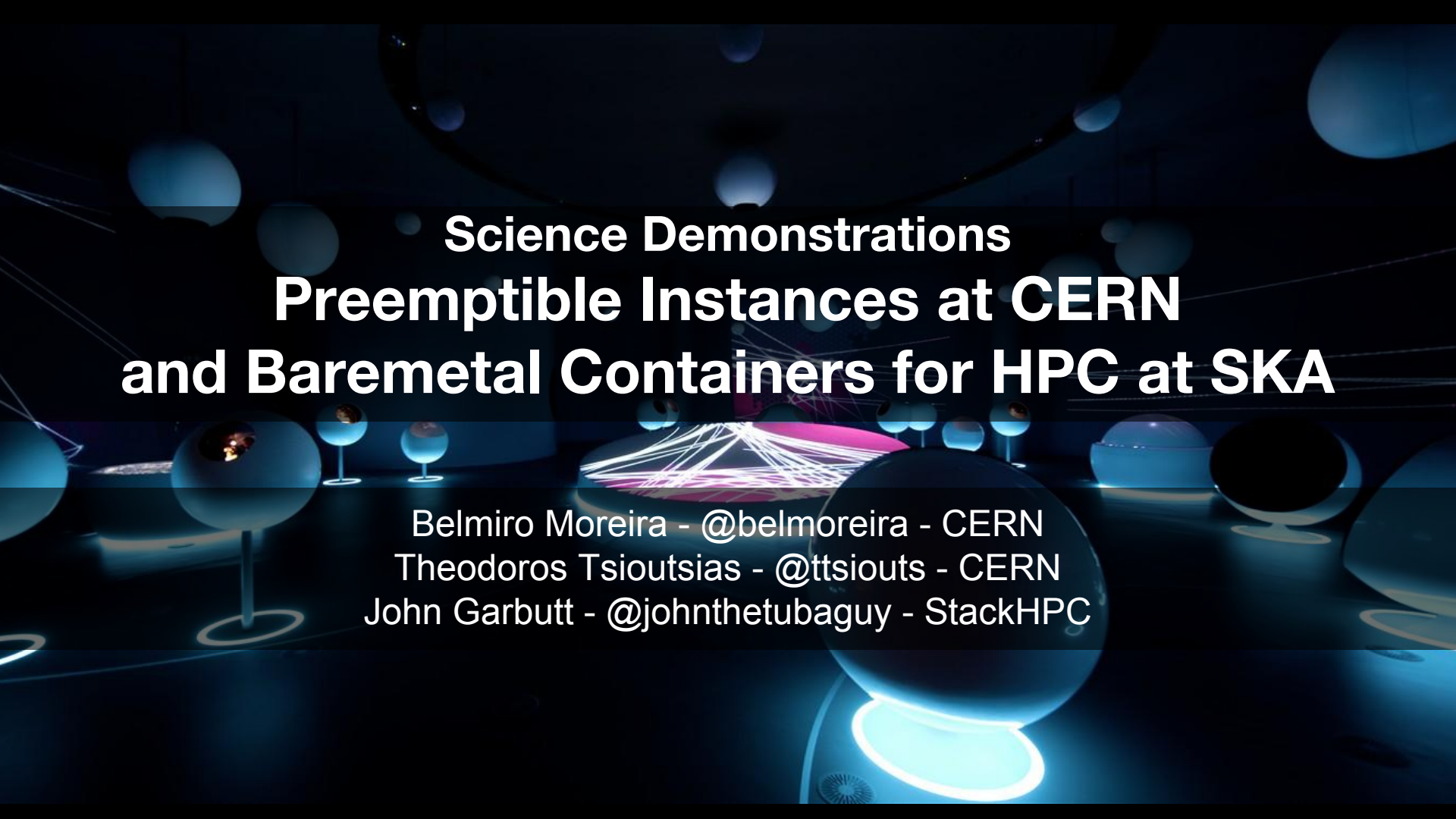
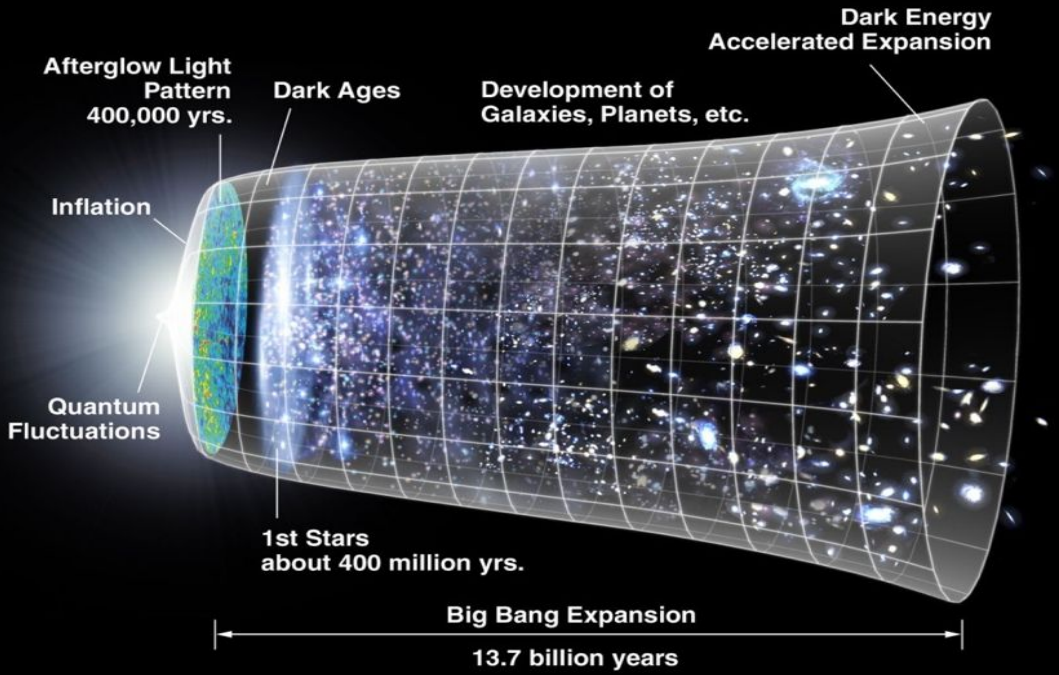
