By Andrew Long (Andrew.Long@pgs.com)

Building a Supercomputer in the Cloud with Google Kubernetes

PGS Chief Enterprise Architect, Louis Bailleul, recently presented at the Google Cloud Next '22 event held in Munich. A related YouTube interview was given on the PGS journey into cloud-based supercomputing with Maciek Różacki, Senior Project Manager, GKE, Google Cloud. Louis discusses the best-practice execution of batch and high-performance computing workloads on Google Kubernetes Engine (GKE) and how PGS used these to replace their 202 000 core Cray supercomputers in Houston. When used at peak capability, the GKE supercomputer running in the Google Cloud Platform (GCP) has a hypothetical 72.02 petaFLOPS capability, equivalent to the 7th largest computer in global rankings. Elements are discussed below of the "move and improve" journey PGS has undertaken from on-premise Cray supercomputers to the GKE supercomputer available today in the cloud.

Our HPC Journey from "Lift and Shift" to "Move and Improve"

Until they were fully decommissioned in May of 2022, PGS owned two Cray XC40 and XC30 supercomputers in Houston with a combined 202 000 CPU cores and access to 7.47 petaflops of peak performance. In June 2015, the XC40 Cray ranked 12 in the global Top 500 list of supercomputers. These platforms, representing "on-premise" resources, also included 12 petabytes of specialized Lustre file systems with the ability to offer up to 257 GB/s of throughput.

PGS used to rely on its Cray supercomputers for about 80% of its total computing capacity, but high-performance computing (HPC) hardware ages quickly. Data size and complexity is constantly growing, and supercomputers are already reaching obsolescence at 4-5 years of age. In 2018, we decided to investigate using the cloud as an extension to on-premise capacity to reduce the future on-premise investments. Correspondingly, our initial plan was to continue using on-premise for the maximum average utilization and use the cloud for excess spikes in computing demand. An agreement was made to adopt the Google Cloud Platform (GCP), and a "lift and shift" was planned for whatever was running on-premise into GCP.

A combination of industry and global economic factors forced a revision of our plans in 2020, and the heavy recurring CAPEX of on-premise computing was no longer feasible. Backed by a couple of years proof-of-concept and prototype successes in the GCP environment, PGS decided that instead of renewing our-premise resources we would instead move the workload into the cloud.

This changed our strategy from "lift and shift" of 20% of our capacity to "move and improve" of 80% of our capacity: necessitating a full re-engineering of our "HPC stack" (the software environment) to be truly cloud-native. The first minimum viable product (MVP) went live in Q1 2021 and by May 2022 we decommissioned the last Cray supercomputer and we had 80% of all the compute running in GCP.

How did we do it? We turned Google Kubernetes Engine (GKE) into a supercomputer.

Kubernetes: The Conductor for an Orchestra of Software Containers in the Cloud

As defined in the **Glossary** at the end of this newsletter, Kubernetes in an open-source software developed by Google that can automatically manage large numbers of server computers and associated programs at the scale of on-premise supercomputers. Referred to as "workloads", applications within Kubernetes run on clusters that can be a combination of both physical and virtual machines. An "instance" is a virtual machine (VM) hosted within the Google Compute Engine (GCE) resource that uses software instead of a physical computer to run programs and deploy apps. A flexible model of VMs with different availability-performance-price settings can be configured to optimize costs during different demand cycles.

On Kubernetes, all programs run in “containers”; packages of software that contain all of the necessary elements to run in any environment. Overall, the Google Kubernetes Engine (GKE) consists of multiple GCE instances grouped together to form a cluster that can support up to 15 000 nodes.

