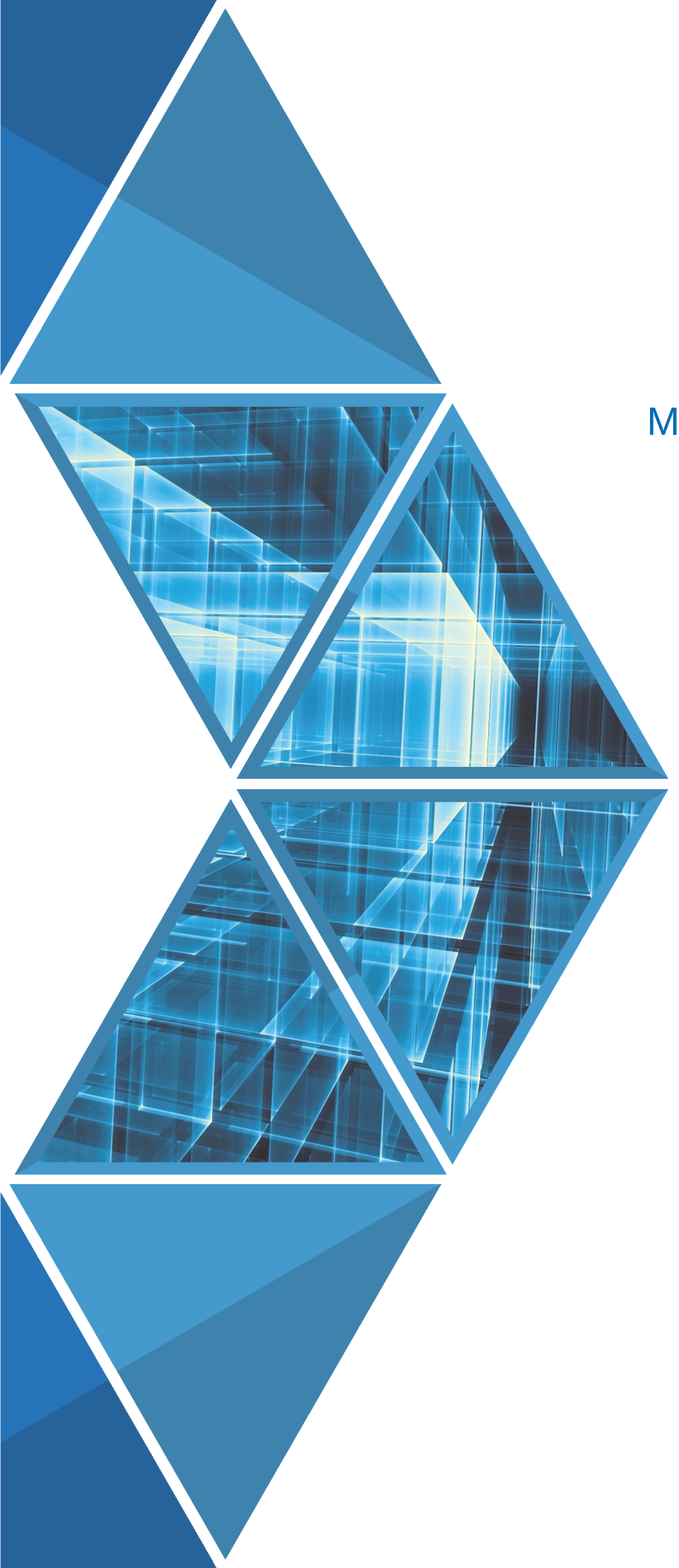

MemVerge

Memory Machine™ Cloud

Quick Start Guide for AWS

First Published: 2022-12-06

Last Revised: 2023-06-06

Contents

MMCloud Quick Start Guide for AWS

Introduction

Memory Machine Cloud (MMCloud) is a software platform that streamlines the way containerized applications are deployed in a Cloud Service Provider (CSP) network, such as AWS. The MMCloud Operations Center (OpCenter) must be running for MMCloud to operate (a license is required). MMCloud includes a built-in job scheduler for deploying containerized applications. Unless the container has a pre-defined ENTRYPOINT, the job scheduler requires a job script that you provide.

This guide provides instructions on how to install the OpCenter in AWS and how to run containerized applications.

Version

This guide is consistent with the OpCenter release shown in the table.

