South Africa’s HPC Investment and Valuable Partnerships

Dell EMC HPC Community Meeting
Austin, Texas

29 March 2017
Background: The CHPC in SA
User Community: South Africa
Background

- CHPC is **national HPC facility** funded by the SA Government
- Department of Science and Technology (DST)
- Administered by the CSIR
- Started operations in **June 2007** and based in Cape Town
- Currently hosts the **largest HPC system** in Africa
- CHPC has total number of ~**40** employees
- CHPC has **Research, Technical** and **Operational** divisions
Background

- NICIS: National Integrated Cyber Infrastructure System
- Framework report\(^1\) issued to DST – made public in June 2014
- Contains recommendations for sustainable NICIS in SA

Investment in HPC Systems

2007: IBM e1350; AMD Opteron; 2.4 GHz; 640 cores; 2.5 TFlops (Linpack)

2008: IBM Blue Gene/P; 0.8 GHz; 4096 cores; 11.5 Tflops (Linpack)

2009: Sun Constellation Cluster; 2.9/3.0 GHz; 2684 cores; 27 Tflops (Linpack); 480 TB Lustre File System

2010: GPU Cluster; 2.4 GHz; 96 cpu cores; 22 GPU's; 16 TFlops (Linpack); 14 TB

2011: Tsessebe Cluster Upgraded 2.9GHz (4032 cores); 61.6 Tflops total (Linpack); 480 TB Lustre File System

4 PB DIRISA Storage Unit
Support for CERN Experiments

- Dedicated cluster for the CERN experiments. Mainly ALICE and ATLAS.
- Currently providing support for 2,400 jobs/day (increase in international bandwidth will increase the # of jobs).
- Plans to move to WACS cable for better international connectivity.
- Officially signed the MOU for a Tier-2 facility in April 2015.
- Plans to expand the service for Tier-1 in future.
## The Road to PetaFLOP

### System Configuration

<table>
<thead>
<tr>
<th></th>
<th>Phase 1</th>
<th>Phase 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dell PowerEdge C6320 Servers:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard Compute nodes 128GB (64GB) / node</td>
<td>1 008</td>
<td>1 368</td>
</tr>
<tr>
<td>2 x Intel Xeon E5-2690 v3 (Haswell) processors (12 Cores Each $\Rightarrow$ 24 cores / node)</td>
<td>24 192</td>
<td>32 832</td>
</tr>
<tr>
<td><strong>Dell PowerEdge R930 servers:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Large Memory Compute Nodes 1024GB / node (FAT nodes)</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>4 x Intel Xeon E7-4850 v3 processors (14 Cores Each $\Rightarrow$ 56 cores / node)</td>
<td>280</td>
<td>280</td>
</tr>
<tr>
<td><strong>Infiniband FDR 2:1 Blocking (56 Gbps)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Parallel Storage (Useable) PB</strong></td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td><strong>Total Number of Racks (including Compute, Login, Management and Storage Nodes)</strong></td>
<td>19</td>
<td>24</td>
</tr>
<tr>
<td><strong>Centos 7.1 with Bright Cluster Manager and Altair PBS Pro</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total Linpack Performance (Tflop/s)</strong></td>
<td>783</td>
<td>1029</td>
</tr>
</tbody>
</table>
Lengau

- 161 on TOP500
- Fastest system in Africa
Resources at the CHPC

**Nodes % occupied**

- **Nov/Dec 2016**
- **Jan/Feb 2017**
Resources at the CHPC

Lengau % Nodes Occupied
Who is using the CHPC?

Non-Academic Public
- CSIR
- SAWS
- ARC
- NZG
- SAAO
- SANBI
- Sugar Institute
- NICD
- NECSA

Research Programmes Distribution
- SA Academic 75%
- SA Public 14%
- SA Industry 7%
- African Academic 4%
Who is using the CHPC?

SA Academic Research Programmes Distribution

- UCT: 17%
- UKZN: 16%
- US: 12%
- UP: 17%
- Other: 4%
- Wits: 7%
- NWU: 6%
- RU: 7%
- UL: 2%
- UV: 2%
- UWC: 2%
- DUT: 2%
- NMMU: 2%
- SA Public: 14%
- SA Industry: 7%
- African Academic: 4%
Who is using the CHPC?

User **Status** on 23 March 2017:

- **207** Registered Research Programmes
- **698** Registered Cluster Users

Research Programmes Distribution

- **SA Academic** 75%
- **SA Public** 14%
- **SA Industry** 7%
- **African Academic** 4%
Who is using the CHPC?

Other African Universities:

- Maseno University (Kenya)
- MMUST (Kenya)
- Moi University (Kenya)
- University of Eldoret (Kenya)
- University of Ghana (Ghana) (Two programs)

Research Programmes Distribution

- SA Academic: 4%
- SA Industry: 7%
- SA Public: 14%
- African Academic: 75%
- Other African Universities: (6 Programs)
Who is using the CHPC?

Science Domain Distribution:  # Users
Resources at the CHPC

Science Domain Distribution: # CPU Hours

- Chemistry: 27%
- Earth Sciences: 21%
- Bioinformatics: 6%
- Health Sciences: 7%
- Computer Science: 3%
- Physics: 5%
- Computational Mechanics: 2%
- Astrophysics: 4%
- Material Science: 25%
Who is using the CHPC?

