



Pexip Infinity and Google Cloud Platform

Deployment Guide

Software Version 38

Document Version 38.a

July 2025

]pexip[

Contents

Introduction	4
Deployment guidelines	5
Deployment options	5
Recommended instance types and call capacity guidelines	6
Security and SSH keys	6
Google Cloud VPN for private/hybrid cloud deployments	6
IP addressing	7
Assumptions and prerequisites	7
Configuring your Google VPC network	8
Google Cloud VPN for private / hybrid cloud deployments	8
Enabling communication between Pexip Infinity nodes	8
Inter-node communication requirements for multiple VPCs	9
Controlling access to the Management Node	9
Controlling access to Conferencing Nodes for installation/provisioning	10
Controlling access to Conferencing Nodes for conference participants	11
Obtaining and preparing disk images for GCE Virtual Machines	13
Obtaining the Pexip disk images	13
Uploading disk images to Google Cloud Storage	13
Preparing custom disk images	13
Prepare a Management Node image	13
Prepare a Conferencing Node image	14
Deploying a Management Node in Google Cloud Platform	16
Initial platform configuration — GCP	20
Accessing the Pexip Infinity Administrator interface	20
Configuring the Pexip Infinity platform	20
Next step	21
Deploying a Conferencing Node in Google Cloud Platform	22
Deploying the VM instance in GCP	22
Generating, downloading and deploying the configuration file	25
Configuring dynamic bursting to the Google Cloud Platform (GCP)	28
Configuring your system for dynamic bursting to GCP	28
Firewall addresses/ports required for access to the GCP APIs for cloud bursting	28
Setting up your bursting nodes in GCP and enabling bursting in Pexip Infinity	28
Configuring a GCP role, permissions and service account for controlling overflow nodes	29
Configuring the bursting threshold	30

Manually starting an overflow node	31
Converting between overflow and "always on" GCP Conferencing Nodes	31
Managing Google Compute Engine VM instances	32
Temporarily removing (stopping) a Conferencing Node instance	32
Reinstating (restarting) a stopped Conferencing Node instance	32
Permanently removing a Conferencing Node instance	32

Introduction

The Google Compute Engine (GCE) service provides scalable computing capacity in the Google Cloud Platform (GCP). Using GCP eliminates your need to invest in hardware up front, so you can deploy Pexip Infinity even faster.

You can use GCP to launch as many or as few virtual servers as you need, and use those virtual servers to host a Pexip Infinity Management Node and as many Conferencing Nodes as required for your Pexip Infinity platform.

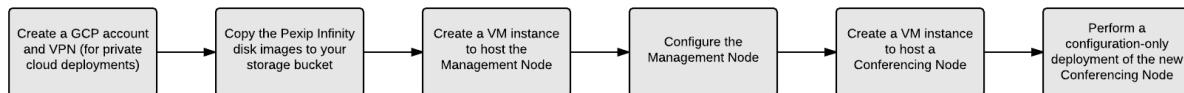
GCP enables you to scale up or down to handle changes in requirements or spikes in conferencing requirements. You can also use the GCP APIs and the Pexip Infinity management API to monitor usage and bring up / tear down Conferencing Nodes as required to meet conferencing demand, or allow Pexip Infinity to handle this automatically for you via its dynamic bursting capabilities.

Pexip publishes disk images for the Pexip Infinity Management Node and Conferencing Nodes. These images may be used to launch instances of each node type as required.

Deployment guidelines

This section summarizes the GCP deployment options and limitations, and provides guidance on our recommended GCP instance types, security groups and IP addressing options.

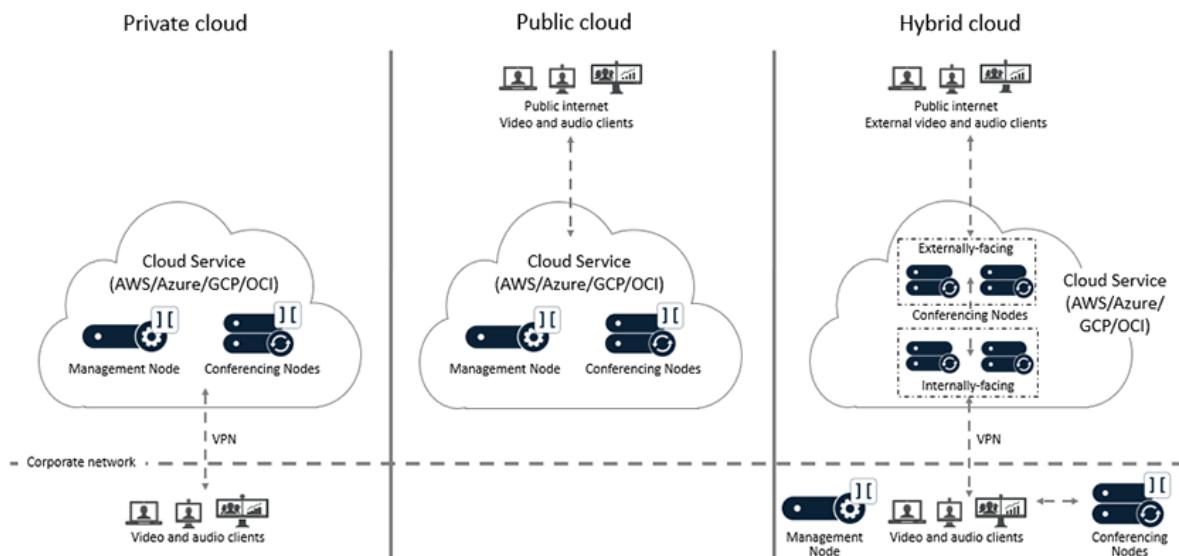
This flowchart provides an overview of the basic steps involved in deploying the Pexip Infinity platform on Azure:



Deployment options

There are three main deployment options for your Pexip Infinity platform when using the Google Cloud Platform:

- **Private cloud:** all nodes are deployed within Google Cloud Platform. Private addressing is used for all nodes and connectivity is achieved by configuring a VPN tunnel between the corporate network and GCP. As all nodes are private, this is equivalent to an on-premises deployment which is only available to users internal to the organization.
- **Public cloud:** all nodes have a private address but, in addition, public IP addresses are allocated to each node. The node's private addresses are only used for inter-node communications. Each node's public address is then configured on the relevant node as a static NAT address. Access to the nodes is permitted from the public internet, or a restricted subset of networks, as required. Any systems or endpoints that will send signaling and media traffic to those Pexip Infinity nodes must send that traffic to the public address of those nodes. If you have internal systems or endpoints communicating with those nodes, you must ensure that your local network allows such routing.
- **Hybrid cloud:** the Management Node, and optionally some Conferencing Nodes, are deployed in the corporate network. A VPN tunnel is created between the corporate network and GCP. Additional Conferencing Nodes are deployed in GCP and are managed from the on-premises Management Node. The GCP-hosted Conferencing Nodes can be either internally-facing, privately-addressed (private cloud) nodes; or externally-facing, publicly-addressed (public cloud) nodes; or a combination of private and public nodes (where the private nodes are in a different Pexip Infinity system location to the public nodes). You may also want to consider dynamic bursting, where the GCP-hosted Conferencing Nodes are only started up and used when you have reached capacity on your on-premises nodes.



All of the Pexip nodes that you deploy in the cloud are completely dedicated to running the Pexip Infinity platform— you maintain full data ownership and control of those nodes.

Recommended instance types and call capacity guidelines

GCP instances come in many different sizes. In general, Pexip Infinity Conferencing Nodes should be considered compute intensive and Management Nodes reflect a more general-purpose workload. Our [Server design recommendations](#) also apply to cloud-based deployments.

If available in your region, we recommend Ice Lake as the minimum CPU platform.

