Project Results

Bob Jones
CERN

25 February 2019

Helix Nebula – The Science Cloud with Grant Agreement 687614 is a Pre-Commercial Procurement Action funded by H2020 Framework Programme

Helix Nebula Science Cloud Joint Pre-Commercial Procurement

Procurers: CERN, CNRS, DESY, EMBL-EBI, ESRF, IFAE, INFN, KIT, STFC, SURFSara
Experts: Trust-IT & EGI.eu

The group of procurers committed

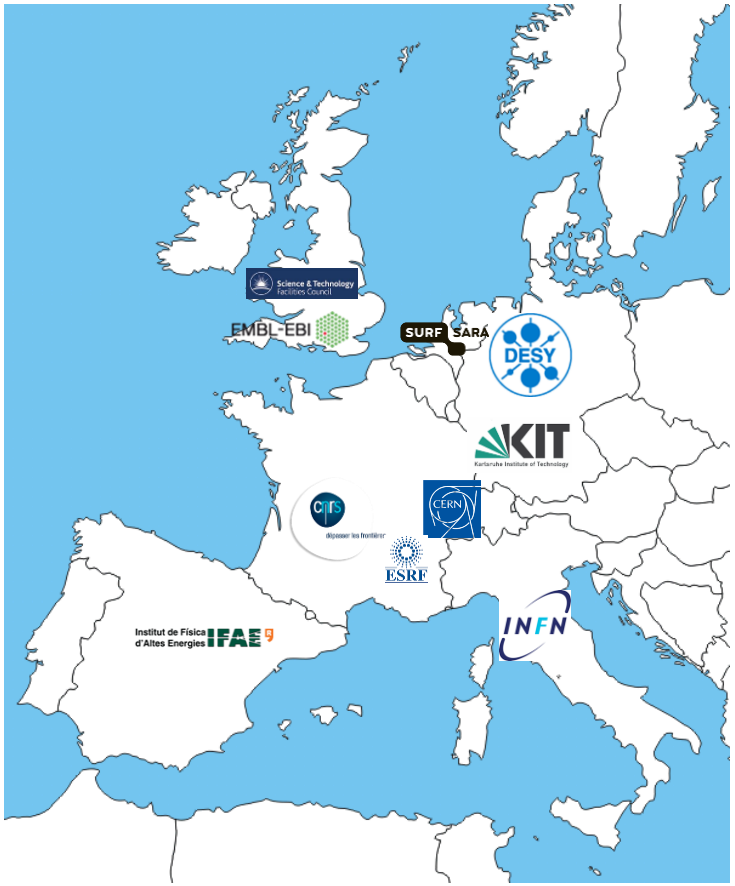
- Procurement funds
- Manpower for testing/evaluation
- Use-cases with applications & data
- In-house IT resources