Rewriting All the Rules

A notable challenge for the PGS developers was to seamlessly integrate the existing on-premise infrastructure and the new GCP environment. Services were built to transfer data from the Lustre filesystem on-premise into GCS using a 40 Gbps interconnect between the main PGS Houston data centre premises and the GCP us-central1 colocation facility. Services were also built to bridge the gap between the cloud and the legacy workflow for the user on-premise: Mainly related to how all the users were submitting jobs and how that are now run in GKE. Within GKE, we leverage node pools in order to run all the workloads and we mostly use preemptible VMs in order to reduce the average cost. Storage was converted to Google Cloud Services (GCS) because Lustre or other parallel filesystems cannot be cost-effectively run at the necessary scales. As the workloads use preemptible nodepools for the processing phase, they store checkpoints on Google Cloud Services (GCS) buckets so they can restart work where they left when they are preempted. We also leverage smaller on-demand node pools for tasks that don’t support preemption.

To meet our ambitions, we have to ability to use up to 12 GKE clusters with their maximum capacity of 15 000 instances (i.e., up to 180 000 instances). Using the n2d-standard-16 virtual machine type (the most common type of instance we are using), this would equate to over 2.8 million virtual CPUs (vCPUs), with 4 GB of system memory per vCPU, and a total of 72.02 petaFLOPS capability (1 petaFLOP = 10^{15} floating point operations per second). The comparison of the fastest supercomputers uses complex metrics to establish rankings, but according to the global Top 500 table based upon the LINPACK Benchmark, at peak capacity this would place 7th on the global table (refer also to **Figure 1**). The actual peak number of instances used today is 52 782, which equates to about 31 petaFLOPS (a mix of mostly n2d-standard-16, n2d-standard-32, and n2d-standard-64, with some n2d-highcpu-32 and n2d-highcpu-64). That would put PGS in the Top 13 in the current list of supercomputers (this is an improvement since the video recorded in August 2022 that stated a figure of 23 petaFLOPS). Storage transfer into GCS is about 500 GB/s.

A 7x capacity increase

	Maximum	Peak
GKE nodes	180000	52782
Total vcpu	2.88 Millions[1]	1,231,385[2]
Performance	72.02 PFlops[1]	~30.79 PFlops[2]
Top500 (June 22)	Top 7	Top 13
GCS bandwidth	1.2 TBps	503 GBps

[1] using n2d-standard-16
[2] using mixed n2d-standard-(16,32,64)

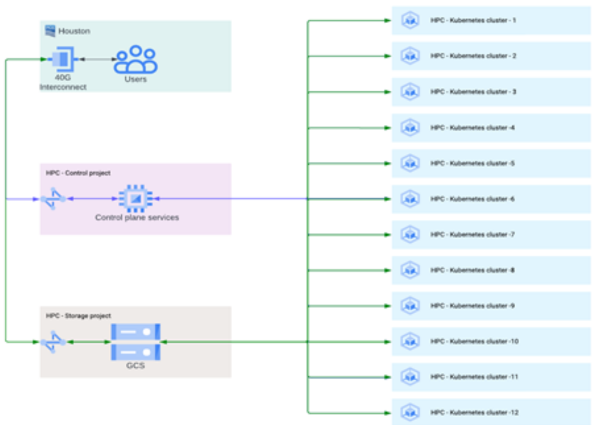


Figure 1. Peak capabilities today for the GKE supercomputer available to PGS.

Finally, the Cray supercomputers were consuming about 3MW of power, so the 97% carbon-free energy powering us-central1 has significant sustainability merits.

Figure 2 shows a recent significant improvement in turnaround, where PGS was able to scale up to 53 417 instances. This workload that would have taken weeks on-premise could be run in days. **Figure 3** shows the preemption rate over a four-day period, where PGS was able to scale up to about 52 782 instances. As noted in the **Glossary**, pre-emption events interrupt jobs and forces them to restart from the last checkpoint. Despite having a fairly constant rate of 500 pre-emptions per minute, and with spikes at 2 000 per minute, the workflows progress smoothly and complete successfully—and significantly faster than would have occurred on-premise.

