The Skinny

Memory Machine™ Cloud

Control of your cloud is radical.



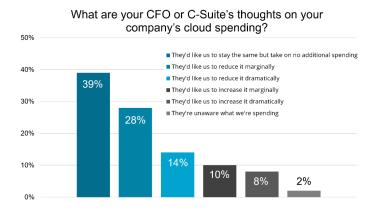




Executive Summary

The massive growth in complexity and cost of cloud estates have the C-suite thinking

Amazon Web Services launched in 2006. Since then cloud computing market revenue has exploded to nearly \$600 billion in 2023 <u>according to Gartner</u>. One consequence of this wild growth is the spiraling cloud costs for many enterprises. The operational complexity of cloud has also become a significant challenge. These issues and others have the C-suite thinking about how to get the cost, performance, availability, and security of their estates under control.



"According to a new study from Wanclouds, 81% of IT leaders say their C-suite has directed them to reduce or take on no additional cloud spending as costs skyrocket and market headwinds worsen."

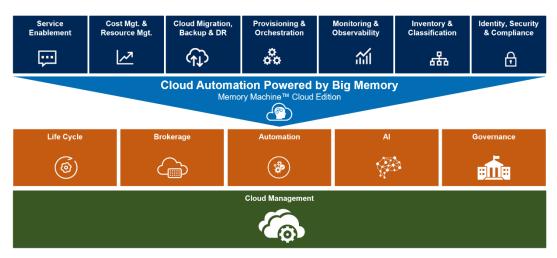
VentureBeat

The industry including MemVerge respond with cloud management

Answering the call to get cloud computing under control, a group of vendors in Cloud Management are delivering 7 categories and 5 sub-categories, as shown below. Memory Machine Cloud from MemVerge is a new segment: Cloud Automation Powered by Big Memory.

Leveraging innovative new Big Memory technology and making use of management technologies such as Infrastructure as Code, Memory Machine uniquely creates the granularity needed to optimize cloud computing during runtime. A few examples of the benefits of this new technology are that long-running workloads can now safely run on low cost spot instances, and resource-intensive workloads can now dynamically migrate to larger or smaller VM instances during runtime for yet another level of cost savings and app acceleration.

A New Class of Cloud Management



Source: MemVerge and Gartner Market Guide for Cloud Management Tooling, 11 April 2022









Introducing Memory Machine™ Cloud



Memory Machine Cloud

Key Features & Benefits

Memory Machine Cloud from MemVerge is cloud automation software with the unique ability to automate and streamline the execution of compute jobs based on patented Big Memory snapshot technology. Key features and benefits include:

Float - A powerful command language with a GUI that allows long complex app deployments to be performed with a few commands or clicks.

AppCapsule – A new type of snapshot that encapsulates an entire application state. The use of AppCapsules allow long-running workloads to recover rapidly from an infrastructure interruption. Never re-start from the beginning again.

SpotSurfer – An application continuity service that uses AppCapsules + cloud automation to make it possible for big stateful workloads to run safely on Spot instances, saving up to 90% in compute cost.

WaveRider - A continuous right-sizing service that allows workloads with short periods of

savings grow. peak usage to float automatically to smaller and larger instances based on their real-time resource needs.

App Library – There is a long and growing list of apps that are qualified for compatibility. Use applications such as Seurat, Cell Ranger, GATK, Sentieon, BWA, SPAdes, Megahit, Supernova, Blast+, and many more, directly from the App Library without needing to install or deploy them.

AWS Qualified Software - Rigorously reviewed and benchmarked against internal and public standards for software quality.

Available in the AWS Marketplace - You can get Memory Machine Cloud today in the AWS Marketplace.

3 ways to get started free - 1. Request a demo and a MemVerge expert will explain how Memory Machine works, then walk you through a demonstration that shows how easy it is to deploy and how powerful the results are. 2. Request a free trial and a MemVerge expert will set you up, or we can provide access for you to play in a Sandbox set-up in AWS. 3. Request Memory Machine WaveRider, a free version you can use as long as you want. It's a lite version of Memory Machine Cloud with automation tools for workload deployment and observability tools that profile your workload to identify opportunities for right sizing.