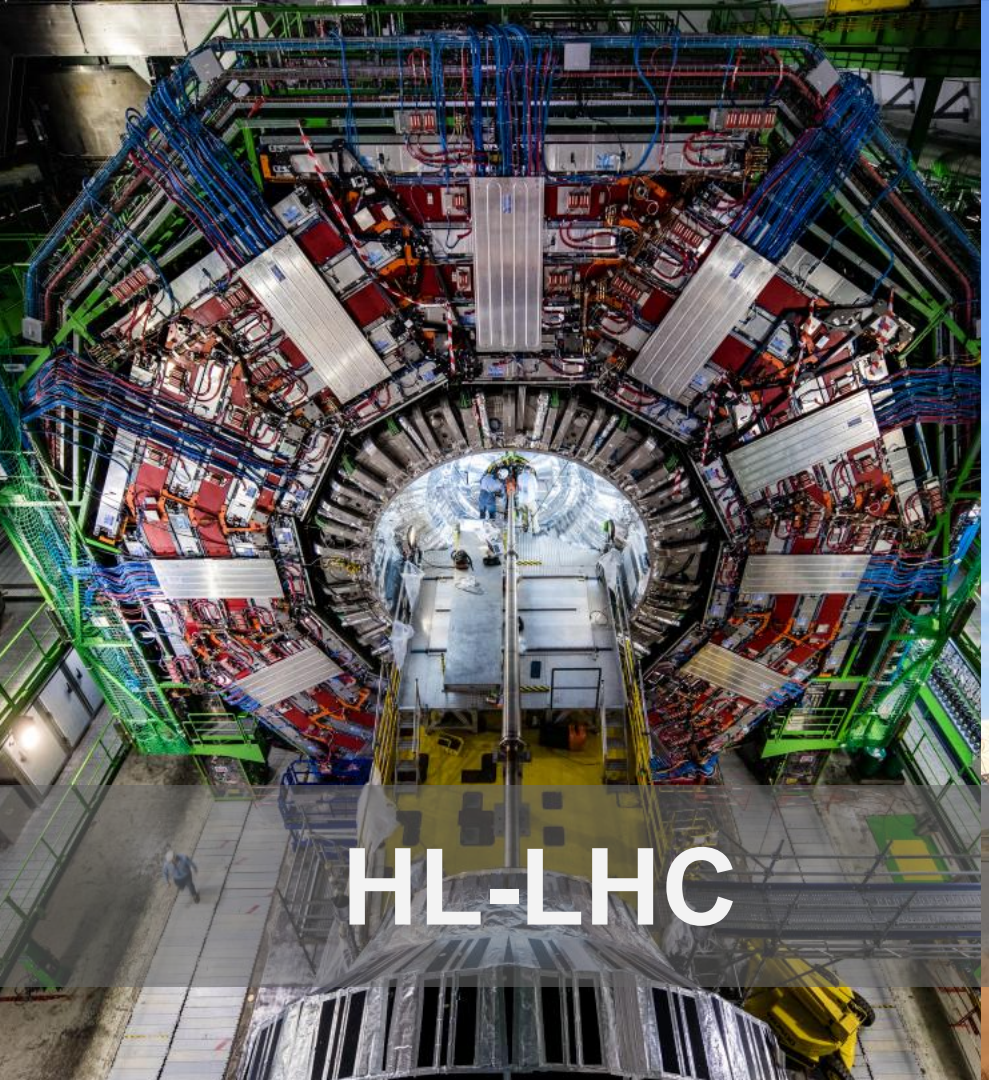


