

The Square Kilometre Array.

The open source way?

catalyst 

open source technologists

/me

Square Kilometre Array

“The world's largest and most powerful telescope”

Biggest of big data problems ever

Introducing...Exascale

How does SKA1 compare with the world's biggest radio telescopes?

SKA1 LOW

Australia

419,000m²
~130,000 antennas



MWA
Murchison Widefield Array, Australia
2,500m²
2048 antennas

LOFAR
Low Frequency Array for Radio astronomy, Netherlands
52,000m²
34,000 antennas

GMRT
Giant Metrewave Radio Telescope, India
48,000m²
30 dishes



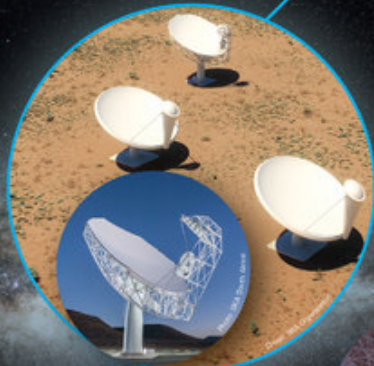
At 110 MHz

LOW FREQUENCIES

SKA1 MID

South Africa

33,000m²
~200 dishes



ASKAP
Australian SKA Pathfinder, Australia
4,000m²
36 dishes

MeerKAT
South Africa
9,000m²
64 dishes

JVLA
Karl G. Jansky Very Large Array, USA
13,200m²
27 dishes

NRT
Nançay Radio Telescope, France
7,000m²
300m x 35m antenna

Lovell
UK
4,500m²
76m dish

Effelsberg
Germany
7,800m²
100m dish

Parkes
Australia
3,200m²
64m dish

GBT
Green Bank Telescope, USA
7,800m²
100m dish

FAST
Five Hundred Meter Aperture Spherical Telescope, China
71,000m²
500m dish

Arecibo
Puerto Rico
42,000m²
305m dish

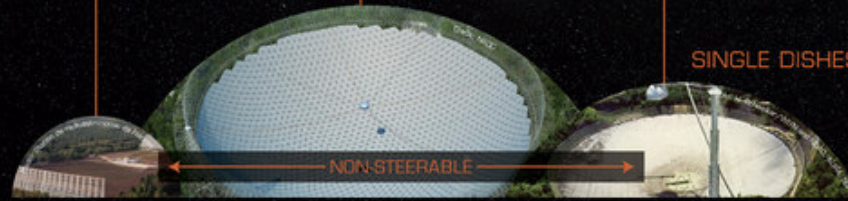
ALMA
Atacama Large Millimeter/ submillimeter Array, Chile
6,500m²
66 dishes

ARRAYS

MID FREQUENCIES

SINGLE DISHES

HIGH FREQUENCIES



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.

A telescope's capacity to receive faint signals - called sensitivity - depends on its collecting area, the bigger the better. But just like you can't compare radio telescopes and optical telescopes, comparison only works between telescopes working in similar frequencies, hence the different categories above.

The collecting area is just one aspect of a telescope's capability though. Arrays like the SKA have an advantage over single dish telescopes: by being spread over long distances, they simulate a virtual dish the size of that distance and so can see smaller details in the sky, this is called resolution.

SOMETIMES I THINK THE
SUREST SIGN THAT INTELLIGENT
LIFE EXISTS ELSEWHERE IN
THE UNIVERSE IS THAT NONE
OF IT HAS TRIED TO CONTACT US.

by WATTERSON



Exascale?

A billion billion calculations per second.

Exascale? $10^{18} \approx$ number of stars in 3 million Milky Way galaxies

How hard can it be?

Power-constrained Exaflop (10^{18} flops per second) computing requires highly efficient systems

'reasonably achievable' COTS Exaflop computing could be 15 to 25 years from now

Moving data costs a tremendous amount of energy

It is likely that the software will ultimately require large scale internationally collaborative development (100s to ~1,000+ staff years), even with reuse of codes and OTS

Martin Fink – CTO HP Enterprise

“HP is doing the project because it believes open source has "won"”.

