



JASMIN Cloud

ESGF and UV-CDAT Conference

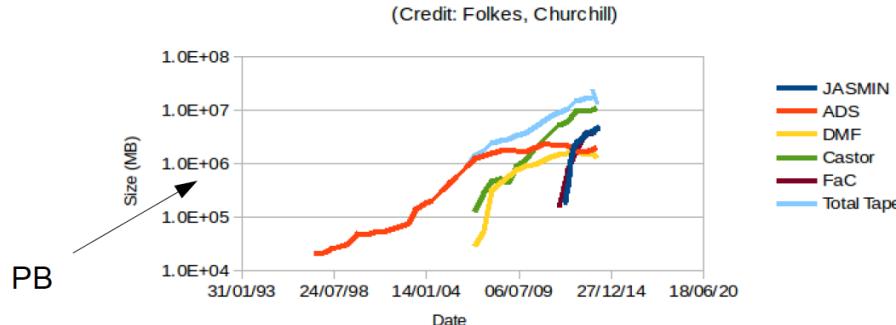
09-11 December 2014

Philip Kershaw (1, 2), Jonathan Churchill (5), Bryan Lawrence (1, 3, 4), Stephen Pascoe (1, 4) and Matt Pritchard (1) Centre for Environmental Data Archival, STFC Rutherford Appleton Laboratory, Didcot, UK

- (1) National Centre for Earth Observation, NERC, UK
- (2) Department of Meteorology, University of Reading, Reading, UK
- (3) National Centre for Atmospheric Science, NERC, UK
- (4) Scientific Computing Department, STFC Rutherford Appleton Laboratory, Didcot, UK

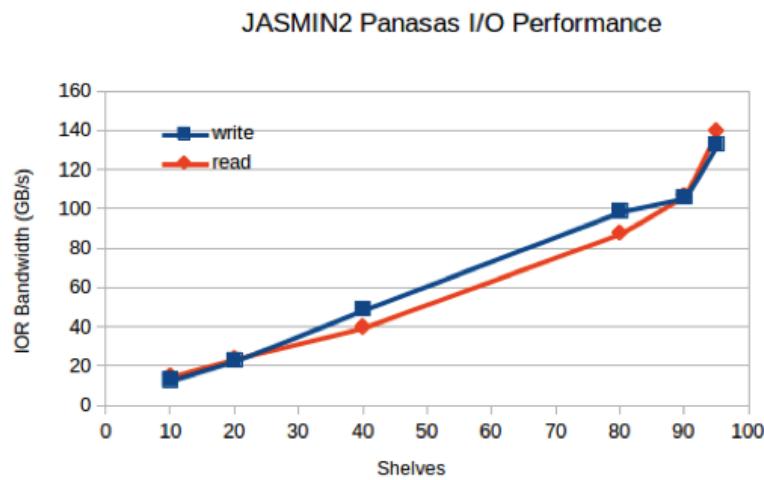


Introduction and Background

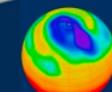


PB

Light blue = total of all tape at STFC
Green = Large Hadron Collider (LHC) Tier 1 data on tape
Dark blue = data on **disk** in JASMIN



130-140GB/s in IOR tests for the new storage



JASMIN 1 Success: UPSCALE

- UPSCALE: UK on PRACE – weather resolving Simulations of Climate for globAL Environmental risk
- Ensembles of global atmospheric climate simulations at weather forecasting resolution
- Required more than 30 times the computing time available to our team on UK supercomputer HECToR
- Successfully applied for a 144 million core hour from PRACE lasting for 1 year on HERMIT in Germany
- Produced more than 400 TB of data over 10 months, which was shipped to JASMIN and the Met Office archives
- Deployment of VMs running custom scientific software, co-located with data