Science Demonstrations

Preemptible Instances at CERN and Baremetal Containers for HPC at SKA

Belmiro Moreira - @belmoreira - CERN
Theodoros Tsioutsias - @ttsiouts - CERN
John Garbutt - @johnthetubaguy - StackHPC


Dark Energy
Accelerated Expansion
Afterglow Light
Pattern
400,000 yrs.
Dark Ages
Development of
Galaxies, Planets, etc.
Inflation
Quantum
Fluctuations
1st Stars
about 400 million yrs.
Big Bang Expansion
13.7 billion years



HL-LHC

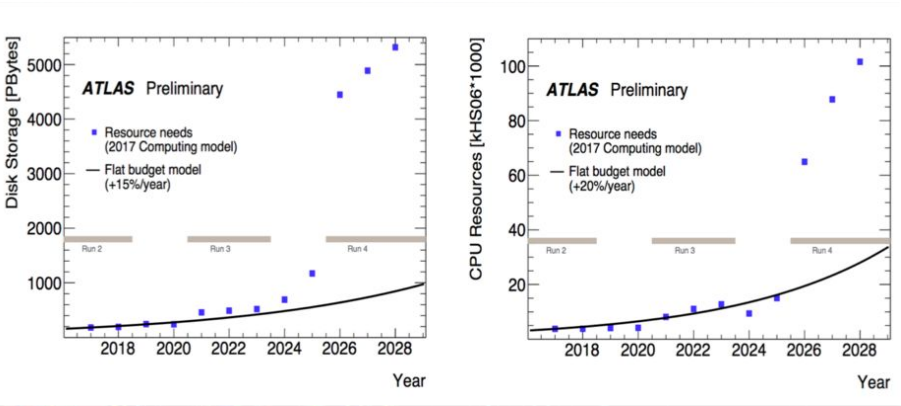
SKA

High Luminosity - LHC

HL-LHC will be a multi-Exabyte challenge

Storage and compute needs x10 above what naive technology extrapolation will bring

Need to drive down costs: focus on performance, efficiency, operations, ...



SKA1 LOW - the SKA's low-frequency instrument

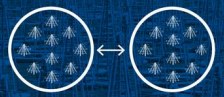
The Square Kilometre Array (SKA) will be the world's largest radio telescope, revolutionising our understanding of the Universe. The SKA will be built in two phases - SKA1 and SKA2 - starting in 2018, with SKA1 representing a fraction of the full SKA. SKA1 will include two instruments - SKA1 MID and SKA1 LOW - observing the Universe at different frequencies.

Location: Australia

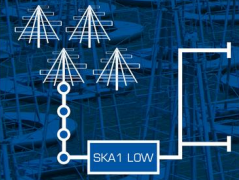
Frequency range:
50 MHz to **350 MHz**

~130,000 antennas spread between **500 stations**

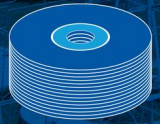
Total collecting area:
0.4km²



Maximum distance between stations:
65km



Total raw data output:
157 terabytes per second
4.9 zettabytes per year



Enough to fill up
35,000 DVDs
every second

5x
the estimated global internet traffic in 2015
(source: Cisco)

Compared to LOFAR Netherlands, the current best similar instrument in the world

25% better resolution
8x more sensitive
135x the survey speed

www.skatelescope.org Square Kilometre Array @SKA_telescope The Square Kilometre Array

SKA1 MID - the SKA's mid-frequency instrument

The Square Kilometre Array (SKA) will be the world's largest radio telescope, revolutionising our understanding of the Universe. The SKA will be built in two phases - SKA1 and SKA2 - starting in 2018, with SKA1 representing a fraction of the full SKA. SKA1 will include two instruments - SKA1 MID and SKA1 LOW - observing the Universe at different frequencies.