MMCloud CLI	OpCenter Release	Date
v2.2.0-1363a1b-ElNido	v2.2.0-1363a1b-ElNido	2023-05-23T14:10:47Z

Summary

To deploy MMCloud, complete the following two steps (the order is not significant).

- Obtain the license that enables you to use OpCenter.

  To request a license, log in to the MemVerge account server (register first if you don't have an account set up). Select the type of license and submit the form that is displayed.
- Create an AWS CloudFormation stack to start an OpCenter instance.

  Using the MemVerge-provided template, the CloudFormation service instantiates your OpCenter. Once the OpCenter is running, the software can be upgraded without needing to re-create the CloudFormation stack. Anyone with login credentials on the OpCenter instance can submit jobs.

Activate the OpCenter license as follows.

- Open a browser and go to the IP address associated with your OpCenter. Use the public (private) IP address if you are connecting from outside (inside) the VPC.
- Log in to the OpCenter web interface with the username admin and default password memverge. You can change the admin password after you have logged in.
- On the top, right-hand side of any screen (except the *Submit Job* screen), click on the license icon. Follow the instructions in the pop-up screen to activate your license.

All interactions with the OpCenter use the MMCloud CLI. There are three ways to access the CLI:

- Use the CLI shell provided by the OpCenter web interface.
- Download the CLI binary and run it in a terminal window on your local machine.
- Log in to the OpCenter server and run the CLI from a secure terminal session (not recommended).

The remainder of this document provides detailed instructions for each of these steps.

Obtaining an OpCenter License

This procedure applies to all Cloud Service Providers.

Step 1 Open a browser and go to the MemVerge account server; enter your credentials at the sign-in screen.

> **Note:** If you are not registered, click on the *Sign Up* link and follow the instructions.

Step 2 Select your preferred MMCloud license plan — the choices are *Essentials* (free), *Pro*, or *Enterprise*.

> **Note:** If you are not ready to choose a license, scroll down the page to locate the link to the live demo environment where you can try out a limited set of MMCloud features. Your credentials are provided under the link.

Step 3 Fill in the pop-up form and click on *Confirm*.

Step 4 Follow the instructions to confirm your license selection.

Deploying OpCenter in AWS

Before You Begin

- Contact your AWS administrator to obtain an AWS account with root privileges.
- You need to select a VPC subnet and enter information on CIDR blocks. If you don't have this information, contact your AWS administrator.

Step 1 Log in to your *AWS Management console*

Step 2 Create a key pair

Step 2.1 From the navigation bar, click on *Services* and select *EC2*.

Step 2.2 On the right-hand side of the navigation bar of the *EC2 Dashboard*, check what region you are in.

If you need to change regions, use the drop-down menu to select a new region.

Step 2.3 From the left-hand panel of the *EC2 Dashboard*, go to *Network & Security*, and click on *Key Pairs*.

The *Key Pairs* panel shows the key pairs that are available. If you need to create a key pair, click on the *Create key pair* button on the right-hand side and follow the instructions. A copy of the key pair file is automatically downloaded to your local machine.

Note: The key pair is unique to a region. If you change regions, you must create a key pair for the new region.

Step 3 Identify the Availability Zone in which each VPC subnet is located

Step 3.1 Using the search tool in the navigation bar, find the *VPC dashboard*.

Step 3.2 On the left-hand panel of the *VPC dashboard*, click on *Subnets*.

The *Subnets* console displays a table that shows each VPC subnet, the VPC the subnet is part of, and the Availability Zone in which the subnet resides.

Step 3.3 Keep this browser tab open.

You need to retrieve information from this table later in the deployment procedure.

Step 4 Locate Memory Machine Cloud in the AWS Marketplace

Step 4.1 Log in to the *AWS Management console*.

Step 4.2 Go the AWS Marketplace and select *Discover products* from the left-hand panel.

Step 4.3 Search for Memory Machine Cloud.

Step 4.4 Click on the result and then proceed to the next step.

Step 5 Subscribe to Memory Machine Cloud

Step 5.1 Click on *Continue to Subscribe*.

Step 5.2 If you agree to the terms and conditions, click on *Accept Terms*, then click on *Continue to Configuration*.

Step 5.3 At the *Configure this software* screen, use the pull-down menu to select MMCE Topology as the fulfillment option. From the Region pull-down menu, select the region in which to deploy OpCenter.