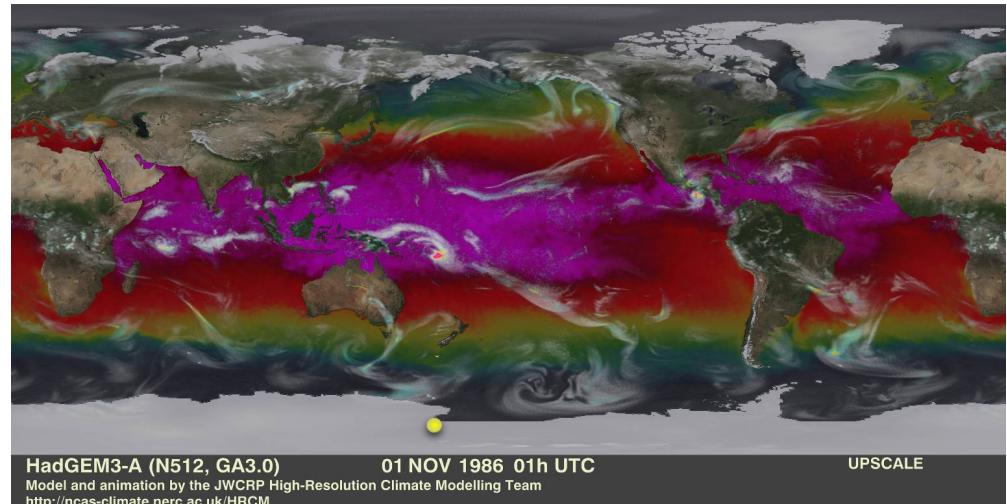


Image: P-L Vidale & R. Schiemann, NCAS

Mizielinski et al (Geoscientific Model Development, submitted)
“High resolution global climate modelling; the UPSCALE project, a large simulation campaign”



Does Cloud Computing fit?

“Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction.” – NIST SP800-145

5 essential characteristics

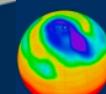
- On-demand self-service
- Broad network access
- Resource pooling
- Rapid elasticity
- Measured service

3 service models

- IaaS (Infrastructure as a Service)
- PaaS (Platform as a Service)
- SaaS (Software as a Service)

4 deployment models

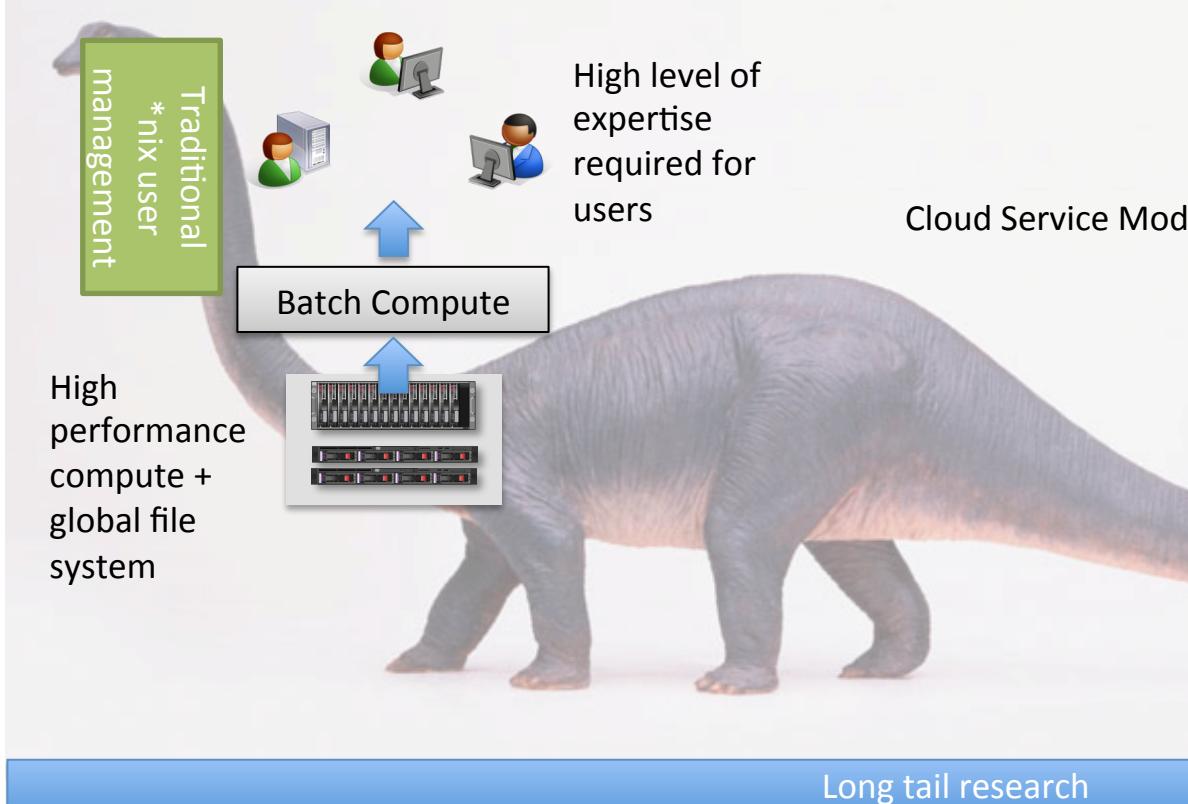
- Private cloud
- Community cloud
- Public cloud
- Hybrid cloud