Resulting made available to end-users from many research communities

Co-funded via H2020 Grant Agreement 687614

Duration: 3 years - Jan 2016 to Dec 2018

Total procurement budget >5.3M€

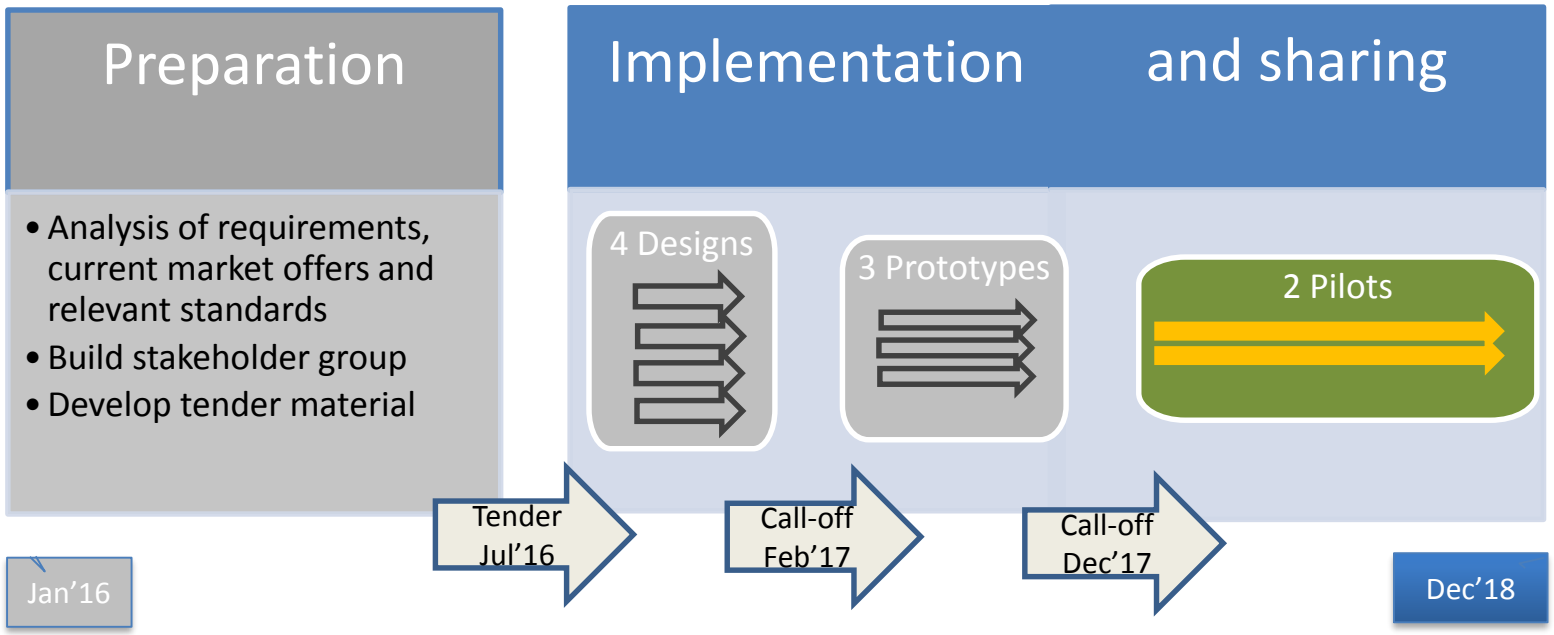


Challenges

Procure R&D for innovative IaaS level cloud services integrated with procurers in-house resources and public e-infrastructure to support a range of scientific workloads