Location: South Africa

Frequency range:
350 MHz to **14 GHz**

~200 dishes (including 64 MeerKAT dishes)

Total collecting area:
33,000m²

or **126 tennis courts**

Maximum distance between dishes:
150km

Total raw data output:
2 terabytes per second
62 exabytes per year

Enough to fill
340,000
average laptops with content **every day**

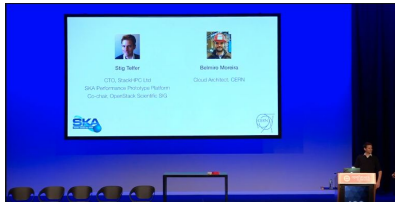
Compared to the JVLA, the current best similar instrument in the world:

4x the resolution
5x more sensitive
60x the survey speed

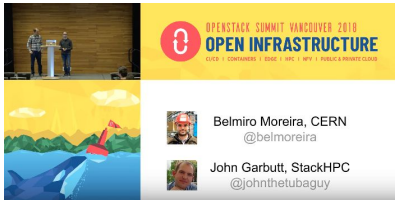
www.skatelescope.org Square Kilometre Array @SKA_telescope The Square Kilometre Array

Preemptible Instances/Containers on Baremetal

- Future Science on Future OpenStack developing next generation infrastructure at CERN and SKA
  - https://www.youtube.com/watch?v=XmQR06Mwd5g



- Containers on Baremetal and Preemptible VMs at CERN and SKA
  - https://www.youtube.com/watch?v=K5N4LYrupSs

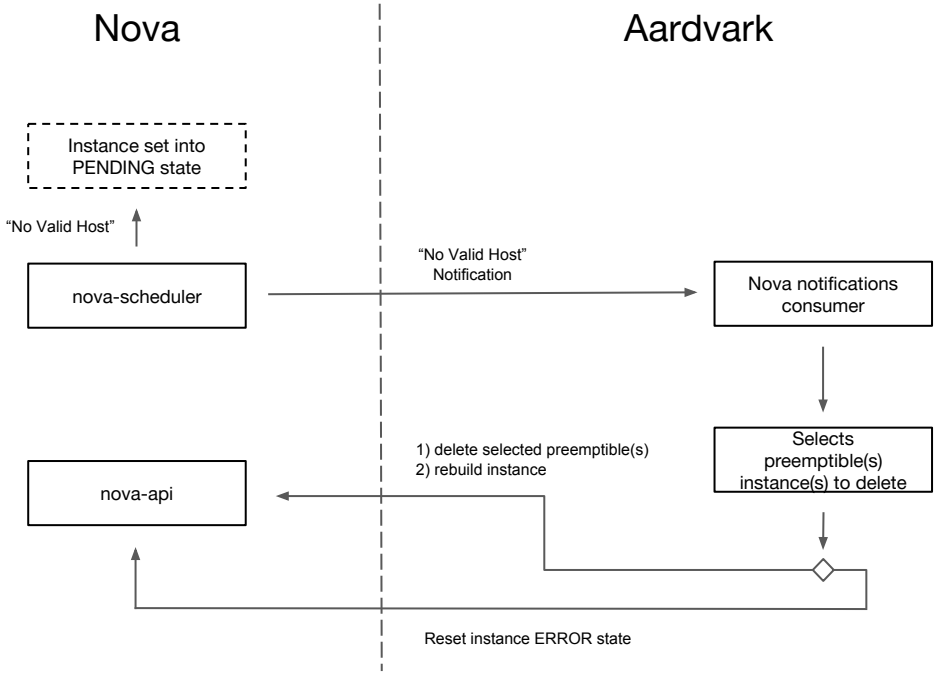


Preemptible Instances

How to Maximize Resource Utilization?

- Introducing the concept of Preemptible Instances:
  - created using idle resources
  - terminated as soon as the resources are needed
- The result:
  - handling the demand for extra resources
  - increasing the cloud utilization
- We need an Orchestrator for Preemptible Instances:
  - Aardvark!!

