



CloudPhysics GCP Cloud Simulator with Sole Tenant Nodes

CloudPhysics GCP Card Analytic Overview and Frequently Asked Questions

CloudPhysics is proud to offer a cloud migration simulation to model on-premises VM and AWS EC2 instances to Google Cloud with Sole Tenant Nodes. This solution offers complex data modeling of workloads based on resource consumption, optimized workload packing, guest OS licenses, and availability on Google Cloud. The VMware to GCP Simulator with Sole Tenant provides new matching criteria to the simulator to enable the latest features and license options available from Google Cloud.

Table of Contents

CLOUDPHYSICS CLOUD SIMULATOR OVERVIEW AND SECTION.....	2
ORIENTING YOURSELF IN THE GCP CLOUD SIMULATOR	2
USING THE SIMULATOR CARD SECTIONS	4
<i>VM Matching Section</i>	4
<i>VM Rightsizing Section</i>	5
<i>VM Pricing and Discounts Section</i>	5
<i>Annual Cost Summary Section</i>	6
FREQUENTLY ASKED QUESTIONS	7

CloudPhysics Cloud Simulator Overview and Section

Orienting Yourself in the GCP Cloud Simulator

The card is broken out into multiple section. Each section offers a unique set of functions that will impact the simulator scope and costs.

The screenshot shows the GCP Cloud Simulator interface. The interface is divided into several sections, each highlighted with a red circle and a number:

- 1. Infrastructure Scope Section:** Located on the left side, it includes filters for VMs and Hosts, a search bar, and options to match any or all VMs.
- 2. VM Matching Section:** Located in the top center, it includes fields for GCP Location, Storage Type, Storage Media, and Match to standard instances only.
- 3. Sole Tenant Nodes Section:** Located in the top right, it includes a checkbox for Sole Tenant Nodes and options for Maximum Oversubscription, Allow Desktop OS/MSDN, BYO Microsoft Licenses, and Exclude Live Migration Nodes.
- 4. VM Rightsizing Section:** Located in the middle left, it includes a dropdown for Rightsizing using and a checkbox for Do not resize below CPU and Memory reservations.
- 5. Pricing and Discounts Section:** Located in the middle right, it includes a dropdown for Term Length and a dropdown for Instance Class.
- 6. Summary Card:** Located on the right side, it displays the GCP Annual Cost (\$2,026,719*), Cloud Environment Size (16 Sole Tenant Nodes, 175 Instances), and Projected Sole Tenant Node Commitment.
- 7. Restore All Defaults Button:** Located in the top right corner of the main content area.
- 8. Instance Class Filter:** Located at the bottom center, it shows a filter for n2-highmem-4 with 12 matches.
- 9. VM List Table:** Located at the bottom, it displays a table of VMs with columns for VM Name, Guest OS, VM Cost / Year, OS Cost Portion, Instance, Sole Tenant Node, Storage Media, vCPU, and vCPU P.

VM Name	Guest OS	VM Cost / Year	OS Cost Portion	Instance	Sole Tenant Node	Storage Media	vCPU	vCPU P
VM-1008711	Microsoft Windows Server 2012 (64-bit)	\$1,244	\$701	n1-custom	Node #16	SSD	3	
VM-101421	Other Linux (32-bit)	\$303	\$0	n1-custom		SSD	1	
VM-1026344	Microsoft Windows Server 2008 R2 (64-bit)	\$2,283	\$1,432	n1-custom	Node #13	SSD	4	
VM-1040005	Other Linux (32-bit)	\$200	\$0	n1-custom		SSD	1	
VM-1043003	Microsoft Windows Server 2012 R2 (64-bit)	\$4,533	\$2,370	n1-custom	Node #12	SSD	6	

The sections are as follows:

- 1. Infrastructure Scope Section:** This area will allow the user to reduce the scope of the VMs or instances within the datacenter to a select group of resources. These filters will allow you to limit your scope by either VMware or AWS criteria.
- 2. VM Matching:** This section allows you to define how you would like to match the VM to GCP cloud services. Options such as GCP Location, Storage Media, Storage Type, and Match to standard instances only all impact matching.