These are not words we hear in New Zealand very often...

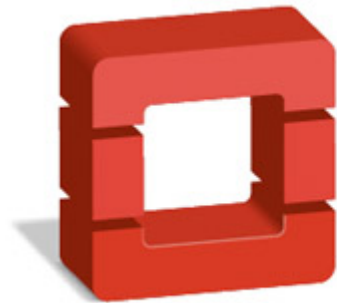
Not in New Zealand?



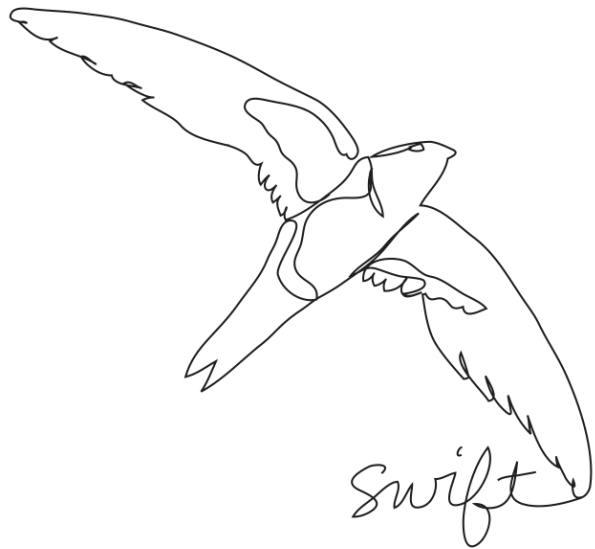
“No open standard for the raw infrastructure, poor resource sharing across organisations, imposes a tool or process on researchers, real time data interaction is a hack, poor isolation (dependency hell), no thought about containers, poor portability.”