Workflow


Nova
Aardvark
Instance set into PENDING state
"No Valid Host"
nova-scheduler
"No Valid Host" Notification
Nova notifications consumer
Selects preemptible(s) instance(s) to delete
1) delete selected preemptible(s)
2) rebuild instance
nova-api
Reset instance ERROR state

DEMO

Current work in Preemptible Instances

- Add instance state PENDING (spec)
  - https://review.openstack.org/#/c/554212/
- Allow rebuild instances in cell0 (spec)
  - https://review.openstack.org/#/c/554218/
- Aardvark repo:
  - https://gitlab.cern.ch/ttsiouts/aardvark

Containers on Baremetal

MAGNUM

an OpenStack Community Project

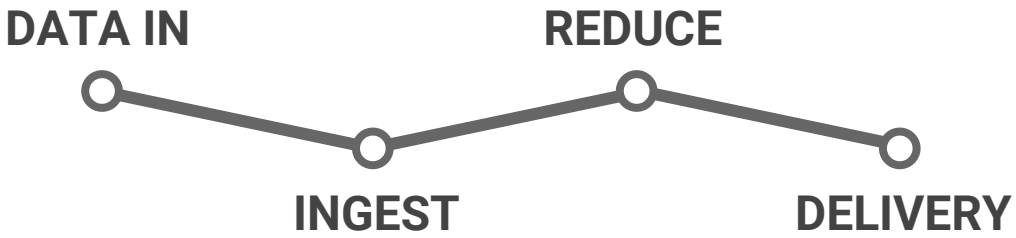
IRONIC

an OpenStack Community Project

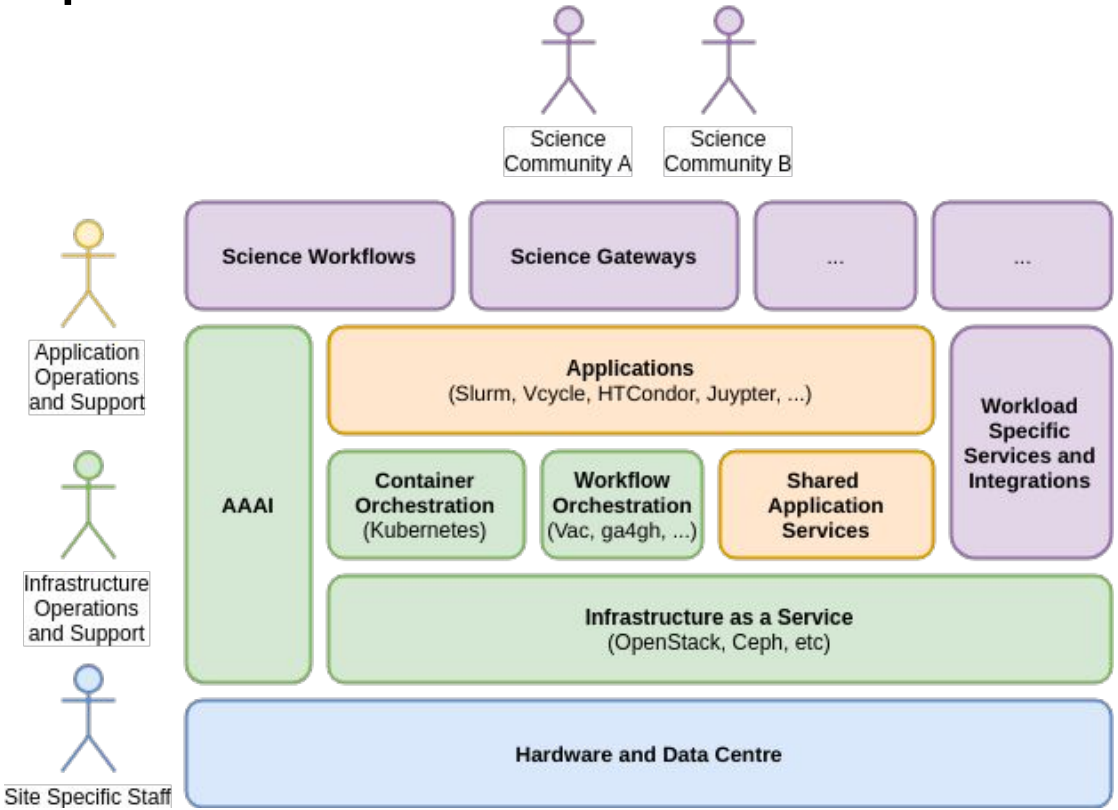
MANILA