The Memory Machine Operations Center is where you register your

workloads, set your policies, monitor activity, and watch your









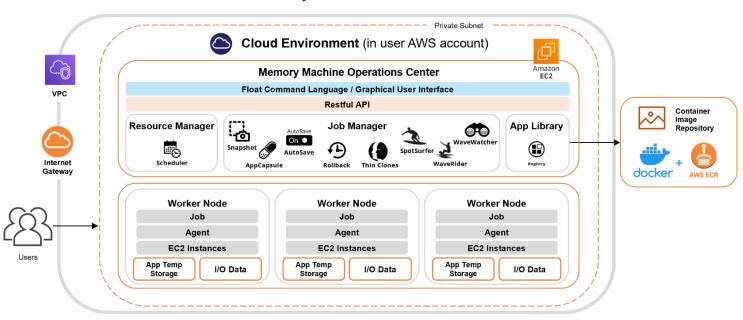
Platform for Job Automation on the Cloud

Memory Machine Cloud is a platform for automated job execution in the cloud, encompassing provisioning, orchestration, and dynamic resource optimization. At the foundation are MemVerge patented AppCapsule technology that allows applications to migrate across multiple VM instances, and the FLOAT command language that simplifies job management in the cloud.

How it works

- Memory Machine Operations Center, deployed directly into your AWS account (100% in Tenant), provisions
 and manages worker nodes that complete the execution of jobs. To ensure privacy, your data and results
 never leave your AWS account.
- Once your application is containerized with Docker or OCI they can be used by Memory Machine.
- The job submission process is fire-and-forget, fully automated from end-to-end.
- Once the job is running, Memory Machine dynamically adapts to the cloud resources used based on realtime workload consumption, and user-defined policies for snapshots, AppCapsules, SpotSurfer, WaveRider and other Job Manager functions.
- The App Library contains a registry of pre-qualified containerized application images. Customers can add to their private library.
- Ideal workloads include bioinformatics, genomics, and metagenomics analyses (Blast+, SPAdes, Seurat),
 MATLAB, Mathematica, NASTRAN, Rendering, Houdini, Maya, R, Python, batch Jobs, and long-running jobs.

Memory Machine Cloud in AWS



Users' set-up their jobs in the Operations Center from the Float command line or Operations Center GUI. Apps are loaded from the App Library or from external registries. Once a job is set-up, it can be redeployed with all its resources in just a few commands or clicks.









Memory Machine™ Cloud Innovation

MemVerge



Float

Command Language and GUI

Memory Machine Cloud is a platform for automated job execution in the cloud. It uses the Float command language, which is designed to manage the end-to-end lifecycle of jobs, define resource policies, and track and analyze job execution metrics. The Memory Machine Cloud <u>user guide</u> includes a complete list Float commands and explains how to use them.

Examples of Float Commands

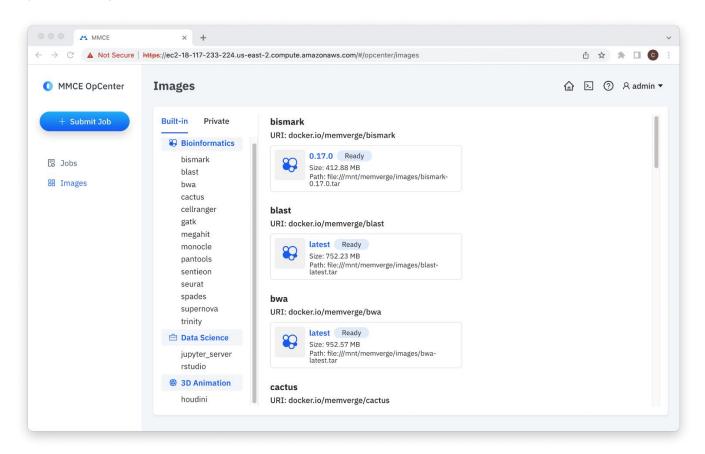
Float submit - This command and subcommands are used to submit a job with VM polices and WaveRider options for automatic migration to larger or smaller VMs.

Float migrate - Manually migrate a running job to a larger or smaller VM.

Float scancel - Manually cancel a running job before it completes.

Float Commands Expressed in a GUI

For customers who prefer to use a graphical user interface instead of a command line, most FLOAT commands can be expressed using a web-based interface.













