

Flex Metal Cloud

# IT Leadership Buyer's Guide



# CONTENTS

<b>Background</b> . . . . .	<b>3</b>
<b>Private Cloud Cores</b> . . . . .	<b>4</b>
How do Flex Metal Clouds work? . . . . .	4
How do I control my Clouds? . . . . .	5
How many VMs will I get in my Cloud? . . . . .	6
How much Storage will I get in my Cloud? . . . . .	7
<b>An Example Deployment</b> . . . . .	<b>8</b>
Example Clusters at 80% Utilization. . . . .	8
How do I grow my Cloud? . . . . .	9
How do I set spending limits? . . . . .	11
How are Networks or IPs handled? . . . . .	11
<b>Support Levels</b> . . . . .	<b>12</b>
What support levels do you offer? . . . . .	12
What skills will my Team need to run the Clouds ourselves? . . . . .	13

This guide is intended for IT Leadership and Corporate Executives. The advent of On-Demand Private Cloud has blurred the line between public and private clouds. This guide provides a framework for this new technology that can increase competitiveness and lower costs.

# FLEX METAL CLOUD BACKGROUND

Use of Open Source Private Cloud has a long history with various benefits, costs, and risks. The risks and costs often outweighed the benefits for all but the largest companies or for hosting providers. This changed significantly with the advent of On-Demand Private Cloud. On a private cloud, time to utilization was **commonly measured in quarters before 2016**, then months in 2016-2017, then weeks in 2018-2019, then in 2020 it fell quickly from weeks to **minutes**.

InMotion Hosting was the first to reach minutes with OpenStack and Ceph. You can now expect Clouds to take about 45 minutes and adding cluster members to take around 20 minutes. This yields a cloud that includes support for Containers/Kubernetes, VMs, Block Storage, Object Storage, Orchestration, and Networking. User Management and Governance models are also standard.



# PRIVATE CLOUD CORES

## How do Flex Metal Clouds work?

First, since it is trivial to spin up and down full clouds, you now have the option to separate workloads into separate clouds based on usage patterns, security models, or geographic considerations. Larger clouds are more efficient, but smaller clouds have many unique benefits to consider.

At the heart of each of your Clouds is the Private Cloud Core (PCC). The PCC is a hyper-converged set of three servers that provide all resources like VMs and Block Storage. It also includes the Control Plane which manages the Cloud itself and provides Networking.

**There are two flavors of Private Cloud Cores, Small and Standard.**

### **SMALL**

Smalls are ideal for testing, training, and smaller production implementations. A good example of a small deployment is a secure environment for critical data that is separated from your main IT infrastructure.

### **STANDARD**

Standards are the typical building block as they provide significant resources as a hyper-converged Cloud and are also ideal as a Control Plane for very large Clouds. It is trivial to lower the workloads on a PCC if it needs to handle more Control Plane work.

# Private Cloud Cores Continued

---

## How do I control my Clouds?

Your top level control panel is known as Flex Metal Central (FMC). From within FMC you can create Clouds with a few clicks. **The Clouds are completely autonomous from each other and have their own hardware, user interfaces, APIs, and public IP space.**

### From Flex Metal Central you can, per Cloud:

- Setup billing restrictions
- Add and remove Cloud Administrators
- Add and remove hardware
- Access Horizon, the Cloud's autonomous User Interface
- Access Billing History
- Contact Support and Account Managers
- Delete the Cloud

### From within each Cloud you will have CLI, API, and GUI access to those Cloud's resources. From within the Cloud you can control:

- Networks including Routers, Switches and access to Public IP Space
- Projects - this provides Self Service access for Groups/Departments/Customers, etc
- Project Users
- Resource limitations for Projects
- Object Storage
- VM Images and Flavors available to Users
- Orchestration

