Calibration of LOFAR data on the cloud

Jose Sabater Montes (jsm at roe dot ac dot uk)
Institute for Astronomy, University of Edinburgh

with S. Sanchez, J. Garrido, J. E. Ruiz, P. Best, L. Verdes-Montenegro and the LOFAR collaboration

Challenges

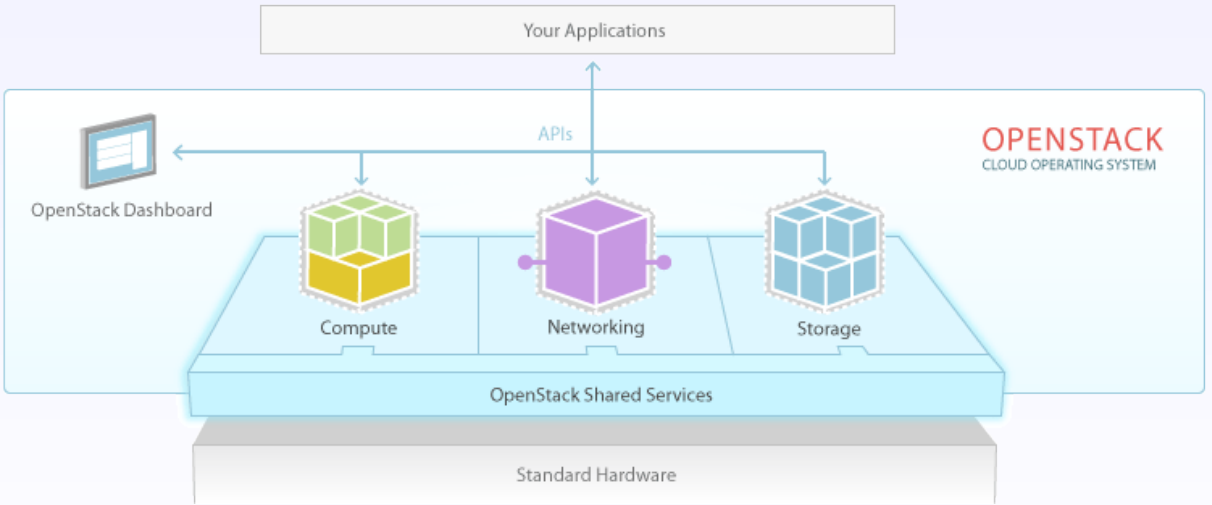
- User data calibration
  - Example:
    - 10 hours full resolution → ~20 TB
    - 2 CPU years to run the calibration
  - Experimental pipeline
- LOFAR calibration software
  - Difficult to install (this is improving quickly)
  - Continuous development

Possible solution

- Parallelizable:
  - Deal with a large amount of data in a reasonable time.
- Flexible:
  - Adapt the infrastructure (“hardware”) to different calibration strategies
  - Deal with quickly changing software