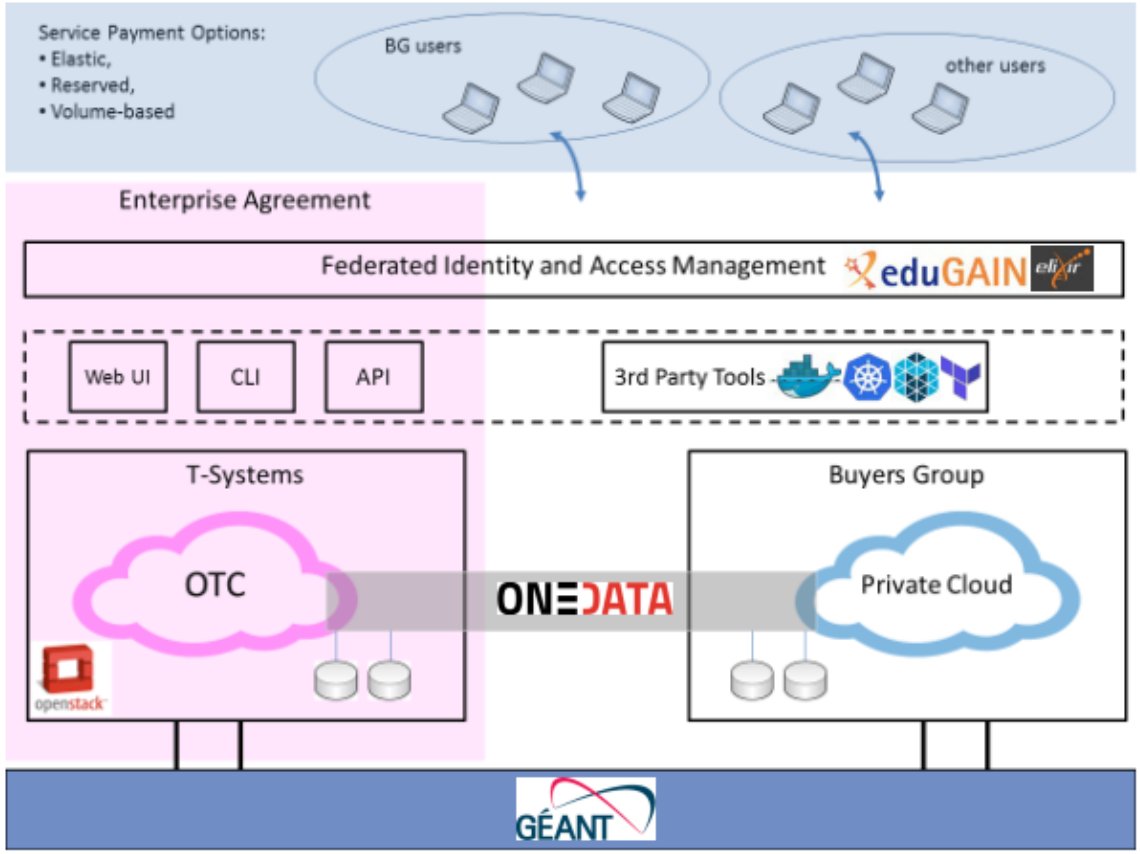
- ***Compute and Storage***
  - support a range of virtual machine and container configurations including HPC working with datasets in the petabyte range
- ***Network Connectivity and Federated Identity Management***
  - provide high-end network capacity via GEANT for the whole platform with common identity and access management
- ***Service Payment Models***
  - explore a range of purchasing options to determine those most appropriate for the scientific application workloads to be deployed

HNSciCloud project phases



Each step is **competitive** - only contractors that successfully complete the previous step can bid in the next