Sole Tenant Nodes: The introduction of Sole Tenant Nodes provides dedicated hosts for affinity of instances and isolation from other customer environment. This option also provided the benefit of allowing customers to bring their own Microsoft Licenses to further reduce their cloud costs.

Google's Details on Sole Tenant Nodes: <https://cloud.google.com/compute/docs/nodes/>

3. **VM Matching:** This section provided criteria that impacts the specific instance class and size through rightsizing. Users can keep workloads as configured or rightsized based on Peak, 99th-percentile, and 95th-percentile performance levels.
4. **Pricing and Discounts:** Users can specify the term length and instance class to impact the discounts applied to each instance. Sustained Usage discounts are applied to On-Demand instances while Committed Use discounts are applied to 1-year and 3-year term commitments. Users who choose On-Demand term lengths can also choose an availability per day average to further reduce costs for dynamic workloads.
Google Pricing for Sole Tenant Nodes: <https://cloud.google.com/compute/sole-tenant-pricing#nodes>
5. **Cost Summary:** This area provides a breakout of Instances or Instances and Sole Tenant Node costs. Colored bars under costs provide details for compute, storage, OS, and Sole Tenant portions of the total cost. Projected Sole Tenant Node Commitment bars provide a visual representation of the resource commitment or oversubscription in Sole Tenant Nodes.
6. **Export and Share Data:** An Export Data and Share Card icon are available for the customer and partner users. Some assessments allow the user or partner to export the data to a CSV file format. The CSV will consist of all data currently limited to the scope. The share data icon will allow an account user to invite an external user to have visibility into only the current card. The invited user will need to create an account and password to view the shared card.
7. **Matched and Unmatched Result Set:** These boxes represent how many VMs match into instances in the Google environment. Clicking on Unmatched will filter the result set to the unmatched workloads that do not have a target instance or are unsupported in the cloud. Additional boxes for each instance class and instance count provide an overview of the result set for each instance class.
8. **Data Table:** The result set of all matched VMs are listed in the data table. This table provides a paginated view of the matching result set along with instance, cost, configuration and performance data. Filters and matching have a direct impact on this result set. This is the data that is exported when Export to CSV is selected. Each column is sortable by the column header in either ascending or descending order.

Using the Simulator Card Sections

Each box on the GCP Cloud Simulator offers a series of related fields that impact matching, rightsizing, or pricing and discounts. These sections have very unique impacts on the simulator matching and pricing logic.

VM Matching Section

- GCP Location - Google offers numerous locations for hosting instances. Not all locations have all services available and will impact matching and cost. All VM's in the simulator will be applied to this geography. You can filter your results to a specific datacenter or VPC for individual location pricing.
GCP Sole Tenant Node Locations: <https://cloud.google.com/compute/docs/regions-zones/#available>
- GCP Storage Type - GCP Offers Local, Zonal, and Regional storage at different price levels with various recovery and performance options.
GCP Storage Types: <https://cloud.google.com/compute/docs/disks/>
- Storage Media - Workloads may not require SSD's for every instance. HDD, SSD allows you to specify the specific disk type. Optimal will match disk type based on capacity, IOPS, and throughputs observed. Optimal provides per instance recommendation for disk type. Optimal will look at the VM size and disk size to determine the allowed IOPS and determine if the HDD will meet the IOPS required or if the workloads should be placed on SSD. If a VM exceeds the IOPS of an HDD, it will automatically be placed on SSD. Note: For Compute Engine, disk size, machine type memory, and network usage are calculated in gigabytes (GB), where 1 GB is 2³⁰ bytes. This unit of measurement is also known as a [gibibyte \(GiB\)](#).
Google Storage Pricing and Configuration Options: <https://cloud.google.com/compute/disks-image-pricing>
- Match Standard Instances Only - Exclude custom instance types from recommendations. This will allow for easy selection of Google based templates without the need for customization and may accelerate cloud migration. Note, Some Instance Restrictions apply to the GCP Instance Matching with Sole Tenant Nodes. These instance class restrictions can be found here: <https://cloud.google.com/compute/docs/nodes/#restrictions>
- Sole Tenant Nodes - Gain additional savings with dedicated compute nodes. This option will introduce a combination of instances and sole tenant nodes to maximize cost savings. Instances on Sole Tenant nodes incur no instance compute cost but require storage and OS (unless BYOL option included). Sole Tenant nodes are not available in all regions. At this time, Sole Tenant nodes have the greatest impact on Microsoft Windows Licenses and as such, Linux nodes are not modeled on to Sole Tenant Nodes. Details on Google BYOL Options can be found here: <https://cloud.google.com/compute/docs/instances/windows/bring-your-own-license/>
- Maximum Oversubscription - Sole Tenant nodes can have a maximum vCPU prescription of 4:1. This option can set a maximum allowable ratio. CloudPhysics will restrict this as the maximum but will never oversubscribe the node GHz available. Optimal will use the ideal ratio.

