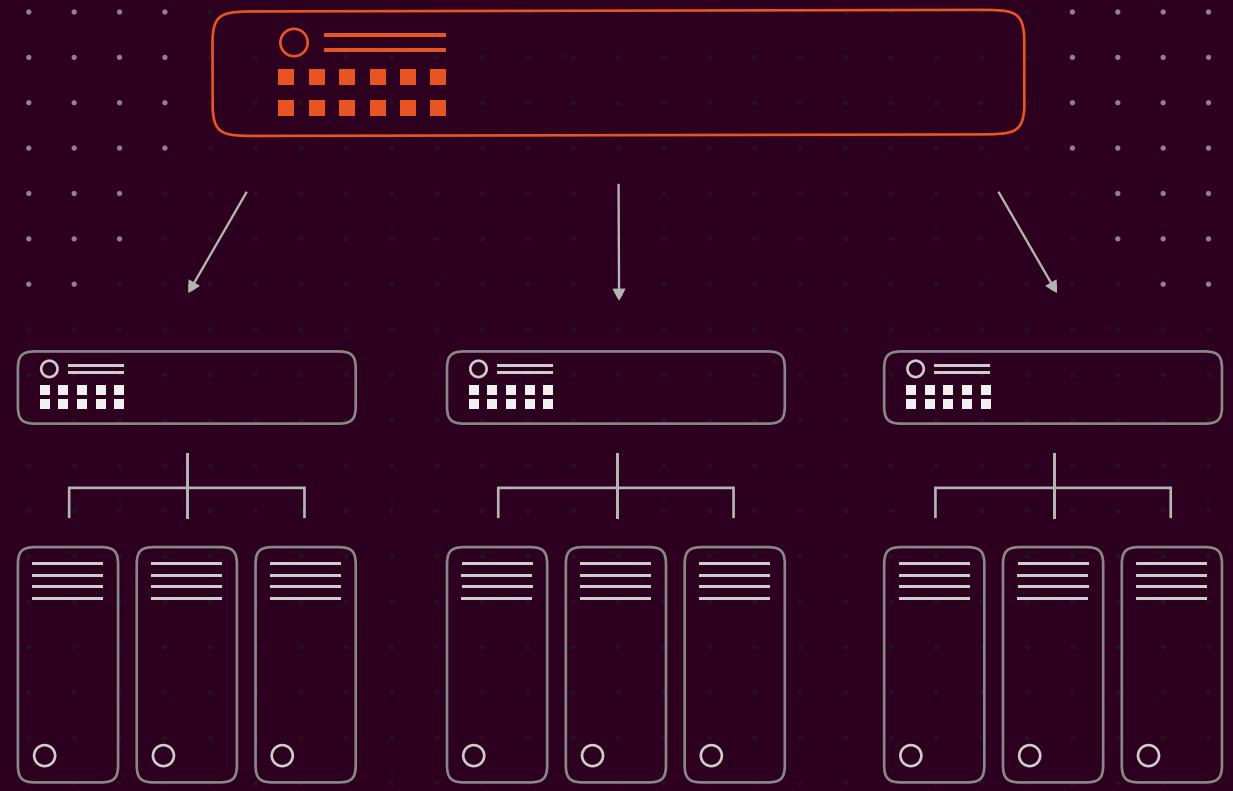
ubuntu
Delivered by Canonical

eBook

Server provisioning: What Network Admins and IT Pros need to know



What you will learn

This document is designed to help system administrators and DevOps focused Organisations to understand bare metal server provisioning, understand its value proposition, and learn about how leading companies are using server provisioning solutions within their hyperscale environments. Canonical addresses these requirements with the open source utility MAAS (Metal as a Service).

This solution helps Organisations to take full advantage of existing hardware investments by maximising hardware efficiency, and a pathway to leverage the performance and security of hardware based solutions with the economics and efficiencies of the cloud.

About the author

Christopher Wilder has domain expertise in Cloud Computing and Infrastructure, the Internet of Things (IoT), machine learning, business Analytics, networking, communications, and software defined infrastructure.

Chris is the author of the book Big Software Has Arrived, and the co-author of the best-seller, Influencing the Influencers. Chris is a frequent contributor to Forbes and TechTarget. He has also published multiple columns on software and technologies in *The New York Times, Boston Globe, CEO Magazine*, and others. He serves on TechTarget's Cloud Advisory board, and is a trusted advisor for dozens of technology companies worldwide.