HNSCICLOUD – T-SYSTEMS SOLUTION



HIGHLIGHTS

HYBRID CLOUD

FEDERATED IDENTITY

CLOUD-NATIVE

LARGE-SCALE DATA

HPC SERVICE

GÉANT ACCESS

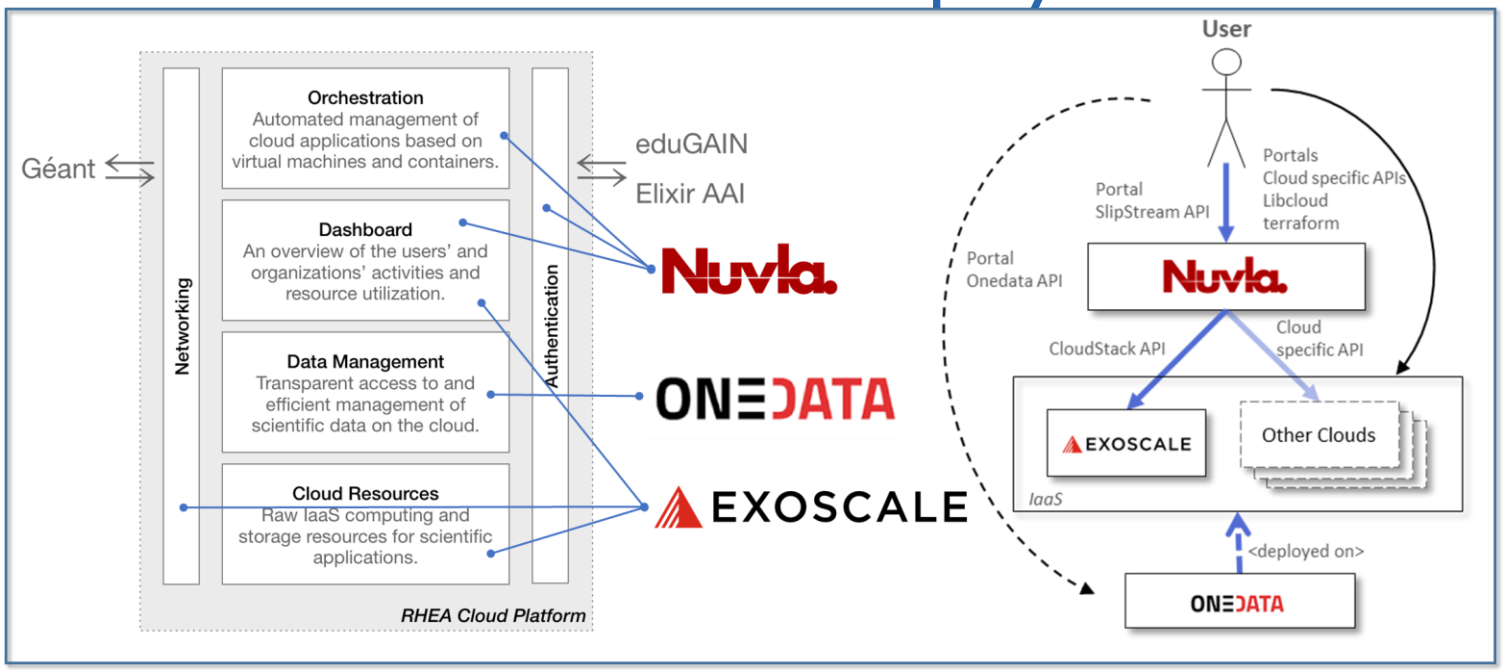
DASHBOARDS

ENTERPRISE SERVICE

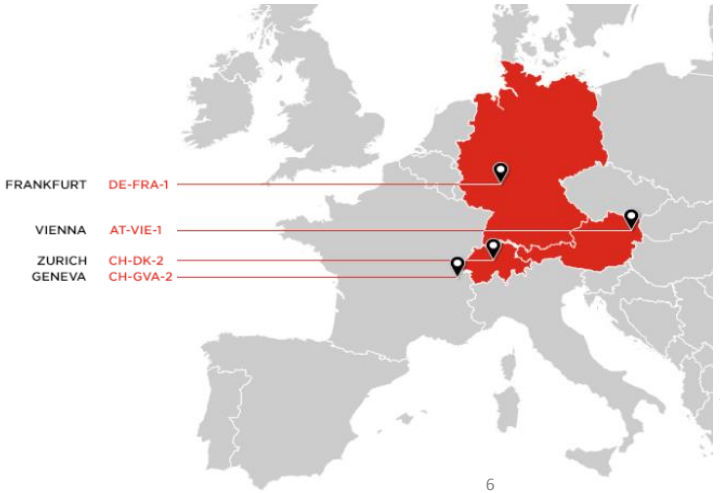
OPENSTACK

T-Systems

HNSciCloud Pilot Deployments



- Four regions: **Geneva**, **Frankfurt**, Zurich, and Vienna
- Based on **Apache CloudStack**
  - Supports CloudStack API, **Libcloud**, **Terraform** and Docker Machine
- **Simple portal** and API to manage workloads
- Powerful **object storage** for cloud native applications
- Géant: aggregate **40 Gb/s bandwidth**