---

THE CLOUDS ARE COMPLETELY  
**AUTONOMOUS FROM EACH OTHER**

---

## Private Cloud Cores Continued

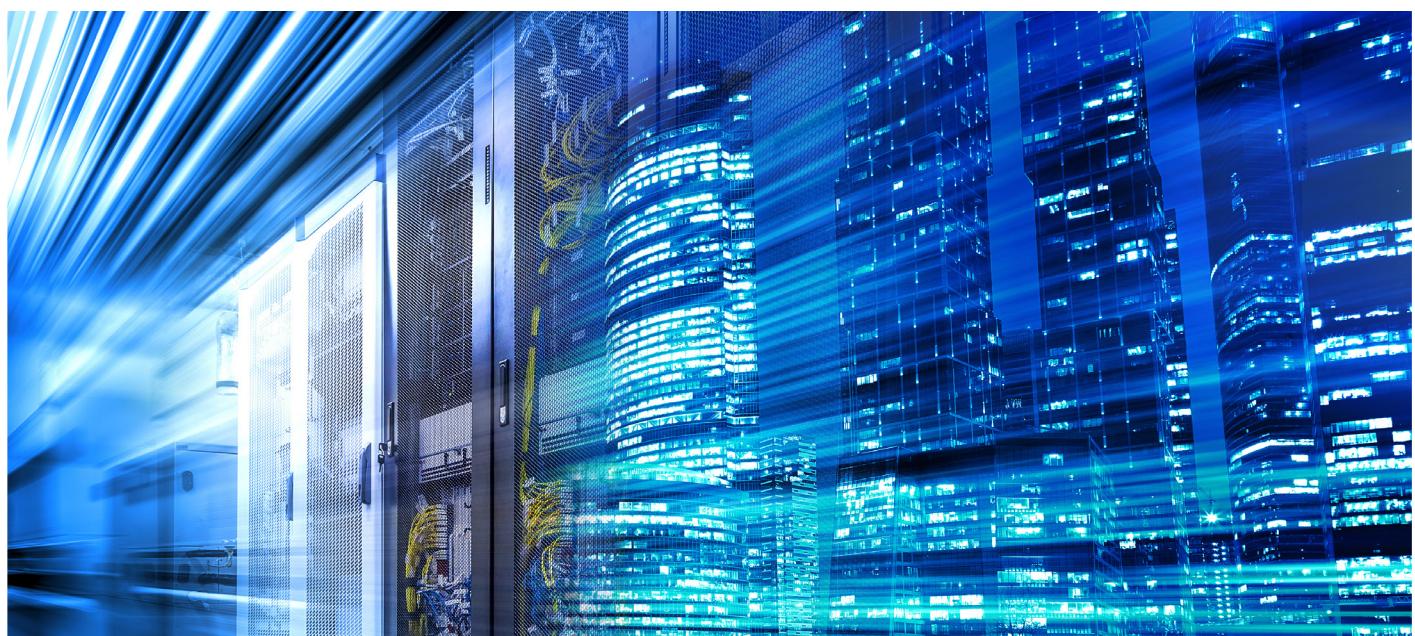
---

### How many VMs will I get in my Cloud?

First, you may be surprised that Public Clouds “oversubscribe” their CPU’s threads by as much as 16X. RAM “oversubscription” ratios of 1.5X also appear to be in use. There is very good logic and real world data backing up this decision for them. For your Cloud, you will control the ratios based on your anticipated workloads. Your Account Manager can help with this. We recommend 12X to start for CPUs and 1.5X for RAM.

To see an estimate of VMs, use our [Pricing Calculator](#). It takes into account if the cluster is Hyper-Converged and thus the Control Plane needs to have resources reserved before RAM and CPU can be used for VMs.

Of note, all servers have small boot drives that are separate from the main drives. To calculate VMs on your own, you take each server’s RAM minus 7GB of RAM for the Control Plane and 2GB of RAM per Ceph OSD (drive or drive partition) then multiply it by 1.5. Next you take each server’s CPU threads and minus 3.5 threads for the Control Plane and then multiply by 12.



## Private Cloud Cores Continued

---

### **How much Storage will I get in my Cloud?**

You will see RAW disk space and Usable disk space within our documentation. RAW is what the actual drives supply. Usable is what can be allocated after data redundancy. We recommend a minimum of “2X Replication” for redundancy. This is akin to RAID 1 and means your data exists in two places at one time. Other common settings are 3X Replication and Erasure Coded 3+2, 4+2, 5+2, etc. Erasure Coding is the most efficient for RAW versus Usable, but requires more resources and is slower than 2X Replication. 3X Replication is common as well and enabled by default unless your storage costs override the data protection benefit of 3X Replication.

To calculate Storage on your own, add together your RAW and divide by your Replication level. For Erasure Coding it is more complex but starts at about a 40% drop with small clusters but goes to as efficient as only a 12% drop for large clusters.



# AN EXAMPLE DEPLOYMENT

For illustration, we will assume you have selected a Standard Private Cloud Core. The PCC Standard is composed of 3 identical servers each with 128GB of RAM, one CPU with 8 cores/16 threads, one 3.2TB PCIeX4 NVMe SSD, one 128GB SSD as a boot drive, and dual 10gbit network uplinks. We will take into account the Subscription Ratios (see below), reserve appropriate RAM and CPU for the Control Plane, and run the Cloud at 80% usage only.

With those requirements, this 3 server cluster can support approximately 100 VMs with 2vCPUs and 4GB RAM with a total of 4.8TB or 3.2TB of redundant HA NVMe storage (9.6 RAW, 2 or 3 Replicas). This starts at only \$44/day or approximately \$1,254/30 days with discounts applied after 1 week of continuous usage. This is roughly \$13/VM/month and backed by ultra high IOPS NVMe. This represents a significant savings over public cloud.

