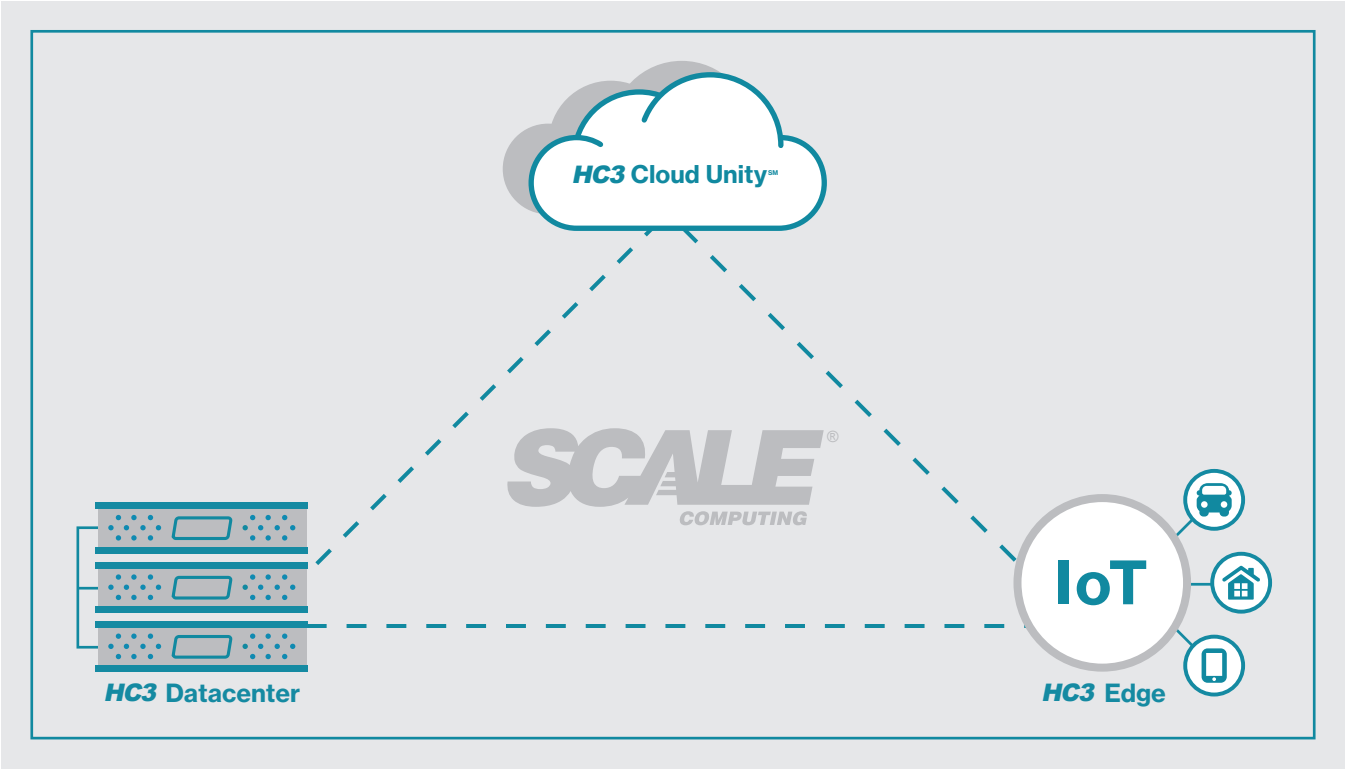
Resources

Scale Computing

- *FAQ: HC3 Cloud Unity DRaaS*
- *HyperCore / SCRIBE Theory of Operations*
- *HyperCore HC3 Capacity, Clone, and Snapshot Management*
- *Snapshot Scheduling Feature Note*
- *HyperCore HC3 Native Replication Feature Note*
- *Blog: HC3 Cloud Unity with Google Cloud Platform*

Google

- *Blog: Introducing Nested Virtualization for Google Compute Engine*



Corporate Headquarters
525 S. Meridian Street
Indianapolis, IN 46225
P. +1 317-856-9959

West Coast Office
360 Ritch Street
Suite 300
San Francisco, CA 94107

EMEA Office
Saunders House
52-53 The Mall
London
W5 3TA
United Kingdom

www.scalecomputing.com

1-877-SCALE-59 (877-722-5359)