Intro to High Performance Computing and its Applications

Prof Catherine Cress
Centre for High Performance Computing, South Africa
University of the Western Cape

CHPC
CENTRE FOR HIGH PERFORMANCE COMPUTING

CSIR
our future through science

The CHPC

- South African National supercomputing facility, started 2007
- 40 staff, 16 researchers supporting >500 users in many fields (>140 active research programmes)
- Over 32 000 cores + 5 large memory machines (1 TB)
  Test-bed facilities including GPUs FPGA, MIC
  Petabyte storage facility
- Training & development

© CSIR 2011

HPC, my view

HPC hardware = many computers linked together with fast networking + storage

HPC = using HPC hardware to do useful research

© CSIR 2011

CHPC CENTRE FOR HIGH PERFORMANCE COMPUTING

CSIR our future through science

The CHPC – applications

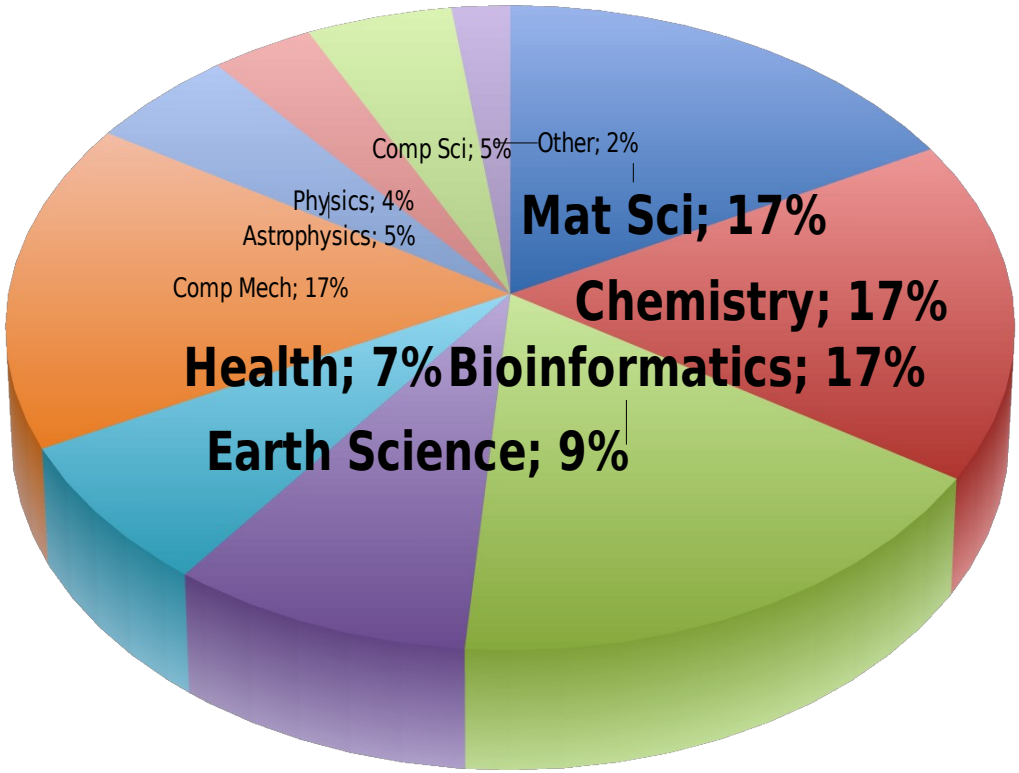
CHPC in South Africa has users in many fields:

- Health
- Mining
- Construction
- Energy
- Weather, Oceans, Climate change
- Film & media
- Finance
- Astronomy