an OpenStack Community Project

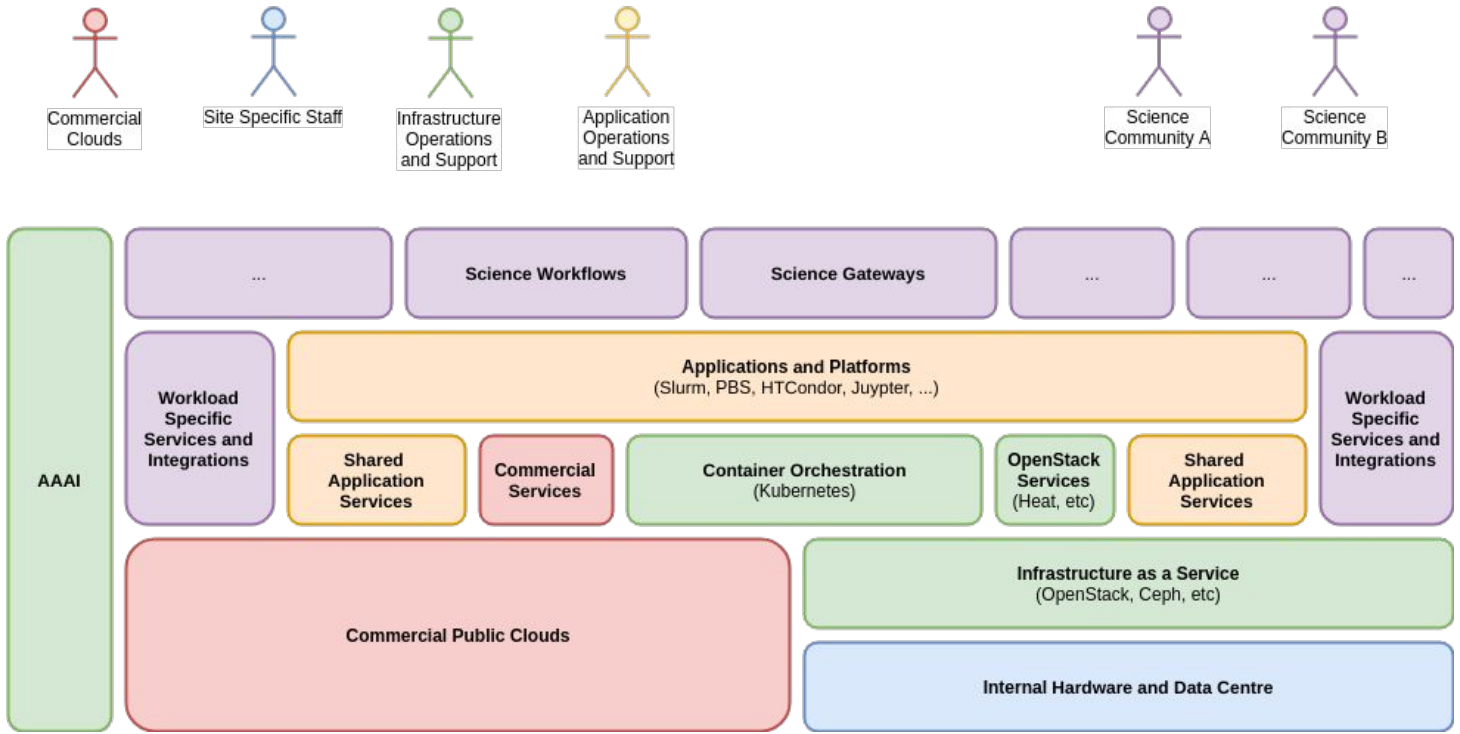
SKA Science Data Processor



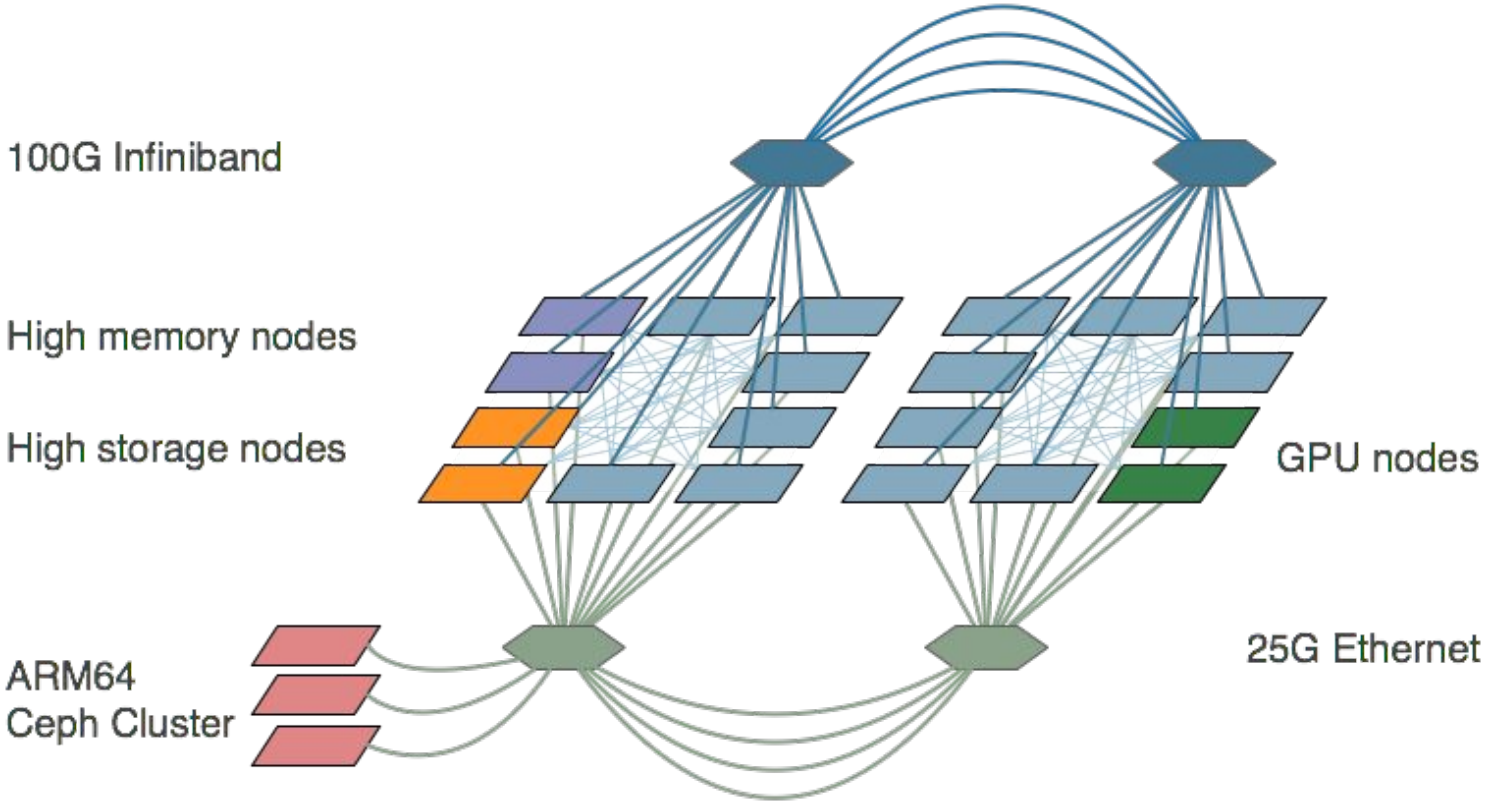
Scientific OpenStack



Scientific OpenStack



P3 AlaSKA


100G Infiniband
High memory nodes
High storage nodes
GPU nodes
ARM64
Ceph Cluster
25G Ethernet

Demo:
Magnum and Ironic

With OpenStack Manilla and Kubernetes

Join the Scientific SIG and...