For deployments of up to 20 Conferencing Nodes, we recommend using:

- **Management Node:** a machine type with 4 vCPUs (or larger), for example:
 - n2-highcpu-4
 - n2-highcpu-8.
- **Transcoding Conferencing Nodes:** a machine type with 8 vCPUs and 8 GB memory (or larger), for example:
 - n2-highcpu-8
 - n2-highcpu-16
 - n2-highcpu-32 (should only be used if Ice Lake CPU platform is available).
- **Proxying Edge Nodes:** a machine type with 4 vCPUs (or larger), for example:
 - n2-highcpu-4
 - n2-highcpu-8.

This should provide capacity for approximately 16 HD / 35 SD / 300 audio-only calls per Transcoding Conferencing Node.

We recommend that you do **not** use AMD instance types.

For all available machine types see: https://cloud.google.com/compute/pricing#predefined_machine_types.

Security and SSH keys

An SSH key must be applied to the VM instance that will host the Management Node (in order to complete the installation) and we also recommend applying SSH keys to your VM instances that will host your Conferencing Nodes. Keys can be applied project wide or for a particular VM instance.

- i** The username element of the SSH key must be "admin" or "admin@<domain>" i.e. the key takes the format:

```
ssh-rsa [KEY_VALUE] admin or  
ssh-rsa [KEY_VALUE] admin@vc.example.com for example.
```

You can create key pairs with third-party tools such as PuTTYgen, or you can use an existing SSH key pair but you will need to format the public key to work in Compute Engine metadata (and ensure the username is modified to "admin"). You can also use other key types than rsa, such as ed25519. For more information about using and formatting SSH keys for GCP, see <https://cloud.google.com/compute/docs/instances/access-overview> and <https://cloud.google.com/source-repositories/docs/authentication#ssh>.

Note that:

- When the Management Node has been deployed, you can assign and use your own SSH keys for the Management Node and any Conferencing Nodes.
- If you are using a Linux or Mac SSH client to access your instance you must use the **chmod** command to make sure that your private key file on your local client (SSH private keys are never uploaded) is not publicly viewable. For example, if the name of your private key file is my-key-pair.pem, use the following command: `chmod 400 /path/my-key-pair.pem`

Google Cloud VPN for private/hybrid cloud deployments

For a private or hybrid cloud deployment, you must configure the Google Cloud virtual private network (VPN) to connect your on-premises network to the Google Cloud Platform.

For full information about how to configure the Google Cloud VPN, see <https://cloud.google.com/compute/docs/vpn/overview>.

IP addressing

All GCE VM instances are allocated a **Primary internal IP** (i.e. private) address. You can optionally also assign a static **External IP** (i.e. public) address to a GCE VM instance. You should assign a public address to all nodes in a public cloud deployment, and to any externally-facing nodes in a hybrid deployment, that you want to be accessible to conference participants located in the public internet.

Pexip Infinity nodes must always be configured with the private IP address associated with its instance, as it is used for all internal communication between nodes. To associate an instance's public IP address with the node, configure that public IP address as the node's **Static NAT address** (via **Platform > Conferencing Nodes**).

The private IP address should be used as the Conferencing Node address by users and systems connecting to conferences from the corporate network (via the Google Cloud VPN) in a private or hybrid cloud deployment. When an instance has been assigned an external IP address and that address is configured on the Conferencing Node as its **Static Nat address**, all conference participants **must** use that external address to access conferencing services on that node.

Assumptions and prerequisites

These deployment instructions assume that within GCP you have already:

- signed up to the Google Cloud Platform
- configured a Google Cloud VPN (for a private or hybrid cloud deployment)

For more information on setting up your Google Cloud Platform Virtual Machines, see

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

Configuring your Google VPC network

All Google Compute Engine (GCE) VM instances belong to a Google Virtual Private Cloud (VPC) network. You need to configure the VPC network to control access to the VM instances that will host your Pexip Infinity nodes in your Google Cloud Platform (GCP) deployment.

Google Cloud VPN for private / hybrid cloud deployments

For a private or hybrid cloud deployment, you must configure the Google Cloud virtual private network (VPN) to connect your on-premises network to the Google VPC network.

- ⓘ Google assigns a default range of private addresses to your VPC regions. You must ensure that the IP address ranges for the VPC regions in which you deploy your VM instances do not overlap with any subnets you use in your corporate network. If you do have overlapping subnets, you can create new subnets for each region in your Google VPC network, and then select that subnetwork when deploying your instance. See <https://cloud.google.com/compute/docs/vpc/#subnet-ranges> for information about the default VPC subnets per region.

For full information about how to configure the Google Cloud VPN, see <https://cloud.google.com/compute/docs/vpn/overview>.

A VPN is not required for public cloud deployments as you can access all of your nodes via their public IP addresses.

Enabling communication between Pexip Infinity nodes

To allow Pexip Infinity nodes to communicate, there must be a firewall rule in place to allow UDP and IPsec ESP protocol traffic between nodes. This applies to all deployment options (private, public and hybrid).

By default, the Google VPC network has a firewall rule called "default-allow-internal". This rule allows TCP, UDP and ICMP traffic between private addresses on the internal network, but it does not allow ESP traffic.

To modify this firewall rule to also allow ESP traffic:

1. From the GCP project console, go to **VPC network > Firewall rules**.
2. Select the **default-allow-internal** rule.
3. Select **Edit**.
4. Change **Other protocols and ports** from "icmp" to "icmp,esp".

Protocols and ports ?

Allow all

Specified protocols and ports

TCP

Ports

0-65535

E.g. 20, 50-60

UDP

Ports

0-65535

E.g. all

Other

Protocols *

icmp,esp

Separate multiple protocols by commas, e.g. ah, sctp

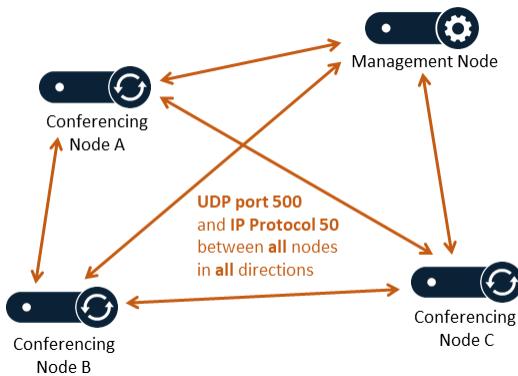
5. Select Save.

Note that this change adds ESP to the existing rule but does not remove or restrict any of the other default protocols and ports. This is because the default-allow-internal rule applies to all instances in your GCP project, and if you have something other than Pexip Infinity running (e.g. a reverse proxy, or something completely unrelated) then you probably want to allow UDP and TCP traffic to work.

Inter-node communication requirements for multiple VPCs

In a basic deployment, your Pexip Infinity platform will be deployed within a single VPC.

In larger deployments you may choose to deploy your Conferencing Nodes across multiple VPCs — in which case there must be a directly routable path (no NAT) between all nodes that allows **UDP port 500 (IKE)**, and **IP Protocol 50 (IPsec ESP)** to pass between all nodes in both directions.



Controlling access to the Management Node

We recommend that you lock down access to the Management Node to just the management stations that will administer your Pexip Infinity platform. This applies to all deployment options (private, public and hybrid), but is particularly important in public cloud

deployments.