© CSIR 2011

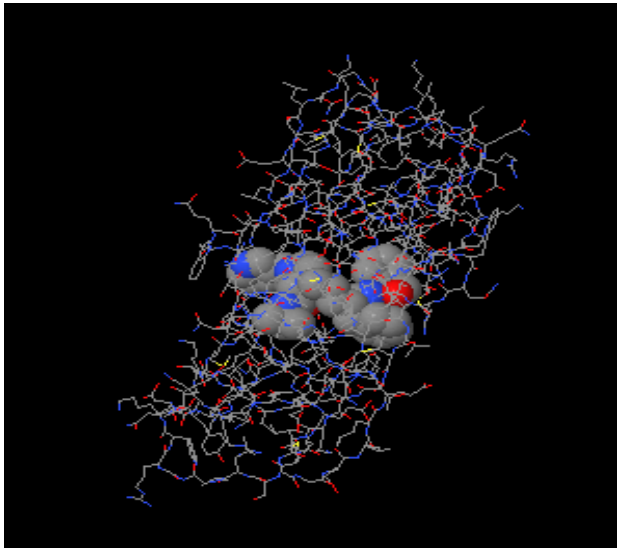
The CHPC – applications

- **Lengau Total # Research Programmes (1 Apr 2016 - 30 Nov 2017): 193**

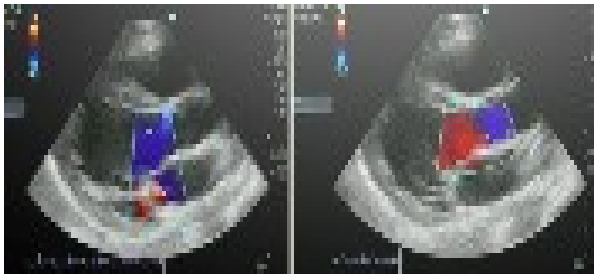


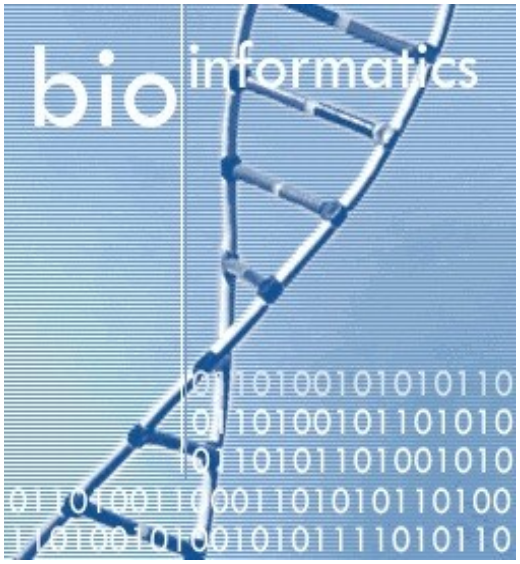
The CHPC - applications

Health → drug design, heart image processing, genetics



GPU + data portal