- All Desktop OS/MSDN - By default, Microsoft Windows desktops will be excluded unless you bring your own Microsoft licenses or MSDN licenses .

https://cloud.google.com/compute/docs/instances/windows/#windows_client

- BYOL Microsoft Licenses - Reduce instance costs by bringing your own Microsoft Software Assurance licenses.

Some restrictions apply to license considerations:

<https://cloud.google.com/compute/docs/instances/windows/bring-your-own-license/>

Google License Mobility FAQ:

<https://cloud.google.com/compute/docs/instances/windows/bring-your-own-license/frequently-asked-questions>

- Exclude Live Migration - Microsoft environments require a dedicated Live Migration Node to ensure availability when upgrading nodes. If you do not require high availability, you may choose to exclude the Live Migration Node to reduce nodes. One node is required for every 20 hosts.

Google Live Migration notes can be found here:

<https://cloud.google.com/compute/docs/instances/live-migration>

VM Rightsizing Section

- Rightsizing Options – Workloads can be rightsized to reflect sizes that better match the workloads performance history. Very large idle workloads may be dramatically reduced.
- Additional Headroom – Users who are concerned with potential performance capacity constraints can request additional headroom. This headroom applies to the current filtered performance level. An instance with 4 vCPU and 50% more headroom would be resized to 6 vCPUs. Options to adhere to reservations and maximums prevent VM's from being sized larger than their current configuration or reduced below reservation levels. This option can help keep VM's with per vCPU licenses from exceeding their current license model.

VM Pricing and Discounts Section

- Term Length - Users can choose a term length of On-Demand, 1-Year, or 3-Year. If the user selects On-Demand, they can adjust the instance class from Regular to Pre-emptive. On-Demand instances qualify for Sustained Usage discounts. This discount will be calculated based on the hours per day specified in the Pre-Emptive option or for 24 hours per day when selecting Regular instance class. Users who choose 1-year or 3-year commitments will receive Committed Use discounts applied to the simulator results. 1-year and 3-year automatically convert to Regular instance classes.

Google Sole Tenant Node Sustained Use Discounts:

https://cloud.google.com/compute/vm-instance-pricing#sustained_use

Google Sole Tenant Node Committed Use Discounts:

https://cloud.google.com/compute/vm-instance-pricing#committed_use

- Instance Classes – allows the user to specify Regular or Pre-Emptive instance classes. Pre-emptive classes can be defined with an Hours per Day that is applied to sustained usage discounts. Regular Instances with On-Demand use a sustained usage discount while Regular instances with annual commitments qualify for committed use discounts.