To create a new firewall rule to restrict access to the Management Node:

1. From the GCP project console, go to **VPC network > Firewall rules**.
2. Select **Create firewall rule**.
3. Complete the following fields (leave all other settings as default):

Name	Enter a name for the rule, for example "pexip-allow-management".
Direction of traffic	Select <i>Ingress</i> .
Action on match	Select <i>Allow</i> .
Targets	Select <i>Specified target tags</i> .
Target tags	Enter a tag name, for example "pexip-management". You will use this name later when you create your Management Node VM instance to associate that instance with these firewall rules (see Deploying a Management Node in Google Cloud Platform).
Source filter	Select <i>IP ranges</i> .
Source IP ranges	Enter the <IP address/subnet> of the management station/browsers that require access to the Management Node. Note that on a corporate network accessing a public cloud deployment, this should be the external public IP address of the corporate network and not the private address of the machine that is hosting the browser.
Protocols and ports	Enter <code>tcp:443</code> Note that you may need to include <code>tcp:22</code> to allow SSH access if you intend to restrict or remove the <code>default-allow-ssh</code> rule.

4. Select **Create**.

Controlling access to Conferencing Nodes for installation/provisioning

We recommend that you lock down access to the provisioning interface on your Conferencing Nodes to just the management stations that will administer your Pexip Infinity platform. This applies to all deployment options (private, public and hybrid), but is particularly important in public and hybrid cloud deployments for nodes with an external IP address.

To create a new firewall rule to restrict access to the provisioning interface of a Conferencing Node:

1. From the GCP project console, go to **VPC network > Firewall rules**.
2. Select **Create firewall rule**.

3. Complete the following fields (leave all other settings as default):

Name	Enter a name for the rule, for example "pexip-allow-provisioning".
Direction of traffic	Select <i>Ingress</i> .
Action on match	Select <i>Allow</i> .
Targets	Select <i>Specified target tags</i> .
Target tags	Enter a tag name, for example "pexip-provisioning". You will use this name later when you create your Conferencing Node VM instances to associate those instances with these firewall rules (see Deploying a Conferencing Node in Google Cloud Platform).
Source filter	Select <i>IP ranges</i> .
Source IP ranges	Enter the <IP address/subnet> of the management station/browsers that require access to the Conferencing Nodes. Note that on a corporate network accessing a public cloud deployment, this should be the external public IP address of the corporate network and not the private address of the machine that is hosting the browser.
Protocols and ports	Enter <code>tcp:8443</code>

4. Select **Create**.

Controlling access to Conferencing Nodes for conference participants

A wider, more general access is typically required to the protocols and ports required to access conferences hosted on your Conferencing Nodes.

To create a new firewall rule to allow access to the conferencing-related ports and protocols of a Conferencing Node:

1. From the GCP project console, go to **VPC network > Firewall rules**.
2. Select **Create firewall rule**.
3. Complete the following fields (leave all other settings as default):

Name	Enter a name for the rule, for example "pexip-allow-conferencing".
Direction of traffic	Select <i>Ingress</i> .
Action on match	Select <i>Allow</i> .
Targets	Select <i>Specified target tags</i> .
Target tags	Enter a tag name, for example "pexip-conferencing". You will use this name later when you create your Conferencing Node VM instances to associate those instances with these firewall rules (see Deploying a Conferencing Node in Google Cloud Platform).
Source filter	Select <i>IP ranges</i> .
Source IP ranges	Enter <code>0.0.0.0/0</code> For a private deployment, the Source IP ranges should be restricted to the corporate intranet IP addresses.
Protocols and ports	Enter <code>tcp:80; tcp:443; tcp:1720; tcp:5060; tcp:5061; tcp:33000-39999; tcp:40000-49999; udp:1719; udp:33000-39999; udp:40000-49999</code> Note that if you have enabled SIP UDP then <code>udp:5060</code> must also be included.

4. Select **Create**.

After you have configured your firewall rules, your ingress rules will look similar to this:

[Ingress](#) [Egress](#)

<input type="checkbox"/>	Name	Targets	Source filters	Protocols / ports	Action	Priority	Network
<input type="checkbox"/>	default-allow-http	http-server	IP ranges: 0.0.0.0/0	tcp:80	Allow	1000	default
<input type="checkbox"/>	default-allow-https	https-server	IP ranges: 0.0.0.0/0	tcp:443	Allow	1000	default
<input type="checkbox"/>	pexip-allow-conferencing	pexip-conferencing	IP ranges: 0.0.0.0/0	tcp:80, tcp:443, 8 more ▾	Allow	1000	default
<input type="checkbox"/>	pexip-allow-management	pexip-management	IP ranges: 81.143.209.108/32	tcp:443	Allow	1000	default
<input type="checkbox"/>	pexip-allow-provisioning	pexip-provisioning	IP ranges: 81.143.209.108/32	tcp:8443	Allow	1000	default
<input type="checkbox"/>	default-allow-icmp	Apply to all	IP ranges: 0.0.0.0/0	icmp	Allow	65534	default
<input type="checkbox"/>	default-allow-internal	Apply to all	IP ranges: 10.128.0.0/9	tcp:0-65535, udp:0-65535, 2 more ▾	Allow	65534	default
<input type="checkbox"/>	default-allow-rdp	Apply to all	IP ranges: 0.0.0.0/0	tcp:3389	Allow	65534	default
<input type="checkbox"/>	default-allow-ssh	Apply to all	IP ranges: 0.0.0.0/0	tcp:22	Allow	65534	default

Obtaining and preparing disk images for GCE Virtual Machines

Pexip publishes Google Compute Engine (GCE) optimized disk images for the Management Node and for Conferencing Nodes.

Before you can use the published Pexip Infinity disk images, you must copy them to your storage bucket in the Google Cloud Platform (GCP). This guide refers to a disk image copied to your storage bucket as a **custom disk image**. All deployment operations use custom disk images.

Obtaining the Pexip disk images

To obtain your disk images, go to <https://dl.pexip.com/infinity/index.html>, select the appropriate directory for your software version, and then download the following files:

- **Pexip_Infinity_v38_GCP_pxMgr_<build>.tar.gz** for the Management Node.
- **Pexip_Infinity_v38_GCP_ConfNode_<build>.tar.gz** for a Conferencing Node.

Uploading disk images to Google Cloud Storage

The Pexip disk image packages must be uploaded to Google Cloud Storage.

1. Create a bucket to store the images:
 - a. From the GCP project console, go to **Cloud Storage > Browser**.
 - b. Select **Create Bucket**.
 - c. Enter a Name (for example, "pexip-v38"), and then select an appropriate **Storage class** and **Location** for your deployment.
For more information about storage buckets, see <https://cloud.google.com/storage/docs/creating-buckets>
 - d. Select **Create**.
2. Upload the Pexip images to the new bucket:
 - a. Select the new bucket e.g. pexip-v38.
 - b. Select **Upload Files**.
 - c. In the dialog that appears, select the Management Node and Conferencing Node tar.gz files that you downloaded from Pexip.
 - d. Select **Open**.

After you have uploaded your files, your bucket will look similar to this:

Buckets / pexip-v20								
	Name	Size	Type	Storage class	Last modified	Public access	Encryption	Retention expiry date
	Pexip_Infinity_v20_GCP_ConfNode_45400.tar.gz	517.77 MB	application/x-gzip	Regional	12/02/2019, 02:36:25 UTC	Not public	Google-managed key	-
	Pexip_Infinity_v20_GCP_pxMgr_45400.tar.gz	1.04 GB	application/x-gzip	Regional	12/02/2019, 02:38:43 UTC	Not public	Google-managed key	-

Preparing custom disk images

You must now prepare a custom disk image for the Management Node and for deploying a Conferencing Node.

Prepare a Management Node image

1. From the GCP project console, go to **Compute Engine > Images**.
2. Select **Create Image**.
3. Enter a Name, for example "pexip-mgr-v38".
4. Select a Source of **Cloud Storage file**.
5. Select **Browse** and select the Management Node image package in your storage bucket e.g. pexip-v38.
6. Select **Create**.

You can now deploy the Management Node in Google Cloud Platform.