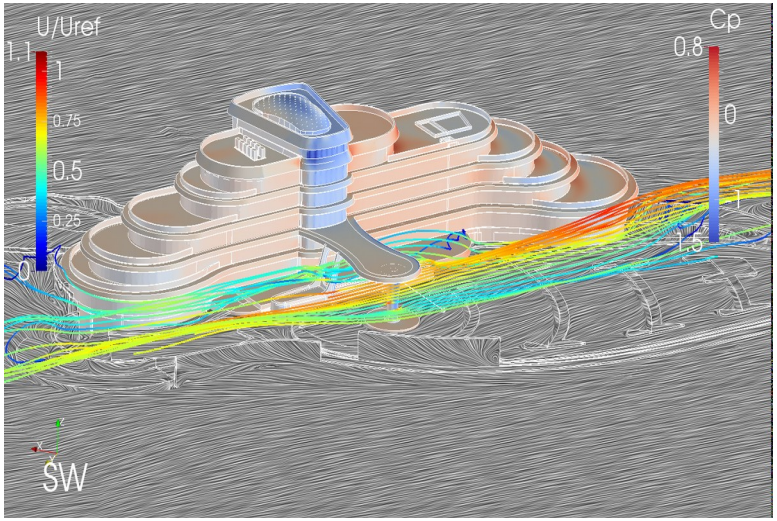


CHPC CENTRE FOR HIGH PERFORMANCE COMPUTING

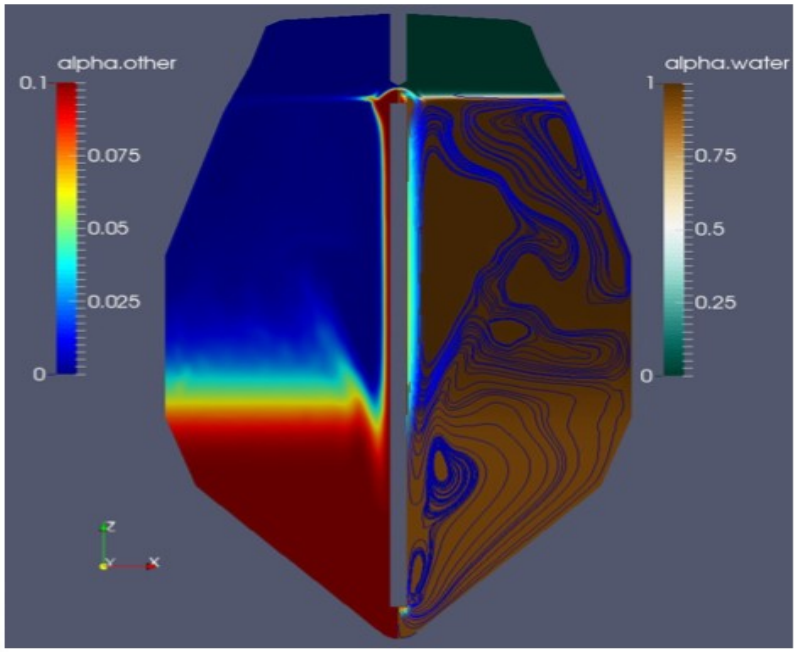
CSIR our future through science

The CHPC - applications

Construction



Water Treatment

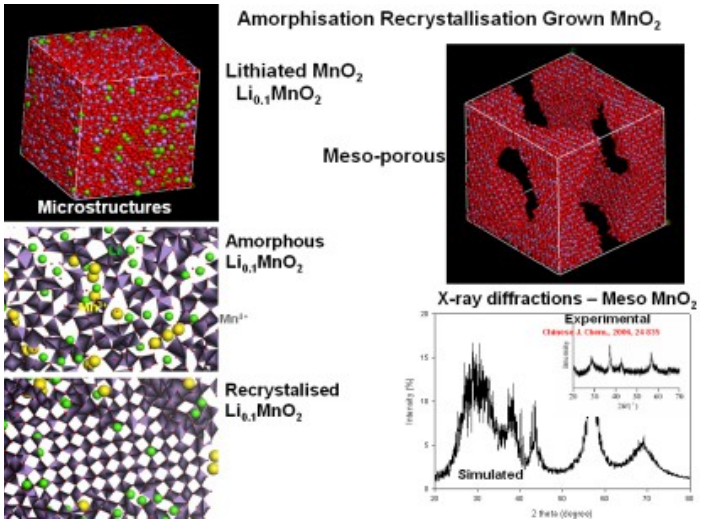


© CSIR 2011

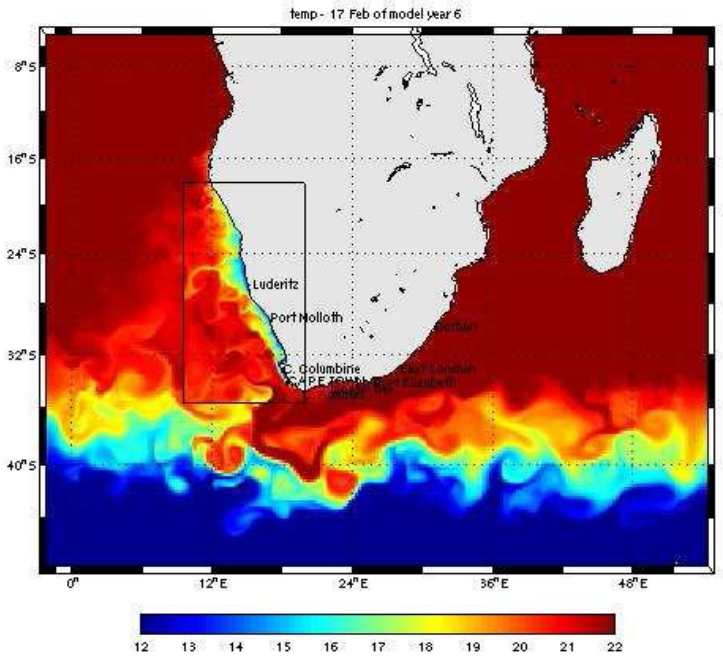
The CHPC - applications

Energy:

