3.3 Horizon 2020 Workshop - Breakout session: Cloud and Data

Date: 6th Dec. 213
Time: 14:00
Minuted by: S. Pinto, N. Ferreira
Session Leader/Convener: David Wallom

Table of Contents

Agenda
 Section 1: Models and technical challenges for the FedCloud
 Section 2: Participation to H2020 calls

Section 1: Models and technical challenges for the FedCloud

Data access and discovery model

?   : This slide seems close to what EUDAT box solution.
Action : We need to understand how EUDAT services could be integrated into the core
?   : In EUDAT we have 2 technical solutions: own cloud (sits in the FS), and PowerFolder.
?   : The idea fits in the overall workflow concept of EUDAT

?   : EUDAT simple store integrates Invenio technology for data publishing
?   : More than be implemented via EUDAT services, this seems to fit the overall workflow EUDAT idea.

?   : What is the situation to process multi-dimensional data?
D. Wallom   : Within the FedCloud we are dealing at lower lavel. We are working at the STaaS level.

?   : It would be good the researcher/universities to have his device with data storage integrated. This could be seen as a storage, the data does not sit only at the data center. This service shall include an easy way to implement this, to see the storage as a simple single entity where the user can put the files.
S. Pinto   : Yes, this was a requirement from the communities not listed. The ENVRI solution we are integrating in the FedCloud satisfy this requirement.

G. Sipos   : Can we support in the FedCloud, legacy data sets? They do not want to copy the data to EGI, they want only to make it available to VM's operated in the cloud.
S. Pinto   : Yes, in the requirement the usage of STaaS is optional, you can use your own data hosting and the Data catalogue and discovery service on top of that.

?   : Data mining, information discovery, post analysis is not listed here as requirement. What do you offer more than just cataloguing or storing of the data?
EUDAT   : Do not provide any service for data mining. We did some tests, with Hadoop, but we do not have a solutions.
J. Gordon : The layer that is running data mining sits on top of this.
Bjorn   : You want to have the data close to where data is processing. Are capacities with processing/data storage available? Do we need staging?
EUDAT   We have data staging services.

Summary
EUDAT services are in early stage and do not seem to provide clear solutions for that. Anyway there is the need to investigate more the EUDAT solutions and see if we can integrate them in the EGI FedCloud.

Personal storage

Patrick   : For this, we have WebDAV, pNFS If you need direct access to the data, and NFS allows to do this.
S. Pinto   : We need to consider that it should be mountable by the user PC, which is in a network who usually does not allow protocols other than HTTP and few others.
?   : WebDAV does not provide well in all situations.

?   : Do we have some instances of these services that we can instantiate on the FedCloud to have this solution on as additional service?

D. Wallom   : There are solutions in the EGI area and these can be moved to the FedCloud.

Summary

Need to port the WebDav personal storage based services available in EGI or the GRNET Pithos+ solution to the fedcloud.

STorage-as-a-Service

Storage accounting:

D. Wallom   : How are we going to do that?

J. Gordon   : Nobody did anything still on the storage. STAR is an initial version, which tells there is a space but not how much it is used. This can be extended also. We need to push for an extension to include the transfer.

S. Pinto   : It is not only of a standard on how to represent that data, but also the middleware need to provide the accounting data. Does it?

J. Gordon   : DPM and dCache do the accounting. STORM doesn't.

S. Pinto   : CDMI is giving an interface that gives many things. Does not defined nothing about authorization at this point.

D. Wallom   : We need to profile the standards together such that they are consistent to provide all the features requested for this service.

S. Pinto   : I think CMDI + Federated AAI + Accounting standards can provide all the features needed.

Summary:
Discussion in the EGI federated cloud need to start on how to account storage

Mountable containers:
D. Wallom: This can be done with Webdav, as done for the personal storage.

Summary: Investigate within the FedCloud if we can export with WebDav the CDMI storage to mount it on the machine.

On the fly processing:
D. Wallom: This should not be a service provided by the STaaS, but should sit on top of the existing services.

Summary: This should not be considered (or put in low priority).

Database-as-a-Service

EUDAT : We do not have any service to do this.

Bjorn   : One way is to do it with templates but more high performance solutions should be preferred. Openstack community is developing something similar using CEMS. It is an in-instance that you can start, so it can work in a federation.

?   : Database as a service is normally provided in a platform as a service.

?   : Plaform as a service would be useful.

Bjorn   : The vm solutions could be a start, because we can easy offer it.

