Ceph Day for Research and Non-profits

Tuesday 17 September 2019 - Tuesday 17 September 2019

CERN

Book of Abstracts

Contents

1

Welcome and Introduction

9

Ad-hoc filesystems for dynamic science workloads

Authors: Stig Telfer[1]; John Garbutt[1]

[1] *StackHPC Ltd*

In this talk, we present recent work supporting the computing demands of the Euclid space mission using resources from across an OpenStack federation for scientific computing. CephFS is used to present a single coherent filesystem drawing upon resources from multiple sites. We present our approach and experiences.

We then present an alternative approach to filesystems on-the-fly using the Data Accelerator project at Cambridge University, currently #1 in the global IO-500 list. We provide an overview of the technologies involved and an analysis of how its high performance levels are achieved.

10

Ceph Supporting Genetic Research at The Wellcome Sanger Institute

Author: Matthew Vernon[1]

[1] *Wellcome Sanger Institute*

The Wellcome Sanger Institute has 18PB in its largest Ceph cluster. This talk will explain how the Sanger used Ceph to build and scale a reliable platform for scientific computing, and enable secure data sharing via S3. And how they got 100GB/s read performance out of their cluster.

Matthew will outline the interesting aspects of the Sanger's Ceph setup, including how the team grew it from a small initial installation, automated deployment management and monitoring, and some of the issues they have encountered along the way. Matthew will also explore some of the good (and less good!) aspects of running Ceph at scale, and supporting scientific workflows.

11

Scale out Sync & Share with Seafile on Ceph

Author: Sönke Schippmann[1]

Co-author: Stefan Wichmann [1]

[1] *Universität Bremen*

Seafile provides an open source solution for sync & share services like ownCloud, but with a much better performance and lower hardware needs. Using Ceph as a S3 storage backend a highly available sync & share cluster can easily be set up.

The talk will focus on practical tips for the implementation, on the fly migration to Ceph for existing installations, and especially on backup and restore scenarios on the multi terabyte scale.

12

MeerKAT Ceph updates from Mzanzi

Author: Thomas Bennett[1]

Co-authors: Martin Slabber [1]; Sean February [1]; Simon Ratcliffe [1]

[1] *SARAO*

MeerKAT, one of the SKA (Square Kilometer Array) precursor telescopes, was inaugurated on the 13th of July 2018 in South Africa. We would like to update the Ceph community with progress and activities relating to the MeerKAT project with a particular focus on MeerKAT data storage.

A number of Ceph RADOS Gateway instances have been implemented for MeerKAT. We will present these use-cases, their current configurations and implementations. We will also discuss the development of bespoke software stacks for data transfer and an end user data access layer.

After two years of using Ceph, the MeerKAT data storage team is also reflecting on what we have learned and where we should focus our efforts w.r.t. Ceph and the Ceph community.

Since our first production cluster, we have been using ceph-ansible for deployment, which has suited or needs. However, looking forward, we have begun developing our own Ceph deployment process.

We have also been through a number of iterations of monitoring and alerting infrastructure for our Ceph production clusters. Our current efforts have been to use a stripped down version of the ceph-metrics for our Prometheus driven Grafana dashboards.

We are also in the process of growing our small Ceph community in South African. Currently this is being driven through meetup events, discussion forums and presentations at local workshops and conferences. Leveraging the perception of MeerKAT we are hoping to reach a wider audience and raise awareness of Ceph.

13

Ceph in Compute Canada

Author: Mike Cave[1]

[1] *University of Victoria*

Compute Canada is the national platform for research computing in Canada. There are five high performance research computing sites across the country offering both traditional HPC and OpenStack cloud resources.

This talk will give an overview of Ceph at the cloud sites and then focus on the specific implementation details of Ceph at, Arbutus. Hosted at the University of Victoria in British Columbia, Canada, Arbutus is the largest non-commercial research cloud in Canada.

Ceph is integral to the success of our cloud deployments because of the versatility, price for performance and scalability. We started with a small 400TB Ceph install which has grown to a 5.3 PB

installation, and in the near future are extending our offering to include CephFS and object storage.

14

CephFS in an HTC cluster and VMs on Ceph RBD with TRIM and differential backups in Bonn

Author: Oliver Freyermuth[1]

Co-author: Peter Wienemann [1]

[1] *University of Bonn (DE)*

CephFS is used as the shared file system of the HTC cluster for physicists of various fields in Bonn since beginning of 2018.
The cluster uses IP over InfiniBand. High performance for sequential reads is achieved even though erasure coding and on-the-fly compression are employed.
CephFS is complemented by a CernVM-FS for software packages and containers which come with many small files.
Operational experience with CephFS and exporting it via NFS Ganesha to users' desktop machines, upgrade experiences, and design decisions e.g. concerning the quota setup will be presented.