Battery materials
Smelting



Oceans & Climate Change



© CSIR 2011

The CHPC - applications

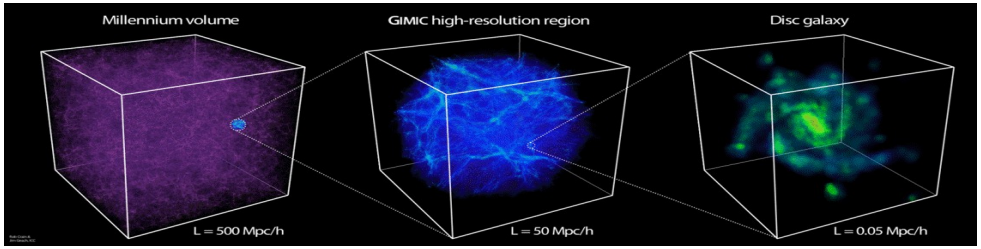
Astronomy

Theoretical calculations (exotic cosmologies)

Sean February

Galaxy Evolution Simulation

Matthew Cawood



Probing Dark Matter & Dark Energy using light from early universe

Data Mining

Israel Tshililo

Measuring expansion rate of universe at half age of universe

Model fitting and Statistics + own data from SA Large Telescope)

Ando Ratsimbazafy - Madagascar

Relating dark matter to luminous matter

Data mining and model fitting

Vijay Chumroo - Mauritius

© CSIR 2011

CHPC CENTRE FOR HIGH PERFORMANCE COMPUTING

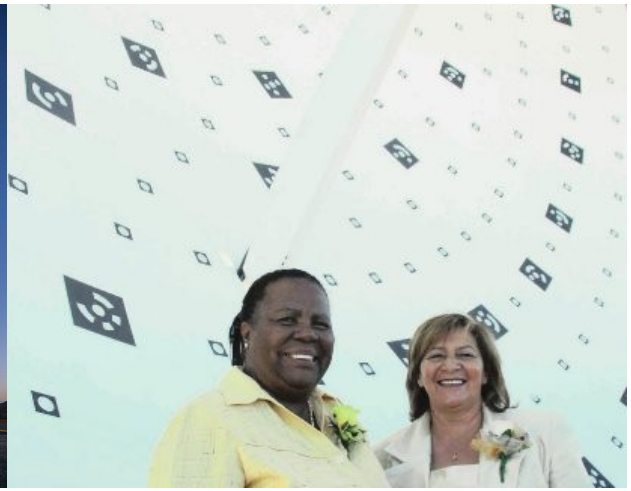
CSIR our future through science

The Square Kilometer

Billion dollar radio telescope for 2020+

Phase 1: 200 dishes in SA + array in Australia

Phase 2: Antennae in 9 African countries including Kenya



© CSIR 2011

CHPC CENTRE FOR HIGH PERFORMANCE COMPUTING

CSIR our future through science

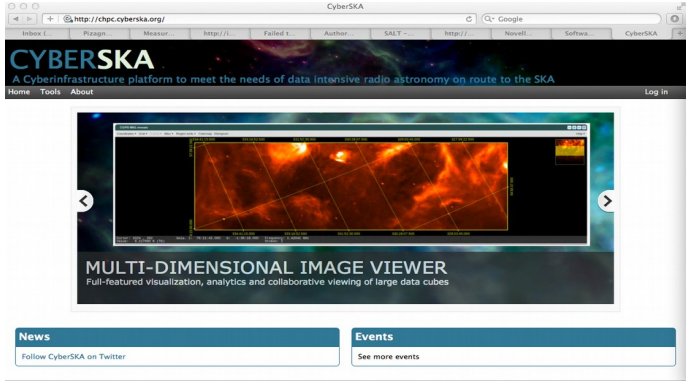
SKA → HPC, Big Data

SKA data more than 10x current global internet traffic

Many levels of data crunching

* Hardware accelerators: FPGA, GPU
* Calibration and Imaging of data → image processing, visualisation
* Detecting sources like radio galaxies in HUGE multidimensional data
* Statistical analysis of resulting catalogs of galaxies
* Comparison of models with data → stats, N-body simulation