Christopher Wilder
Content Marketing, Canonical

Contents

Executive summary

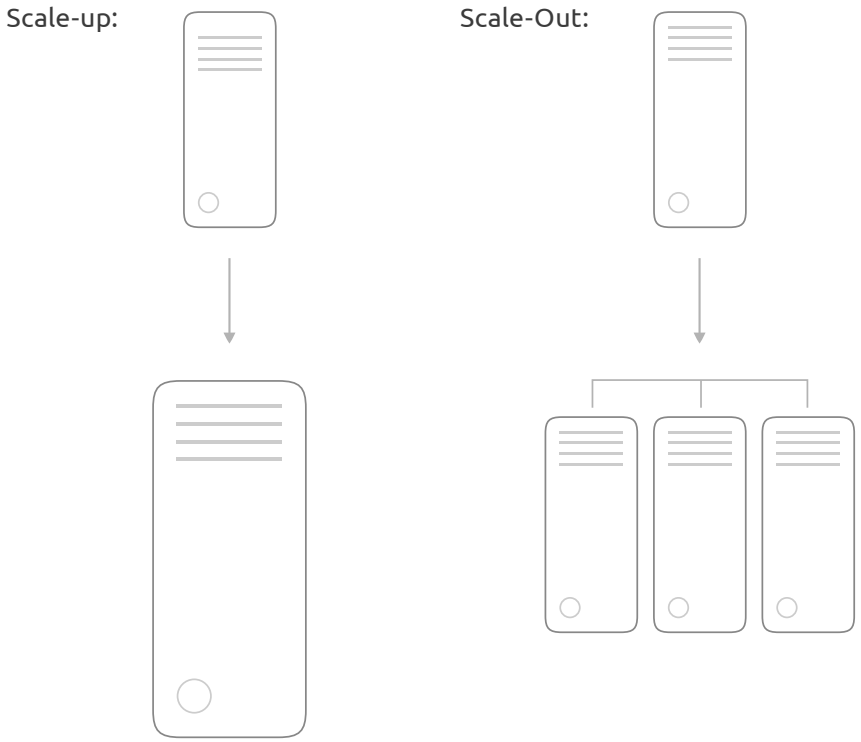
As Larry Ellison, founder of Oracle, once famously said of cloud "All it is, is a computer attached to a network." Larry and Oracle have since embraced cloud technologies such as OpenStack yet the basic premise that it starts with a physical server and a network still holds true. Organisations wishing to run a cloud on premises need to master bare metal servers and networking and this is causing a major transition in the datacenter.

Big Software, IoT (Internet of Things), and Big Data are changing how operators must architect and deploy and managed servers and networks. The traditional Enterprise scale-up models of delivering monolithic software on a limited number of big machines are being replaced by scale-out solutions that are deployed across many environments on many servers.

This shift has forced data centre operators to look to alternative methods of operation that can deliver huge scale while reducing costs.

As the pendulum swings, scale-out represents a major shift in how data centres are deployed today. This approach presents a more agile and flexible way to drive value to cloud deployments while reducing overhead and operational costs. Scale-out is driven by a new era of software (web, Hadoop, Mongodb, ELK, NoSQL, etc.) that enables Organisations to take advantage of hardware efficiencies whilst leveraging existing or new infrastructure to automate and scale machines and cloud-based workloads across distributed, heterogeneous environments. This next generation of software brings new automation and deployment tools, efficiencies, and methods for deploying distributed systems in the cloud.

Big Software is the key driver forcing Organisations to focus on tools and models to deploy and manage workloads and applications spread across scaled-out environments, including disparate data centre and cloud environments. All the while, optimising components within distributed, often hyper-converged hardware environments.



Scale-up vs. Scale-out

However, no matter what infrastructure you have, there are bare metal machines under it, somewhere. When rolling out data centre deployments, companies need a tool that can provision everything they need, while working with the infrastructure they have. For private infrastructure to thrive in the cloud era, it must be agile and efficient.

In the data centre, Organisations had significant friction with the onboarding, provisioning, and management of their physical hardware. This is, in large part, why Virtual Machines (VMs) became popular.

