Edge Cloud Infrastructure Specification

Version 2.0 – March 2021

Abstract

OpenNebula Edge Cloud is a distributed architecture composed of edge clusters that can run **any workload**—both virtual machines and application containers— **on any resource**—bare metal or virtualized— **anywhere**—on-premises and on a cloud provider. It enables true hybrid and multi-cloud computing by combining public and private cloud operations with workload portability and unified management of IT infrastructure and applications. Based exclusively on solid open source technologies, our Edge Cloud Architecture implements enterprise-grade cloud features for performance, availability and scalability with a very simple design that avoids vendor lock-in and reduces complexity, resource consumption and operational costs.

Our Edge Cloud Architecture design has been defined to be much simpler than traditional cloud computing architectures, which are usually composed of complex general-purpose software systems. Our alternative is based on storage and networking software that already exist in the Linux operating system and modern storage hardware that is available from existing public cloud and edge providers. This guide discusses infrastructure specifications and the tools and services needed to set up the infrastructure software, including the foundation edge clusters and the front-end. It also covers the configuration details that are needed when the front-end or the clusters run on public cloud resources or on the edge.

Contents

Glossary

AZ	Availability Zone
HA	High Availability
VM	Virtual Machine
VPC	Virtual Private Cloud

1. What is the OpenNebula Edge Cloud Architecture?

OpenNebula is a **powerful, but easy-to-use, open source solution to build and manage Enterprise Clouds**. It combines virtualization and container technologies with multi-tenancy, automatic provision and elasticity to offer on-demand applications and services.

OpenNebula provides a single, feature-rich and flexible platform with **unified management of IT infrastructure and applications** that **avoids vendor lock-in** and **reduces complexity, resource consumption and operational costs**. OpenNebula manages:

- **Any Application**: Combine containerized applications from Kubernetes and Docker Hub ecosystems with virtual machine workloads in a common shared environment to offer the best of both worlds: mature virtualization technology and orchestration of application containers.
- **Any Infrastructure**: Unlock the power of a true hybrid, edge and multi-cloud platform by combining private cloud with infrastructure resources from third-party virtual and bare-metal cloud providers such as AWS and Packet (Equinix Metal).
- **Any Virtualization**: Integrate multiple types of virtualization technologies to meet your workload needs, including VMware and KVM virtual machines for fully virtualized clouds, LXC system containers for containers clouds, and Firecracker microVMs for serverless deployments.
- **Any Time**: Add and remove automatically new clusters in order to meet peaks in demand, or to implement fault tolerant strategies or latency requirements.

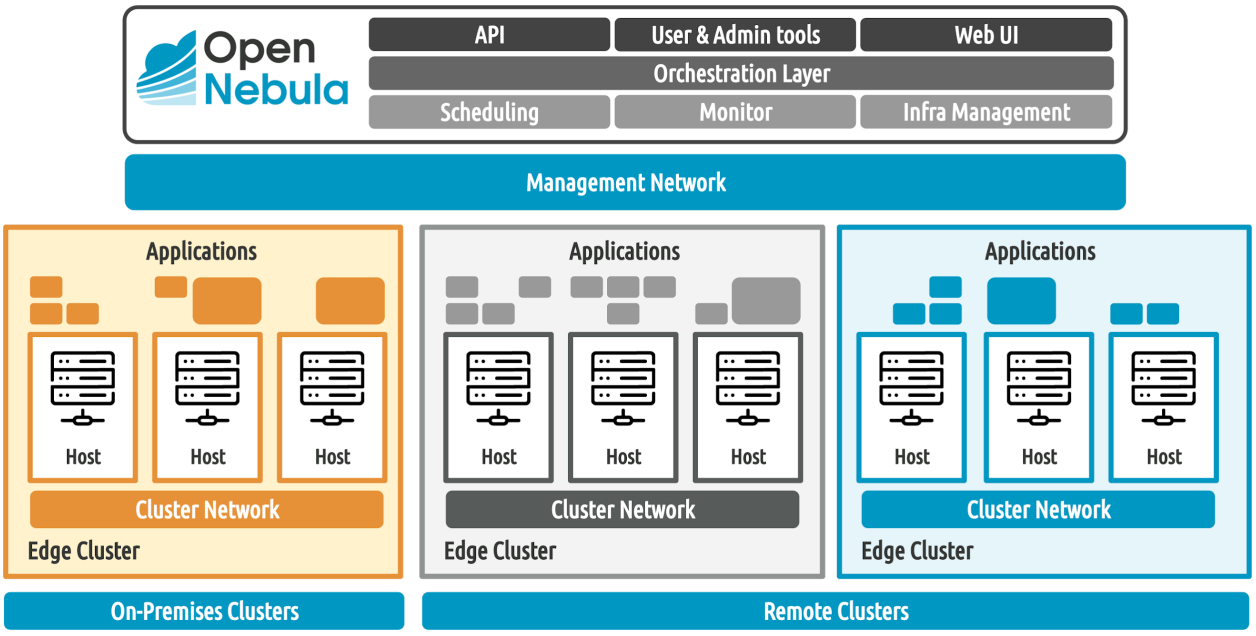

Figure 1. Main components of our Edge Cloud Architecture.