Step 5.4 Click on *Continue to Launch*.

Step 5.5 Under *Choose Action*, use the pull-down menu to select *Launch CloudFormation*.

Step 5.6 Click on *Launch*.

You are now at the *CloudFormation* > *Stacks* > *Create stack* screen. The left-hand panel shows the four-step procedure for creating a new stack. You are at the first step.

Step 5.7 Click on *Next*.

Step 6 Specify the *CloudFormation* stack details

Step 6.1 Enter a unique stack name, using the allowed characters.

Step 6.2 Fill in the required parameters as follows (seek guidance from your AWS administrator if needed).

- *0mvOpCenterType*: Accept the default or use the pull-down menu to change the size of the VM to run OpCenter.
- *1mvFloatKeyName*: Use the pull-down menu to select key pair name you created in step 2.
- *2mvFloatVPCID*: Select a VPC from the pull-down menu. (Usually, there is only one VPC; if there are multiple, consult your AWS administrator for guidance.)
- *3mvFloatSubnetID*: Select a VPC subnet from the pull-down menu. (Usually, VPC subnets are equivalent; consult your AWS administrator for guidance.)
- *4mvAvailabilityZone*: Go to the browser tab you opened in step 3 to determine the Availability Zone in which your VPC subnet resides. Then return here to use the pull-down menu to select the appropriate Availability Zone.
- *5mvFloatPublicCIDR*: Provide the range of IP addresses allowed to access server. To allow access from any address, enter `0.0.0.0/0`

- *6mvFloatSSHCIDR*: Provide the range of IP addresses allowed to access server using **ssh**. To allow **ssh** access from any address, enter **0.0.0.0/0**
- *7mvFloatInternalCIDR*: Provide the range of IP addresses for internal communication among OpCenter instances. To allow communication from any address, enter **0.0.0.0/0**

Step 6.3 Click on *Next*.

Step 7 Configure stack options

Keep the default options and click on *Next*.

Step 8 Review

Step 8.1 Check the box at the bottom of the page to acknowledge that you are aware *CloudFormation* may create IAM resources.

Step 8.2 Click on *Create stack* to proceed.

Wait until the process completes successfully.

Step 9 Check your stack

Step 9.1 Go to the *CloudFormation* > *Stacks* screen and select your stack.

Step 9.2 From the *CloudFormation* > *Stacks* > *"your_stack_name"* screen, select the *Resources* tab.

Step 9.3 Click on *Physical ID* associated with *mvOpCenter*.

This takes you to the *Instances* screen of the EC2 console.

Step 9.4 Select the OpCenter instance.

This screen displays OpCenter status as well as private and public IP addresses associated with the OpCenter.

Step 9.5 If you are outside (inside) your organization's virtual private cloud, open a browser and go to the OpCenter public (private) IP address, respectively.

If the OpCenter is running, the OpCenter web interface displays the version of OpCenter software installed.

Using the OpCenter Web Interface

OpCenter provides a web interface to activate the license and to submit and manage jobs.

Accessing the OpCenterWeb Interface

- Open a browser and go to the IP address (public if you are outside the VPC, private if you are inside the VPC) associated with your OpCenter. The landing page allows you to do the following.
  - See the version of software installed
  - Log in to OpCenter
  - Download the CLI tool for your local operating system (Windows, Linux or macOS)
- Log in to the OpCenter by entering your username and password, and then clicking on the *Log In* button.

  The first time you log in, enter the username admin and default password memverge. After you log in, you can change the admin password.

Navigating the OpCenter Web Interface

The left-hand panel displays the available screens. The following screens are available.

- Submit Job
- Cost Summary (only visible if you log in as *admin*)
- Jobs
- App Library

At the top, right-hand side are a series of icons that allow you to perform the following actions.

- Open the web CLI shell
- Manage the OpCenter license
- Access documentation at docs.memverge.com
- Log out of the OpCenter

Applying the OpCenter License

- From any screen (except the *Submit Job* screen), click on the license icon (top, right-hand side).
- On the pop-up form, enter your login credentials for *account.memverge.com* and click on *Confirm*. If the license is successfully applied, the pop-up display shows the license ID and issue date.

Note: You can log in to a different account at account.memverge.com by clicking on the license icon and then clicking on the *Change* button in the pop-up screen.