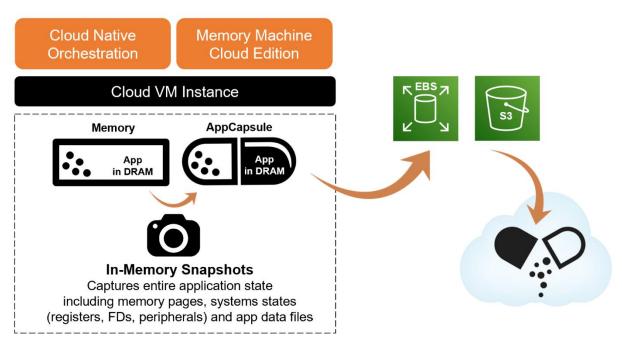
AppCapsule

A new type of snapshot that encapsulates the running states of the application, and that can be transported or "floated" to other servers

Storage snapshots save the states of data in the storage devices, but do not capture the states of memory pages, or systems states (registers, FDs, peripherals), that are crucial components of the running state of the application. As a result, the storage snapshots alone are insufficient to checkpoint and restore a running application.

AppCapsules are patented MemVerge technology and a new type of snapshot that encapsulates an entire application state and saves it in persistent storage. Memory Machine AppCapsules deliver checkpoint and restore services for stateful workloads to recover rapidly from a crash or an infrastructure interruption, to float workloads to larger and smaller compute instances, and soon, to float workloads across clouds.

Application checkpoint / restore service using AppCapsules



AppCapsules are generated and stored on persistent storage, such as EBS on AWS. Recovery is quick because the entire application state is restored from the AppCapsule.











SpotSurfer

Automated checkpointing and recovery service for Spot

The cloud, especially a Spot instance, is cost-effective for fault-tolerant, cloud native apps. Not necessarily so for thousands of stateful, long-running, non-fault-tolerant apps.

While spot instances can save up to 90%, there is one big catch. The cloud service provider can take them away any time they need the capacity. Customers get a notification ranging from only 30 seconds to 2 minutes before the instances disappear. The result is spot instances are not conducive to applications that cannot handle this kind of infrastructure volatility. Imagine a long-running app almost getting to the finish line but getting the spot eviction notification and having to re-start from the beginning. That's why cloud service providers recommend running only stateless or fault-tolerant apps on spot instances.

SpotSurfer is an application continuity service that uses the AppCapsule checkpointing and restore service + cloud automation to allow stateful workloads to run safely on Spot instances, saving up to 90% in compute cost.

It's simple. Register your app container in the Memory Machine Operations Center, submit your job and the SpotSurfer service takes it from there. After receiving a spot eviction notice from the Cloud, Memory Machine will generate an AppCapsule. It will then coordinate with the cloud service to allocate a new spot instance, restore the AppCapsule on the new instance and the workload will continue. The entire process is transparent to the application and fully automated.

How SpotSurfer Works

Submit job on Spot instance

Š.

Create AppCapsule snapshot



Spot Reclaimed





Coordinate with cloud services and schedulers



Restore AppCapsule on new Spot













WaveRider

Continuous Right Sizing Service

Another service that uses the AppCapsule checkpointing and restore service + cloud automation is called WaveRider. It continuously right-sizes workloads with short periods of peak usage. WaveRider automatically floats them to smaller and larger instances based on their real-time resource needs.

If you don't know the peak CPU or Memory usage of your workloads, you can choose policies for memory and CPU usage thresholds that will automatically trigger a migration.

How WaveRider works



AppCapsules floated from a smaller instance to larger instance after a period of peak usage, then back to a smaller instance when less resources are needed. When the job is complete, WaveRider terminates the instance to make sure your meter does not keep running.











App Library

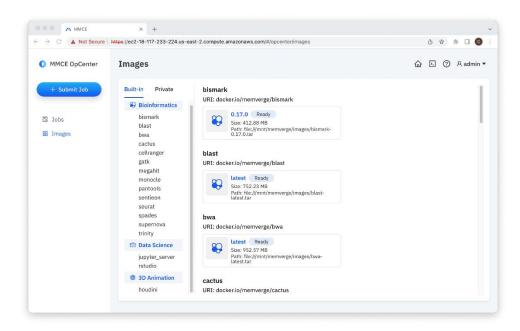
Qualified App Registry

Memory Machine Cloud edition makes deploying apps easier by making qualified apps accessible from the App Library. The growing list of apps in the library include Seurat, Cell Ranger, GATK, Sentieon, BWA, SPAdes, Megahit, Supernova, Blast+. You can create your own private library so your team can deploy apps with a simple Float command or from the Memory Machine GUI.