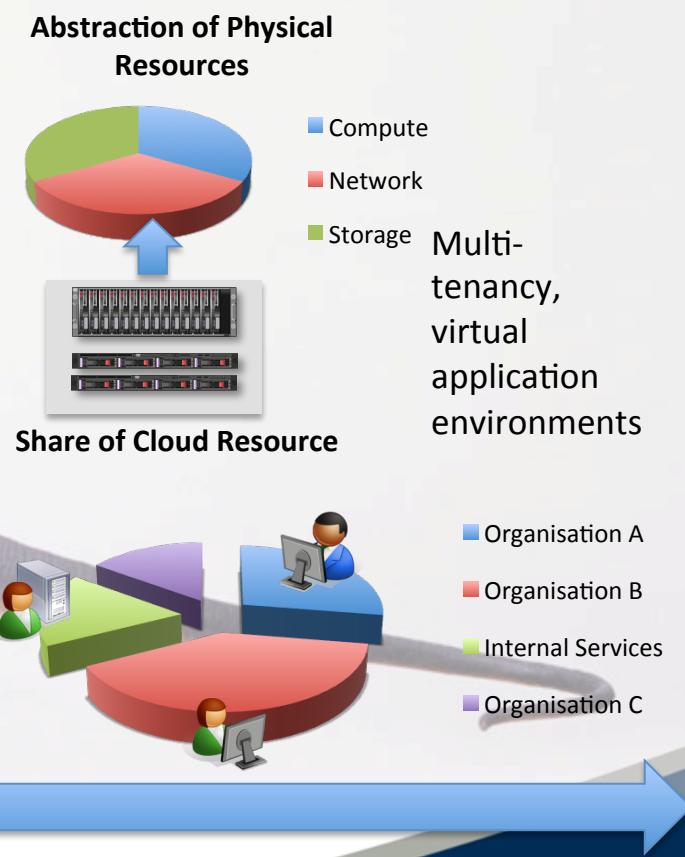


Lessons Learnt from JASMIN I and The long tail

Raw infrastructure power (data available all the time, next to the compute) but more constrained service model