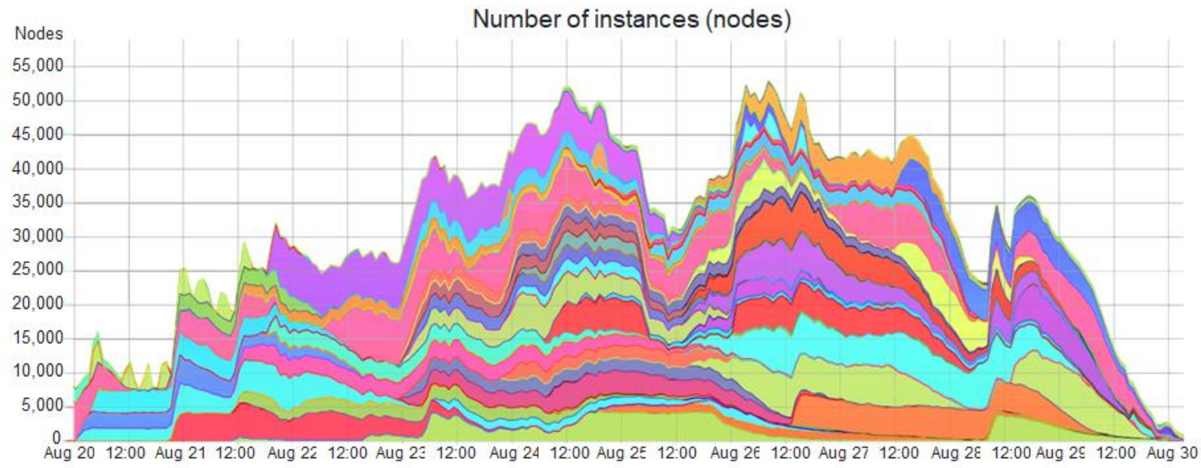


*Figure 2. A large workload run over 10 days with up to 53 417 instances. The 12 colors represent the 12 GKE clusters in use. By comparison to **Figure 3**, this peaked at about 855 536 vCPUs because a larger proportion of instances had a low vCPU count*.

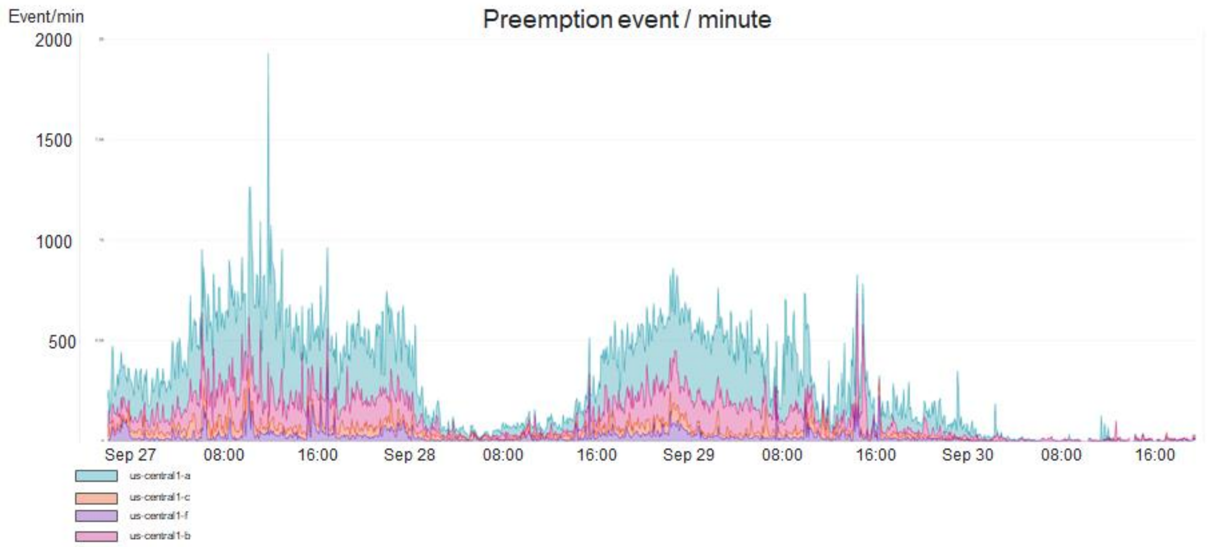
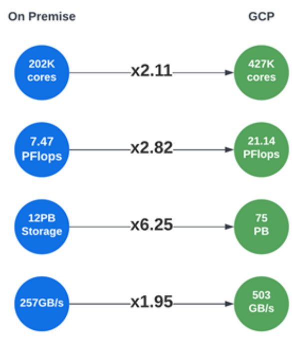


*Figure 3. Preemption rate over a four-day period demonstrates how robust and fault-tolerant the workloads are. This workload peaked at about 52 782 instances and 1 231 385 vCPUs. A large proportion of instances had a higher vCPU count than the example in **Figure 2***.