Data Portals: large international collaborations

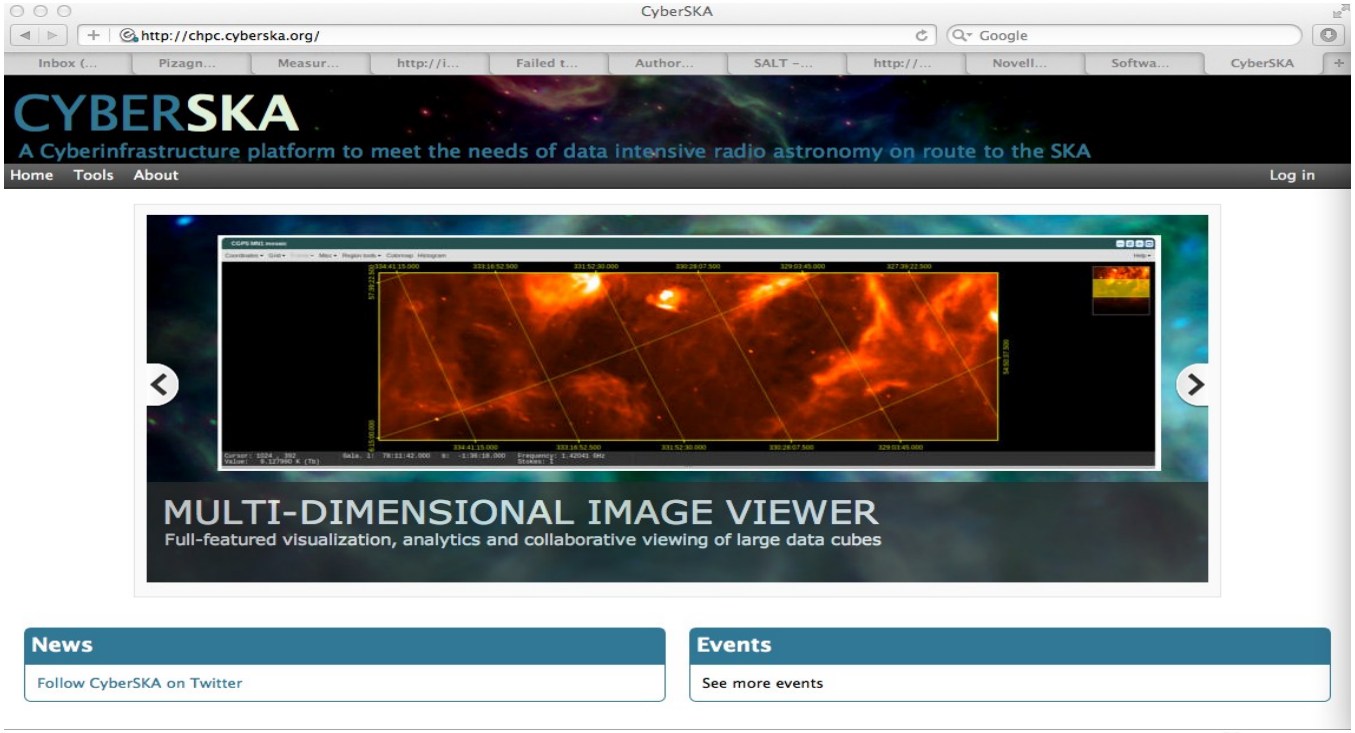


CENTRE FOR HIGH PERFORMANCE COMPUTING

CSIR our future through science

SKA → HPC, Big Data

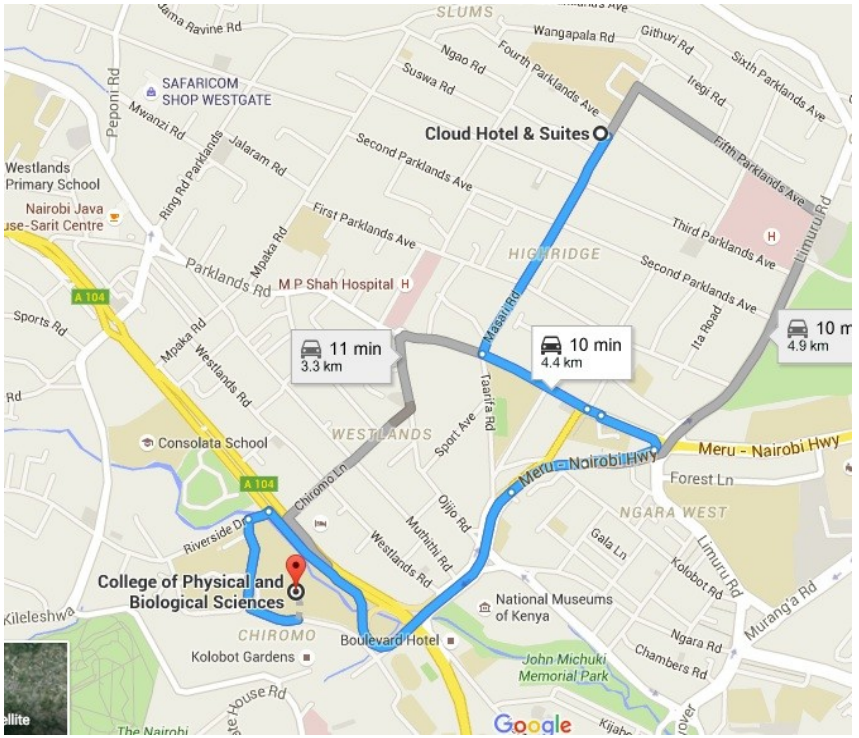
Needs for large global teams to collaborate
CyberSKA data portal