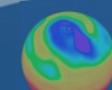
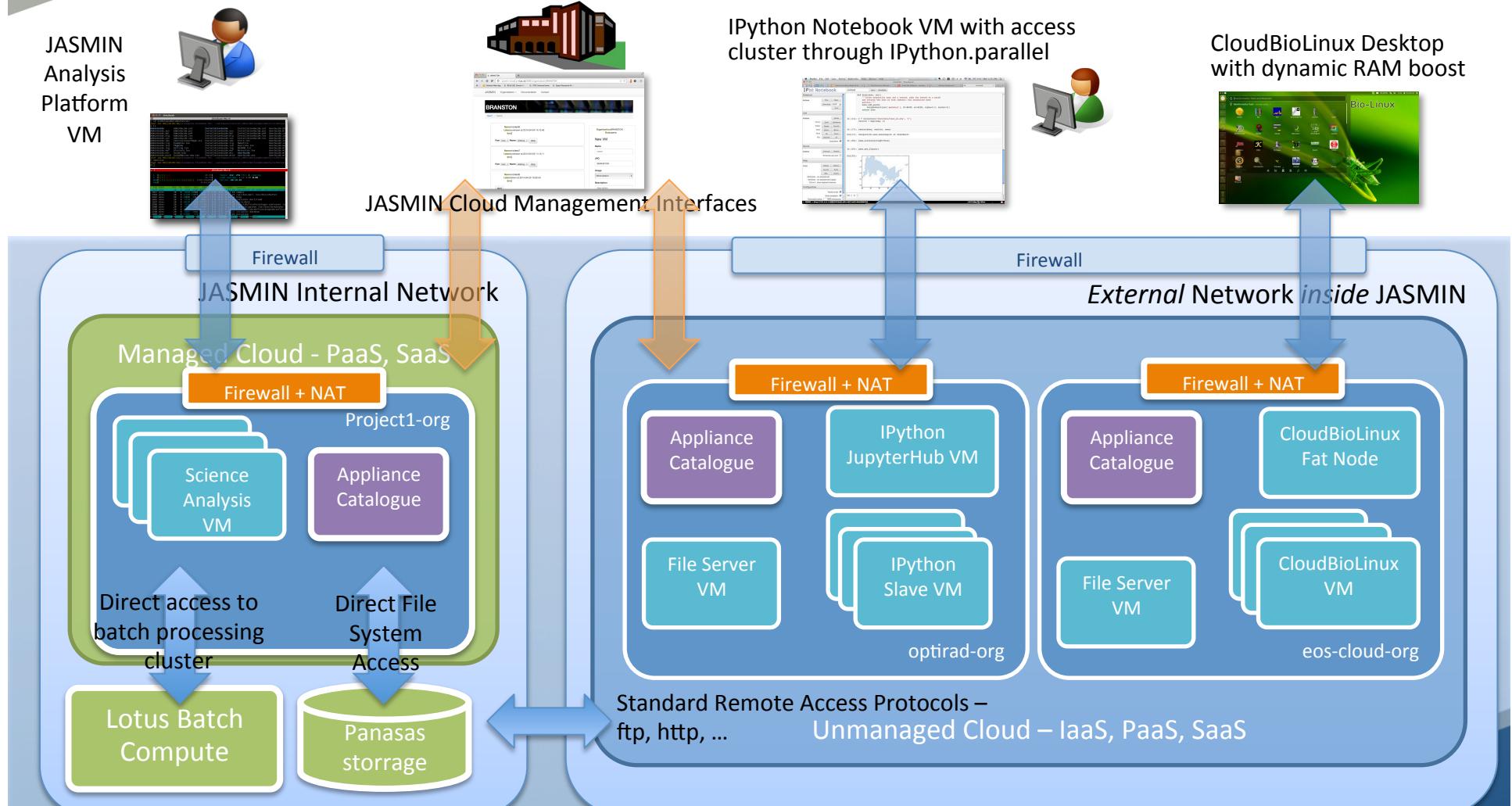