As VMs appeared and evolved within the data centre, enterprises moved from purpose-configured applications designated for specific hardware configurations to more general solutions designed to work within a virtual machine environment.

Hardware has always been an expensive and difficult resource to deploy within a data centre, but is unfortunately still a major consideration for any organisation moving all or part of their infrastructure to the cloud.

To become more cost-effective, Organisations hire teams of developers to cobble together software solutions that solve functional business challenges while leveraging existing legacy hardware in the hopes of offsetting the need to buy and deploy more hardware-based solutions. VMs require a hypervisor, which enables streamlined operations through software, but managing the hardware itself remains a painful journey through proprietary APIs and often incompletely-implemented specifications like Distributed Management Task Force's IPMI (intelligent platform management interface).

Organisations are looking for more efficient ways to balance their hardware and infrastructure investments with the efficiencies of the cloud. Canonical's MAAS (Metal As A Service) is such a technology. MAAS is effectively a hardware API that turns bare metal servers into its own cloud, without virtual machines. MAAS is designed for devops at scale, in places where bare metal is the best way to run your app. Big data, private cloud, PAAS and HPC all thrive on MAAS.

MAAS is trusted by

Cloud speed with bare metal reliability and efficiency

MAAS allows operators to deploy physical hardware as opposed to virtual environments. Within the service there are common technologies like PXE (preboot execution environment) and IPMI to ensure interoperability and support for a range of hardware. MAAS makes it easy to provision physical servers as easily as deploying a virtual machine in the cloud with full programmatic control over the hardware and its capabilities. Further, MAAS works across all vendors and operating systems including Windows, Ubuntu, CentOS, RedHat and Suse.

MAAS is the fastest way to deploy operating systems

MAAS isn't a new concept, but demand and adoption rates are growing because many enterprises want to combine the flexibility of cloud services with the raw power of bare metal servers to run high-power, scalable workloads. MAAS, however, is a new way of thinking about physical infrastructure and how Organisations can leverage the best of all worlds. This is especially true for compute, storage, and networking as they have become commodities in the virtual world. MAAS lets enterprises treat farms of servers as malleable resources for dynamic allocation to specific areas within the ecosystem.

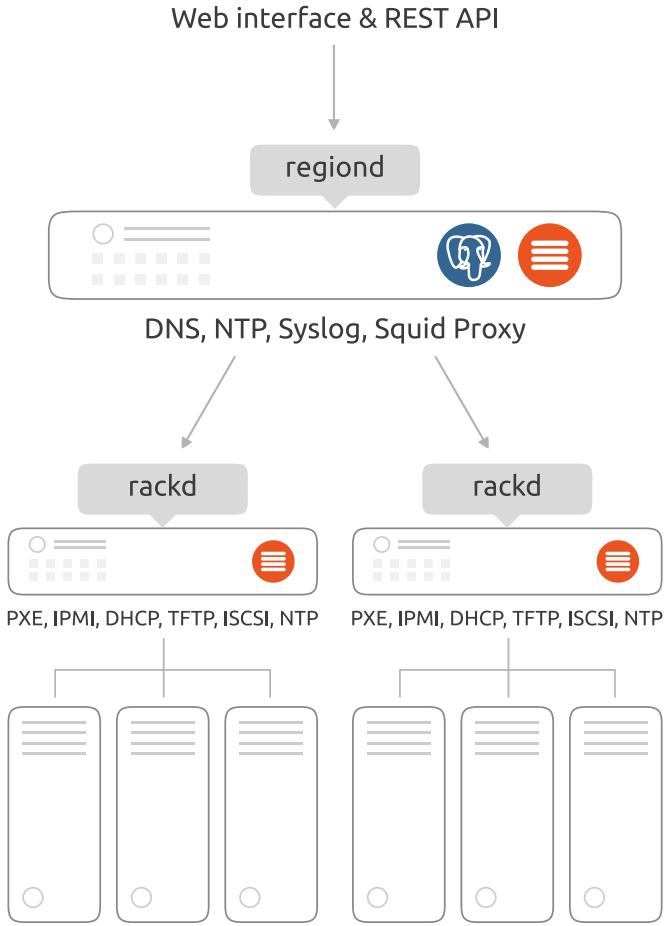
MAAS is much like any XaaS business model dedicated to a specific tenant, but the main difference is customers choose the type of compute configuration they want in their servers (e.g. x86, single dual, quad core processors) combined with applicable storage, memory, and other functionality. Applications and workloads are deployed onto servers that have the sufficient compute power, storage, and an operating system that allows for optimal performance and efficiency.