## Example Clusters at 80% Utilization

Servers	2 VCPU/4GB VMS Available Storage	Flex Metal Cloud (30 Day Discounted, Year Commit)	Public Cloud (On-Demand/ Year Commit)
PCC Small (3 Servers)	46 VMS, 1.4TB HA Sata SSD	\$597.60 Or \$459.00	\$1,700+/Mo Or \$1,100+/Mo
PCC Standard (3 Servers)	102 VMS, 4.8TB HA NVMe SSD	\$1353.60 \$1,101.60	\$6,500+/Mo Or \$4,800+/Mo
PCC Standard + 3 Converged Standards	210 VMS, 9.6TB HA NVMe SSD	\$2649.60 \$2,203.20	\$13,400+/Mo Or \$8,900+/Mo
PCC Standard + 3 Compute + 3 Storage Blended	552 VMS 10.6TB HA NVMe SSD 9.6TB LVM NVMe SSD 76.8TB HA HDD	\$7034.40 \$5,930.28	\$39,600+/Mo Or \$31,900+/Mo
PCC Standard + 6 Compute + 4 Storage Blended	1002 VMS 12.5TB HA NVMe SSD 19.2TB LVM NVMe SSD 115.2TB HA HDD	\$10,540 30 Days \$8,910.72	\$62,100+/Mo Or \$47,900+/Mo

\* These examples are based on 80% usage of the CPU and RAM. Going above is possible but can lead to undesired when approaching 100%. These are examples and do not guarantee exact performance.

\*\* Pricing varies by location and other factors. The Official Rate Cards available in Flex Metal Central are the sole official binding pricing.

\*\*\* Estimated based on research in Feb 2021. This is not an official price from any specific Public Cloud but a mixture of the 3 mega-clouds.

### Pricing

Pricing varies by Data Center location. Billing is usage based, hourly, and is discounted after 7 days of continuous usage. Additional discounts are available for Term Agreements ranging from 1 year to 5 years. Contact your Account Manager for details. Sign up for Flex Metal Central for current pricing.

# An Example Deployment Continued

---

## How do I grow my Cloud?

It is always important to note that it is easy to create Clouds so sometimes the need for more resources can best be met with another Cloud.

For an existing Cloud, it can be grown one of three ways. First, you can grow your Private Cloud Core by adding additional Converged Nodes. These Converged Nodes are spec matches for your PCC and allow Ceph to use these additional drives within the existing Storage Cluster. Thus, you will grow your Compute and Storage at the same time.

Second, you can select Compute Nodes. These nodes will come with various CPUs and Storage but will not be joined to your existing Ceph by default. The drives on the Compute Nodes are not part of an HA cluster and can either be local LVMs or Ephemeral (it will not survive a VM being deleted). When using storage on a Compute Node the application must provide for data protection as both LVM and Ephemeral drives are single points of failure. The upside for storage is NVMe and it is incredibly fast.

The third way to grow your cloud is when you have determined you want specific Compute or Storage options different from your current cluster's performance. In this case, you will need another set of servers with the performance characteristics you want.

For Storage, clusters have a minimum size of 3 servers for data replication and to be highly available.



## An Example Deployment Continued

---

For Compute, you can add one server at a time but it is recommended to add 3 or more servers at one time to allow proper performance if one server must be removed from the cluster for any reason. For example, if you are growing from 3 Small Compute servers, you can add one Large to it, but it is better to add 3 Standard Compute servers, then migrate your VMs to the Standards and return the 3 Smalls.

Adding or removing nodes can be done within the Flex Metal Central Interface and, in a future release, via an API call. Once that server is within your network, you will decide what current VMs will be migrated to the new Compute.

---

### **HOW FAST WILL MY STORAGE BE?**

BLAZING! OR AS FAST AS YOU NEED.

---

We have three types of Storage - NVMe for ultra high performance, Sata SSD for high performance, and spinning HDD for large data stores at relatively slow performance. Large spinner based servers, like our 12 X 12TB option, also come with a set of NVMe drives to accelerate the storage pool to a medium performance level. Reference Architectures with FIO performance stats are available from your Account Manager.

When using highly available storage, like a Ceph Cluster, there are tradeoffs. More replicas take longer to complete writes. A 2 Replica will be faster than a 3 Replica which will be faster than Erasure Coding. Traversing the network lowers throughput. Most deployments will mix how storage is accessed based on the applications.

For another example, drives (excluding boot drives) on a hardware node that are not part of your Storage Cluster can be made available directly to a VM running on that node. For example, high performance MySQL services are often configured to use a locally mounted NVMe drive with MySQL asynchronously replicating across the network to another VM with local NVMe. Thus HA is accomplished via the software.

This is more performant than using the Ceph's highly available block device mount by 3-4X. It does require additional knowledge but that configuration is common.