- SA Industry
  - Mintek
  - Eskom
  - Transnet
  - De Beers Marine
  - Johnson Matthey
  - Hatch
  - Roche
  - Inqaba Biotec
  - MTech
  - eScience

Research Programmes Distribution

- SA Academic: 75%
- African Academic: 4%
- SA Public: 14%
- SA Industry: 7%
Industry Support

PetroSA Eclipse Benchmark Case

Dimensions 149X118X128 - 2.25 Million cells

[Graph showing performance of different configurations for runs per hour vs. processes]
**SKA Readiness Programme**

**CHPC Contribution to Square Kilometre Array (SKA):**

- Appointment of key **staff** to support the SKA activities
  - Currently there are 2 astronomy research scientists

- **SKA Project** Focus Areas:
  - Design of the radio-astronomy **dishes**
  - Storing **KAT-7** data
  - Development of MeerKAT data centre
  - Science Data Processing (**SDP**) Consortium
  - **CyberSKA** – Cloud based web front-end for radio data

- Support for the **SKA Readiness Strategy** in developing big data capabilities within the **8 African** partner **countries**
  - Dedicated **funding** from **DST** for 3 year period (2016-2018)
User Community Support

HPC Ecosystems Project

- Strategy to repurpose HPC systems for local compute or training needs
- HPC Ecosystem initiative → Provide mid-range HPC systems to universities
- One of key strategic initiatives of the CHPC
- CHPC initiated the RANGER Project – Support to institutions
- Received resources from other collaborators, like Cambridge.
  → Do not have HPC facilities
  → Need of HPC resources for learning
## User Community Support

### HPC Ecosystems Project

### Local SA Sites:

<table>
<thead>
<tr>
<th></th>
<th>Operational</th>
<th>Used</th>
<th>Needs Support Equipment To Operate</th>
<th>Power &amp; cooling</th>
</tr>
</thead>
<tbody>
<tr>
<td>University of Fort Hare</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>University of Kwa-Zulu Natal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>University of Venda</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>University of Witwatersrand</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stellenbosch University</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sol Plaatjie University</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>North West University</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Courtesy: Bryan Johnston (CHPC)
### User Community Support

**HPC Ecosystems Project**

#### African Sites:

**SKA Partner Countries**

<table>
<thead>
<tr>
<th>Site ID</th>
<th>Hardware</th>
<th>SysAdmin Training</th>
</tr>
</thead>
<tbody>
<tr>
<td>Botswana</td>
<td>Shipped</td>
<td>Ready</td>
</tr>
<tr>
<td>Namibia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zambia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Madagascar</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mauritius</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mozambique</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kenya</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ghana</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---
SKA Computing Options

2 Tbps (SKA SA)

Karoo Array Processor Building (CSP + SDP + some archive)

20 Gbps (SANReN)

CHPC MeerKAT archive (10-12 racks & some access to CPUs)

10 Gbps

Other Tier 2 Centres

IDIA Tier 2 storage & science analysis support
African Data Intensive Research Cloud (based on federated OpenStack s/w)

- UK Cloud Node
- NL Cloud Node
- SA Other (mid scale instance)
- N.Cape Sol Plaatje (small scale instance)
- CHPC (large scale instance)
- IDIA (mid scale instance)

SKA African Partner 1 (small scale instance)
SKA African Partner 2 (small scale instance)
SKA African Partner 3 (small scale instance)
Objectives of ADIRC

- The African Data Intensive Research Cloud aims to support data intensive radio astronomy among collaborating partners in South Africa and SKA partner countries.
- Infrastructure As A Service Cloud using OpenStack Middleware are deployed at a number of sites as Proof of Concept for future large scale deployments.
- The project will deploy a mix of scales of systems, with users bursting to larger facilities as needed.
- Software to support data distribution, analysis and visualisation of astronomy data is being deployed.
- The infrastructure will be extended to other science domains, such as health, earth sciences and material science.

Currently there are 3 prototypes at UCT, NWU and CHPC.
- 21 compute nodes with 256 GB RAM each.
- Over 400 TB of storage (CEPH and Block storage).
- Future to provide POSIX and LUSTRE file systems.
- Systems running on Ubuntu
- LSCS – Large Scale Compute and Storage
- RPoP – Regional points of Presence critical for training and local processing.
- Connectivity provided by a minimum 1 Gb/s through SANReN.
OpenStack Dashboard for CyberSKA
Remote Visualisation
Student Cluster Competition (SCC): Training undergraduate students in HPC

Three stage program:

(1) Winter school: **22 teams**

(2) CHPC National Meeting: **10 Teams**

(3) ISC: **1 Team** representing SA
Partnerships in technology testing

- In partnership with Dell and CoolIT testing the DLC on C-Series Servers.
- Intel provides evaluation Servers for KNL, FPGAs, OPA and Xeon.
- Extend the support to universities.
- Dell Academy to train entrepreneurs committed for the next 10 years through Khulisa Program.
2017 CHPC National Meeting: Durban

4 – 8 December 2017

www.chpccconf.co.za
Thank You...

https://users.chpc.ac.za

www.chpc.ac.za

hsithole@csir.co.za