The screenshot shows the Google Cloud Platform Compute Engine interface. On the left, a sidebar lists various Compute Engine resources: VM instances, Instance groups, Instance templates, Sole tenant nodes, Disks, Snapshots, and Images. The 'Images' option is selected and highlighted in blue. The main area is titled 'Create an image' and contains fields for Name (pexip-mgr-v20), Family (Optional), Description (Optional), and Labels (Optional). Below these, there's an 'Encryption' section with options for Google-managed key (selected), Customer-managed key, and Customer-supplied key. The 'Source' dropdown is set to 'Cloud Storage file'. Under 'Cloud Storage file', it shows a selected file: pexip-v20/Pexip_Infinity_v20_GCP_pxMgr_45400.tar.gz. A note states: 'You will be billed for this image. Compute Engine pricing ↗'. At the bottom are 'Create' and 'Cancel' buttons.

Prepare a Conferencing Node image

1. From the GCP project console, go to Compute Engine > Images.
2. Select Create Image.
3. Enter a Name, for example "pexip-node-v38".
4. Select a Source of *Cloud Storage file*.
5. Select Browse and select the Conferencing Node image package in your storage bucket e.g. pexip-v38.
6. Select Create.

You can now deploy a Conferencing Node in Google Cloud Platform.

Compute Engine [Create an image](#)

VM instances
Instance groups
Instance templates
Sole tenant nodes
Disks
Snapshots
Images
TPUs
Committed use discounts
Metadata
Health checks
Zones
Network endpoint groups
Operations
Security scans
Settings

Name

Family (Optional)

Description (Optional)

Labels (Optional) + Add label

Encryption
Data is encrypted automatically. Select an encryption key management solution.
 Google-managed key
No configuration required
 Customer-managed key
Manage via Google Cloud Key Management Service
 Customer-supplied key
Manage outside Google Cloud

Source
Cloud Storage file Your image source must use the .tar.gz extension and the file inside the archive must be named disk.raw. [Learn more](#)
 pexip-v20/Pexip_Infinity_v20_GCP_ConfNode_45400.tar.gz

You will be billed for this image. [Compute Engine pricing](#)

Create **Cancel**

Deploying a Management Node in Google Cloud Platform

As with all Pexip Infinity deployments, you must first deploy the Management Node before deploying any Conferencing Nodes. In a hybrid cloud deployment the Management Node may be deployed in the corporate network or in GCP. This section describes how to deploy the Management Node in GCP.

To deploy a Management Node on a Google Compute Engine VM:

1. Prepare a Management Node disk image. For more information on this, see [Obtaining and preparing disk images for GCE Virtual Machines](#).
2. From the GCP project console, go to Compute Engine > VM Instances.
3. Select **Create Instance**.
4. Complete the following fields (leave all other settings as default):

Name	Enter a unique name for the instance, for example "pexipmanager".
Zone	Select an appropriate Zone. Typically you should choose a zone that is geographically close to the location from where it will be administered.
Machine type	Select 4 vCPUs (n1-standard-4) .
Boot disk	Select the Management Node custom disk image: <ol style="list-style-type: none">Select Change.Select Custom images.Select your GCP project.Select the Management Node custom disk image, e.g. "pexip-mgr-v38".Select a Boot disk type of SSD persistent disk.Select Select.
Identity and API access	For Service account, select No service account .
Networking:	Expand the Advanced options section and open the Networking section.
Network tags	Assign a Network tag to the instance, for example "pexip-management". This is the tag that should be applied to your Management Node firewall rule (see Controlling access to the Management Node).
Networking:	You must decide whether or not to assign an external IP address to the instance.
External IP	You must assign a static public/external IP address to the Management Node if you have a public cloud deployment. On a private or hybrid deployment you will typically access the Management Node over a VPC network, or host it on-premises. <ol style="list-style-type: none">Expand the Advanced options section and open the Networking section.In the Network interfaces field, select the default interface to open the Network interface dialog.Select a Subnetwork if appropriate (e.g. if it is a private/hybrid deployment and you have created new subnets to avoid overlapping addresses in your corporate network).Select an appropriate External IP:<ul style="list-style-type: none">■ None: no external IP address will be assigned. Use this where the instance does not need to have a publicly-accessible IP address.■ Create IP address: select this option to create a static external address. You can enter a Name for the address and GCP will allocate a static IP address.■ <external address>: you can select a specific static external address if you have already created one in advance.
Do not select Ephemeral — if you stop and restart the instance a new address will be assigned.	

SSH keys An SSH key must be applied to the Management Node instance if you are not already using a project-wide key for all of the instances in your project.

The username element of the SSH key must be "admin" or "admin@<domain>". To apply an instance-level key:

- a. Open the **Security** section and then open the **Manage Access** section.
- b. Select **Add item** to add your own, existing SSH key. This produces a text box. Copy the contents of your public SSH key file and paste them into the text box. Modify the username element at the end of the key to "admin" or "admin@<domain>" if necessary.

See [Security and SSH keys](#) for more information.

The screenshot shows the 'Create an instance' dialog for Google Cloud Platform's Compute Engine. The left sidebar lists options like VM instances, Instance groups, Instance templates, Disks, Snapshots, Images, Committed use discounts, Metadata, Health checks, Zones, Operations, Quotas, and Settings. The main form is titled 'Create an instance' and contains the following fields:

- Name:** pexipmanager
- Zone:** europe-west1-d
- Machine type:** 4 vCPUs, 15 GB memory, Customize (button)
- Boot disk:** New 99 GB standard persistent disk, Image: pexip-mgr-v20, Change button
- Identity and API access:** Service account: No service account, Access scopes: Select a service account to enable API access
- Firewall:** Add tags and firewall rules to allow specific network traffic from the Internet.
 - Allow HTTP traffic
 - Allow HTTPS traffic
- Networking:** Management, Disks, Networking (selected), SSH Keys
- Network tags:** pexip-management
- Network interfaces:** default default (10.132.0.0/20), Add network interface button, Note: To create another network interface you need to have a new network first.
- Less** link
- Your Free Trial credits, if available, will be used for this instance**
- Create** and **Cancel** buttons

5. Select **Create** to create the instance.
6. You must now connect over SSH into the Management Node instance to complete the installation of Pexip Infinity:
 - a. Use an SSH client to access the Management Node by its IP address, supplying your private key file as appropriate.
 - b. At the "Enter new UNIX password:" prompt, enter your desired password, and then when prompted, enter the password again.

This will then log you out and terminate your SSH session.

7. Reconnect over SSH into the Management Node instance and continue the installation process:

- Log in again as *admin*.

You are presented with another login prompt:

```
[sudo] password for admin:
```

- Enter the UNIX password you just created.

The Pexip installation wizard will begin after a short delay.

- Complete the installation wizard to apply basic configuration to the Management Node:

IP address	The IP address must match the internal IP address allocated by GCE (the default should be correct).
------------	---

Network mask	i The Network mask must be 255.255.255.255 .
--------------	--

Gateway	The Gateway must be the internal IP address of the gateway for the GCE region (the default should be correct).
---------	--

Hostname Domain suffix	Enter your required Hostname and Domain suffix for the Management Node.
---------------------------	---

DNS servers	Configure one or more DNS servers. You must override the default values if it is a private deployment.
-------------	--

NTP servers	Configure one or more NTP servers. You must override the default values if it is a private deployment.
-------------	--

Web administration username Password	Set the Web administration username and password.
--	---

Enable incident reporting	Select whether or not to Enable incident reporting.
---------------------------	---

Send deployment and usage statistics to Pexip	Select whether or not to Send deployment and usage statistics to Pexip.
--	---

i The DNS and NTP servers at the default addresses are only accessible if your instance has a public IP address.

The installation wizard will fail if the NTP server address cannot be resolved and reached.

After successfully completing the wizard, the SSH connection will be lost as the Management Node reboots.

