## PRP PACIFIC RESEARCH PLATFORM

# ICECUBE Nautilus, ICECUBE IceCube and LIGO

by Igor Sfiligoi, UCSD/SDSC isfiligoi@sdsc.edu UC San Diego

Nautilus



#### Nautilus



- This being GRP, I am assuming you all know what Nautilus is!
- Just for offline reading, I mostly refer to the distributed Kubernetes cluster
  - And all the other great supporting services around it

https://nautilus.optiputer.net



#### IceCube

- IceCube is a Neutrino experiment at the South Pole
  - Using natural ice as a detector media
- Premier compute application is simulation workload running photon propagation on GPUs
  - Essential for proper calibration of the detector



#### https://icecube.wisc.edu





#### LIGO



Gravitational Wave Observatories

1 sec. time observable by LIGO

Operationa Under Con Planned

••••••

- The Laser Interferometer Gravitational-Wave Observatory (LIGO) is mainly designed to detect gravitational-waves
  - Through the use of laser interferometry
  - Several detectors built to aid with detection confirmation
- LIGO has a high noise to signal ratio

https://www.ligo.caltech.edu

- Requiring significant compute power to filter it out
- Some parameter fitting workloads (e.g. RIFT) great match for GPUs

#### Focus: RIFT

The Global Research Platform Workshop.

- Likelihood eval: timeseries manipulations & matrix operations: numpy → cupy
- CPU likelihood eval: 407s
  - GPU likelihood eval: 21s

#### **Using OSG portal into Nautilus**

#### **Open Science Grid**

- IceCube and LIGO both see Nautilus just as another Open Science Grid (OSG) resource
  - Nothing Nautilus/PRP/TNRP specific in their setup
- Delegated trust
  - Nautilus trusts "the OSG personnel" and gives resources to "OSG"
  - OSG then deals with authenticating, authorizing and allocating resources to IceCube and LIGO
- OSG deployed three layers as Kubernetes pods to make this work
  - A HTCondor batch system, with all execute nodes being purely opportunistic
  - A CVMFS (CSI) driver to allow for efficient and transparent on-demand software distribution
  - A OSG portal to the HTCondor batch system, known as a HTCondor-CE
  - The rest of the OSG infrastructure then treats Nautilus/PRP as a "normal Grid site"

#### **Opportunistic use**

- IceCube and LIGO run purely opportunistically
  - i.e. only if no regular users are requesting those resources
- They demand nor expect no guaranteed time from Nautilus
  - Worker nodes (i.e. pods) can, and will be evicted without any advance warning
  - OSG infrastructure helps them seamlessly recover
- They thus can recover most of the compute cycles that would else go to waste
  - Minus a small amount wasted due to preemption
  - Including unused interactive-focused resources, like the SunCave



#### **Opportunistic GPU use**

IceCube and LIGO • the largest users of Nautilus GPUs

> **Being flexible** • pays off



250

100

50

https://grafana.nautilus.optiputer.net/d/dRG9q0Ymz/ k8s-compute-resources-namespace-gpus ?orgId=1&var-namespace=osggpus



#### **Opportunistic CPU use**

 Opened access to CPUs more recently

https://grafana.nautilus.optiputer.net/d/KMsJWWPiz/cluster-usage

 IceCube in particular made good use of them

Vautilus





#### **CPUs mostly for IceCube**

- Almost all of the CPU cycles went to IceCube
  - Only a minor fraction went to LIGO
- PRP accounted for about 10% of total IceCube CPU capacity
  - The largest in the US, but still small compared to EU contributions
  - Note that it was larger than NERSC



SDSC-PRP

SDSC-PRP: 2.069 Mil (8.95%)

NERSC

WIPAC

Cedar

2.069 Mil

1.804 Mil

1.513 Mil

1.495 Mil

1.161 Mil 983 K

Open Science Gri https://gracc.opensciencegrid.org/dashboard/db/payload-jobs-summary?orgId=1&var-Facility=SDSC-PRP The Global Research Platform Workshop, Sept 2019

### **Multi-level containerization**

- As most OSG users, IceCube and LIGO run their workloads inside containers
- But OSG is running "generic HTCondor worker node pods" in Nautilus
  - How do we launch another container from inside a running container?
  - Can we do it without requiring any privileges?
- Turns out, it is quite trivial with recent Linux Kernels (4.15+)
  - Singularity can run from inside an (unprivileged) Docker container like in Nautilus k8s
  - Most Nautilus nodes have the updated Kernel
- Note: Singularity recently fixed a bug that would prevent the use of GPU containers



**HTCondor** 

Singularity

**User jobs** 

#### **Data and Networking**

- Open Science Grid (OSG) operates a Content Delivery Network (CDN) inside Nautilus
  - Called StashCache
  - Based on xrootd technology
- Data Caches strategically positioned close to major compute centers
  - Clients use GeoIP to pick the closest one
- Greatly reduces latency and origin throughput for applications making repeat use of large inputs

- IceCube and LIGO jobs inside Nautilus do not use the OSG StashCache CDN
  - Each job has unique input data
- Many LIGO CPU jobs in both OSG and other Grid infrastructures do make extensive use of the CDN
  - But current CVMFS k8s setup does not play well with LIGO's authenticated use-case
  - One of the reasons for LIGO's modest use of Nautilus CPUs

Please come see my poster this afternoon, too.

#### IceCube and CEPH

- OSG is currently operating a gridFTP server into CEPH for IceCube
  - Used for handling input and output data for jobs running on Nautilus/PRP
- Currently intended only as a temporary solution
  - Need arisen due to IceCube's internal storage procurement delays
  - Currently using about 65TB one of the largest users
- Has been performing to our satisfaction
  - Even though it is fronted by a single gridFTP server pod at UCSD

1			
•	Mame edex	<sup>size</sup> 86.26 TB	
r i	Name cecube	<sup>size</sup> 65.98 TB	
1	Name Becewcsng	<sup>size</sup> 26.7 TB	
1	Name Srip19-pointcloud	<sup>size</sup> 24.51 TB	
ľ	Name ucsd-haosulab	<sup>size</sup> 20.38 TB	
P r	Name mas-medical	<sup>size</sup> 19.73 TB	
P N	Name Viscompfs	<sup>size</sup> 11.95 TB	
ľ	Name ucicompvis	<sup>Size</sup> 10.54 TB	
P C	Name Shei-ml	<sup>size</sup> 9.97 TB	
			1 - T

#### Acknowledgements

- Most of the work going in this presentation has been funded through NSF grants OAC-1826967 and OAC-1841487.
- I also kindly thank
  - The NSF and other funding agencies for their support for IceCube and LIGO.
  - The support from Internet2 in hosting and operating several of the nodes comprising the OSG CDN.
  - The NSF for all the other grants funding the infrastructure and services needed for Nautilus and the PRP.
  - The various other funding agencies that are contributing hardware and effort to Nautilus.