MAAS is the fastest way to deploy operating systems

Search for more supported hardware on our partner portal.

For example, when a new server needs to be deployed, MAAS automates most, if not all, of the provisioning process. Automation makes deploying solutions much quicker and more efficient because it allows tedious tasks to be performed faster and more accurately without human intervention.

Even with proper and thorough documentation, manually deploying server to run web services or Hadoop, for example, could take hours compared to a few minutes with MAAS. This is why IT Pros are looking at MAAS as a way to make the most effective use of their team's precious resources and time. Moreover, MAAS provides a uniform way to provide a hyperscale environment for admins and users to load applications onto servers via their preferred automation tool, i.e. Chef, Puppet, Ansible, Juju etc.



...
MAAS Region Controller (regiond)

Get the most out of your hardware investment

Every IT department has made significant investments in hardware. However, as the cloud has disrupted traditional business models, IT Pros needed to find a way to combine the flexibility of the cloud with the power and security of their bare metal servers. Canonical's MAAS solution allows IT Organisations to discover, commission, and deploy physical servers within any cloud environment.

As new services and applications are deployed, MAAS can dynamically re-allocate physical resources to match cloud-based workload requirements. This means Organisations can deploy both virtual and physical machines across multiple architectures and virtual environments, at scale.

MAAS was designed to make complex hardware deployments faster, more efficient, and with more flexibility. While there are many use cases, below are a few segments that have found success.

High Performance Computing (HPC):

HPC relies on aggregating computing power to solve large data-centric problems in subjects like health care, engineering, business, science, etc. Many large Organisations are leveraging MAAS to modernise their OS deployment toolchain (a set of tool integrations that support development, deployment, and operations tasks) and lower server provisioning time.

These Organisations found their tools were outdated thereby prohibiting them from deploying large numbers of servers. Server deployments were slow, modular/monolithic, and could not integrate with tools, drivers, and APIs. By deploying MAAS they were able to speed-up their server deployment times as well as integrate with their orchestration platform and configuration management tools like Chef, Ansible, and Puppet, or software modeling solutions like Juju.

Smart Data Centers

Servers installed within a data centre typically serves a single purpose for the duration of its life. Smart data centres enable the full utilisation of hardware, thus improving the total cost of ownership (TCO). With MAAS, smart data centre operators like Walmart and Box can quickly power off a server and install a different OS for a few hours to perform different tasks. MAAS enables multi purpose server usage which improves efficiency and doesn't let servers go underutilised.

For example, banks typically use full server power during their normal work hours, taking in requests from customers (e.g. web banking). During low volume traffic, unutilised server power can be redeployed dynamically to do perform other tasks e.g. fraud detection, batch processing, etc. To make this process completely automatic an orchestration tool is required, but MAAS ensures the reallocation is done quickly.

Hybrid Cloud

In a hybrid cloud environment, which is an environment that leverages on-premise or private cloud infrastructure with public cloud utilising orchestration tools or service modeling solutions, MAAS optimises and unifies operations. MAAS exposes bare-metal server provisioning operations and an API (application programming interface) that can be consumed by service modeling solutions like Canonical's Juju as the building blocks for an optimised hybrid cloud.

As an example, many large enterprises that rely on transactions as a major part of their business model (retail, airlines, etc.) manage their infrastructure via private cloud. However, during peak demand times they require extra support via public cloud providers like Amazon Web Services, Microsoft Azure, and Google Cloud Platform. Canonical's Juju works seamlessly between each environment to ensure communications between public cloud APIs and the organisation's private cloud (i.e OpenStack). In some cases, the private cloud needs to run from bare metal servers (i.e. Hadoop). In those cases, the only possible way to interface at that level is with MAAS, which provides an API that allows administrators and users to provision solutions like a VM.

Each of these examples demonstrates how forward-thinking Organisations are using MAAS and other technologies to take full advantage of their infrastructure investments.