8. After a few minutes you will be able to use the Pexip Infinity Administrator interface to access and configure the Management Node (remember to use https to connect to the node if you have only enabled access to `tcp:443` in your firewall rule, as shown in our example "pexip-allow-management" firewall rule). You can configure your Pexip Infinity platform licenses, VMRs, aliases, locations etc. as described in [Initial platform configuration — GCP](#) before you go on to add Conferencing Nodes.

Initial platform configuration — GCP

After you have run the installation wizard, you must perform some preliminary configuration of the Pexip Infinity platform before you can deploy a Conferencing Node.

This section lists the configuration required, and provides a summary of each step with a link to further information.

All configuration should be done using the Pexip Infinity Administrator interface.

- ❶ **No changes** should be made to any Pexip VM via the terminal interface (other than as described when running the initial Pexip installation wizard) unless directed to do so by Pexip support. This includes (but is not limited to) changes to the time zone, changes to IP tables, configuration of Ethernet interfaces, or the installation of any third-party code/applications.

Accessing the Pexip Infinity Administrator interface

The Pexip Infinity Administrator interface is hosted on the Management Node. To access this:

1. Open a web browser and type in the IP address or DNS name that you assigned to the Management Node using the installation wizard (you may need to wait a minute or so after installation is complete before you can access the Administrator interface).
2. Until you have uploaded appropriate TLS certificates to the Management Node, your browser may present you with a warning that the website's security certificate is not trusted. You should proceed, but upload appropriate TLS certificates to the Management Node (and Conferencing Nodes, when they have been created) as soon as possible.
The Pexip Infinity Conferencing Platform login page will appear.
3. Log in using the web administration username and password you set using the installation wizard.

You are now ready to begin configuring the Pexip Infinity platform and deploying Conferencing Nodes.

As a first step, we strongly recommend that you configure at least 2 additional NTP servers or NTP server pools to ensure that log entries from all nodes are properly synchronized.

It may take some time for any configuration changes to take effect across the Conferencing Nodes. In typical deployments, configuration replication is performed approximately once per minute. However, in very large deployments (more than 60 Conferencing Nodes), configuration replication intervals are extended, and it may take longer for configuration changes to be applied to all Conferencing Nodes (the administrator log shows when each node has been updated).

Brief details of how to perform the initial configuration are given below. For complete information on how to configure your Pexip Infinity solution, see the Pexip Infinity technical documentation website at docs.pexip.com.

Configuring the Pexip Infinity platform

This table lists the Pexip Infinity platform configuration steps that are required before you can deploy Conferencing Nodes and make calls.

Configuration step	Purpose
1. Enable DNS (System > DNS Servers)	Pexip Infinity uses DNS to resolve the hostnames of external system components including NTP servers, syslog servers, SNMP servers and web proxies. It is also used for call routing purposes — SIP proxies, gatekeepers, external call control and conferencing systems and so on. The address of at least one DNS server must be added to your system. You will already have configured at least one DNS server when running the install wizard, but you can now change it or add more DNS servers.

Configuration step	Purpose
2. Enable NTP (System > NTP Servers)	Pexip Infinity uses NTP servers to obtain accurate system time. This is necessary to ensure correct operation, including configuration replication and log timestamps. We strongly recommend that you configure at least three distinct NTP servers or NTP server pools on all your host servers and the Management Node itself. This ensures that log entries from all nodes are properly synchronized. You will already have configured at least one NTP server when running the install wizard, but you can now change it or add more NTP servers.
3. Add licenses (Platform > Licenses)	You must install a system license with sufficient concurrent call capacity for your environment before you can place calls to Pexip Infinity services.
4. Add a system location (Platform > Locations)	These are labels that allow you to group together Conferencing Nodes that are in the same datacenter. You must have at least one location configured before you can deploy a Conferencing Node.
5. Upload TLS certificates (Certificates > TLS Certificates)	You must install TLS certificates on the Management Node and — when you deploy them — each Conferencing Node. TLS certificates are used by these systems to verify their identity to clients connecting to them. All nodes are deployed with self-signed certificates, but we strongly recommend they are replaced with ones signed by either an external CA or a trusted internal CA.
6. Add Virtual Meeting Rooms (Services > Virtual Meeting Rooms)	Conferences take place in Virtual Meeting Rooms and Virtual Auditoriums. VMR configuration includes any PINs required to access the conference. You must deploy at least one Conferencing Node before you can call into a conference.
7. Add an alias for the Virtual Meeting Room (done while adding the Virtual Meeting Room)	A Virtual Meeting Room or Virtual Auditorium can have more than one alias. Conference participants can access a Virtual Meeting Room or Virtual Auditorium by dialing any one of its aliases.

Next step

You are now ready to deploy a Conferencing Node — see [Deploying a Conferencing Node in Google Cloud Platform](#) for more information.

Deploying a Conferencing Node in Google Cloud Platform

After deploying the Management Node and completing the initial platform configuration you can deploy one or more Conferencing Nodes in GCP to provide conferencing capacity.

Creating a new Conferencing Node is a two-step process:

1. Deploying a new VM instance in GCP.
2. Configuring the VM with the details of the specific Conferencing Node being deployed, using a file generated from the Pexip Infinity Management Node.

Deploying the VM instance in GCP

To deploy a Conferencing Node on a Google Compute Engine VM:

1. If you have not already done so, prepare a Conferencing Node disk image. For more information on this, see [Obtaining and preparing disk images for GCE Virtual Machines](#).

Note that if you have upgraded your Pexip Infinity software, you need a Conferencing Node disk image for the software version you are currently running.

2. From the GCP project console, go to Compute Engine > VM Instances.
3. Select Create Instance.
4. Complete the following fields (leave all other settings as default):

Name	Enter a unique name for the instance, for example "pexipnode-europe-1".
Zone	Select an appropriate Zone. Typically you should choose a zone that is geographically close to the location from where users will connect to it.
Machine type	Select 8 vCPUs (n2-highcpu-8) . We recommend selecting a minimum CPU platform. Select Customize and then select the most modern platform available that does not incur a surcharge, typically <i>Intel Broadwell or later</i> . For more information see Recommended instance types and call capacity guidelines .
Boot disk	Select the Conferencing Node custom disk image: <ol style="list-style-type: none">Select Change.Select Custom images.Select the Conferencing Node custom disk image, e.g. "pexip-node-v38".Select Select. We strongly recommend SSDs for Conferencing Nodes. General VM processes (such as snapshots and backups) and platform upgrades will be faster with SSDs.
Identity and API access	For Service account, select No service account .
Networking:	Expand the Advanced options section and open the Networking section.
Network tags	Assign Network tags to the instance, for example "pexip-provisioning pexip-conferencing". These are the tags that should be applied to your Conferencing Node firewall rules (see Controlling access to Conferencing Nodes for installation/provisioning and Controlling access to Conferencing Nodes for conference participants).

Networking:	You must decide whether or not to assign an external IP address to the instance.
External IP	You must assign a static public/external IP address to the Conferencing Node if you want that node to be able to host conferences that are accessible from devices in the public internet. <ol style="list-style-type: none">Expand the Advanced options section and open the Networking section.In the Network interfaces field, select the default interface to open the Network interface dialog.Select a Subnetwork if appropriate (e.g. if it is a private/hybrid deployment and you have created new subnets to avoid overlapping addresses in your corporate network).Select an appropriate External IP:<ul style="list-style-type: none">■ None: no external IP address will be assigned. Use this where the instance does not need to have a publicly-accessible IP address.■ Create IP address: select this option to create a static external address. You can enter a Name for the address and GCP will allocate a static IP address.■ <external address>: you can select a specific static external address if you have already created one in advance.
	<p>Do not select Ephemeral — if you stop and restart the instance a new address will be assigned.</p>
SSH keys	We recommend applying an SSH key to the Conferencing Node instance if you are not already using a project-wide key for all of the instances in your project. <p>The username element of the SSH key must be "admin" or "admin@<domain>". To apply an instance-level key:</p> <ol style="list-style-type: none">Open the Security section and then open the Manage Access section.Select Add item to add your own, existing SSH key. This produces a text box. Copy the contents of your public SSH key file and paste them into the text box. Modify the username element at the end of the key to "admin" or "admin@<domain>" if necessary.