Submit Job Screen

On the left-hand side of the *Submit Job* screen, you can fill in the fields in a form and then submit a job. The right-hand side shows the equivalent CLI commands. After completing the form, submit the job by clicking on the *Submit* button at the bottom, on the right.

Cost Summary Screen

The *Cost Summary* screen summarizes your cloud spending in a single dashboard. View cloud costs and savings by week, month, year, application, or user.

When you instantiate a VM, the CSP begins charging (in one second increments) for the compute instance and any storage devices you configure. You must configure disk space for the root volume (where the operating system is stored) and normally you configure additional disk space for your data (the data volume). VMs instantiated by OpCenter include two additional disk volumes: the image volume (to store the container image) and the snapshot volume (to store snapshots).

The default size for the root volume is 40 GB. If this volume is used for an entire month, the charge would be about USD 4.

Cloud cost for a single job is calculated as follows.

Cloud cost = Instance cost + data volume cost + image volume cost + snapshot volume cost

where the Instance can be an On-demand Instance or a Spot Instance. The root volume cost is *not* included.

Net savings is calculated by comparing what the cost would be if the job were run on an On-demand Instance, not using OpCenter, with the *Cloud Cost*. That is,

Net savings = On-demand cost - Actual instance cost - image volume cost - snapshot volume cost

Cloud Cost and *Net Savings* for each job are aggregated to arrive at the numbers reported per week, per month, etc.

Jobs Screen

Using the *Jobs* screen, you can view and manage current jobs, and you can view archived jobs. You can search for jobs and filter the displayed results. Click on a job to display detailed information about the job, including logs and graphs of resource utilization over time.

To migrate, modify or cancel a currently running job, identify the job by ID or by name. In the *Actions* column are three icons corresponding to migrate, modify, and cancel, respectively. Clicking on an icon brings up a dialog box. Fill in the dialog box to complete the action.

To view the logs associated with a current or archived job, click on the job ID. In the table that is displayed, click on *Attachments* to list the logs associated with the job. Click on the *Preview* icon to view the contents of a log.

To view a graphical display of resource utilization associated with a current or archived job, click on the job ID and then click on the *WaveWatcher* tab in the displayed table.

App Library Screen

The *App Library* screen shows the container images that are currently loaded in to the OpCenter image library. The display shows images that are automatically loaded (*Built-in*) as well as images that you add (*Private*). To add an image, go to the *Private* tab and click on *Add Image*.

Images with the status listed as *Ready* are currently loaded in to the OpCenter's image cache.

Using the MMCloud CLI

Step 1 Select your method for invoking MMCloud CLI commands.

Choose one of the following:

- Log in to the OpCenter web interface and open the *Web CLI* shell.

  When you log in for the first time, enter username *admin* and default password *memverge*. From any screen (except *Submit Job*), click on the terminal icon on the top, right-hand side to open the *Web CLI* shell.

  A subset of CLI commands is available via the *Web CLI* shell. Type **`-h`** to see what commands are supported. Skip to step 3.

- From the OpCenter web interface landing page, go to *Download command line tool for* displayed at the bottom of the page on the right-hand side. Click on the link for your operating system (Windows, Linux, or macOS)

  Use **`chmod`** to make the downloaded **`float`** binary executable and add path to the **`float`** binary to your PATH variable.

  > **Note:** From macOS 10.15 (Catalina) onwards, zsh is the default shell. In zsh, float is a reserved word. To use the **`float`** CLI with macOS, either change the shell to bash or alias the word *float* to the **`float`** binary, as follows:
  >
  > ```
  > alias float = /path/to/float_binary/float
  > ```
  >
  > where **`/path/to/float_binary/`** is the path to where you placed the **`float`** binary.

- Establish a secure terminal session (using **`ssh`**) on the OpCenter server. The **`float`** CLI is automatically installed on this server.

> **Note:** The *web CLI* shell is a web interface that provides a **`float`** prompt, that is, every entry you type automatically has **`float`** pre-pended. The other two methods provide a terminal with a shell prompt — you must type **`float`** followed by subcommand and options.

Step 2 Log in to the OpCenter as the *admin* user by typing the following:

```
float login -u admin -p memverge -a <OpCenter_ip_address>
```

where **`<OpCenter_ip_address>`** is the OpCenter's private IP address if you are within your organization's virtual private cloud (VPC), or public IP address if you are outside the VPC.