Google Instance Classes and Machine Types:

<https://cloud.google.com/compute/docs/machine-types>

Pre-Emptive Instances

<https://cloud.google.com/compute/docs/instances/preemptible>

- Uptime per Day – This number allows user to define usage levels of dynamic workloads. This value is used for sustained usage discounts.
- BYOL Linux – This option can be selected if the customer intends to bring their own Linux licenses to the cloud. Checking this option will eliminate all Linux OS costs from guest OS on both Sole Tenant nodes and general fleet instances. Google License Mobility FAQ: <https://cloud.google.com/compute/docs/instances/windows/bring-your-own-license/frequently-asked-questions>

Annual Cost Summary Section

- Annual Costs – A quick estimate of the Google GCE cost for Instances and Sole Tenant Nodes broke by resource type. Hover your mouse pointer over a colored bar for additional cost details by resource.
- Environment Size – This will be the total number of Sole tenant nodes or Instances in the environment.
- vCPU Constraints – A visual representation of the oversubscription of the Sole Tenant Nodes in GHz
- Memory Constraints – A visual representation of the memory consumption in the sole tenant nodes. Memory is never oversubscribed in the packing models used by CloudPhysics.
- Compare Clouds – Once you have modeled your result set, take it to the next level and compare your result with other clouds.

Frequently Asked Questions

What is a General Fleet Instance?

According to Google, “Compute Engine instances can run the public images for Linux and Windows Server that Google provides as well as private custom images that you can create or import from your existing systems. You can also deploy Docker containers, which are automatically launched on instances running the Container-Optimized OS public image.” These are considered General Fleet Instances if they are not currently running on a Sole Tenant Node.

Google Cloud Instances: <https://cloud.google.com/compute/docs/instances/>

What is a Sole Tenant Node?

According to Google, “A sole-tenant node is a physical Compute Engine server that is dedicated to hosting VM instances only for your specific project. Use sole-tenant nodes to keep your instances physically separated from instances in other projects, or to group your instances together on the same host hardware.” This option allows for multiple instances to be placed on a dedicated host ensuring isolation from other companies as well as guaranteeing placement of workloads on the same host for performance or compliance reasons. You can specify node affinities between your nodes and the instances that you want to run on those nodes. Use these affinities to either group several workloads together on the same nodes or isolate your workloads from one another on different nodes to meet your data compliance requirements. To use sole-tenant nodes, you create node groups, which contain one or more nodes, rather than creating individual nodes. The following illustration shows one node that has instances from multiple customer projects and another node that has instances exclusively from one customer project.

Google Sole Tenant Node Overview: <https://cloud.google.com/compute/docs/nodes/>

How large is a Sole Tenant Node?

Currently, the only available node type is *n1-node-96-624* node type which has 96 vCPUs and 624 GB of memory, available in multiple zones. Nodes of this size can accommodate VM instances up to 96 vCPUs and 624 GB of memory, but you can also fill the node with multiple smaller VM instances with various sizes including custom machine types and instances with extended memory.

Node Types: <https://cloud.google.com/compute/docs/nodes/#types>

How many instances can I run on a Sole Tenant Node?

A sole tenant node can host as many instances that utilize the resources of the server with a maximum vCPU oversubscription of 4:1. This oversubscription applies to defined vCPU's, not the GHz available in the server. The GHz available is fixed, but underutilized instances can release their unused vCPU commitment to other workloads to increase the density of the host. Since the minimum workload size is 2 vCPU, the maximum logical instances available would be 192 or however many instances consume the total GHz or GB RAM available to the server. It is not advised to oversubscribe GHz or RAM on a server as this results in CPU contention, memory swapping, and potentially unresponsive instances. The Optimal ratio will be the total number of instances that use the GHz and RAM optimally across the day without oversubscribing the 4:1 ratio.

Which is better, General Fleet Instances or Sole Tenant Nodes + Instances? When should I choose one over another?

While all workloads are still run within an instance, instances on Sole Tenant nodes can be oversubscribed to reduce costs and increase the resource usage per host. A maximum vCPU oversubscription of 4:1 is permitted. This allows resources to be used more optimally. General Fleet Instances do not benefit from overcommitment as the user pays for the full instance size regardless of location. Since Sole Tenant Nodes are paid for at a hardware level, instances on the node incur no cost for computing or memory but must adhere to limits on size defined by the Sole Tenant requirements.