Additionally, Ceph RBD is used as backend for a libvirt/KVM based virtualisation infrastructure operated by two institutes replicated across multiple buildings.
Backups are performed via regular snapshots which allows for differential backups using open-source tools to an external backup storage.
Via file system trimming through VirtIO-SCSI and compression of the backups, significant storage is saved.
Writeback caching allows to achieve sufficient performance. The system has been tested for resilience in various possible failure scenarios.

17

CephFS: looking for the Swiss Army knife of POSIX filesystems

Author: Mattia Belluco[1]

Co-authors: Enrico Favero [2]; Filippo Stenico [2]

[1] *Univerisity of Zurich*

[2] *University of Zurich*

At the University of Zurich we strive to offer our researchers the best solutions to store and access their data. Last year we deployed a new Ceph cluster exclusively devoted to CephFS to replace both the traditional NFS boxes and the RBD-images-exported-over-NFS ones. The ultimate goal is to use CephFS everywhere POSIX compatibility is required, including in our (small) HPC cluster instead of a traditional parallel filesystem.
We will share the benchmarks we took and the bumps we hit during the journey, navigating between releases with different maturity levels, experimental features, and performance hiccups.

18

Utilising Ceph for large scale, high throughput storage to support the LHC experiments

Author: Tom Byrne[1]

[1] *STFC*

Our large erasure coded ceph cluster is used by the four large LHC experiments for scientific data storage, providing 30PB of usable storage and averaging a 30GB/s read rate to the analysis cluster. In this talk I will talk about the architecture of the system, and how we have optimised it to allow us to reliably support a large transfer rate. I will also discuss some of the issues, and solutions, surrounding transfer performance monitoring in our architecture.

19

Ceph storage for openstack in a security context

Author: Etienne Chabrerie[None]

"Minister of the Interior" (France) has implemented a cloud for internal customers in a complex security environment.
Our specific activity requires to have a scalable, reliable, and highly available storage with moderate operating expenses.
A private Openstack Cloud has been deployed 2 years ago, and more and more internal customers are interested in using it, consequently increasing the cpu, memory and storage usages.
So far, SAN storage was used for instances volumes, and scalability was complicated. Swift object storage was also hard to extend with this implementation.

Our objective was to implement more scalable storage system with higher performance for the MI cloud along with getting a better monitoring.
To achieve this, 2 types of ceph clusters were defined:
- one for block storage dedicated for openstack instances
- one for object storage with 2 services: swift and s3
An automated deployment method has been designed with cobbler, ansible, salt, jenkins.
The Support team is in charge of the system's commissioning and maintaining in operational condition

To conclude, the ceph solution provides a full compability with openstack, and offers s3 service (like amazon s3). For robust high-availability, the described architecture works also on a multi-site environment with asynchronous replication, which we are using today.

20

Ceph at the Flatiron Institute

Author: Andras Pataki[1]

[1] *Flatiron Institute*

The Flatiron Institute, a division of the Simons Foundation, is a privately funded non-profit organization in Manhattan with a mission to advance scientific knowledge via computational methods.

Operating in a variety of disciplines, from astrophysics to biology, quantum physics and mathematics, the breadth of computational problems our researchers tackle present unique challenges to our infrastructure. We are early adopters of Ceph and CephFS from the Hammer days, and now run close to 30PB of Ceph storage that serves our HPC environment. The open source development model of Ceph enabled us to make customizations and apply patches both for early fixes as well as for custom enhancements specific to our environment. This talk will give an overview of our over four year journey with Ceph, highlighting choices we made for our setup, the unique issues we face, some of the tools/patches we are working on for our environment and disasters that Ceph successfully saved us from over the years.

21

Storage for High Energy Physics

Author: Andreas Joachim Peters[1]

[1] *CERN*

Welcome to CERN and summary talk about data storage in high energy physics.

22

Applications of Ceph @ CERN

Authors: Dan van der Ster[1]; Jakob Blomer[1]; Roberto Valverde Cameselle[1]; Theofilos Mouratidis[1]

[1] *CERN*

23

Ceph at NASA Atmosphere SIPS

Author: Kevin Hrpcek[1]

[1] *NASA*

The NASA VIIRS Atmosphere SIPS, located at the University of Wisconsin, is responsible for assisting the Science Team in algorithm development and production of VIIRS Level-2 Cloud and Aerosol products. To facilitate algorithm development, the SIPS requires access to multiple years of satellite data occupying petabytes of space. Being able to reprocess the entire mission and provide validation results back to the Science Team in a rapid fashion is critical for algorithm development. In addition to reprocessing the SIPS is responsible for the timely delivery of near real time satellite products to NASA. To accomplish this task the Atmosphere SIPS has deployed a seven petabyte Ceph cluster employing numerous different components such as librados, EC-Pools, RBD, and CephFS. This talk will discuss choices we made to optimize the system allowing for rapid reprocessing of years of satellite data.

27

Ceph Upstream News

28

Ceph Community Talk

29

Speakers AMA

Ask the speakers anything