View of Memory Machine App Library from CLI

NAME	URI	TAGS	ACCESS USER
-spades	docker.io/lmwbio/metaspades	v3.15.5	
bismark	docker.io/memverge/bismark	0.17.0	opcenter
blast	docker.io/memverge/blast	latest	opcenter
bwa	docker.io/memverge/bwa	latest	opcenter
cactus	docker.io/memverge/cactus	latest	opcenter
cellranger	docker.io/memverge/cellranger	latest	opcenter
gatk	docker.io/memverge/gatk	v2	opcenter
houdini	docker.io/memverge/houdini	latest	opcenter
jupyter_server	docker.io/memverge/jupyter_server	v3.5.0, latest	opcenter
loadgen	docker.io/memverge/loadgen	latest	opcenter
megahit	docker.io/memverge/megahit	1.2.9	opcenter
monocle	docker.io/memverge/monocle	latest	opcenter
pantools	docker.io/memverge/pantools	v3.4	opcenter
python	docker.io/bitnami/python	latest	
rstudio	docker.io/memverge/rstudio	latest, 2022.12.0-353	opcenter
sentieon	docker.io/memverge/sentieon	latest	opcenter
seurat	docker.io/memverge/seurat	4.2.1	opcenter
spades	docker.io/memverge/spades	v3.15.5	opcenter
supernova	docker.io/memverge/supernova	2.1.1, 2.0.1	opcenter
trinity	docker.io/memverge/trinity	2.14.0	opcenter

View of Memory Machine App Library from GUI













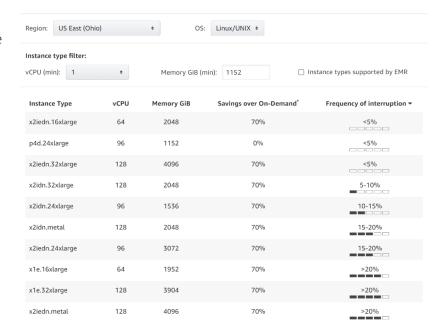
Use Cases

Control of your cloud is radical.



#1 Compute costs are high because big data, long-running workloads can't take advantage of low-cost Spot instances

The chart on the right shows the Spot savings over On-Demand for large memory instances are high, but so are the frequency of interruption. With customers getting a notification ranging from only 30 seconds to 2 minutes, that's not enough time for big memory apps to save their application state so they can recover quickly and gracefully. For long-running apps, it could mean almost getting to the finish line, but having to restart from the beginning.



The Memory Machine SpotSurfer service allowed this Cell Ranger workload to run safely on Spot instances and save 57%

Savings with SpotSurfer

A 6-hour Cell Ranger workload is run 100 times per week. The workload was run on m6i.8xlarge on-demand and Spot EC2 instances which included 32 CPUs and 128GB of RAM.

After factoring in the cost of Memory Machine with its SpotSurfer service, the move to Spot resulted in a savings of 57%.

EC2 Instance: m6i.8xlarge	On-Demand	Spot	Spot w/ SpotSurfer	Savings
Cost Per Week	\$921.60	\$216.54 - \$433.08	\$392.81	-57%
Completion Time	600 Hours	600-1200 Hours*	600 Hours	-











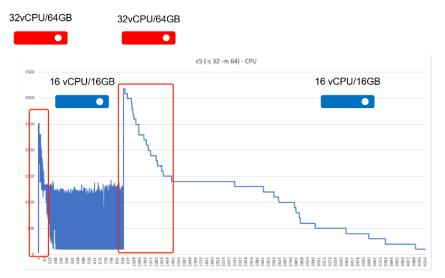
#2 - Compute costs are high because large compute instances are needed for short periods of peak resource usage

Basic Local Alignment Search Tool (BLAST) finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance. The BLAST workload of this Memory Machine user had short periods of peak usage but required them to configure the whole run with expensive resources for those short periods.

The Memory Machine WaveRider feature automatically right-sizes cloud resources DURING the workload execution by floating the workload to the appropriate VM instances

AppCapsules floated from larger instances to smaller instances after a period of peak usage, then back to a larger instance when more resources are needed. When the job is complete, WaveRider terminated the instance to make sure your meter does not keep running.