General Fleet Instances cannot utilize BYOL Microsoft license model as Microsoft imposes limits on the migration of workloads. When using Sole Tenant Nodes, users can maximize their Microsoft license model. In addition, BYOL Microsoft instances can leverage a Live Migration node to cluster management to ensure high availability of workloads.

If you require application affinity, host-based licenses, or dedicated hosts for compliance reasons, Sole Tenant may be a requirement. Sole Tenant nodes also gains the benefit of dramatic cost savings due to oversubscription and BYOL Microsoft options.

General Fleet Instances have the advantage of more configuration options as well as mobility. General Fleet instances can support small shared core instances as well as very large memory and vCPU configurations. These instances can also support solutions with GPU's,

What happens if my Sole Tenant Node needs maintenance?

If your node's host system requires maintenance, the node and all of the instances on the node continue to operate while they live to migrate to updated host hardware. Live Migration Nodes are required if you require continuous availability of instances during a maintenance window. One Live Migration Node per twenty hosts will be required for Live Migration.

Google Live Migration: <https://cloud.google.com/compute/docs/instances/live-migration>

Do I need a Live Migration Node?

Live Migration Nodes are used for high availability during physical server maintenance events. In the event a server needs to be serviced, all instances can be moved to a Live Migration Node to ensure availability. This is a requirement for Microsoft Windows Licenses. To keep compliance. If high availability is not a requirement and you are willing to take a service outage during a maintenance window, you can eliminate the Live Migration node from your estimate.

How many hosts does a Live Migration Node support?

You can have up to 20 hosts per Live Migration Node. A Live Migration node is not required if you do not require a high availability of workloads during server maintenance.

How does rightsizing impact cost?

VM Rightsizing will reduce the size of an instance based on actual utilization observed. If a VM or instance is defined for a large vCPU and vRAM size, but the actual utilization is significantly less, we will map the VM or Instance to a smaller sized instance that better represents the workload resource requirements observed. This will reduce the spend on instance sizes that are significantly greater than required.

What is the difference between As Configured, Peak, 99th percentiles and 95th Percentiles?

When we observe a workload over a period of time, we capture the actual performance characteristics of vCPU, vRAM, and storage performance. The defined size of a VM or Instance would be the “As Configured” size of the resource. A VM defined for 32 vCPU and 64GB RAM would be the “As configured” size.

When we collect usage data over time, we can quickly identify at what levels the resources are being consumed. These granular performance points represent the performance and usage characteristics of a workload over the observed time. These performance data points create a histogram of data over a day, week, and month.

The Peak would be the highest level of utilization observed for any resources. This may be represented as a percentage of total available or a unit of performance (GHz, GB, IOPS, etc.) For VMware Environments, CloudPhysics collects data from vCenter at 20-second granularity. The Peak would be the peak observed resource consumption during that 20 second period. Over the course of a day or week, the Peak would be the largest observed number for that resource during the extended period.

When looking at data over extended periods of time with multiple data points, we can derive a 99th Percentile and 95th percentiles. If all data points were to be sorted from lowest to highest and placed into 100 equal units, the 99th percentile would be in the 99th position representing all workloads at this level and below. This point would be higher than 99% of the other workloads observed in the time frame with 1% of the workloads observed being greater than this point. This is meaningful in that while a VM may experience a peak of 100% CPU usage, it may run at 25% usage 99% of the time of less meaning that significant savings could be gained by rightsizing to the 99th Percentile vs keeping the workloads sized at As Configured or Peak.

How does Rightsizing impact Sole Tenant Nodes

CloudPhysics only rightsizes to Peak for Sole Tenant Nodes. This is because in Sole Tenant nodes, resources are shared, and any extra capacity defined in the instance size is unlikely to impact the host performance and has no impact on costs on Sole Tenant Nodes. Keeping instances “As Configured” will consume the same vCPU performance but introduce contention in the CPU Scheduler for very large workloads. Rightsizing to Peak offers equal performance and reduces scheduler contention.

Does CloudPhysics Rightsize vDisks?