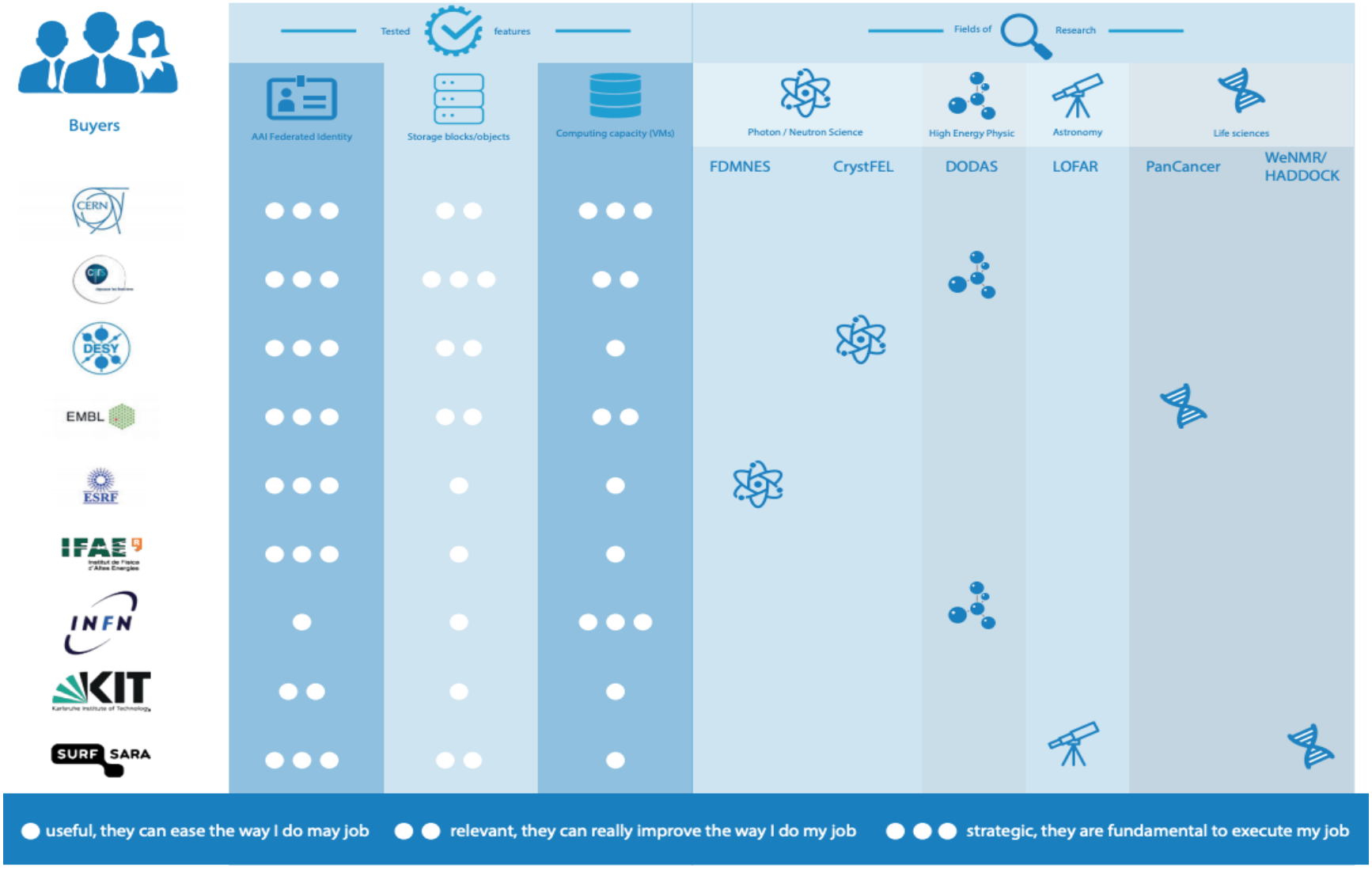


Test suite

Test/Deployment completion	74.07%	Last Update:						
Test/Deployment on-going	25.93%	2/6/2018						
TEST/Deployment NAME	PROCURER	STATUS	Expected Results	Summary of results obtained	Running Provider	Score if applicable	USE OF PROCURER'S DATA?	Additional Comments
Batch Service Deployments	CERN	Completed	Use of 50% of the quota on both providers to deploy WLCG workloads		Both		No	
TOTEM Deployment test		On-going	Fully functional data analysis software stack. Scaling out the installation and testing a complex analysis example	Functional deployment OK. Now scaling out.	TSystems		Yes but can be deleted	Test to be completed before December 19th 2018
Security test challenge		Completed		Security challenge executed in both providers	Both		No	
PerfSONAR tests		Completed	Avg Latency (in/out): Reference value for DCs located in Europe: 30ms Latency will be averaged across the 10 Buyers Institutes: 0 - Latency > 60 ms (Failed) 1 - 60 > Latency > 50 (Failed) 2 - 50 > Latency > 45 (Failed) 3 - 45 > Latency > 40 (Passed) 4 - 40 > Latency > 30 (Passed) 5 - Latency < 30 ms (Passed) Avg. Bandwidth (in/out): Reference value observed was ~ 5Gbps max. end to end. Bandwidth will be averaged across the 10 Buyer Group Institutes: 0 - Bandwidth < 100 Mbps 1 - 100 Mbps < Bandwidth < 500 Mbps 2 - 500 Mbps < Bandwidth < 1 Gbps 3 - 1 Gbps < Bandwidth < 2 Gbps 4 - 2 Gbps < Bandwidth < 5 Gbps 5 - Bandwidth > 5Gbps	Network performance was satisfactory for both providers. TSystems avg latency across all buyers group was 26.9 ms (max for PIC at 46.1), avg throughput across all buyers was 2.23 Gbps (min was for PIC at just 108 Mbps). Exoscale avg latency was 24.98ms (max for PIC at 31ms), avg throughput across all buyers was 1.38 Gbps (min was for PIC at just 103 Mbps).	Both	Passed	No	Dedicated perfSONAR on public network was recently established by TSystems and results from it will be taken into account if available.