From the experience of working with hundreds of users and client engagements, OpenNebula Systems has defined an Edge Cloud Architecture that implements enterprise-grade cloud features for performance, availability and scalability with a very simple design that avoids vendor lock-in and reduces complexity, resource consumption and operational costs. Its architecture design is described in our **Edge Cloud Reference Architecture**.[1] This guide specifically discusses infrastructure specifications and the tools and services needed to deploy the infrastructure software, including the foundation edge clusters and the front-end. It also covers the configurations that are needed when the front-end or the clusters run on cloud resources. Performance, reliability and scalability depend on the underlying hardware specifications.

[1] https://support.opennebula.pro/hc/en-us/articles/360050302811

2. Front-end Deployment

The maximum number of virtual instances and servers (virtualization hosts) that can be managed by a single OpenNebula front-end instance strongly depends on the performance and scalability of the underlying cloud architecture and platform infrastructure, mainly the storage subsystem.

The capacity specifications and configurations described in this section have been tested for installations running up to 20,000 VMs on 1,250 hosts.[2] The differences in the underlying hardware and performance tuning can result in varying capabilities even between similar configurations. Contact us if you need larger scalability. The general recommendation is that no more than 2,500 servers should be managed by a single front-end.

You can scale down your resources to the requirements of your target workload. For example when running the front-end in a cloud resource, you can use `t1.small.x86`[3] in Packet (Equinix Metal) or `t2.medium`[4] in AWS if you are planning to deploy medium-scale clouds.

The front-end sustains the Image Datastores of the distributed cloud, and it stores the catalog of the apps available in the cloud. The storage area of the front-ends needs to be dimensioned accordingly.

