EGI Federated Cloud Task Force – PlugFest – 12-13th July 2012

This two day event was organised to bring the TF together with the user representatives to review progress with the use case prototypes being developed for the demonstrator.

# Programme

Thursday - 12 July:

11:00 - 11:15 Welcome

11:15 - 11:45 Introduction to the plug fest goals and objectives

11:45 - 12:30 Use case: BNCweb/CLARIN: presentation (Martin Wynne) + discussion

12:30 - 13:45 Lunch

14:45 - 14:15 Use case: Presentation + discussion

14:15 - 15:00 Use case WeNMR: Presentation Marc Van Dijk + discussion

15:00 - 15:15 Coffee break

15:15 - 16:00 Use case: P-Grade: presentation (Peter Kacsuk) + discussion

16:00 - 16:45 Use case: Ngram Viewer (Music): presentation (Vladimir Viro) + discussion

16:45 - 18:00 Panel: FedCloud TF and use cases

19:00 - 21:00 Dinner

Friday - 13 July:

09:00 - 10:15 Path towards the integration of use cases into the test bed

10:15 - 11:00 Use case: GAIA Space (Astro): Presentation (Nic Walton) + discussion

11:00 - 11:30 Coffee break

11:30 - 12:30 TF third phase: integration into the EGI infrastructure
Attendees

## Present in Amsterdam:

Stephen Brewer (Chair); Karolis Eigelis; Nuno Ferreira; Jhon Masschelein; Richard McLennan; Alison Packer; Gergely Sipos; Peter Solagna; Jelena Tamuliene; Matteo Turilli; Marc van Dijk

## Online:

Bjoern Hagemeier; Peter Kacsuk; Gabor Kecskemeti; Josef Pacula; Alvaro Simon; Kostas Koumantaros; Roberto Rosende, Fraser Barnsley, Stuart Kenny, Vladimir Viro, Nicholas Walton, Guy Rixon, Victor Mendez, Roberto Dopazo

Individuals will be referred to by their initials in these notes, the Task Force itself as FC-TF

## Notes:

## 11:00 - 11:15 Welcome

The participants were welcomed and introduced as they joined the call.

## 11:15 - 11:45 Introduction to the plug fest goals and objectives – SB chair, and MT leading discussions.

MT explained that the goal for the day was to learn more about the details of the usecases currently embedded in the FC-TF. This would involve presentations from the various participating user communities and the capturing of their requirements.

There was a question about whether more user communities were needed. MT responded that no more user communities were needed for now.

11:45 - 12:30 Use case: BNCweb/CLARIN: presentation (Martin Wynne) + discussion
MW was unable to dial-in so MT gave the talk and shared the slides.

As a result of presenting this use case there was a discussion about various aspects of the provision and monitoring of VMs.

1. EGI is essentially offering consultancy in delivering VMs – we cannot know what happens inside the VM
2. Should we support ‘cold suspension’? ie. no resource consumption = no cost? No, suspension has a cost which must be measured and accounted for.
3. So, we need a full economic model which must come from ‘the heaven of EGI’ – is this a policy issue? Yes, it probably is.
4. SARA has experience of this – AP also has knowledge of Accounting URs
5. FC-TF should deliver multiple options for accounting and budgeting of resources
6. Avoid overcommitting – this would not be efficient planning of resources
7. Accounting – SARA would say 100% irrespective of actual usage
8. FC-TF does not have to deal with accounting and billing (out of scope), but we should advise on, or at least signpost, the issues and opportunities.
9. 3 ways to offer VMS: 1) cold suspension – no resources used, 2) hot suspension – resources used, and 3) permanent services – continuous (SARA just have 2)
10. VM/resource delivery model – we should compare each scenario
11. Action NF to talk to Jhon - Glossary needed to clarify terminology – update or extend EGI
12. Auto span out to other providers? Not now because of technological feasibility.
13. Access provision: certificates should be avoided where possible – username and password is preferable to many in many scenarios.

Questions for MT to Martin:

What information do you need to persist – so that we know what data we need to maintain as persistent storage? Availability? – local installation, hot migration – a different thing? Ability for users to move to different locations.