## An Example Deployment Continued

---

### **How do I set spending limits?**

Spending limits are set directly in Flex Metal Central per Cloud. Once set, if a Server is requested that will put the daily bill over the maximum, the requests will fail and indicate a budget has been reached.

### **How are Networks or IPs handled?**

All Clouds support private networks including on the fly creation of Routers, Switches, and Subnets. VxLAN technology is used for separation of Projects that are VM based. From an application level VLANs created with VxLAN are indistinguishable from a VLAN. Hundreds or even thousands of VLANs are supported in this way. For Bare Metal, you are supplied a total of 6 VLANs with each Cloud and can access additional if required.

You can grow or shrink from within each Cloud. For public IP addresses, we supply you with a /28 to start. Additional IPs are available for monthly per IP fee. We also support assigning our routers to broadcast your IP blocks (/24 and larger via SWIP) and route them to your deployment.



# SUPPORT LEVELS

## What support levels do you offer?

Many IT departments and IT leadership are looking for flexible Support Levels as they navigate this transition to off premise hybrid cloud solutions. There are two standard offerings plus we can provide tailored offerings.

- **Hardware Level Ticket Based Break/Fix** - This is included with all Clouds. Your team works with us using Standard Response Tickets and Emergency Response Tickets with appropriate SLAs for each. This is the lowest cost level and is intended for Teams that either already managed Clouds or will be training up to take over the Clouds.
- **Managed Private Cloud** - This includes the Hardware and Software Level Break/Fix plus management of the OpenStack and Ceph.
- **Co-Management of Your Deployment** - Starts at \$5,000/mo. Includes a lead IT professional responsible for your organization's needs and the health of your deployment. Access to your lead and subject matter experts are directly within a shared Slack channel or private Slack or other communications channel if necessary. Planning and execution of migrations, team training, and more can be included.

## Support Levels Continued

---

### What skills will my Team need to run the Clouds ourselves?

The complexity in private clouds comes primarily from the architecting stage, not the operational stage. As your team will have access to already running Clouds, much of the complexity has been eliminated. Though running a Cloud is much easier than ever before, the following factors will help you navigate this decision.

- 1.** It is important to remember that individual Clouds can be spun up in a matter of minutes. Testing, training, and dry runs become simple and fast in a safe environment. **This is very empowering for your team and is one of the most important attributes of on-demand private cloud.**
- 2.** OpenStack and Ceph are mature technologies. There are vibrant communities around both with extensive documentation and training. Both certainly are complex but configurations have been supplied that work for the majority of workloads.
- 3.** We provide a Cloud Operators Training Manual that can be used as your base training for all Cloud Operations Staff. For System Administrators that have been running large deployments, the tools and processes will be familiar. Previous knowledge of Ansible will speed up training. For experienced System Admins, the Cloud Operators Manual is designed to take between 30 and 60 hours to complete.
- 4.** We can provide Tier 3 Support in case your Team has situations they can not solve.
- 5.** You can elect to use Managed Private Cloud to get started and transition after your team reaches sufficient expertise.

## HOW DO I GET STARTED?

Speak with our Cloud Solutions Team to get help navigating the world of complex infrastructure decisions.

- » [1-888-470-FLEX \(3539\)](tel:1888470FLEX)
- » [privatecloudsolutions@inmotionhosting.com](mailto:privatecloudsolutions@inmotionhosting.com)
- » [inmotionhosting.com/flex](http://inmotionhosting.com/flex)

## Get \$200 in Free Cloud Credits

**Get started with free cloud credits for an On-Demand Hosted Private Cloud. Use the credits to test your Proof of Concept (PoC) before going into production or to get familiar with OpenStack.**

**CLAIM YOUR FREE CREDITS**

### \*Detailed Ratio Guide

When creating a Compute cluster there are macro settings available to help maximize the underlying hardware when dividing it up into Virtual Machines. A critical number to understand, and a number used by Public Cloud and Private Cloud providers alike and based on a tremendous amount of data, is that when using a “standard 2vCPU 4GB” VM, most VMs will only use 1/16th of the allocated CPUs. This can vary, but with a large number of VMs in a cluster this average is very common. In addition, RAM is also underused on average and the industry has settled to 1.5x the hardware RAM can be allocated to VMs. Your particular applications will be different and care should be taken if you are putting specific workloads versus general workloads for your company. To start, we recommend a slightly more conservative ratio than the industry for CPU and have set 1:12 for CPUs and we use the industry standard of 1:1.5 for RAM. Our calculations use these ratios within our documentation.



# *inmotion* cloud

555 S Independence Blvd  
Virginia Beach, VA 23452

**PHONE:** 888.470.FLEX (3539)

**EMAIL:** [privatecloudsolutions@inmotionhosting.com](mailto:privatecloudsolutions@inmotionhosting.com)

**WEB:** [www.inmotionhosting.com/flex](http://www.inmotionhosting.com/flex)