Big Data Impact

Example:

Google Maps:
Real time navigation,
Incredible detail



© CSIR 2011

Astronomy → Data Scientists

McKinsey: By 2018, shortage of over 150000 data scientists.

Astronomers/Physicists often data scientists

Data Science courses:
Statistical Analysis
Modelling
Visualisation
Machine learning
HPC & parallelisation
Databases and SQL
Security & governance

© CSIR 2011

CHPC CENTRE FOR HIGH PERFORMANCE COMPUTING
CSIR our future through science

SKA Computing development in Africa

* Zambia:
hardware up and running, workshops run

* Mauritius, Madagascar, Namibia
hardware shipped, projects underway

* Botswana
hardware in place

* Kenya
Negotiation on hardware, projects initiated

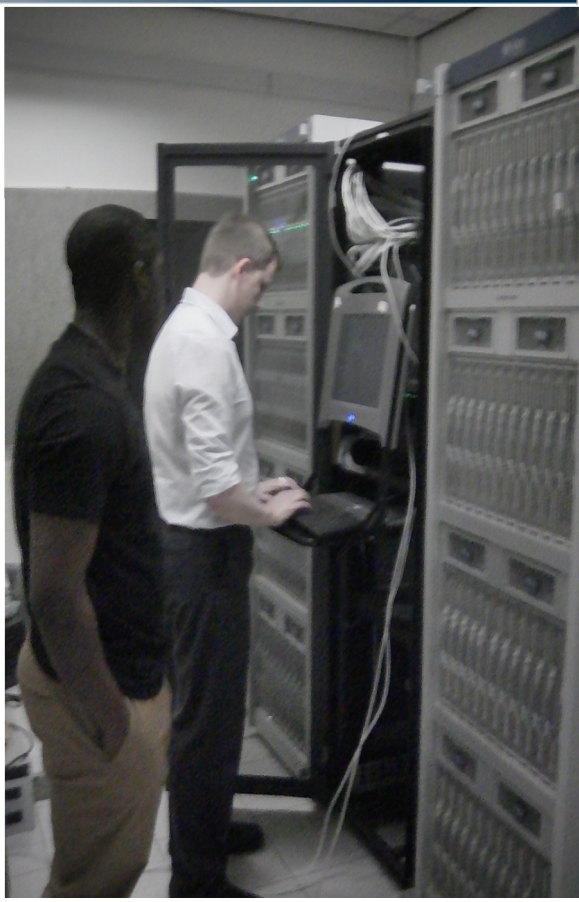
* Ghana
AVN, existing computer infrastructure

* Mozambique
Initial discussions

ACCESS TO CHPC

© CSIR 2011

eg. Botswana's



Used hardware donated by Texas ACC
4 racks
rack has 4 chassis
chassis has 12 nodes
node: 4 quad-cores = 16 cores

16x12x4x4= 3072 cores
Only power-up a fraction of these (2x192?)

Users?:
Atmospheric Physics, Chem/Phys:
eg Catalysis, Bioinformatics, Geophysics (eg seismology)

CHPC Computing development in Africa

Twelve visits (22 trips funded)
Over 200 contacts in database

	SysAdmin	User ID (staff)	User Workshops	Visit Dates
Botswana	30	10UB+3 (BIUST)	20+15	8Jul16, 15-22Oct16
Namibia		3		25Jun16
Zambia	3	4CBU+5UZAM	50+20+20	23-30Sept15
Madagascar	3	4UA+5	20+30	14-26 Nov15, 20-31Aug16
Mauritius		3		1-10Jul13
Mozambique		3	50+20	16-23Mar16
Kenya		6	30+30	6Jul15, 5-14Jun16
Ghana				21-22Nov13

CHPC Computing development in Africa

Priorities:

1. User development
   - people on the ground – in Astronomy and other fields
   - need projects in useful fields eg agric, health, mining, tourism
   - view to jobs & commerce → data science business?: ML etc.