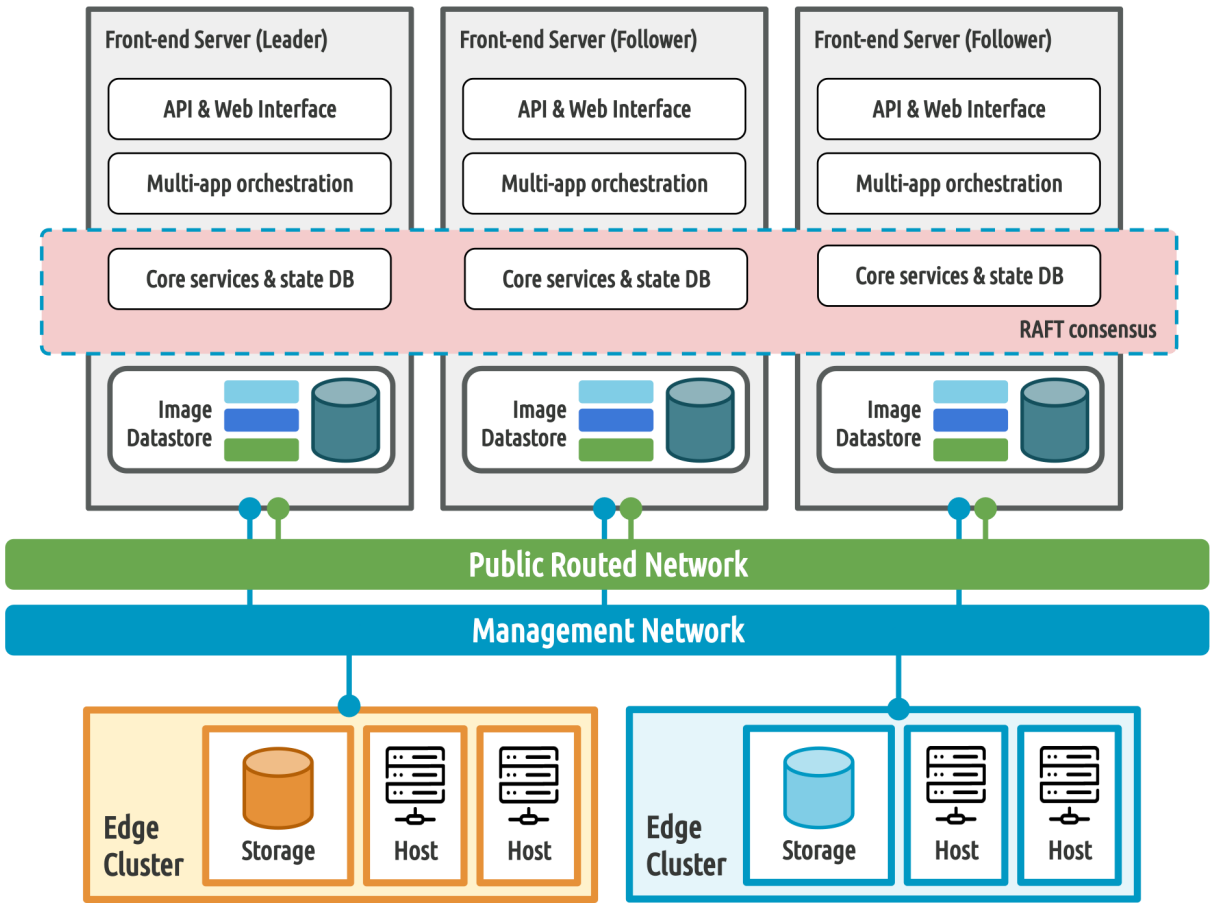

Figure 2. Front-ends in a HA configuration and interconnection with edge clusters.

[2] http://docs.opennebula.io/5.12/deployment/references/one_scalability.html
[3] https://www.packet.com/cloud/servers/t1-small/
[4] https://aws.amazon.com/ec2/instance-types/

ON-PREMISES FRONT-END CONFIGURATION	
Number of Servers	1 (3 servers needed in HA configuration)
Server Specs	CPU: 1 CPU with 2 physical cores @ 2.2 GHz RAM:16GB Disk: SSD 500 GB (must be scaled to accommodate the Image Datastore) NICs: 2x10Gb/s
Operating System	RHEL/CentOS 8
Networking	VLAN (802.1Q linux bridging) • 1x10Gb/s VLAN for Management Network • 1x10Gb/s VLAN for Public Routed Network
Connectivity Requirements (inbound)	Traffic in Public routed network • 443 Sunstone Web UI • 2633 oned XMLRPC API • 5030 OpenNebula Gate • 2474 Flow • 29876 VNC connections to instances Traffic in Management network • 22 SSH • 4124 Monitor and status messages from clusters

PACKET (EQUINIX METAL) - FRONT-END CONFIGURATION	
Number of Servers	1 (3 servers needed in HA configuration)
Instance Specs	`c3.small.x86`[5] metal (disk size must be scaled to meet Image Datastore requirements)
Location	Any (3 different AZ recommended for HA configuration)
Operating System	RHEL/CentOS 8
Networking	L2 configuration with multiple VLANs
Connectivity Requirements (inbound)	Traffic in Public Routed Network • 443 Sunstone Web UI • 2633 oned XMLRPC API • 5030 OpenNebula Gate • 2474 Flow • 29876 VNC connections to instances Traffic in Management Network • 22 SSH • 4124 Monitor and status messages from clusters

[5] https://www.packet.com/cloud/servers/c3-small/

AWS - FRONT-END CONFIGURATION	
Number of Servers	1 (3 servers needed in HA configuration)
Instance Specs	`t2.xlarge`[6] virtual (disk size must be scaled to meet Image Datastore requirements)
Location	Any (3 different AZ recommended for HA configuration)
Operating System	RHEL/CentOS 8
Networking	VPC with Internet Gateway and subnets
Connectivity Requirements (inbound)	Traffic in Public routed network • 443 Sunstone Web UI • 2633 oned XMLRPC API • 5030 OpenNebula Gate • 2474 Flow • 29876 VNC connections to instances Traffic in Management network • 22 SSH • 4124 Monitor and status messages from clusters

3. Edge Cluster Deployment

The edge cluster capacity specifications and configurations have been defined and tested to hold within each cluster up to 25 nodes, and a single front-end to manage up to 50 clusters, achieving an overall total of 1,250 nodes per cloud instance. The differences in the underlying hardware, performance tuning and distance between clusters can result in varying capabilities even between similar configurations.