CloudPhysics does not rightsize storage as this resource grows with an environment. Many CloudPhysics tools can address storage reclamation of thin-provisioned disks and dead space, snapshots, and unused volumes but vDisks are never rightsized on migration simulators. CloudPhysics will always use the defined disk capacity size for all attached vDisks, not the thin provisioned disk volume size.

Why is there no price associated with VM's on a Sole Tenant node in the CloudPhysics Data Table?

Sole Tenant nodes are priced based on the physical host of the resource and not at a per instance level. Sole Tenant nodes allow for an oversubscription of a maximum of 4:1 vCPU to physical CPU. CloudPhysics will oversubscribe vCPU but never oversubscribes memory on a host node as this would lead to disk swapping and performance degradation.

Does Sustained Use or Committed Use discounts apply to Sole Tenant Nodes?

Yes. Sustained use discounts and committed use discounts reduce the costs of your sole-tenant nodes.

Google Sole Tenant Node Sustained Use Discounts:

https://cloud.google.com/compute/vm-instance-pricing#sustained_use

Google Sole Tenant Node Committed Use Discounts:

https://cloud.google.com/compute/vm-instance-pricing#committed_use

How do Sole Tenant Nodes work with VPC networks?

VPC networks work with instances running on sole-tenant nodes the same way that they work with normal VM instances. You can use VPC networks to establish network connections between sole-tenant instances and normal VM instances.

Google VPC: <https://cloud.google.com/vpc/docs/vpc>

What types of Instances are supported on Sole Tenant Nodes?

Predefined instances and custom instances are supported on Sole Tenant Nodes. Use custom machine types or predefined machine types to create instances on your sole-tenant nodes. Because you already pay for the vCPUs and memory of the node itself, you do not pay extra for these instances. All instances must have a minimum of 2 vCPU. Instances cannot exceed the total RAM available in the sole tenant node.

What are the restrictions on Sole Tenant Nodes?

You cannot start VMs with machine types that have fewer than two vCPUs. This includes:

- Shared-core machine types:
 - f1-micro
 - g1-small
- The n1-standard-1 machine type
- Custom machine types with only 1 vCPU

Sole-tenant nodes are only available in select zones. Review the Regions and Zones page (<https://cloud.google.com/compute/docs/regions-zones/#available>) at Google to learn which zones currently support sole-tenant nodes.

The following features are unavailable on sole-tenant nodes at this time:

- GPUs
- Local SSDs
- Memory-optimized machine types

Instance restrictions: <https://cloud.google.com/compute/docs/nodes/#restrictions>

How does a BYOL Microsoft option impact my costs?

If a customer has Software Assurance from Microsoft, they may qualify for a license transfer to the cloud. This license transfer would allow the customer to run their current Microsoft instances in the cloud on their own Microsoft Windows Server Licenses without paying Google for access to a Microsoft Windows Server licenses on each instance in a Sole Tenant node. Since Microsoft restricts the mobility of a license across hosts, BYOL is ideal for Sole Tenant Nodes where workloads will be associated with a dedicated host. The additional allowance of Live Migration Nodes ensures availability during

maintenance events. BYOL Microsoft option cannot be applied to general fleet instances or any instance not on a sole tenant node.

Google License Mobility FAQ:

<https://cloud.google.com/compute/docs/instances/windows/bring-your-own-license/frequently-asked-questions>

Google Compute and Bring Your Own License overview:

<https://cloud.google.com/compute/docs/instances/windows/bring-your-own-license/>

How does a BYOL LINUX option impact my costs?

This option would allow you to bring your own SUSE, RedHat, or other Linux variant license to the cloud to run within your cloud environment without paying Google for a Linux license. Linux licenses do not incur the same restrictions for live migration nodes or hosts as Microsoft and as a result, can be applied to all instances in the Google Cloud. This license can apply to both general fleet instances and sole tenant node instances.

Google License Mobility FAQ:

<https://cloud.google.com/compute/docs/instances/windows/bring-your-own-license/frequently-asked-questions>