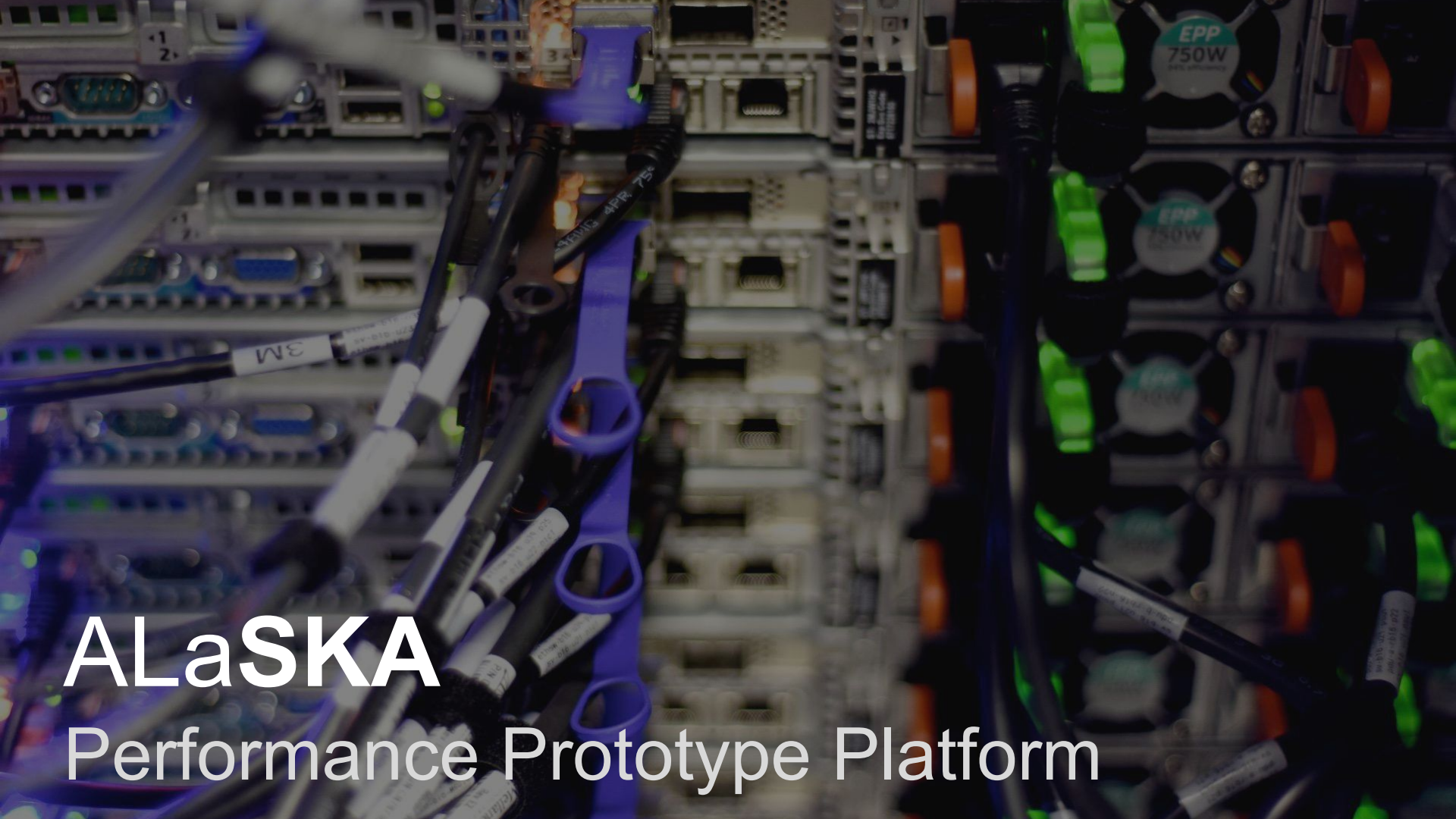
Get involved!

https://www.openstack.org/science/

Theodoros Tsioutsias, CERN
@ttsiouts

Belmiro Moreira, CERN
@belmoreira

John Garbutt, StackHPC
@johnthetubaguy



ALaSKA

Performance Prototype Platform

Join the Scientific SIG and...

Get involved!

https://www.openstack.org/science/

Why Containers on Baremetal?

- Baremetal Performance
- 30 seconds to switch workflow
- Easier to describe the SKA SDP Platform Architecture

Antennas

Digital Signal Processing (DSP)

Transfer antennas to DSP
2020: 5,000 PBytes/day
2030: 100,000 PBytes/day

Over 10's to 1000's kms

To Process in HPC
2020: 50 PBytes/day
2030: 10,000 PBytes/day

Over 10's to 1000's kms

High Performance Computing Facility (HPC)

HPC Processing
2023: 250 PFlop
2033: 25 EFlop

Magnum with Ironic

- Magnum used extensively at CERN
- Docker Swarm and Kubernetes are supported
- Historically a separate driver for baremetal, badly maintained
- Queens moves to using Fedora Atomic for VM and baremetal

Kubernetes and Manila

- OpenStack Provider
- Automatically create Manila share
- Use existing Manila share
- Helped with SKA SDP Platform Architecture

How to Maximize Resource Utilization

- “Spot Instances”

Resource Management

- Private Clouds
  - Usually is based in project quotas
  - Prevent resources being exhausted
  - Prevent “over-committing” resources/quota
  - Manage individual projects requirements
  - Reserve resources for operations with higher priority
  - Scientific Clouds
    - Projects have different funding models
    - They expect a predefined number of resources available
    - But not always these resources are used full time
- Public Clouds give the illusion of infinite capacity
  - Users pay for resources that they use

Problem

- Free resources that cannot be allocated
- Low **resource utilization**