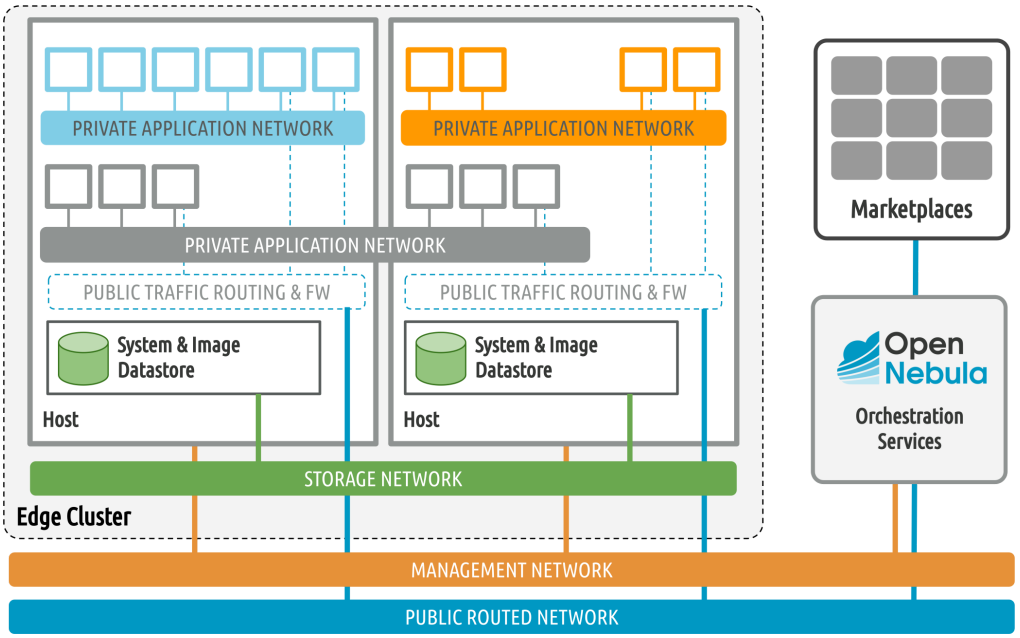

Figure 3. Overview of the networks used in the Edge Cloud Architecture.

[6] https://aws.amazon.com/ec2/instance-types/

A key task when defining a cloud infrastructure is to correctly dimension the virtualization nodes according to the expected workload, that is, the characteristics of the VMs that are going to be run in the cloud. The specifications below have been selected so a host can run comfortably 16 medium-size instances (2 virtual CPUs and 16 GB RAM) or up to 500 small-size instances (1 virtual CPU and 512 MB RAM). Alternatively, a host with these specifications can manage up to 1500 microVMs (Firecracker/KVM).

ON-PREMISES - EDGE CLUSTER CONFIGURATION	
Number of Nodes	Up to 25 (it is recommended to configure 2 of them as replica hosts with larger storage capacity for cluster failover)
Server Specs	CPU: 2 CPUs with 32 Physical Cores @ 2.2 GHz RAM: 256GB Disk: 500 GB SSD (2 TB SSD for replica hosts) NICs: 4x10Gb/s
Operating System	RHEL/CentOS 8
Hypervisor	QEMU-KVM / Firecracker-KVM / LXC
Networks	VLAN (802.1Q linux bridging) with VLAN trunk support • 1x10Gb/s VLAN for Storage Network • 1x10Gb/s VLAN for Private instances Network • 1x10Gb/s VLAN for Management Network • 1x10Gb/s VLAN for Public routed Network
Connectivity Requirements (inbound)	Traffic in Management Network • 22 SSH for hypervisors • 5900 and upwards for VNC ports

PACKET (EQUINIX METAL) - EDGE CLUSTER CONFIGURATION	
Number of Nodes	Up to 25 (it is recommended to configure 2 of them as replica hosts with larger storage capacity for cluster failover)
Instance Specs	`m3.large.x86`[7] metal
Location	Any
Operating System	RHEL/CentOS 8
Networking	L2 configuration with multiple VLANs • Elastic Public IPs • VLANs for instances • Storage VLAN
Connectivity Requirements (inbound)	Traffic in Public Network • 22 SSH for hypervisors • 5900 and upwards for VNC ports

[7] https://www.packet.com/cloud/servers/m3-large/

AWS - EDGE CLUSTER CONFIGURATION	
Number of Nodes	Up to 25 (it is recommended to configure 2 of them as replica hosts with larger storage capacity for cluster failover)
Instance Specs	`m6gd.metal`[8] metal or `m6gd.16xlarge`[9] virtual
Location	Any
Operating System	RHEL/CentOS 8
Networking	VPC with Internet Gateway and subnets • Elastic Public IPs • VPC subnet for storage network • Internal bridging networking for instances
Connectivity Requirements (inbound)	Traffic in Public Network • 22 SSH for hypervisors • 5900 and upwards for VNC ports

[8] https://aws.amazon.com/ec2/pricing/on-demand/
[9] https://aws.amazon.com/ec2/pricing/on-demand/

4. Conclusions

This document outlines the recommended capacity specifications and configurations of the infrastructure components needed to deploy OpenNebula's Edge Cloud Architecture, including the foundation edge clusters and the front-end. Note that the final performance, reliability and scalability of the cloud depends on the underlying hardware specifications and configuration tuning.

OpenNebula Systems offers Enterprise support for the complete true hybrid cloud software stack through its OpenNebula **Software Subscription** and offers managed cloud services through a new OpenNebula **Managed Subscription** so your team can forget about infrastructure and focus on business workloads.

This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement ONEedge 880412

ONEedge is an OpenNebula project developing innovative features to bring private cloud computing to the Edge (ONEedge.io)

Sign up for updates at OpenNebula.io/getupdated

Rev2.0_20210302