?   : There is a platform as a service which can run gears and support mysql and other.

S. Pinto : Do we need to provide this service or not? Because it was not explicitly requested for the user, who set up himself a machine for the database.

?   : We had from the public administration this request, so there are customers for this.

?   : We need to think on how persistent this database should be. What happens when you shutdown the machine?

?   : It would be interesting to have a sort of SDK to develop solutions which works in the federated cloud? There are many ways to do that. We engaged with SlipStream.

?   : Why only relational databases, why not noSQL?

S. Pinto : No users requested this.

Summary: We should implement that with basic templates (ex. taking the database template from the OpenStack project).

SQS

D. Wallom: This is an high level service and should be a core service to have from the EGI community.

Summary: Skipped, we already do that in the Grid. We should redirect the users to the Grid community.

HTC-as-a-service

Dynamic deployment:

D. Wallom: We have HTC as a service that will come in the next couple of months from different projects.

K Koumantaros: From okeanos, we have tools to create hadoop clusters on demand.

?   : You should use condor to do the dynamic deploying?

?   : Would be interesting to turn the FedCloud to make this puppet enabled? Not

only services?

Security:
? : For encrypted storage the problem is to store the key for the VM to be used. You need the key in clear text to decrypt the drive. You need to store the key and send the key to the VM and this is a big security issue.

Summary: Check how the okeanos Hadoop cluster solutions can be turned into a production service. For security, security on HTC for processing cannot be offered now, we have no technology to do that.

Infrastructure-as-a-Service

Contextualization:
D. Wallom: We are already working with this and we are at a late stage.

Infrastructure broker:
S. Pinto: That is a work in progress because SlipStream and the other technologies the Helix Nebula community tested did not fit all their requirements. Thus, we are open to new technologies.

High-Availability:
D. Whallom   : We need to understand the definition of this.
S. Pinto   : This is about providing 99.99-100% SLA to the users. This is not about service high avaliability but about infrastructure level high availability, it is like hardware high avaliability.
D. Whallom   : We do not want to push this to our RP's. We should provide consultancy to the community.
S. Pinto   : It is hard to provide this as a consultancy, a lot of times you need to rewrite the entire code of the service to make it work in HA.
?   : From an hardware point of view, this can be done, but it has an high cost.
S. Newhouse: This should be a service that the user should pay for it.

Block Storage:
D. Wallom : This is a basic solution and we should have this with high priority.
Bjorn   : The problem is that we have this in OCCI but it is not yet implemented in rOCCI and not all the middlewares support this feature.

Single marketplace:
D. Wallom: We have the solution under late development and we are almost there.

On-demand-scaling:
D. Wallom: We should have on demanding scaling also in our IaaS features portfolio.

Summary: The High Availability this should be done but as a pay-for-it service. The resources requested here are too much.
Block storage is very important and should be a top priority to implemente.

Section 2: Participation to H2020 calls

Call4: skills for infrastructures

D. Wallom : we can do an onramp for cloud in europe, bringing the expertise we have in europe. This is also to train people.
?   : Is this relevant to EGI? How much budget is allocated?
Luis   : It is in the work program because there is the need to give support to scientists for the infrastructure. EC does not expect e-infrastructures to apply here, this is for universities. But EGI.eu has a lots of universities as members.
Luis   : The overall idea of the call is to train new professions, by having universities training courses.

Call3

D. Wallom : I think here we can target multiple areas with one big proposal.
Luis   : You can target multiple areas, but you should go for a proposal for each area. You can do a proposal for the multiple areas, but is not recommended. You should go for a proposal on one topic, targeting only some of the sub topics or all the sub topics.

D. Wallom : There will be a task force within the EGi Federated Cloud to write the papers, FedCloud members who wants to join are invited to send email to Michel Drescher (at) egi.eu

Summary of the breakout session by D. Wallom

For the service models presented by S. Pinto, what we saw is that there are solutions available for almost any of the features requested by the users, which we need to integrate. We have resource providers who already provides services which we need to expand.

Looking inside the Infrastructure as a Service, we need to be sure that the market place is on top of the priorities.

Also, we saw that many of the wishes of the user community are high level services.

For the H2020 calls, we concluded that the expertise of the FedCloud would be very valuable for call 4. For Call 3 there are a number of different areas where we can expand. There can be large flagship proposal on the topic covering almost all the multiple subtopics. There will be meetings

in December/January to define the context of future collaborations.