See [Security and SSH keys](#) for more information.

Compute Engine ← Create an instance

VM instances

Name: pexipnode-europe-1

Zone: europe-west1-d

Machine type: 8 vCPUs, 7.2 GB memory, Customize

Boot disk: New 49 GB standard persistent disk, Image: pexip-node-v20, Change

Identity and API access:

Service account: No service account

Access scopes: Select a service account to enable API access

Firewall: Add tags and firewall rules to allow specific network traffic from the Internet

Allow HTTP traffic
 Allow HTTPS traffic

Management Disks **Networking** SSH Keys

Network tags (Optional): pexip-provisioning, pexip-conferencing

Network interfaces:

Network interface

Network: default

Subnetwork: default (10.132.0.0/20)

Primary Internal IP: Automatic

Show alias IP ranges

External IP: node1address (35.187.39.104)

IP forwarding: Off

Done Cancel

5. Select **Create** to create the instance.
6. On the **VM Instances** page, make a note of the "Internal IP" address, and the "External IP" address (if appropriate) that have been assigned to the new instance / Conferencing Node.
7. After the instance has booted, perform a configuration-only deployment on the Management Node to inform it of the new Conferencing Node as described below.

Generating, downloading and deploying the configuration file

1. From the Pexip Infinity Administrator interface, go to **Platform > Conferencing Nodes** and select **Add Conferencing Node**.
2. You are now asked to provide the network configuration to be applied to the Conferencing Node, by completing the following fields:

Option	Description
Name	Enter the name to use when referring to this Conferencing Node in the Pexip Infinity Administrator interface.
Description	An optional field where you can provide more information about the Conferencing Node.
Role	This determines the Conferencing Node's role: <ul style="list-style-type: none">○ Proxying Edge Node: a Proxying Edge Node handles all media and signaling connections with an endpoint or external device, but does not host any conferences — instead it forwards the media on to a Transcoding Conferencing Node for processing.○ Transcoding Conferencing Node: a Transcoding Conferencing Node handles all the media processing, protocol interworking, mixing and so on that is required in hosting Pexip Infinity calls and conferences. When combined with Proxying Edge Nodes, a transcoding node typically only processes the media forwarded on to it by those proxying nodes and has no direct connection with endpoints or external devices. However, a transcoding node can still receive and process the signaling and media directly from an endpoint or external device if required.
Hostname Domain	Enter the hostname and domain to assign to this Conferencing Node. Each Conferencing Node and Management Node must have a unique hostname. The Hostname and Domain together make up the Conferencing Node's DNS name or FQDN. We recommend that you assign valid DNS names to all your Conferencing Nodes.
IPv4 address	Enter the IP address to assign to this Conferencing Node when it is created. This should be the GCE Internal IP address of the new VM instance.
Network mask	Enter the IP network mask to assign to this Conferencing Node. For GCP you should always enter 255.255.255.255 Note that IPv4 address and Network mask apply to the eth0 interface.
Gateway IPv4 address	Enter the IP address of the default gateway to assign to this Conferencing Node. This is the default gateway address for the region in which the node is deployed. See https://cloud.google.com/compute/docs/vpc/#subnet-ranges for a table of regions and default gateway addresses. Note that the Gateway IPv4 address is not directly associated with a network interface, except that the address entered here lies in the subnet in which either eth0 or eth1 is configured to use. Thus, if the gateway address lies in the subnet in which eth0 lives, then the gateway will be assigned to eth0, and likewise for eth1.
Secondary interface IPv4 address	Leave this option blank as dual network interfaces are not supported on Conferencing Nodes deployed in public cloud services.

Option	Description
Secondary interface network mask	Leave this option blank as dual network interfaces are not supported on Conferencing Nodes deployed in public cloud services. Note that Secondary interface IPv4 address and Secondary interface network mask apply to the eth1 interface.
System location	Select the physical location of this Conferencing Node. A system location should not contain a mixture of proxying nodes and transcoding nodes. If the system location does not already exist, you can create a new one here by clicking  to the right of the field. This will open up a new window showing the Add System Location page.
Configured FQDN	A unique identity for this Conferencing Node, used in signaling SIP TLS Contact addresses.
TLS certificate	The TLS certificate to use on this node. This must be a certificate that contains the above Configured FQDN. Each certificate is shown in the format <subject name> (<issuer>).
IPv6 address	The IPv6 address for this Conferencing Node. Each Conferencing Node must have a unique IPv6 address.
Gateway IPv6 address	The IPv6 address of the default gateway. If this is left blank, the Conferencing Node listens for IPv6 Router Advertisements to obtain a gateway address.
IPv4 static NAT address	Configure the Conferencing Node's static NAT address, if you have assigned a public/external IP address to the instance. Enter the External IP address allocated by GCE for the VM instance.
Static routes	From the list of Available Static routes, select the routes to assign to the node, and then use the right arrow to move the selected routes into the Chosen Static routes list.
Enable distributed database	This should usually be enabled (checked) for all Conferencing Nodes that are expected to be "always on", and disabled (unchecked) for nodes that are expected to only be powered on some of the time (e.g. nodes that are likely to only be operational during peak times).
Enable SSH	Determines whether this node can be accessed over SSH. <i>Use Global SSH setting:</i> SSH access to this node is determined by the global Enable SSH setting (Platform > Global Settings > Connectivity > Enable SSH). <i>Off:</i> this node cannot be accessed over SSH, regardless of the global Enable SSH setting. <i>On:</i> this node can be accessed over SSH, regardless of the global Enable SSH setting. Default: <i>Use Global SSH setting</i> .
SSH authorized keys	You can optionally assign one or more SSH authorized keys to use for SSH access. From the list of Available SSH authorized keys, select the keys to assign to the node, and then use the right arrow to move the selected keys into the Chosen SSH authorized keys list. Note that in cloud environments, this list does not include any of the SSH keys configured within that cloud service.
Use SSH authorized keys from cloud service	When a node is deployed in a cloud environment, you can continue to use the SSH keys configured within the cloud service where available, in addition to any of your own assigned keys (as configured in the field above). If you disable this option you can only use your own assigned keys. Default: enabled.

3. Select Save.

4. You are now asked to complete the following fields:

Option	Description
Deployment type	Select <i>Generic (configuration-only)</i> .
SSH password	Enter the password to use when logging in to this Conferencing Node's Linux operating system over SSH. The username is always <i>admin</i> . Logging in to the operating system is required when changing passwords or for diagnostic purposes only, and should generally be done under the guidance of your Pexip authorized support representative. In particular, do not change any configuration using SSH — all changes should be made using the Pexip Infinity Administrator interface.

5. Select **Download**.

A message appears at the top of the page: "The Conferencing Node image will download shortly or click on the following link".

After a short while, a file with the name **pexip-<hostname>.<domain>.xml** is generated and downloaded.

Note that the generated file is only available for your current session so you should download it immediately.

6. Browse to <https://<conferencing-node-ip>:8443/> and use the form provided to upload the configuration file to the Conferencing Node VM.

If you cannot access the Conferencing Node, check that you have allowed the appropriate source addresses in your ingress firewall rules for management traffic. In public deployments and where there is no virtual private network, you need to use the public address of the node.

The Conferencing Node will apply the configuration and reboot. After rebooting, it will connect to the Management Node in the usual way.

You can close the browser window used to upload the file.