How the smartest IT Pros let software do the work

MAAS acts as an abstraction layer between the management layer and the underlying physical hardware. MAAS also can discover existing hardware resources and automate many management tasks, including:

- Installing, configuring, and monitoring bare metal hardware. Including but not limited to, servers, switches, power distribution units (PDU)/mains distribution units (MDU), and Data Acquisition Engines (DAE), etc.
- Install and upgrade firmware, patches, and updates
- Automated server utilisation and re-utilisation based on need
- Discovery of compute, network, capabilities, and storage based on server
- Power on and off servers as needed

By automating these functions MAAS eliminates the extensive manual process required for traditional server operations and allows Organisations to become more operationally efficient.

Making MAAS Work

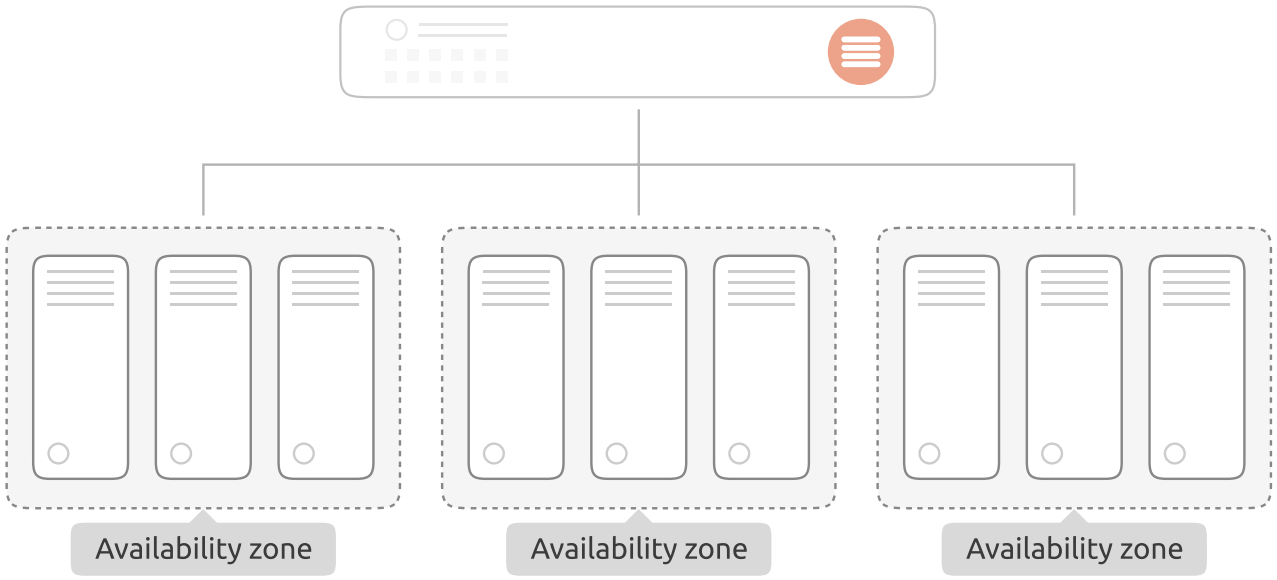
MAAS has a tiered architecture with a central postgres database backing a 'Region Controller (regiond)' that deals with operator requests. Distributed Rack Controllers (rackd) provide high-bandwidth services to multiple racks. The controller itself is stateless and horizontally scalable, presenting only a REST API.

Rack Controller (rackd) provides DHCP, IPMI, PXE, TFTP and other local services. They cache large items like operating system install images at the rack level for performance but maintain no exclusive state other than credentials to talk to the controller.

Physical availability zones

In keeping with the notion of a 'physical cloud' MAAS lets you designate machines as belonging to a particular availability zone. It is typical to group sets of machines by rack or room or building into an availability zone based on common points of failure. The natural boundaries of a zone depend largely on the scale of deployment and the design of physical interconnects in the data centre.

Nevertheless the effect is to be able to a scale-out service across multiple failure domains very easily, just as you would expect on a public cloud. Higher-level infrastructure offerings like OpenStack or Mesos can present that information to their API clients as well, enabling very straightforward deployment of sophisticated solutions from metal to container.



