Geophysical Research Abstracts
Vol. 16, EGU2014-10820-3, 2014
EGU General Assembly 2014


# The JASMIN Cloud: specialised and hybrid to meet the needs of the Environmental Sciences Community

Philip Kershaw (1,2), Bryan Lawrence (1,3,4), Jonathan Churchill (5), Matt Pritchard (1,4)

(1) Centre for Environmental Data Archival, STFC Rutherford Appleton Laboratory, Didcot, UK, (2) National Centre for Earth Observation, UK, (3) Department of Meteorology, University of Reading, Reading, UK., (4) National Centre for Atmospheric Science, UK, (5) Scientific Computing Department, STFC Rutherford Appleton Laboratory, Didcot, UK

Cloud computing provides enormous opportunities for the research community. The large public cloud providers provide near-limitless scaling capability. However, adapting Cloud to scientific workloads is not without its problems. The commodity nature of the public cloud infrastructure can be at odds with the specialist requirements of the research community. Issues such as trust, ownership of data, WAN bandwidth and costing models make additional barriers to more widespread adoption.

Alongside the application of public cloud for scientific applications, a number of private cloud initiatives are underway in the research community of which the JASMIN Cloud is one example. Here, cloud service models are being effectively super-imposed over more established services such as data centres, compute cluster facilities and Grids. These have the potential to deliver the specialist infrastructure needed for the science community coupled with the benefits of a Cloud service model.

The JASMIN facility based at the Rutherford Appleton Laboratory was established in 2012 to support the data analysis requirements of the climate and Earth Observation community. In its first year of operation, the 5PB of available storage capacity was filled and the hosted compute capability used extensively. JASMIN has modelled the concept of a centralised large-volume data analysis facility. Key characteristics have enabled success: peta-scale fast disk connected via low latency networks to compute resources and the use of virtualisation for effective management of the resources for a range of users.

A second phase is now underway funded through NERC's (Natural Environment Research Council) Big Data initiative. This will see significant expansion to the resources available with a doubling of disk-based storage to 12PB and an increase of compute capacity by a factor of ten to over 3000 processing cores. This expansion is accompanied by a broadening in the scope for JASMIN, as a service available to the entire UK environmental science community.

Experience with the first phase demonstrated the range of user needs. A trade-off is needed between access privileges to resources, flexibility of use and security. This has influenced the form and types of service under development for the new phase. JASMIN will deploy a specialised private cloud organised into "Managed" and "Unmanaged" components. In the Managed Cloud, users have direct access to the storage and compute resources for optimal performance but for reasons of security, via a more restrictive PaaS (Platform-as-a-Service) interface. The Unmanaged Cloud is deployed in an isolated part of the network but co-located with the rest of the infrastructure. This enables greater liberty to tenants – full IaaS (Infrastructure-as-a-Service) capability to provision customised infrastructure - whilst at the same time protecting more sensitive parts of the system from direct access using these elevated privileges.

The private cloud will be augmented with cloud-bursting capability so that it can exploit the resources available from public clouds, making it effectively a hybrid solution. A single interface will overlay the functionality of both the private cloud and external interfaces to public cloud providers giving users the flexibility to migrate resources between infrastructures as requirements dictate.