After deploying a new Conferencing Node, it takes approximately 5 minutes before the node is available for conference hosting and for its status to be updated on the Management Node. Until it becomes available, the Management Node reports the status of the Conferencing Node as having a last contacted and last updated date of "Never". "Connectivity lost between nodes" alarms relating to that node may also appear temporarily.

When the node is up and running you can optionally remove the "pexip-provisioning" Network tag from the instance (or whichever tag you have associated with your provisioning firewall rule as described in [Controlling access to Conferencing Nodes for installation/provisioning](#)) as it is no longer required. Note, do not delete the firewall rule or remove the "pexip-conferencing" tag.

Configuring dynamic bursting to the Google Cloud Platform (GCP)

Pexip Infinity deployments can burst into the Google Cloud Platform (GCP) cloud when primary conferencing capabilities are reaching their capacity limits, thus providing additional temporary Conferencing Node resources.

This provides the ability to dynamically expand conferencing capacity whenever scheduled or unplanned usage requires it. The GCP cloud Conferencing Nodes instances are only started up when required and are automatically stopped again when capacity demand normalizes, ensuring that GCP costs are minimized.

For complete information about dynamic bursting, see [Dynamic bursting to a cloud service](#).

Configuring your system for dynamic bursting to GCP

These instructions assume that you already have a working Pexip Infinity platform, including one or more primary (always on) Conferencing Nodes in one or more system locations. These existing Conferencing Nodes can be deployed using whichever platform or hypervisor you prefer.

Firewall addresses/ports required for access to the GCP APIs for cloud bursting

Access to the GCP APIs for cloud bursting is only required from the Management Node.

The Management Node connects to www.googleapis.com over HTTPS port 443.

Setting up your bursting nodes in GCP and enabling bursting in Pexip Infinity

You must deploy in GCP the Conferencing Nodes that you want to use for dynamic bursting, and then configure the Pexip Infinity location containing those nodes as the overflow destination for the locations that contain your primary (always on) Conferencing Nodes:

1. In Pexip Infinity, configure a new system location for media overflow e.g. "GCP burst", that will contain your bursting Conferencing Nodes.
(Note that system locations are not explicitly configured as "primary" or "overflow" locations. Pexip Infinity automatically detects the purpose of the location according to whether it contains Conferencing Nodes that may be used for dynamic bursting.)
2. In GCP, set up the service account, roles and permissions that the Pexip Infinity Management Node will use to log in to GCP to start and stop the node instances.
See [Configuring a GCP role, permissions and service account for controlling overflow nodes](#) for more information.
3. Deploy in GCP the Conferencing Nodes that you want to use for dynamic bursting. Deploy these nodes in the same manner as you would for "always on" usage (see [Deploying a Conferencing Node in Google Cloud Platform](#)), except:
 - a. Apply to each cloud VM node instance to be used for conference bursting a label with a Key of pexip-cloud and an associated Value set to the Tag value that is shown in the Cloud Bursting section on the Platform > Global Settings page (the Tag value is the hostname of your Management Node).
This tag indicates which VM nodes will be started and shut down dynamically by your Pexip system.
 - b. When adding the Conferencing Node within Pexip Infinity:
 - i. Assign the Conferencing Node to the overflow system location (e.g. "GCP burst").
 - ii. Disable (uncheck) the Enable distributed database setting (this setting should be disabled for any nodes that are not expected to always be available).
 - c. After the Conferencing Node has successfully deployed, manually stop the node instance on GCP.
4. In Pexip Infinity, go to Platform > Global Settings > Cloud Bursting, enable cloud bursting and then configure your bursting threshold, minimum lifetime and other appropriate settings for GCP:

Option	Description
--------	-------------

Enable bursting to the cloud	Select this option to instruct Pexip Infinity to monitor the system locations and start up / shut down overflow Conferencing Nodes hosted in your cloud service when in need of extra capacity.
Bursting threshold	The bursting threshold controls when your dynamic overflow nodes in your cloud service are automatically started up so that they can provide additional conferencing capacity. When the number of additional HD calls that can still be hosted in a location reaches or drops below the threshold, it triggers Pexip Infinity into starting up an overflow node in the overflow location. See Configuring the bursting threshold for more information.
Tag name and Tag value	These read-only fields indicate the tag name (always pexip-cloud) and associated tag value (the hostname of your Management Node) that you must assign to each of your cloud VM node instances that are to be used for dynamic bursting. In some circumstances the Tag value may auto populate as "unknown" instead of the hostname, in which case you should also use "unknown" on your cloud VM node instances.
Minimum lifetime	An overflow cloud bursting node is automatically stopped when it becomes idle (no longer hosting any conferences). However, you can configure the Minimum lifetime for which the bursting node is kept powered on. By default this is set to 50 minutes, which means that a node is never stopped until it has been running for at least 50 minutes. If your service provider charges by the hour, it is more efficient to leave a node running for 50 minutes — even if it is never used — as that capacity can remain on immediate standby for no extra cost. If your service provider charges by the minute you may want to reduce the Minimum lifetime.
Cloud provider	Select GCP .
GCP private key	The PEM-formatted private key for the Google Cloud Platform service account that the Pexip Infinity Management Node will use to log in to GCP to start and stop the node instances. If you have created and then downloaded a JSON private key file from the GCP service account, open the file in a plain text editor and copy everything between the double quotes following "private_key" : including the -----BEGIN PRIVATE KEY----- and -----END PRIVATE KEY----- header and footer, and then paste it into this field. After you have saved the settings, the private key will be stored in an encrypted format and subsequently displayed as asterisks.
GCP service account ID	The Service account ID of the service account you have set up in GCP, for example, bursting@example.com.iam.gserviceaccount.com.
GCP project ID	The ID of the GCP project containing the bursting nodes.

5. Go to **Platform > Locations** and configure the system locations that contain your "always on" Conferencing Nodes (the nodes/locations that initially receive calls) so that they will overflow to your new "GCP burst" location when necessary. When configuring your principal "always on" locations, you should normally set the **Primary overflow location** to point at the bursting location containing your overflow nodes, and the **Secondary overflow location** should normally only point at an always-on location.

ⓘ Nodes in a bursting location are only automatically started up if that location is configured as a **Primary overflow location** of an always-on location that has reached its capacity threshold. This means that if a bursting location is configured as a **Secondary overflow location** of an always-on location, then those nodes can only be used as overflow nodes if they are already up and running (i.e. they have already been triggered into starting up by another location that is using them as its **Primary overflow location**, or you have used some other external process to start them up manually).

We recommend that you do not mix your "always on" Conferencing Nodes and your bursting nodes in the same system location.

Configuring a GCP role, permissions and service account for controlling overflow nodes

Within GCP you must set up a service account to be used by Pexip Infinity to start up and shut down the Conferencing Node overflow instances.

First, you need to create a GCP role and its associated permissions:

- From the GCP project console, go to **IAM & Admin > Roles**.
- Select **+ Create Role**.

3. Configure a Title and ID e.g. "Bursting".
4. Select + Add Permissions.
5. From the list of permissions, select `compute.instances.list`, `compute.instances.start`, `compute.instances.stop` and `compute.zoneOperations.list`.
These permissions allow the service account to stop and start VM instances.
6. Select Create.

Next, you need to create a service account and apply the role to the account:

1. From the GCP project console, go to **IAM & Admin > Service Accounts**.
2. Select + Create Service Account.
3. Configure a Service account name e.g. "bursting" and add a Description.
The Service account ID is automatically created, based on the name you enter, and is what you need to configure in the GCP service account ID field in Pexip Infinity.
4. Select Create.
5. For Service account permissions, select the "Bursting" role you created above.
6. Select Continue.
7. On the next page you can optionally grant users access to the service account.
8. Select + Create Key to create a private key for the account.
9. On the Create key page, select a key type of **JSON**.
10. Select Create.
A JSON file that includes your private key is generated and saved to your local computer.
Select Close to dismiss the message saying that the private key has been saved to your computer.
11. Select Done at the bottom of the Create service account page.