2. “last mile” networking

3. Online community development – courses together, data science

BIG DATA AFRICA FELLOWSHIP PROGRAM

© CSIR 2011

CHPC user support

Over 500 users on system – not all active
Over 150 different research programs

Each program has Principal Investigator(PI) + team
PI = permanent staff member with publication record from South Africa or SKA partner country
Team = students, postdocs, foreign members
Each program allocated research scientist at CHPC

User database, resource allocation
Wiki
Training
National conference – www.chpcconf.co.za
Flagship projects

Do I need CHPC?

Scenario 1:

Student:
“my code runs to too slowly on my laptop, I need HPC”

Supervisor:
“I can't/won't code – contact the CHPC”

HPC not necessarily the answer
But possibly support for code improvement from CHPC staff

Do I need CHPC?

HPC:

1. Many computers linked together + fancy networking
2. One computer with many cores, large shared memory (1TB)

(3. not really at CHPC: GPUs)

Before considering HPC:

1. Can I improve my code so it runs faster on my laptop?
   Excessive I/O? Unnecessary loops? code profiler? Compiler?
2. How can I break down my problem into smaller pieces?
   Run many instances at same time? Which parts parallelize?
3. Do I have huge files that need to be open in memory?

CHPC CENTRE FOR HIGH PERFORMANCE COMPUTING
CSIR our future through science

Do I need CHPC?

HPC vs Big Data

HPC used to be focussed on maximising number of operations possible
Now need more thought about large I/O → Big Data

CHPC vs Cloud services eg Amazon etc
No GPUs on production cluster (future?) → machine learning?

CHPC CENTRE FOR HIGH PERFORMANCE COMPUTING
CSIR our future through science

Do I need CHPC?

At CHPC:

Main cluster
1 node has 24 cores, 128 GB mem (360nodes have 64GB)
1368 nodes => ~32000 cores

'Fat nodes'
56 cores, 1 TB each

You will need to rewrite code to run in parallel
Eg. MPI4Py

Do I need CHPC?

Scenario 2:

I want to store large amounts of data (many TB) for long periods

→ CHPC does not provide storage for long periods

→ Contact DIRISA (another part of SA CyberInfrastructure)