1. Brokering - needed if there is a need to manage where VM lands, by criteria
2. Scalability – VM instances not interconnected so this is really availability or capacity
3. Would you expect to span across other partners?
4. Federated authentication – possibly with help from the EGI technical outreach team

## 12:30 - 13:30 Lunch

13:30 – 14:45 – Continued discussion

1. Gergely’s team can promote the FC-TF use cases, handle the activity and build up a portfolio of resources and support documentation
2. Contrail project discussion with Jens and Bjorn – internal group
3. Set up Requirement Tracker channels for the different use cases

## 14:15 - 15:00 Use case WeNMR: Presentation Marc Van Dijk + discussion

|  |
| --- |
| **Background information from Alexandre Bonvin (WeNMR):**1. abstract submitted to EGI-TF12 together with the WNoDeS team for a demo that, among others, will be using the VCing Virtual Machine within WNoDeS- G.Vuister's team, a partner of WeNMR, ensured they'll do their best (they have limited manpower) to provide a VCing image from where the licence-sensitive programs are stripped out
2. this VCing image has to be converted into KVM format as required by WNoDeS
3. an input dataset for the demo (of ~0.4 MB) to be validated and recalculated by VCing and the validation script to be executed on VCing are already available
4. once the KVM VCing image is ready (hopefully before September) it will be deployed on the WNoDeS testbed at INFN-CNAF and we'll start to understand how things work and to define a demo that makes sense
5. after that, once they have gained enough experience with VCing and the cloud, when an execution model with WNoDeS cloud has been implemented and all related issues understood, maybe they might think of uploading VCing on the Federated Cloud marketplace and see how things can work there.
 |

* VM currently running at SARA.
* Each student gets their own VM for a few days
* Virtual Cing. Pilot machine creates multiple instances of VMs – students interact with VM, others don’t, there is a portal. Typically around 30 students at a time
* Licences – restrictions on VMs – cannot share freely
* Should be get involved in this or restrict eg. Who can download
* Web interface appreciated, CLI not an obstacle
* Ganglia for each VM – too much work/incompatible with several VM delivery models
* May want to extend monitoring FC-TF activities in 6months to monitor state of the VM (up/down/suspended/volumes attached)
* We have two types of users: VM managers and users/consumers of the VM services
* MT - We may be closer to solving the training scenario than we first realised
* Commercial issues: BitBrain commercial partner handles part of this
* MT - It will be necessary to convert the VMware instance to KVM for our demonstrator, using OCCI as management interface.
* ACTION – Marc to discuss with AB – SB to follow up
* Tech questions – does CPU speed matter? Do you care or just any? Intel v. AMD? Does pilot Manager care? No. (Swarm) MT – This is a well-matched scenario
* WeNMR will have to work on developing the connector (OCCI) – check this - SB

## 15:00 - 15:15 Coffee break

15:15 - 16:00 Use case: P-Grade: presentation (Peter Kacsuk) + discussion

* PK presents slides on SCIBUS
* Cloud Broker service will be ready in a couple of weeks
* Very thorough slides indicating offerings, needs, issues and suggested plans including project (available on Indico)

Discussion:

* MT explained that we are not focussed on StratusLab solutions – we have a broad spectrum usage.
* No decision by FC-TF on a particular repository – this can evolve as availability and needs evolve.
* Action: MT and PK to have further 1-1 discussion. Possibly to discuss OCCI after both have two weeks holiday.
* Action SB/GS to talk to MT about AppDB evolution

16:00 - 16:45 Use case: Ngram Viewer (Music): presentation (Vladimir Viro) + discussion
Vladimir described the Peachnote NGram viewer

* Images currently stored on Google Storage
* Annotations persist on Google App Engine – ‘information not as extensive as Wikipedia’
* Images more interesting – take some traffic
* Google money, VV is not even an employee of University, so this funding is independent and owned by VV.
* Where could we help? Processing of music sheet images – 1 sheet take 30 seconds x a few million items – this is an important background task that is independent of the online service.
* Data sets are constantly growing
* 1000 scores a week are being added – typically 13 pages each
* Only 30% of one library done so far
* Also, the database where these are stored is managed using Hadoop
* Index creation has recently been improved
* 1TB current total for data with some breathing space
* Remember Hadoop stores data redundantly – 3TB
* Stability important – including communication

Discussion:

* MT – We cannot currently help with the production service. But we can help with the processing – is this of interest?
* We need an architecture where you use our services when they are available
* VV- Hadoop not suitable for this test but…
* A small library or collection needs to be prepared – a few GB of images supplied and need processing asap
* VV and GS to follow up – to finalise requirements within the Requirement Tracker
* What is needed is a simple machine for VV to upload image and fire up and spin down, update image occasionally
* 16:45 - 18:00 Panel: FedCloud TF and use cases
MT Three interesting user communities
* We saw how we can work with them to gather requirements
* We will have to deal with how we move from prototype to production facilities
* Table comparing requirements across use cases would be a good idea - put this on wiki
* Hadoop? This is not a usecase for now
* Use of testbed as middle ware backend good idea
* Heterogeneous set of use cases – they will need to speak OCCI – could be a challenge
* Another issue could be how users get access to facilities – certificate? Try and avoid cert.
* Contrail might work out of the box