You can now use this service account and private key when configuring bursting in Pexip Infinity (**Platform > Global Settings > Cloud Bursting**).

Configuring the bursting threshold

When enabling your platform for cloud bursting the most important decision you must make is the level at which to set the bursting threshold:

- The bursting threshold controls when your dynamic overflow nodes in your cloud service are automatically started up so that they can provide additional conferencing capacity. When the number of additional HD calls that can still be hosted in a location reaches or drops below the threshold, it triggers Pexip Infinity into starting up an overflow node in the overflow location.

For example, setting the threshold to 5 means that when there are 5 or fewer HD connections still available in a location, an overflow node will be started up.

- When an overflow location reaches the bursting threshold i.e. the number of additional HD calls that can still be hosted on the Conferencing Nodes in the overflow location reaches the threshold, another overflow node in that location is started up, and so on.

Note that the current number of free HD connections in the original location is ignored when deciding if the overflow location needs to overflow further — however, new calls will automatically use any available media resource that has become available within the original principal location.

- The bursting threshold is a global setting — it applies to every system location in your deployment.
- Note that it takes approximately 5 minutes for a dynamic node instance to start up and become available for conference hosting. If your principal deployment reaches full capacity, and the overflow nodes have not completed initiating, any incoming calls during this period will be rejected with "capacity exceeded" messages. You have to balance the need for having standby capacity started up in time to meet the expected demand, against starting up nodes too early and incurring extra unnecessary costs.

Manually starting an overflow node

If you know that your system will need additional capacity at a specific time due to a predictable or scheduled spike in demand, but do not want to wait for the bursting threshold to be triggered before starting up the overflow nodes, you can manually start up any of your overflow nodes.

- ⓘ Do not manually start an overflow node too early. If you manually start up a node more than the **Minimum lifetime minutes** before the node is needed, it will most probably get automatically stopped again before it is used.

You can start overflow nodes via the management API or via the Administrator interface:

- **Via the management API:** the `cloud_node` status resource can be used to list all of the available overflow nodes, the `cloud_monitored_location` and `cloud_overflow_location` resources retrieve the current load on the primary locations and any currently active overflow locations respectively, and the `start_cloudnode` resource can be used to manually start up any overflow node. This means that a third-party scheduling system, for example, could be configured to start up the overflow nodes via the management API approximately 10 minutes before a large conference is due to start.

For example, let's assume that you have:

- ⓘ a regular spike in conferencing capacity demand at 9:00am every morning
- ⓘ an even usage of about 20% of that spike level during the rest of the day
- ⓘ a 30:70 ratio between your "always on" capacity and your overflow cloud capacity

we would recommend:

- ⓘ configuring a low bursting threshold, such as 10-20% of your "always on" capacity (i.e. if your "always on" capacity is 80 HD calls, then set the bursting threshold to 12)
- ⓘ getting your scheduling system to call the API to manually start up all of your overflow cloud nodes at 8:50am on weekdays.

- **Via the Pexip Infinity Administrator interface:** go to Status > Cloud Bursting and select Start for the required nodes (the Start option is in the final column of the Cloud overflow nodes table).

Converting between overflow and "always on" GCP Conferencing Nodes

If you need to convert an existing "always on" GCP Conferencing Node into an overflow node:

1. In GCP:
 - ⓘ Apply to the instance a label with a Key of `pexip-cloud` and an associated Value set to the Tag value that is shown in the **Cloud bursting** section of the **Platform > Global Settings** page.
 - ⓘ Manually stop the node instance on GCP.
2. In Pexip Infinity:
 - ⓘ Change the system location of the Conferencing Node to the overflow system location (e.g. "GCP burst").
 - ⓘ Disable the node's **Enable distributed database** setting.
 - ⓘ You should avoid frequent toggling of this setting. When changing this setting on multiple Conferencing Nodes, update one node at a time, waiting a few minutes before updating the next node.

If you need to convert an existing GCP overflow Conferencing Node into an "always on" node:

1. In GCP:
 - ⓘ Remove the label with a Key of `pexip-cloud` from the GCP instance.
 - ⓘ Manually start the node instance on GCP.
2. In Pexip Infinity:
 - ⓘ Change the system location of the Conferencing Node to a location other than the overflow system location.
 - ⓘ Enable the node's **Enable distributed database** setting.
 - ⓘ You should avoid frequent toggling of this setting. When changing this setting on multiple Conferencing Nodes, update one node at a time, waiting a few minutes before updating the next node.

Managing Google Compute Engine VM instances

This section describes the common maintenance tasks for [stopping](#), [restarting](#) and [permanently removing](#) Conferencing Node VM instances on the Google Cloud Platform (GCP).

Temporarily removing (stopping) a Conferencing Node instance

At any time you can temporarily remove a Conferencing Node instance from your Pexip Infinity platform if, for example, you do not need all of your current conferencing capacity.

To temporarily remove a Conferencing Node instance:

1. Put the Conferencing Node into maintenance mode via the Pexip Infinity Administrator interface on the Management Node:
 - a. Go to **Platform > Conferencing Nodes**.
 - b. Select the Conferencing Node(s).
 - c. From the **Action** menu at the top left of the screen, select **Enable maintenance mode** and then select **Go**.
While maintenance mode is enabled, this Conferencing Node will not accept any new conference instances.
 - d. Wait until any existing conferences on that Conferencing Node have finished. To check, go to **Status > Live View**.
2. Stop the Conferencing Node instance on GCP:
 - a. From the GCP project console, select **Virtual Machines** to see the status of all of your instances.
 - b. Select the instance you want to shut down.
 - c. At the top of the VM instances page, select **Stop**.

Reinstating (restarting) a stopped Conferencing Node instance

You can reinstate a Conferencing Node instance that has already been installed but has been temporarily shut down.

To restart a Conferencing Node instance:

1. Restart the Conferencing Node instance on GCP:
 - a. From the GCP project console, select **Virtual Machines** to see the status of all of your instances.
 - b. Select the instance you want to restart.
 - c. At the top right-hand of the page, select **Start** to restart the instance.
2. Take the Conferencing Node out of maintenance mode via the Pexip Infinity Administrator interface on the Management Node:
 - a. Go to **Platform > Conferencing Nodes**.
 - b. Select the Conferencing Node.
 - c. Clear the **Enable maintenance mode** check box and select **Save**.

After reinstating a Conferencing Node, it takes approximately 5 minutes for the node to reboot and be available for conference hosting, and for its last contacted status to be updated on the Management Node.

Permanently removing a Conferencing Node instance

If you no longer need a Conferencing Node instance, you can permanently delete it from your Pexip Infinity platform.

To remove a Conferencing Node instance:

1. If you have not already done so, put the Conferencing Node into maintenance mode via the Pexip Infinity Administrator interface on the Management Node:
 - a. Go to **Platform > Conferencing Nodes**.
 - b. Select the Conferencing Node(s).
 - c. From the **Action** menu at the top left of the screen, select **Enable maintenance mode** and then select **Go**.

- While maintenance mode is enabled, this Conferencing Node will not accept any new conference instances.
- d. Wait until any existing conferences on that Conferencing Node have finished. To check, go to **Status > Live View**.
 2. Delete the Conferencing Node from the Management Node:
 - a. Go to **Platform > Conferencing Nodes** and select the Conferencing Node.
 - b. Select the check box next to the node you want to delete, and then from the **Action** drop-down menu, select **Delete selected Conferencing Nodes** and then select **Go**.
 3. Delete the Conferencing Node instance on GCP:
 - a. From the GCP project console, go to **Compute Engine > VM Instances**.
 - b. Check the instance you want to permanently remove.
 - c. Select **Delete** to remove the instance.