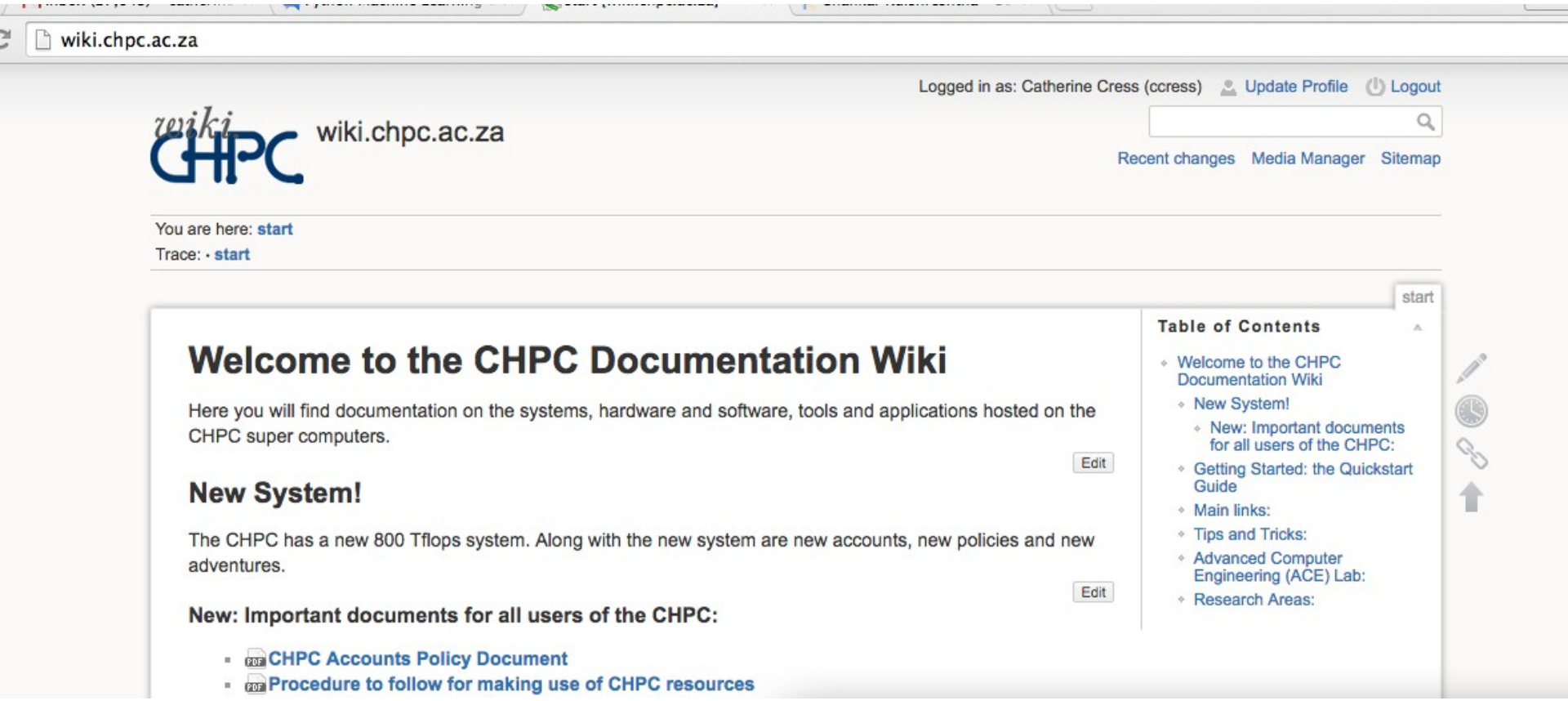
CHPC user support - database

Principal Investigator (PI) : permanent staff member at SA university/institute, with PhD & publications

PI applies, once approved, creates research project then adds team members (students etc)

Amolo, George	gamolo	georgeamolo862@gmail.com	University of Eldoret	Properties of Materials for Green Energy Harnessing
Anand, Krishnan	kanand	organicanand@gmail.com	Durban University of Technology	Combined kinetics,quantum-chemical investigation of reaction mechanisms involving catalysts
Anandjiwala, Rajesh	ranandi	ranandjiwala@nmmu.ac.za	Council for Scientific and Industrial Research	Mechanics of Flexible Fibrous Assemblies
Anderson, Matthew	None	andersmw@indiana.edu	0 Other	
Andeve, Calvin	candeve	215082608@stu.ukzn.ac.za	University of KwaZulu-Natal	IN SILICO EVALUATION OF NANO DRUG DELIVERY SYSTEMS
Andriambelaza, Noeliarinala Felana	None	arinala.f@gmail.com	University of Pretoria	
Andriambeloson, Joely	jandriambeloson	andyjoely@gmail.com	Stellenbosch University	Radio Frequency Interference Measurement, Monitoring and Mitigation for MeerKAT and SKA

CHPC user support - wiki



CHPC training

- Intro to Linux and python → data science
- Parallel programming
- Applications courses (usually at december conference)
- Research Scientist help

- **Student Cluster Competition:**
  Students compete to make best compute cluster
  S.African team won 3 times! (vs Chinese,Germans etc)

CHPC CENTRE FOR HIGH PERFORMANCE COMPUTING

CSIR our future through science

Cloud at the CHPC

Cloud for HPC Project
Virtual Clusters
- HPC Software Evaluation and R&D
- Clusters for training courses
- Support users with special computing needs
- Hybrid Clouds
- Clusters for the Student Cluster Competition

Public Cloud Prototype
Exploring OpenStack further
- New features, Docker integration, GPUs

Summary

Centre for High Performance Computing in South Africa:
> 500 users: chem, materials, climate, fluid engineering, film

Involved in SKA project: Astronomy driving tech dev → HPC, Big Data
Astrophysics projects: Nature of dark matter and dark energy?
Data, simulation, theory → Data science

Computing development in African SKA partner countries
old hardware in some places – few users

Who should use HPC?

→ User support at CHPC → training, in-house researchers, wiki, slack