Another problem – stability, how to ensure this

Flexible accounting system needed

Not a strong need for interconnecting services at this point in time

* Roadmap? Plan next steps
* Choose a few do-able challenges and move forward
* Use EGI Requirements Tracker to capture and monitor progress with requirements
* Identify some resource providers, get them working and put in demo
* How does these activities relate to the accounting work groups etc? Clarify
* Separate things. But we need to demonstrate that science can be done with this technology
* Development will be user-informed

19:00 - 21:00 Dinner

Friday - 13 July:
Apologies – PS

09:30 - 10:15 Path towards the integration of use cases into the test bed
Now an official task – update on SN’s email to the FC-TF

Welcome and recap of yesterday – actions:

1. Glossary – update EGI Glossary so that we are using terms consistently
2. Requirements Tracker – use to capture and progress requirements
3. Some actions for each usecase needed
	* CLARIN/BNCweb
	* WeNMR – VCing, teaching next iteration
	* Peachnote – backend processing
	* PGRADE – further discussions
4. Clear focus on demo – not yet on production services

10:15 - 11:00 Use case: GAIA Space (Astro): Presentation (Nic Walton with Guy Rixon) + discussion
Full presentation from NW and GR with Taverna demo – see slides on Indico site.

Discussion

* NW where do we go for definitive info – Answer: we will create single wiki page with appropriate links ACTION, on whom?
* MT: we need more details
	+ Face to face meeting – in Cambridge – when? Before November
	+ ½ day last day of September – 26th?
	+ Interested in the prospective charging model – compared to Amazon for example – we can do calculations. Evolving – but definitely measuring usage – cost models are changing
* MT – Do you move a lot of data? We can, so yes and no
* Quite a bit of local processing but typically remote
* Some work requires generation of models and hence data exract
* Combinatorial stage generates output
* Number of scenarios
* Trying to avoid shifting data
* File i/o costly (eg Amazon)
* MT – are you interested in Amazon model? NW Yes, in theory
* Much science requires on demand access to data for exploratory investigations (perhaps two weeks) so need to call the process etc. to achieve this
* Now setting up infrastructure to support the release of the GAIA data
* Enable people to upload apps to run this against data in VMs
* Must be easy to set up
* Two ways of looking at this - GR
1. Scientist run something in cloud
2. Pre-existing service eg Taverna, host that in the cloud, load balance, greater usage.

Thanks and goodbye – NW to send slides

## 11:00 - 11:30 Coffee break

## 11:30 - 12:30 TF third phase: integration into the EGI infrastructure (post TF12 for 6m)

MT – review

Make a case for our usefulness

Transition from prototypes to production phase

This is a big task – hence has become an EGI-InSPIRE task

Now need to brainstorm – especially from the work groups – for next 6 months

AP - accounting

* In next 3 months – put what we do have on Accounting portal with help from CESGA guys
* Sub-step – we don’t yet have storage accounting in place (separate from cloud work) important for some centres. Is there a roadmap for this? Not really, no records have been kept so far, there has been no need for records so far. AP volunteered to move on from this situation by starting to keep records so that they can be investigated.
* Used v. allocated resource records

AS – accounting

* Do we use Nagios?
* MT monitoring and brokering is coming – this is needed by the user groups, and also single entry point
* MT🡪 AS please capture the needs

Bjorn – AAI

* What have we learned in last two days?
* AAI is separate thing from user aspects, although they will need something
* We need to manage carefully access to the market place especially with respect to competitive providers
* CONTRAIL will be of interest – may be able to reuse

Stuart – Marketplace

* StratusLab instance – images contextualised for StratusLab
* Not meant to be a full set of grid m/w
* Images which are endorsed – endorsers need to know their responsibilities
* SL will become OS project, discuss with SN what options are
* SL instance running at TCD who want to continue – although it may move to a different instance
* EGI instance is run from Greece

MT/SB

Wrap up and close meeting

Note – SARA has monitoring screen to show available resources - can this be integrated with Nagios?

# Notes:

The meeting was hosted from Amsterdam with two different video-conferencing technologies. The video suite in Nikhef H3.31 worked quite well but not everyone managed to connect despite using WebEx. However the Webex meeting on the Friday in the non-video room worked very well using WebEx. NW was able to make his presentation using screen share and participate through his video link. An external webcam at the Amsterdam end would have been preferable to laptop cameras but that worked ok.

We stuck closely to the agenda so that external participants knew when to join. Most users just joined for their section with the exception of MvD from WeNMR who attended most of the meeting.

The meeting was chaired (and hence time-managed) by SB with MT leading the technical discussions and Richard managing the technical communication channels. This breakdown of roles is useful to run a long distributed meeting efficiently and successfully.