Deep Learning with GPUs		Completed	Parallel training of Generative Adversarial Networks. Scaling experiments and comparison to HPC clusters. Expect linear (strong) scaling.	TSystems: initial run on 2 VMs (NVIDIA P100). Scaling benchmark on 8 NVIDIA V100 (on a BM.S). Exoscale: Scaling benchmark using container, kubernetes/kubeflow on 18 GPUs	Both		No	Extend to larger number of VMs as soon as they become available. Test scheduling with SLURM.
Deployment of Lightweight WLCG sites		On-going	Deployment of a fully functional and containerized WLCG site (CreamCE, Torque Batch System, Torque Worker nodes and later, HTCondor as well) using SIMPLE Grid Framework.		RHEA/Exoscale		No	Awaiting release of SIMPLE Framework. (Tentative date: 1st week of December 2018)
IAAS access via BluGrid+ and local accounts	CNR 8	Completed		In Advania we had issue with the openstack network configuration. We could not assign EIP on our VM. the issue solved after intervention of the provider. Our model works well for medium number of VM O(100) (with low IO network demand) and we can deploy with efficient manner on CSPs (OTC and Exoscale). The availability of Public IP facilitates a lot the deployments and the operation mode.	Both		No	The Test consists of a set subtests (IAM subtests and subtests for IAAS capabilities as are described in the initial document of the test cases) 0 Subtest Failed 1 Subtest Partial failed 2 Subtest Partial Succeed 3 Subtest Almost Succeed 4 Subtest Full Succeed 5 Subtest Full Succeed and recovers status bey and the expected requirements. The final score will be the average of the average of subtests for IAM subtests and average of subtests for IAAS
Onedata Wave 1: HDFS_IO	DESY	Completed		marked as completed but was not able to reach 50MiB/s. Should be return as if problems with network performance are solved	Both		No	