MAAS physical availability zones

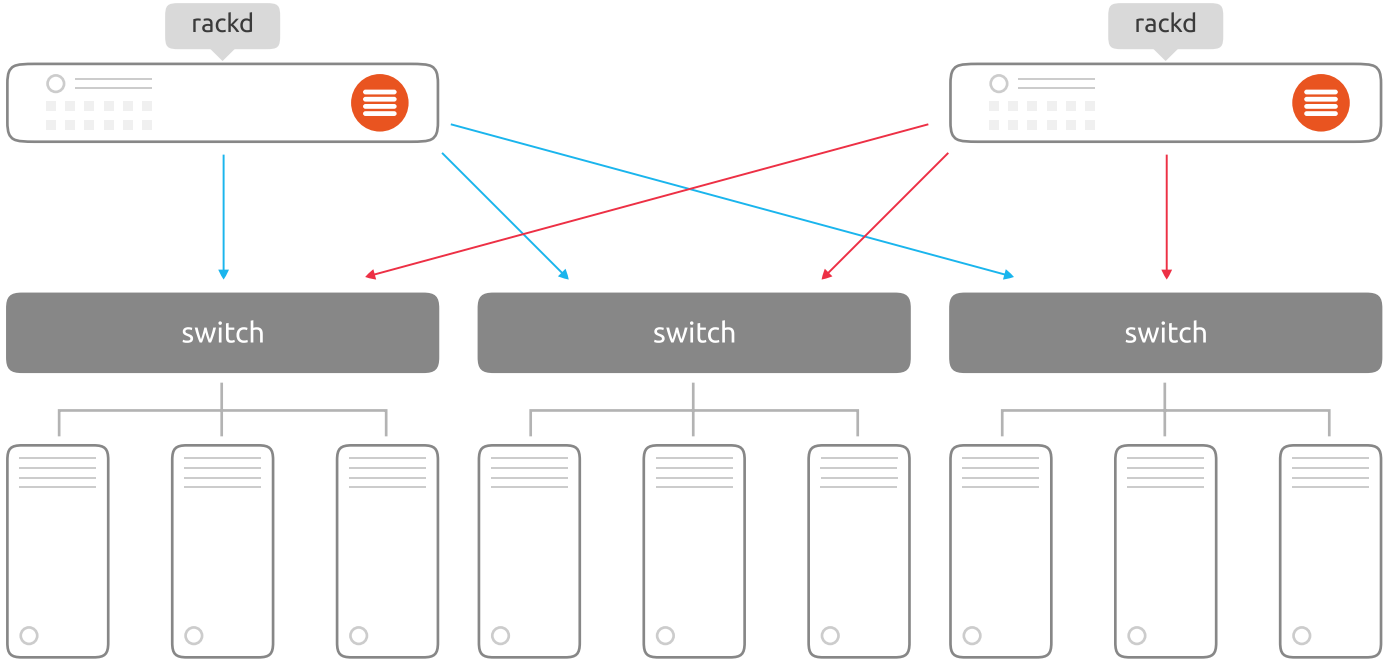
The MAAS API allows for discovery of the zones in the region. Chef, Puppet, Ansible and Juju can transparently spread services across the available zones. Users can also specifically request machines in particular Availability Zones.

There is no forced correlation between a machine location in a particular rack and the zone in which MAAS will present it, nor is there a forced correlation between network segment and rack. In larger deployments it is common for traffic to be routed between zones, in smaller deployments the switches are often trunked allowing subnets to span zones.

By convention, users are entitled to assume that all zones in a region are connected with very high bandwidth that is not metered, enabling them to use all zones equally and spread deployments across as many zones as they choose for availability purposes.

The node lifecycle

Each machine (“node”) managed by MAAS goes through a lifecycle – from its enlistment or onboarding to MAAS, through commissioning when we inventory and can setup firmware or other hardware-specific elements, then allocation to a user and deployment, and finally they are released back to the pool or retired altogether.



MAAS high availability

New

New machines which PXE-boot on a MAAS network will be enlisted automatically if MAAS can detect their BMC parameters. The easiest way to enlist standard IPMI servers is simply to PXE-boot them on the MAAS network.

Commissioning

Detailed inventory of RAM, CPU, disks, NICs, specific models, serial numbers, and accelerators like GPUs itemised and usable as constraints for machine selection. It is possible to run your own scripts for site-specific tasks such as firmware updates.

Ready

A machine that is successfully commissioned is considered “Ready”. It will have configured BMC credentials (on IPMI based BMCs) for ongoing power control, ensuring that MAAS can start or stop the machine and allocate or (re)deploy it with a fresh operating system.