Rich and flexible service model but sacrifice in performance



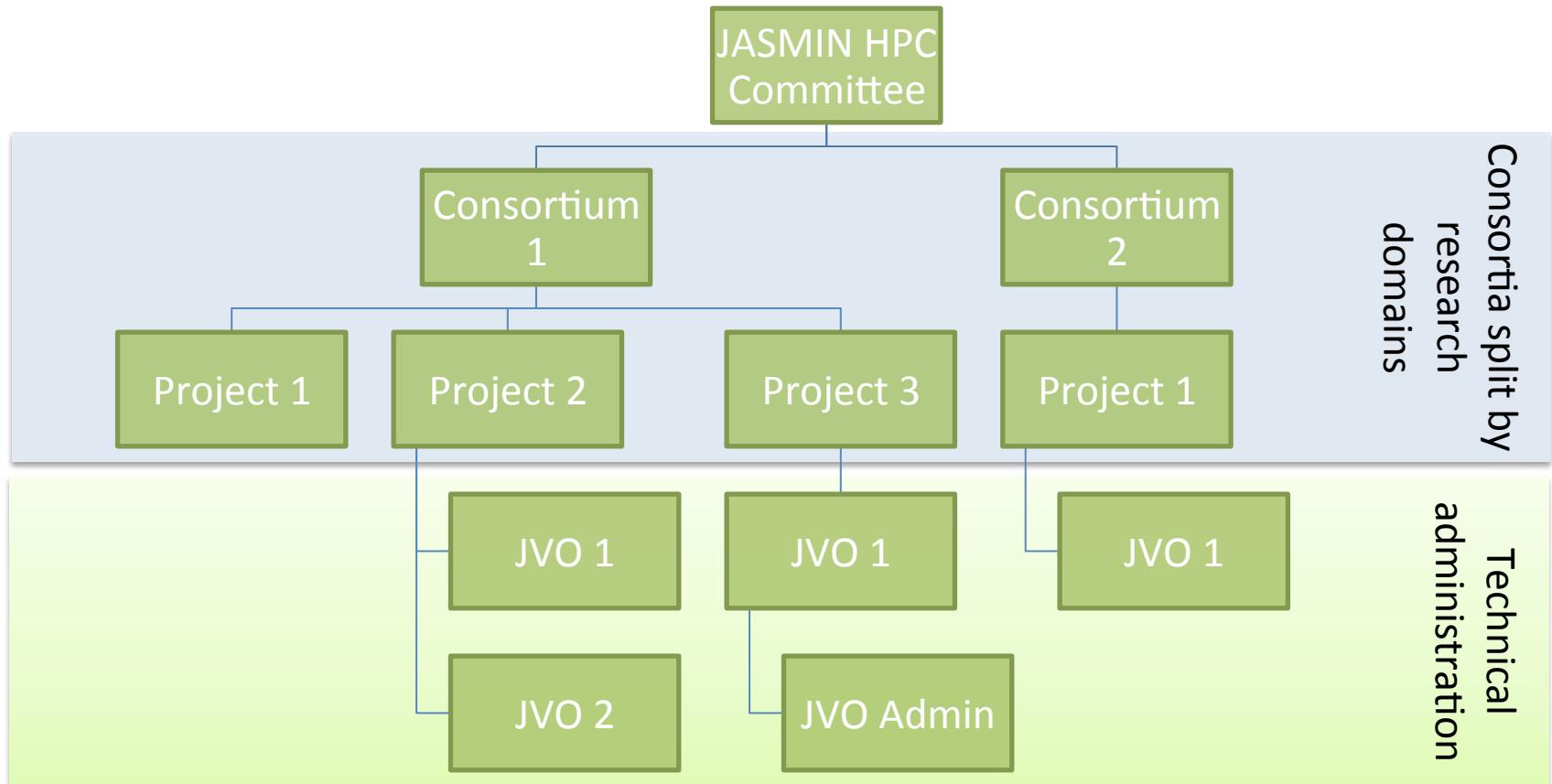


JASMIN Cloud Architecture





Hierarchical Organisation and Governance Structure

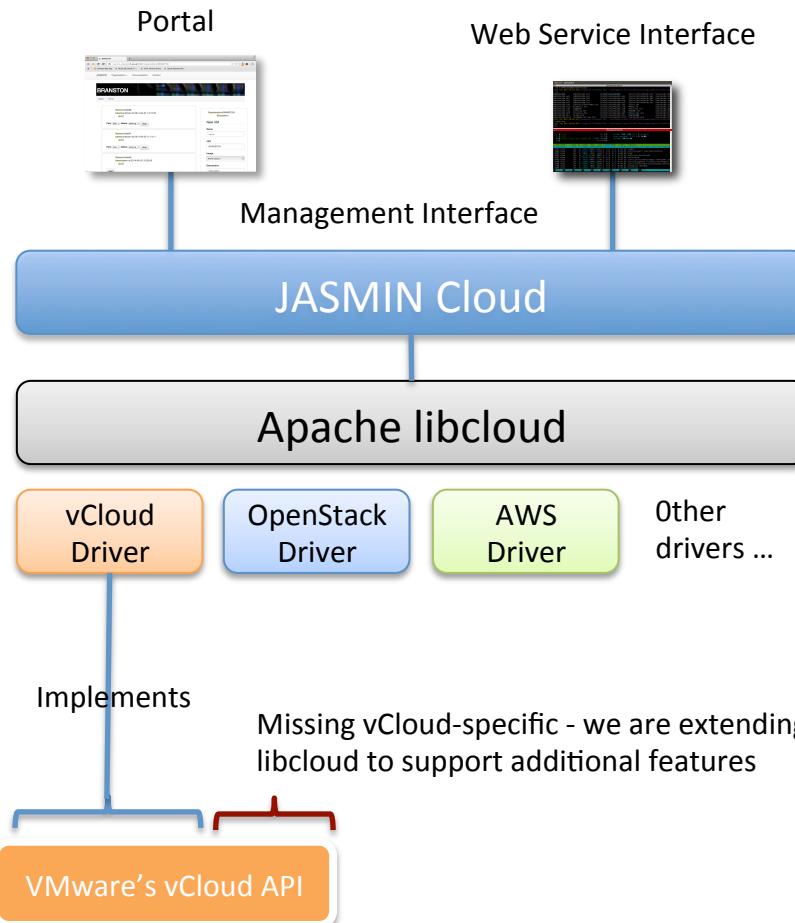


JVO = JASMIN Virtual Organisation

Manage Complexity with Documented Processes

- Analysis
 - The meeting of requirements with the application of the available technology
 - Defined a set of processes
 - Each is documented
 - A combination of manual and automated steps with fallback to manual steps where needed
- Project negotiation
 - Match proposal with a JASMIN Consortium
 - Proposal sponsor negotiates with consortium manager(s) – amount and type of resources – number of Virtual Orgs, compute, network and storage quotas
- JVO (JASMIN Virtual Organisation) *Onboarding*
 - Create a virtual organisation tenancy
 - Register responsible administrator(s)
 - Induction + training
- Management of a *Catalogue of Appliance Templates*