Step 3 Add a user by entering the following command:

```
float user add <user_name> --passwd <password> --group <group_name> --create
```

where **`<user_name>`**, **`<password>`**, and **`<group_name>`** are the user name, password, and group, respectively. Repeat the command with different parameters to add other users.

Step 4 Change the default password for the *admin* user by typing the following command:

```
float user passwd admin --passwd <admin_password>
```

where **<admin_password>** is the new password for the *admin* user.

Step 5 From a terminal session (not the *web CLI*), log in to the OpCenter as the *admin* user by typing the following:

```
float login -u admin -p <admin_password> -a <OpCenter_ip_address>
```

where **<OpCenter_ip_address>** is the OpCenter's private IP address if you are within your organization's virtual private cloud (VPC), or public IP address if you are outside the VPC.

Step 6 Upgrade to the latest OpCenter version.

Step 6.1 To see what releases are available, type the following command:

```
float release ls
```

Step 6.2 If there is a later release, upgrade by typing the following command:

```
float release upgrade -r <release_name>
```

where **<release_name>** is the latest release shown in the previous step. After the upgrade completes, log back in.

Step 6.3 Synchronize the **float** version with the OpCenter version by typing the following command:

```
float release sync
```

Submit Your First Job Using the Web Interface

The web interface provides a form that you complete to submit a job.

Step 1 Open the OpCenter web interface and click on *Submit Job*.

Step 2 In the form that pops up, fill in the fields as follows.

- *Name* (optional): Enter a name to identify your job, for example, HelloWorld. If left blank, the OpCenter creates a name for you.
- *Image* (required): Use the drop-down menu to select a container image to run your application, for example, cactus.
- *Job Script*: Enter a location to reach your job script, such as an S3 bucket or a URL, or you can paste the commands into the window. Here is an example job script:

```
#!/bin/bash
echo "Congratulations! You have submitted your first job"
for(( c=1; c<3; c++))
do
        if [[ $(($c % 3)) == 1 ]]; then
                echo "Hello World!"
        else
                echo "Your next job will be more interesting"
        fi
                sleep 20s
done
echo "Job complete"
```

Note: Some containers have a built-in ENTRYPOINT, that is, a command that runs automatically when the container starts. These containers do not require a job script.

- Under *Instance Preferences*, select *CPU & Memory*, then choose, for example, 2 for *vCPU (min)* (required) and 4 for *Memory (min GiB)* (required).
- Under *Storage Volumes*, select *New volume*, 10 for the *Size*, and enter /data for the *Mount Point*, for example.
- Ignore the *Advanced* settings.

Note: The generated CLI command string is shown on the right-hand side.

Step 3 Click on the *Submit* button at the bottom of the right-hand side.

Step 4 To view the status of the job you just submitted, click on *Jobs* in the left-hand panel.

Step 5 Click on your job and then go to the *Attachments* tab.

Step 6 Click on the *Preview* icon next to the log named *stdout.autosave* to view the output of the example job you ran.

Submit Your First Job Using the CLI

Use the MMCloud CLI to submit a job.

Step 1 On the computer where you installed the MMCloud CLI, create a file called helloworld.sh and insert the following contents.

```
#!/bin/bash
echo "Congratulations! You have submitted your first job"
for(( c=1; c<3; c++))
do
        if [[ $(($c % 3)) == 1 ]]; then
                echo "Hello World!"
        else
                echo "Your next job will be more interesting"
        fi
                sleep 20s
done
echo "Job complete"
```

Step 2 Save and then close the file.

Step 3 Open a terminal on this computer and go to the directory where you created helloworld.sh.

Step 4 Log in to the OpCenter by entering:

```
float login -a <op_center_ip_address> -u admin -p <password>
```

where **<op_center_ip_address>** is the public (if you are outside the VPC) or private (if you are inside the VPC) IP address associated with the OpCenter. Use the default value (memverge) for **<password>** if you have not changed it.

Step 5 Submit a job by entering:

```
float submit -i cactus -j helloworld.sh -c 2 -m 4 --dataVolume [size=10]:/data
```

Step 6 Query the job status by entering:

```
float squeue
```

The entry in the first column (labeled ID) is the identifier for your job.

Step 7 View the output of your example job by entering:

```
float log cat stdout.autosave -j <job_id>
```

where **<job_id>** is the identifier associated with your job.