ORACLE for Research + Carnegie Mellon University

Technical Brief

Carnegie Mellon University studies rodent brain activity to bolster neurological disease treatment

Mapping the neural underpinnings of behavior is crucial for both improving our knowledge of natural behaviour, and better understanding what might be happening in neural diseases. The research team from Carnegie Mellon University (CMU) led by Eric Yttri set out to determine and create a dataset answering this question, making use of the technology of Oracle Cloud Infrastructure (OCI).

Using the cloud for a scalable and flexible computing environment

This project team is no stranger to large computational workloads, and in past projects had relied on local clusters and on-prem solutions. However, they faced challenges when relying on local solutions such as being limited by the available software dependencies and having to wait in lengthy queues to run jobs. To make this ambitious project work, Eric and team turned to the cloud as it offered a customizable environment and a flexible, on-demand access to computational resources.

To develop a dataset mapping neural underpinning of observed behaviour, the team would have to process large amounts of neural and video data. Both required GPUs, and each 24 hours of data generated around 3TB of videos.

For GPU options within OCI, they turned to virtual machine compute shapes with NVIDIA V100s. Using multiple *VM.GPU3.1* (which has a single V100 GPU and 16GB GPU Memory) and *VM.GPU3.2* (with two V100 GPU with 32GB GPU Memory) enabled their processing to keep up with their data collection, and do so in a cost-effective way. This was important as the team didn't want to have pay for unused-GPUs on larger shapes, nor did they want to waste unnecessary time waiting for runs to complete.

To handle the sizable amount of data, Block Volumes holding both the input data and processed output were created and attached directly to the GPU instances. Keeping with the flexibility of the cloud, block volumes can have a customized (and editable) size and attachment performance. They can be formatted into filesystems familiar to anyone whose worked with attached storage on-prem, and therefore be immediately useful when lifting their software and workloads to OCI.

CMU's cloud migration path

Standing up a processing pipeline within OCI was largely a lift-and-shift exercise; that is to say, they were moving familiar tools/software to run in a cloud infrastructure environment.

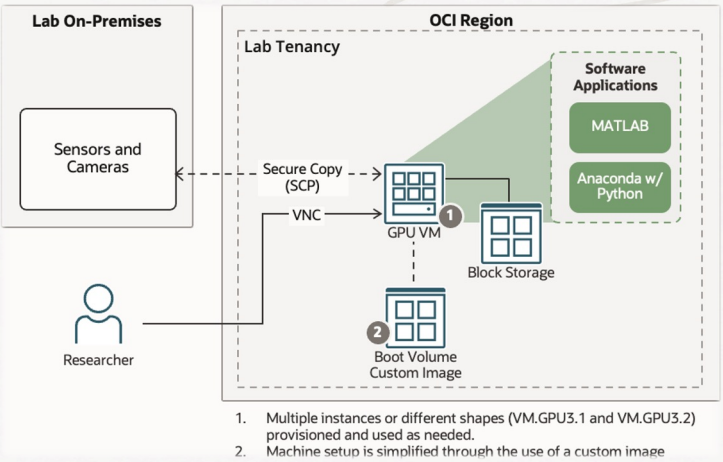
The pipeline comprised of MATLAB and Python through Anaconda applications, and we're to be run on Linux-based GPU compute instances. To make this seamless and get up and running faster, the team turned to Virtual Network Computing (VNC for short). VNC allows for remote operation of instances through a graphical user interface. This means that rather than relying on the out-of-the-box command line interface and adapting their workload to fit, the team was able to work with MATLAB and Anaconda just as they do locally.

Because data was processed when it was collected, the team opted to spin up and down GPU instances on-demand and when needed. To prevent having to setup instances anew each time, they took advance of Custom Images. In this pattern, an image of a instance's boot volume is captured *after* the machine is first setup with all necessary customizations, configuration and software. This image is then used as a 'base' for future instances, saving any repeating of setup steps.

With compute and applications setup in the cloud, all that remained was uploading data collected in the lab. As the team was using block volumes, the SCP (secure copy) utility made sense as it 1) has a variety of local clients supporting command-line, GUI and various operating systems and 2) is easily supported by the remote instance. For any project looking to upload data, it's worth pointing out that there is no one-size-fits-all solution to storage and upload/download, and other options are given here.

ORACLE
for Research

OCI Solution Architecture



The OCI services in use are:

- Oracle Cloud Infrastructure
- Compute
- Storage
- Networking
- Security, Identity, Compliance

Capturing terabytes daily with the cloud

With this setup in place the Yttri lab was able to successful run their research program, analyzing multiple days of video and corresponding neural data to generate a novel dataset with high utility.

Their success was made possible by having full configurability over their compute environment, easy setup, quick scaling up and down to meet requirements by using custom images and finally ample and flexible space for storage.

oracle.com/research