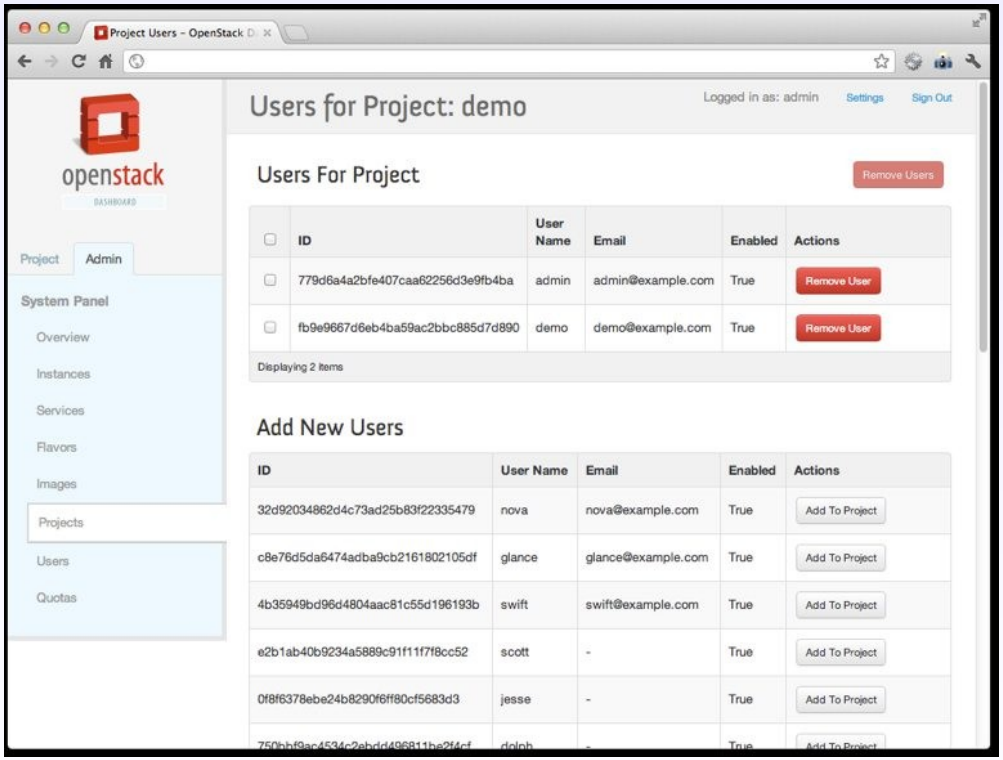
Cloud computing

- Infrastructure as a Service (IaaS)
- Examples:
  - Amazon Web Services (EC2, S3, etc), Google Compute Engine, RackSpace...
  - Eucaliptus, OpenStack...

