Release Notes for Memory Machine™ Cloud Edition 1.1

This document announces the release of Memory Machine Cloud Edition 1.1 and highlights key features.

Summary

Memory Machine Cloud Edition (CE) is a new software product from MemVerge that provides automated resource management, job scheduling, and workload mobility for cloud computing.

Based on user-supplied policy, Memory Machine CE instantiates cloud resources on behalf of the user. Memory Machine CE includes a built-in job scheduler, so users can deploy containerized applications across a group of virtual machines in a public cloud.

Memory Machine CE includes AppCapsule, MemVerge's innovative checkpoint/restore (C/R) capability. The AppCapsule is a moment-in-time snapshot of the application instance, including in-memory state and relevant files. AppCapsule can be used to support workload mobility and workload continuity. Workload mobility means that a job can be moved from one virtual machine to another, for example, to a more powerful virtual machine that is a better fit for the next stage of execution. Workload continuity is the automated version of workload mobility — if the underlying virtual machine is withdrawn, the workload automatically moves to a new virtual machine and resumes running, making the disruption transparent to the end-user.

Most of the time, hyperscale Cloud Service Providers (CSPs) have excess virtual machine capacity, which they offer as Spot Instances at varying discounts, sometimes as much as 90%. The downside is that any Spot Instance can be reclaimed with only nominal warning (typically, two minutes or less). Memory Machine CE's AppCapsule feature is triggered automatically when the CSP signals that it is reclaiming the Spot Instance. Job execution pauses and then resumes on a new Spot Instance, allowing users to take advantage of the deep discounts offered on Spot Instances without incurring the risk of losing intermediate results before the job has run to completion.

Supported Clouds

Although Memory Machine CE is designed to work on any cloud infrastructure, the initial release is supported on AWS and Alibaba Cloud only.

Features

Platform

- Self-service license registration via the Memory Machine CE license portal
- Memory Machine CE CLI for submitting and managing workloads, and for maintaining the Memory Machine CE platform

◦ One-click download and installation of CLI binary for Linux, macOS, or Windows clients

◦ Web-based console for invoking CLI commands without needing to download the CLI binary

- User and group management — add, update, or delete users or groups

- Global login session (eliminates per job login)

- On-line Memory Machine CE software upgrade using CLI commands

   ◦ Synchronization of CLI client version with Memory Machine CE version

- Support for multiple data volume types, including local file system mount, EBS volume, and remote NFS mount

Cloud Resource Management

- Per job virtual machine (VM) allocation policy

- Customizable VM allocation policy, for example, try Spot Instance first, then On-demand instance

- Price limits for allocating Spot Instances

- Downloadable usage and cost reports

Security

- Extra security group attached per job

- Limited authority for IAM roles to provide job insulation

- Validation of important CLI commands such as "float submit job"

- Double confirmation on critical CLI commands, for example, when upgrading software or canceling jobs

- Selection of virtual private cloud (VPC) in which to deploy Memory Machine CE (not limited to default VPC)

Container Images

- Built-in library of container images

- Image pull from popular container registries, such as DockerHub, GitHub Container Registry, quay.io (public repositories only), and Amazon Elastic Container Registry (public repositories only)

- Image upload from local server

Workload Management

- Runtime service configuration to customize execution environment

- Pre-defined environment variables for inclusion in job scripts

- Detailed information on running jobs available using CLI

- Real-time metrics for running jobs
- Autosave of stdout and stderr logs
- Customizable tags for each job, for example, for charge-back
- Port mapping between container and container host for interactive sessions
- Job migration between different VM configurations
- Automatic job migration for Spot Instances
- Support for Cromwell workflow manager
- Customizable filters to expedite job search
- Archival of completed jobs

Known Issues and Limitations

- Application memory footprint limitation for Spot Instance use case

  If the cloud provider signals that the Spot Instance will be reclaimed, the checkpoint image must be created and written to persistent storage before the Spot Instance terminates when the warning period expires. For AWS and AliCloud, this warning period is two minutes and five minutes, respectively. This limits the size of the application's in-memory state to 120 GB for AWS and 256 GB for AliCloud.
- Single node applications only — clustered applications are not supported.
- CPU-only VMs — VMs with GPUs or FPGAs are not supported.

Document Revisions

- Initial release: 12-01-2022
- Revised: 12-02-2022