Onedata Wave 2: Scaling HDFS_IO		Completed	100TB shared data (50MB average file size, min 7MB, max 4GB), more than 100 (parallel) instances (of HDFS_IO) on 20 nodes minimum. Single scientific app performance of 400MB/sec (10 mins peak, 100MB/sec total average). In case DESY needs to base the performance metrics of the local (at OTC) caching infrastructure, it's expected to be less that 5% overhead in terms of I/O through OneClient when compared to the direct access to the storage system. In case of the 'final' configuration (reference scenario I) DESY expects the following for data not on cache (only one local DESY copy exists, application runs at OTC): max. file detecting delay (file could be opened in OneData - known to space - after local creation): 100 seconds same max delay is expected for the way back (to DESY storage system) Metadata scan rate (from DESY storage into OneData Space and the reverse direction): 50 Hz File (initial) access rate not less than 85% of the slowest network segment (subpath)	Currently trying to run tests with multiple files. Experience problems with importing data fro DESY to OneData - some files are only partially imported. Need to investigate further. Update 06.09.2018 - Release 10 - problem is still there. Access to OneProvider on our local instance is provided to Cyfronet (although RHEA experts seeing similar problem on their environment as well. TSystem experts did not perform scaling tests with a significant amount of files). Update 26.11.2018 - Release 13. Functionality is ok now. Can run all the tests. File detecting ok OK. On OTC: VM with High Performance flavor and NVMe based cash storage reach 180MiB/s for read, 80MiB/s for write per each application (out of 25 tested), but the storage is limited (scales with number of VMs). Storage which scales is based on standards SSD and does not allow to reach the specified bandwidth. On Exoscale - only S3 is really scales, but it does not allow to reach desired bandwidth (just several MiB/s per application)	Both		No	
SLURM Integration		Completed		Works, still need to check updated LUM A version and replication scripts Update 26.11.2018 - works as expected, score 5 for OTC, 3 for Exoscale	Both		No	
								0 - Cannot execute job with Docker in the cloud