Allocated

Ready machines can be allocated to users, who can configure network interface bonding and addressing, and disks, such as LVM, RAID, bcache or partitioning.

Deploying

Users then can ask MAAS to turn the machine on and install a complete server operating system from scratch without any manual intervention, configuring network interfaces, disk partitions and more.

Releasing

When a user has finished with the machine, they can release it back to the shared pool of capacity. You can ask MAAS to ensure that there is a full disk-wipe of the machine when that happens.

Making hardware investments work

Systems must be configured to ensure maximum throughput and service delivery. Because each application has different demands and resource utilisation, many Organisations tend to over-build to compensate for peak-load, or they will over-provision VMs to ensure enough capacity years out. With MAAS, today's IT Pros no longer have to perform capacity planning five-years out. Instead, they can develop strategies for creating differently configured hardware and cloud archetypes to cover all classes of applications within their current environment and existing IT investments.

MAAS makes it possible for Organisations to make the most of their hardware by enabling hardware to reprovision systems for the needs of the data centre. For example, a server used for transcoding video 20 minutes ago is now a Kubernetes worker node, later a Hadoop Map-Reduce node, and tomorrow something else entirely.

One of the often overlooked components to scale-out are the tools and techniques for leveraging bare metal servers within the environment. What happens in the next 3-5 years will determine how end-to-end solutions are architected for the next several decades. OpenStack has provided an alternative to public cloud. Containers have brought new efficiencies and functionality over traditional VM models, and service modeling brings new flexibility and agility to both enterprises and service providers, while leveraging existing hardware infrastructure investments to deliver application functionality more effectively.

Further, by complementing MAAS with Juju, IT Organisations can leverage bundles of Charms (sets of encapsulated code for deploying and managing services) to automatically deploy and configure various server software stacks and applications functionality. Juju integrates seamlessly with MAAS, making it possible to centrally deploy software to the hardware nodes in a server cluster. Using MAAS and Juju together can significantly reduce the difficulty deploying an OpenStack private cloud, thereby increasing time to market.

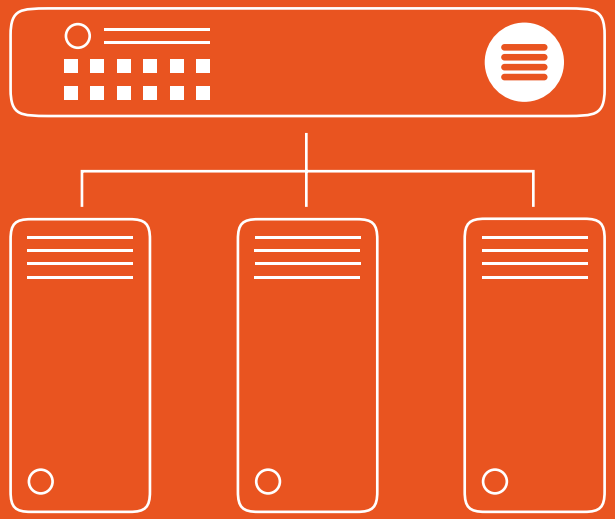
Conclusion

The industry is at a pivotal period, transitioning from traditional scale-up models of the past to scale-out architecture of the future where solutions are delivered on disparate clouds, machines, and environments simultaneously. IT customers need to have the flexibility of not ripping and replacing their entire infrastructure to take advantage of the opportunities the cloud offers. This is why new architectures and business models are emerging. Canonical's MAAS is a mature solution to help Organisations to take full advantage of their cloud and legacy hardware investments.

Get started with MAAS

To download and install MAAS for free please visit **ubuntu.com/download/server/provisioning**

Or to talk to one of our scale-out experts about deploying MAAS in your datacenter contact us at **ubuntu.com/about/contact-us/form**



About Canonical

At Canonical, we are passionate about the potential of open source software to transform business. For over a decade, we have supported the development of Ubuntu and promoted its adoption in the enterprise.

By providing custom engineering, support contracts and training, we help clients in the telecoms and IT services industries to cut costs, improve efficiency and tighten security with Ubuntu and OpenStack. We work with hardware manufacturers like HP, Dell and Intel, to ensure the software we create can be delivered on the world's most popular devices. And we contribute thousands of man-hours every year to projects like OpenStack, to ensure that the world's best open source software continues to fulfil its potential.