Bruno Lago – GM Catalyst Cloud

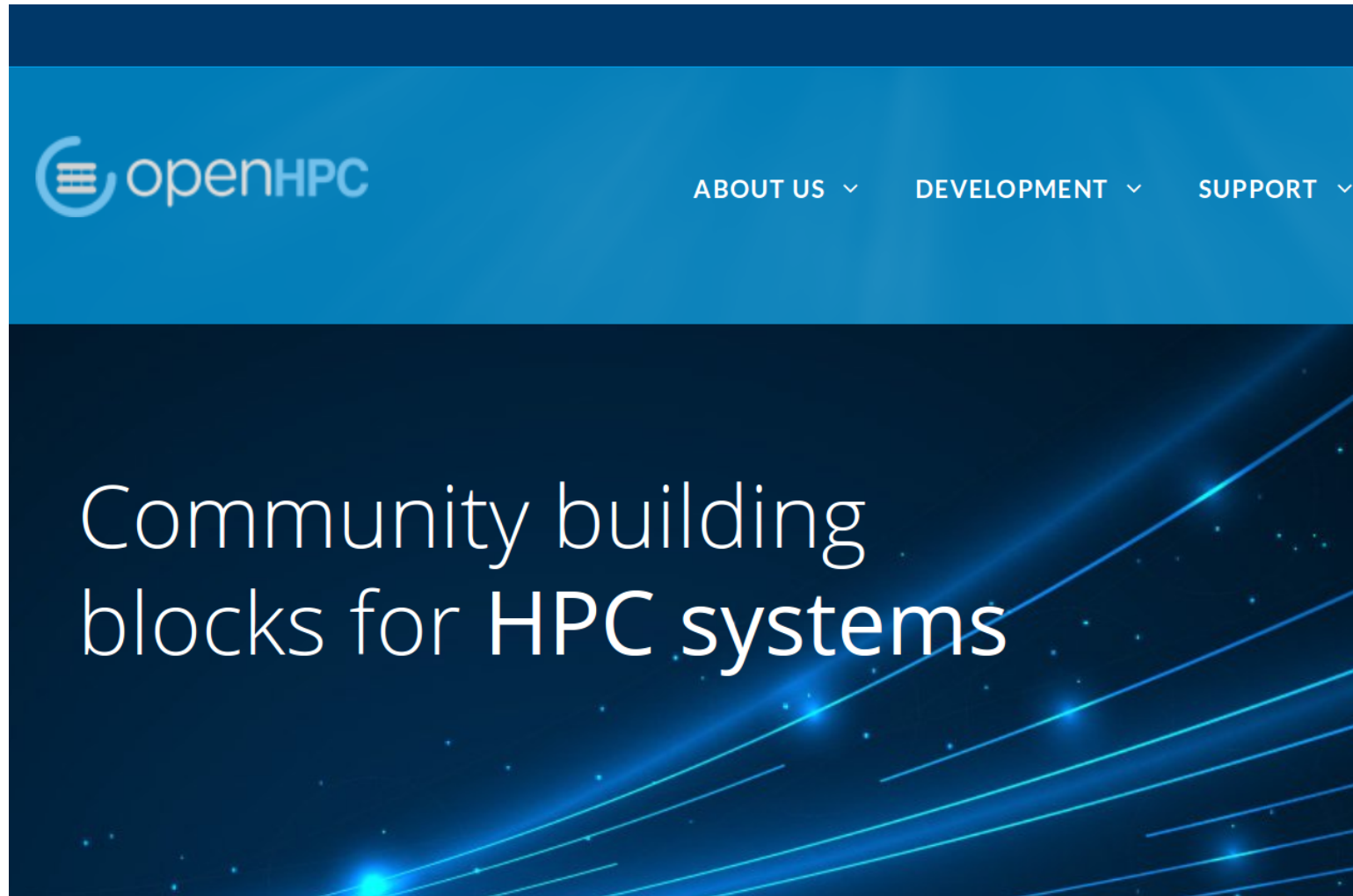
Collaborate, be open, be free



openstack™
CLOUD SOFTWARE



Introducing OpenHPC



HPC4Health Technical Director / Systems Architect



The HPC4Health's mission is to make high-performance computing accessible to health-care providers. Together we are building the engine that will help make personalized medicine and diagnostics a reality.

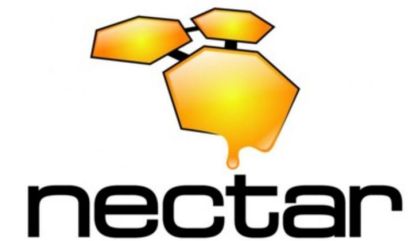
“We use cutting-edge HPC, virtualization, and networking technologies such as OpenStack, Infiniband, and Docker to provide a secure computational environment for biomedical research to our members.

This position would be working on the development, deployment, and support of the HPC4Health cloud environment.”

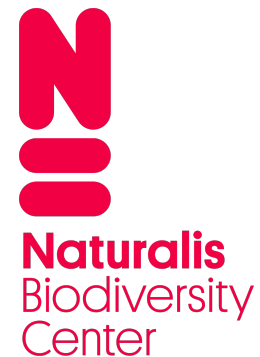
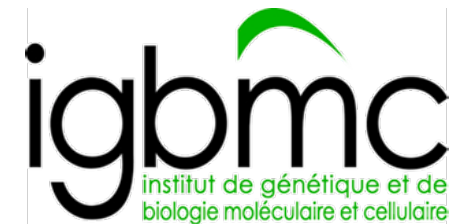
Who is using OpenStack?



University of
Zurich ^{UZH}



Massachusetts
Institute of
Technology



NASA and Rackspace - 2010

10,000 Members

200 sponsors

Catalyst a core contributor

Covering all layers?