 - Create new appliance
 - Update appliance
 - Propagation to organisations
 - Control for Managed and Unmanaged environments, stem cells
- Networking
 - Manage quota of public IPs for tenancy

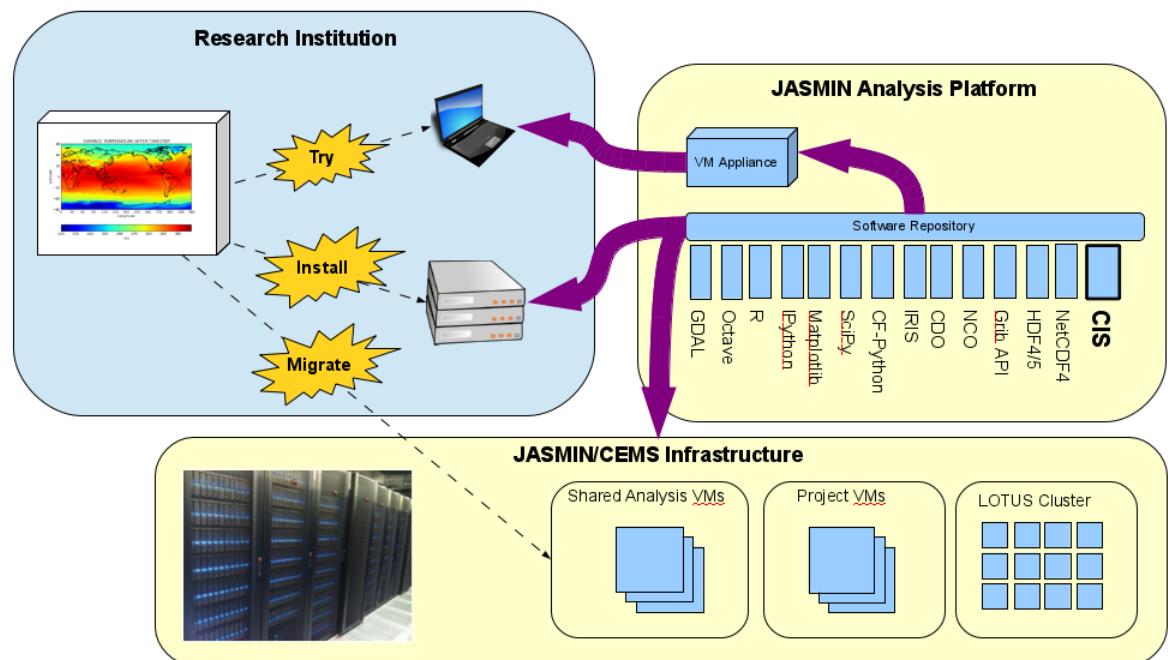
Wrapping vendor-specific Cloud APIs



- VMware vCloud Director was selected as the private cloud implementation:
 - For its maturity vs. Open Source alternatives at the time of JASMIN I
 - Capital funding favoured it vs. increased recurrent of Open Source
- Cloud broker technologies provide an insulation layer from cloud provider-specific APIs
 - Apache libcloud (Python) selected
- This allows
 - Future use of an alternative private cloud implementation e.g. OpenStack
 - Implementation of *Cloudbursting* – augment private cloud with access to Public cloud providers via APIs supported in libcloud
- Its not a panacea:
 - Libcloud provides a common subset of the features of each cloud provider supported