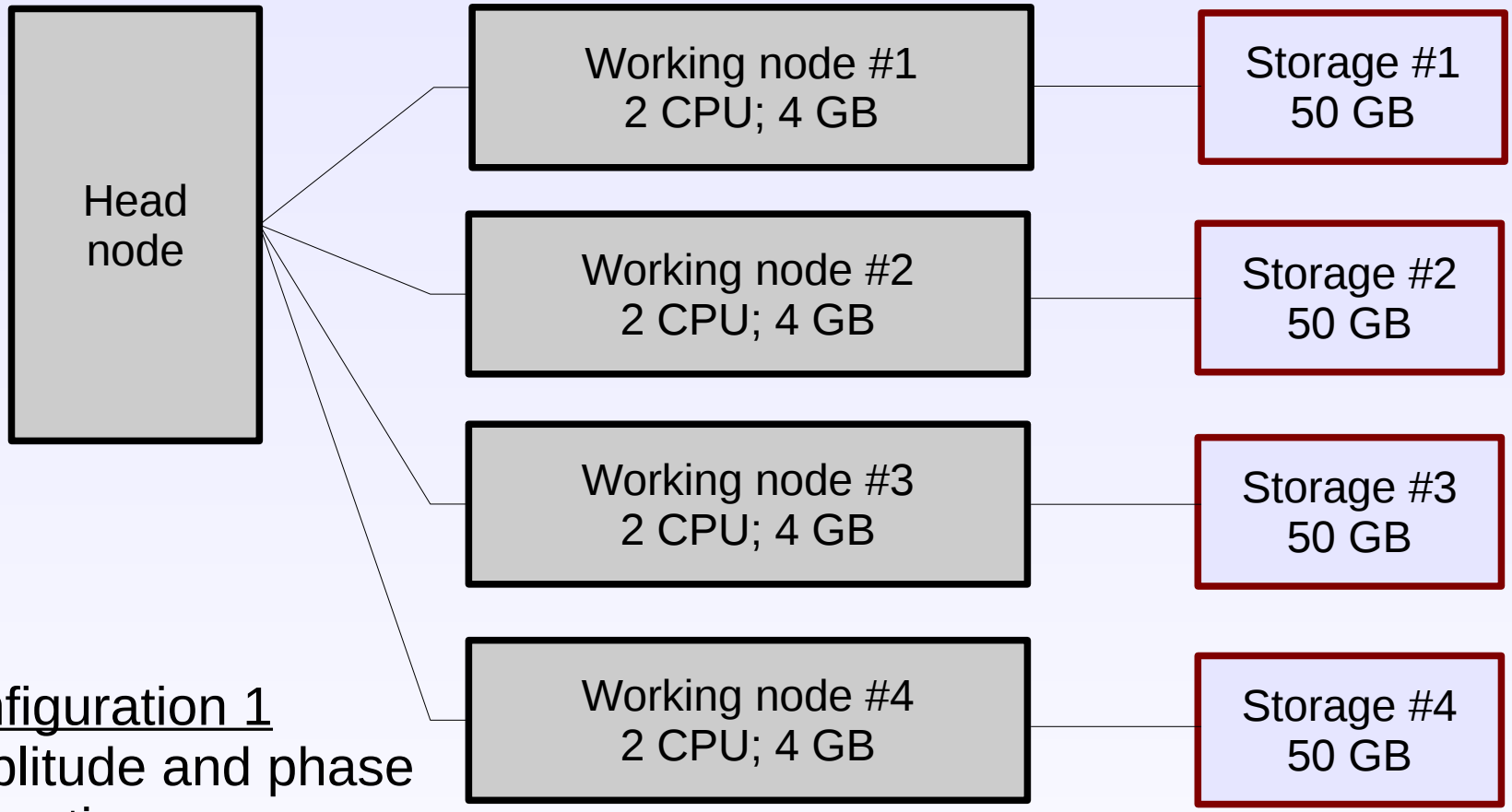


Tests on IBERCLOUD

- Cloud infrastructure provided by Ibergrid
- Implemented with OpenStack



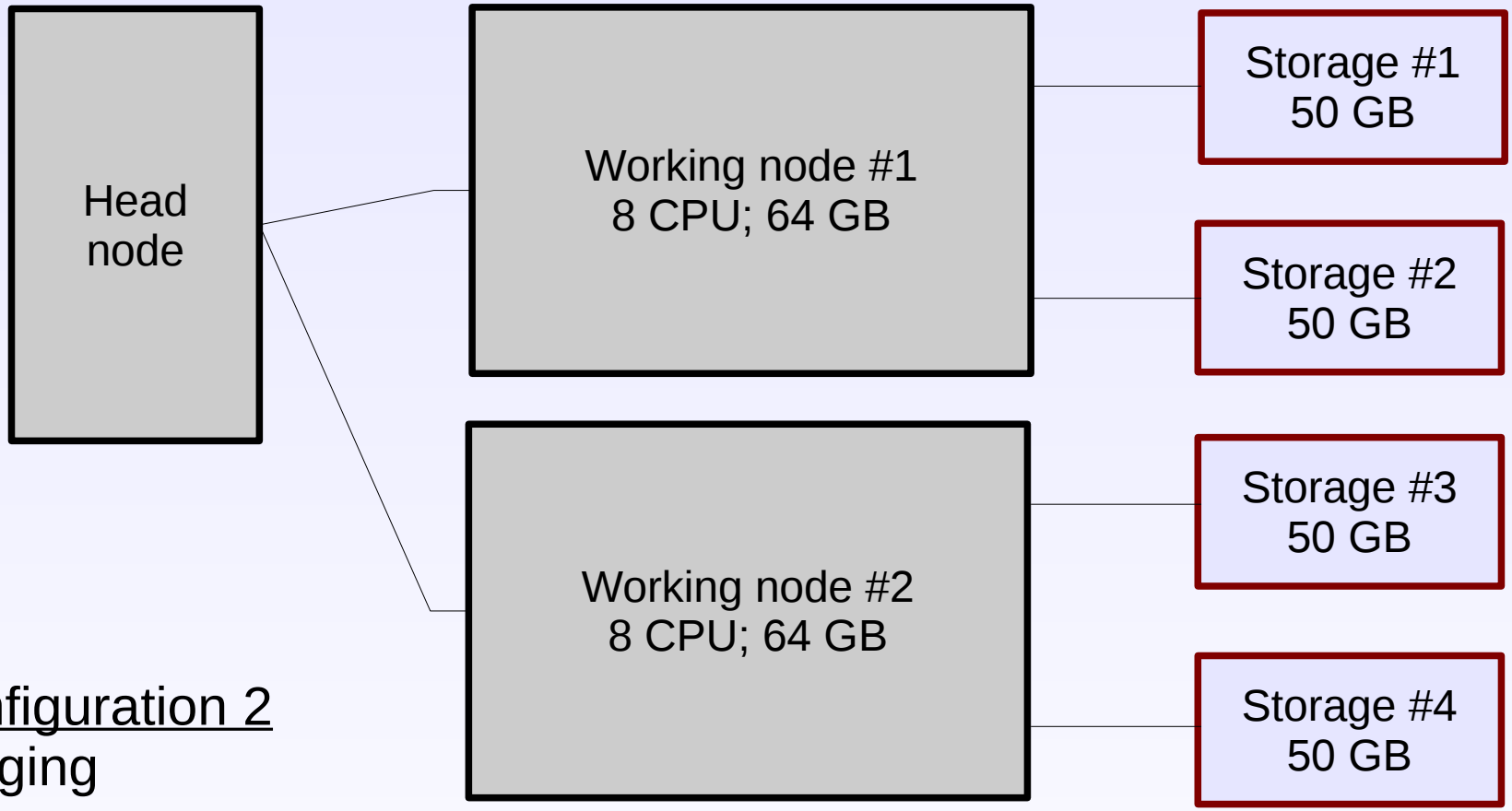
LOFAR pilot in Ibercloud



Configuration 1
Amplitude and phase calibration

(...)

LOFAR pilot in Ibercloud



Configuration 2
Imaging

(...)

Current status

- Current status:
  - LOFAR instance images created
  - Virtual on-demand cluster working
  - IPython parallel used to orchestrate the calibration on the nodes
- Next steps:
  - Additional testing
  - Use in production
  - Adapt to the new European Grid Infrastructure (EGI) Federated Cloud service.

Conclusions

- Cloud infrastructure to calibrate LOFAR data:
  - Elastic on-demand resource consumption
  - Parallellization – Ability to deal with big data
  - Flexibility - Quick development of innovative strategies