Figure 4. Milestones achieved to date by moving from on-premise to GCP: number of cores, petaFLOPS, storage, and bandwidth.



Summary

Overall, PGS and GCP have managed to meet all project objectives, with further improvements in performance and capacity expected in future. Adapting for the cloud is the only way to run cost efficiently but it is a significant engineering undertaking. We are running our high-end business without having to reinvest in on-premises hardware, and continually passing new milestones in workload instances and capacity.

Our transition to the cloud began with the most compute-intensive technologies, such as FWI (full waveform inversion) and RTM (reverse time migration). Efforts are now underway to complete the migration of the complete PGS software stack to the GKE environment.

Acknowledgements

The entire PGS Eos team (named after the Greek goddess of dawn and new beginnings) are thanked for bringing this project to fruition, and our partners at Google Cloud. Thanks also to Louis Bailleul for his assistance with this newsletter.

Glossary

This section draws heavily on Google Cloud Platform (GCP) resources to define the terminology used in this article.

Virtual Machine (VM): A Google Compute Engine (GCE) resource that uses software instead of a physical computer to run programs and deploy apps. One or more virtual “guest” machines run on a physical “host” machine. Each virtual machine runs its own operating system and functions separately from the other VMs, even when they are all running on the same host.

An instance is a virtual machine hosted on Google's infrastructure.

A Preemptible VM (PVM) and Spot VMs are GCE VM instances that are heavily discounted (up to 91%) as long as the customer accepts that the instance can be preempted at any time. GCE might preempt PVM/Spot VMs if it needs to reclaim those resources for other tasks. PVMs/Spot VMs are excess capacity, so their availability varies based on GCE usage. Spot VMs do not have a minimum or maximum runtime whereas PVMs have a maximum runtime of 24hrs. PVMs/Spot VMs are able to cost-effectively complete batch processing tasks provided the workload is fault tolerant.

Kubernetes is software first developed by a team at Google, and later donated to the open-source Cloud Native Computing Foundation (CNCF). It manages many server computers and runs a large number of programs across those computers (01:21 mins:secs movie). On Kubernetes, all programs run in “containers” so that they can be isolated from each other and be easy to develop and deploy.”

Containers are packages of software that contain all the necessary elements to run in any environment. In this way, containers isolate applications and dependencies allowing to run anywhere, from a private facility to the public cloud. Containers are granted shares of CPU, memory, storage, and network resources at the operating system level, providing developers with a view of their application logically isolated from other applications.

Pods are the smallest deployable units of computing that you can create and manage in Kubernetes. A Pod is a group of one or more containers, with shared storage and network resources, and a specification for how to run the containers.

A Workload is an application running on Kubernetes. Whether the workload is a single component or several that work together, on Kubernetes it runs inside a set of pods. Kubernetes apps run on a cluster, which is made up of at least one control plane and machines called Nodes. A node may be a virtual or physical machine, depending on the cluster. Each node contains the services necessary to run Pods. There are typically several nodes in a cluster, although some learning or resource-limited environments might have only one node.

Indeed, a Cluster is the foundation of Google Kubernetes Engine (GKE).

When using GKE, Nodes run the services supporting the containers that make up your workloads. The control plane then decides what runs on those nodes, including scheduling and scaling. The GKE environment consists of multiple GCE instances grouped together to form a cluster. GKE is a solution to automatically deploy, scale, and manage Kubernetes, and supports up to 15 000 nodes in a single cluster.

Further Reading

- Kubernetes on Vessels, with Louis Bailleul (2022): Kubernetes Podcast with Google, Episode #194.