JASMIN Analysis Platform (JAP)

- Multi-node infrastructure requires a way to install tools quickly and consistently
- The community needs a consistent platform where ever they need them.
- Users need help migrating analysis to JASMIN.
- JAP provides RPMs and pre-built images based on CentOS



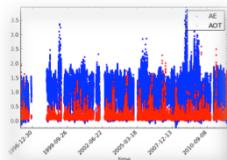
<http://proj.badc.rl.ac.uk/cedaservices/wiki/JASMIN/AnalysisPlatform>



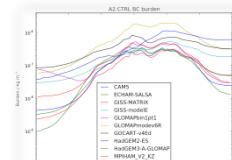
Community Intercomparison Suite (CIS)

- A command line interface and Python APIs for ...

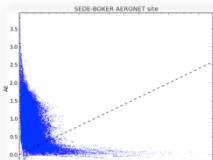
Time-series



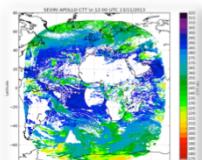
Line plots



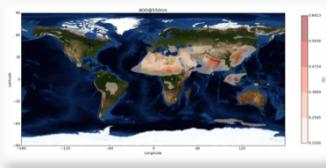
Scatter plots



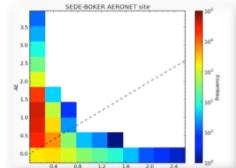
Global plots



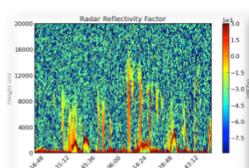
Overlay plots



Histograms



Curtain plots

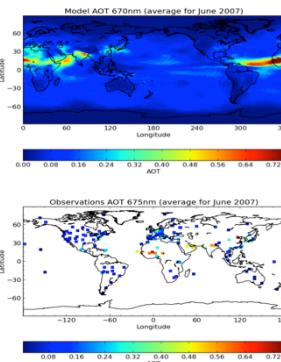


Co-location

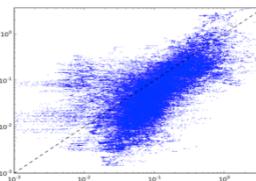
Model gives global output every 3 hours for a full month

Observations are day-time site measurements, every 15 min for a full month

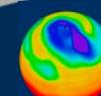
Sampling



Collocation



Dataset	Format
AERONET	Text
MODIS	HDF
CALIOP	HDF
CloudSAT	HDF
AMSRE	HDF
TRMM	HDF
CCI aerosol & cloud	NetCDF
SEVIRI	NetCDF
Flight campaign data	RAF
Models	NetCDF



Early Results for JASMIN II

- Development and rollout a huge challenge
 - Manage complexity with well documented processes
- User community
- Model emerging of PaaS (Managed Cloud) + IaaS (Unmanaged Cloud)
 - PaaS with access to HPC and global file system
 - IaaS for application and dissemination of outputs to wider community
- Fellow tenants helping each other out and developing common solutions
- IPython Notebook with JupyterHub
- NERC Environmental Workbench
 - Docker
- JASMIN Analysis Platform
 - Try out JASMIN-like env in VirtualBox

Further information

- JASMIN website:
 - <http://jasmin.ac.uk/>
- Centre for Environmental Data Archival
 - <http://www.ceda.ac.uk>
- JASMIN paper (Sept 2013)
 - http://home.badc.rl.ac.uk/lawrence/static/2013/10/14/LawEA13_Jasmin.pdf
 - Follow-up paper on cloud work planned
- JASMIN Analysis Platform
 - <http://proj.badc.rl.ac.uk/cedaservices/wiki/JASMIN/AnalysisPlatform>
- Apache libcloud
 - <https://libcloud.apache.org>