- **Infrastructure** - OpenStack, Kubernetes
- **Storage** - CEPH, HDFS, Swift, Nyriad
- **Data Warehousing** - Trove, ImpalaDB, Hive, HBase, InfluxDB, ElasticSearch
- **Extract, Transform and Load (ETL)** - Sqoop, Logstash, Fluentd, PDI, MAPR, Spark, Oozie, OpenRefine, OpenCPU
- **Analysis, Visualisation, and Data Product Delivery** - RStudio, Shiny, Jupyter, DKAN
- **Advanced Analytics** - Spark MLlib, scikit-learn, R, Weka

Portability and reproducibility



Portability and reproducibility



Bare metal



Virtual Machines



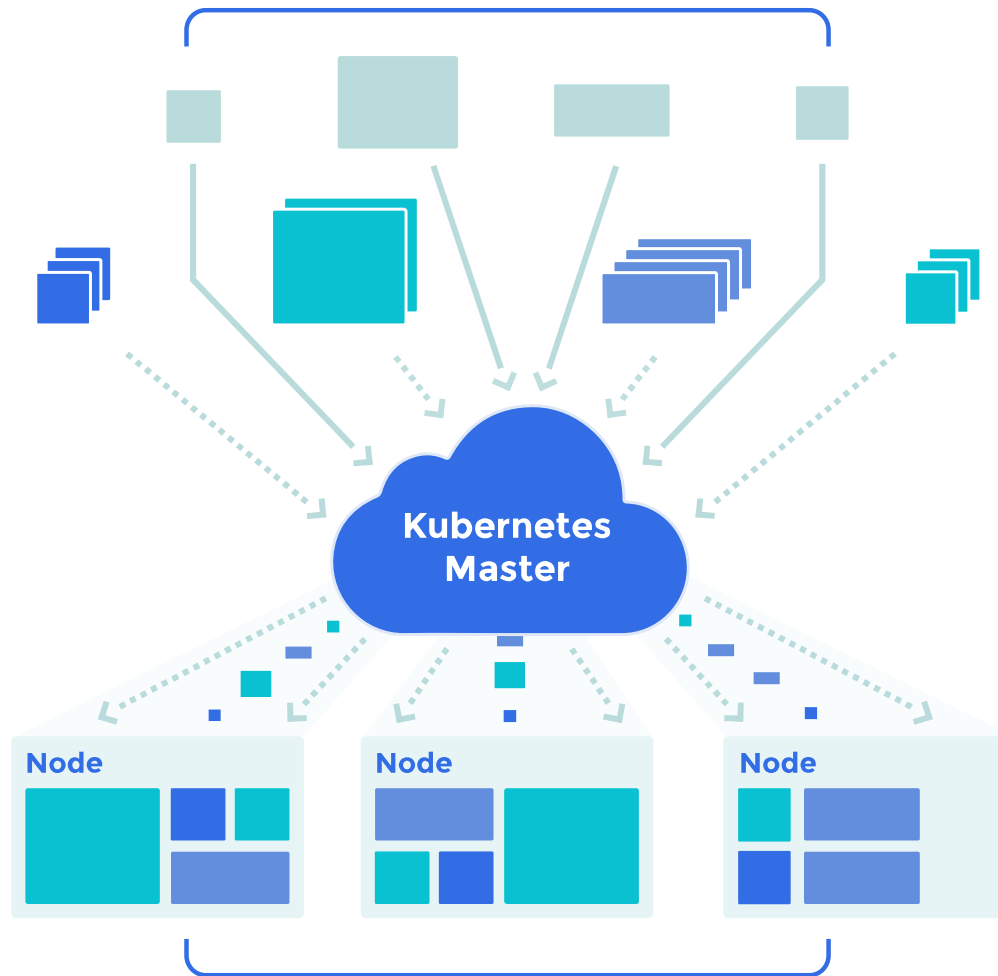
Linux Containers
(Docker)

Portability and reproducibility



Orchestration

An ocean of user containers



Scheduled and packed dynamically onto nodes

Ansible - Automation

“Ansible is a radically simple IT automation engine that automates cloud provisioning, configuration management, application deployment, intra-service orchestration, and many other IT needs.”

Now with OpenStack support