In this example, some of the savings from moving to Spot were used to configure a more powerful compute instance. The results were a savings of 91% and a job completion time that was 31% faster.



BASELINE 32 vCPU 64 GB VM

Save money and accelerate completion time with WaveRider

Test cases	Time	Delta	Cost	Delta
OnDemand - ecs.c5.8xlarge (32 vCPU, 64GB)	11h 44m	Fastest	\$17.22	Most expensive
OnDemand - ecs.ic5.4xlarge (16 vCPU, 16GB)	18h 4m	53% slower	\$12.63	27% less expensive than 32vCPU
MMCE (Spot) ecs.c5.8xlarge	11h 44m	Fastest	\$1.93	89% less expensive than 32vCPU
MMCE (Spot) ecs.ic5.4xlarge	18h 4m	53% slower	\$1.48	91% less expensive than 32vCPU
MMCE (WaveRider) ecs.c5.8xlarge =>	12h 23m	5% slower than 32vCPU	\$1.17	93% less expensive than 32vCPU
ecs.c5.8xlarge => ecs.ic5.4xlarge		31% faster than 16vCPU		91% less expensive than 16vCPU







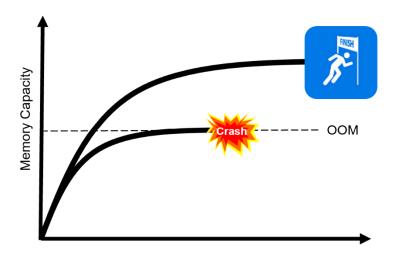




#3 – Long running jobs don't finish and must be re-started from the beginning

A common occurrence for big data analytic pipelines is Out of Memory (OOM) errors that crash the job. Jobs can run for days, crash near the end, and must be re-started from the beginning. This can be costly in terms of schedule or time and is always frustrating.

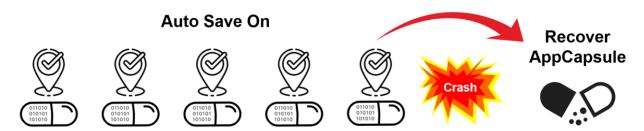
Out of Memory (OOM) errors often force long-running jobs to be re-started from the beginning



The Memory Machine AppCapsule feature in Auto Save mode allows long-running jobs to recover quickly and re-start from the last AppCapsule generated

Just set your AppCapsule policies for Auto Save, and the checkpointing service will happen automatically. In the event of a crash, your AppCapsule can be recovered automatically and re-started from a recent point-in-time. Never start from the beginning again.

Frequent AppCapsule Snapshots allow jobs to be recovered and re-started from a recent point-in-time.













#4 – Re-deploying apps and resources is complicated and takes a long time

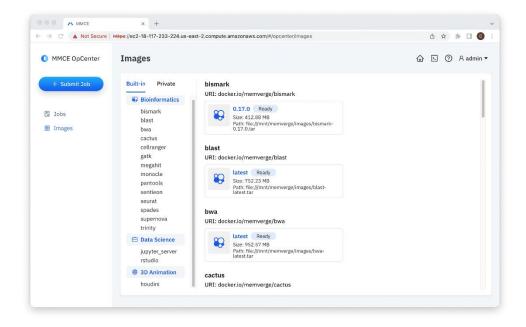
Setting an app for deployment in the cloud involves dozens of steps and a single entry done wrong can delay the deployment or result in a costly mistake. Repeating this process for multiple runs is inefficient which can be frustrating.

Just a few examples of cloud services and their resources that must be configured



Use Float commands and App Library to re-deploy apps with a few commands or clicks

Re-deploy your compute environment and app with a few clicks on the Memory Machine GUI.













Available in the AWS Marketplace

You can get Memory Machine Cloud today in the AWS Marketplace.

Get Started Free

<u>Request a demo</u> and a MemVerge expert will explain how Memory Machine works, then walk you through a demonstration that shows how easy it is to deploy and how powerful the results are.

Request a free trial and a MemVerge expert will set you up, or we can provide access for you to play in a Sandbox set-up in AWS.

Resources

Memory Machine Cloud Landing Page

Memory Machine Documentation Page

Memory Machine Cloud Installation Demo

Installing RStudio on Memory Machine Cloud Demo

Memory Machine Cloud Daily Operations Demo