Buyers Group assembled a suite of approx. 30 tests and executed them against commercial cloud provider solutions

Use cases deployed



HEP Deployments

- WLCG
  - ALICE, ATLAS, CMS and LHCb
    - Daniele Spiga (INFN) talk about DODAS, Track 7:
      - https://indico.cern.ch/event/587955/contributions/2937198/
    - Matthias Schnepf (KIT) talk about Dynamic Integration of resources, Track 8:
      - https://indico.cern.ch/event/587955/contributions/2937900/

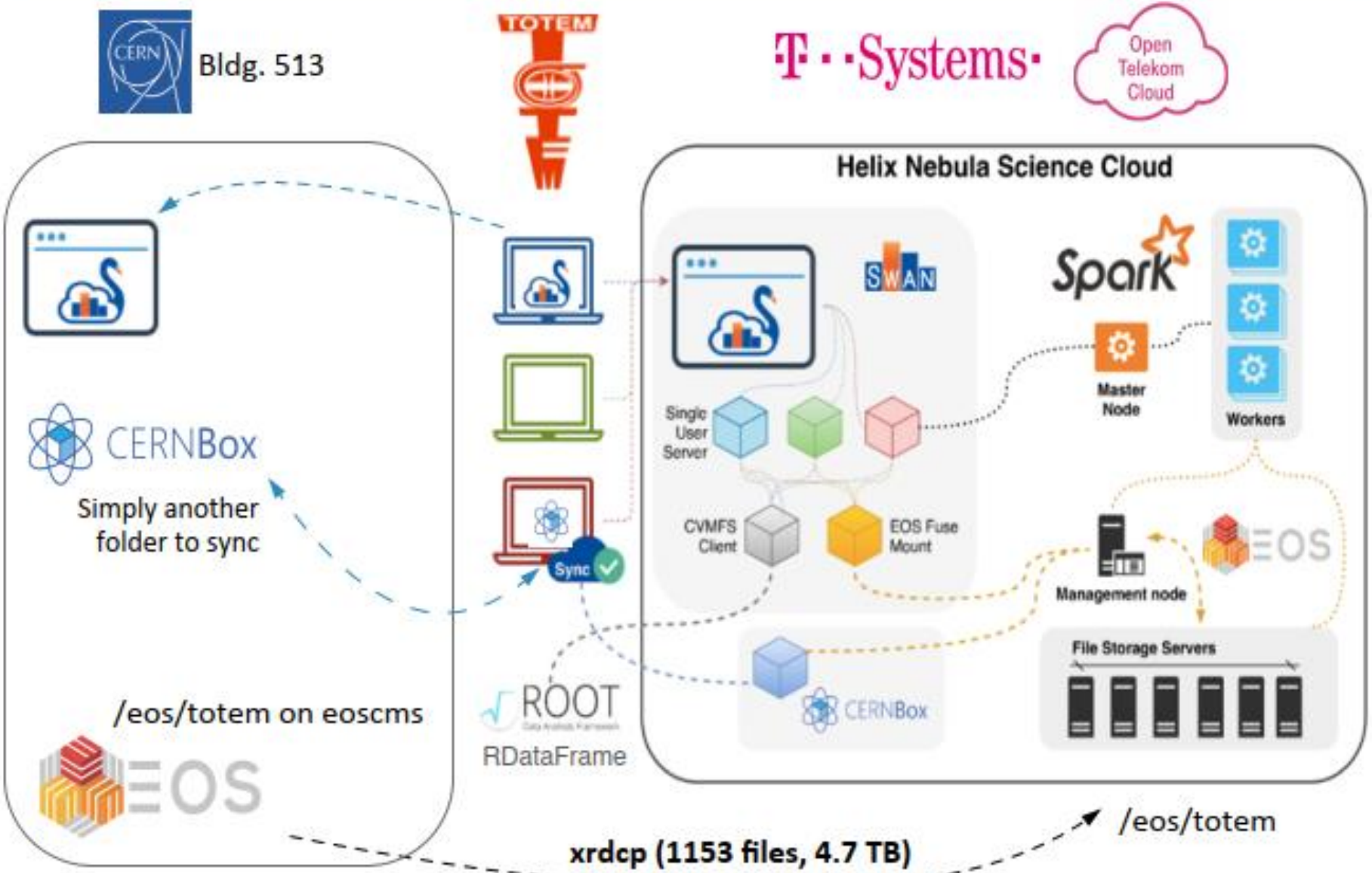
- CERN Batch Service
- Container Federation
  - Openstack Summit, reference talk:
    - https://www.openstack.org/summit/vancouver-2018/summit-schedule/events/20768/cern-experiences-with-multi-cloud-federated-kubernetes

- Belle II
  - Silvio Pardi (INFN) poster about experience of Belle II with commercial clouds, Track 7:
    - https://indico.cern.ch/event/587955/contributions/2937060/

- Interactive Analysis for End Users for TOTEM
  - https://indico.cern.ch/event/727193/contributions/3039091/attachments/1667076/2674030/TotemTest_HNSciCloud.pdf

- Machine Learning/Deep Learning for Fast Detector Simulation using GPUs
  - Sofia Vallecorsa (CERN openlab), Track 2 and Jean-Roch Vlimant (CMS), Track 6:
    - https://indico.cern.ch/event/587955/contributions/2937595/
    - https://indico.cern.ch/event/587955/contributions/2937513/

Deployment (June-December 2018)



Total Cost of Ownership study

- Understand the costs of using commercial cloud services as part of a hybrid cloud model
- 2 use-cases selected with different requirements
  - **ALICE**: single core jobs, up to 50.000 at any time (monte-carlo, reconstruction, analysis)
  - **PANCANCER**: burst pattern, with minimal resources constantly used (few VMs) and periods of ramp-up (up to 400 VMs)

TCO results: T-Systems

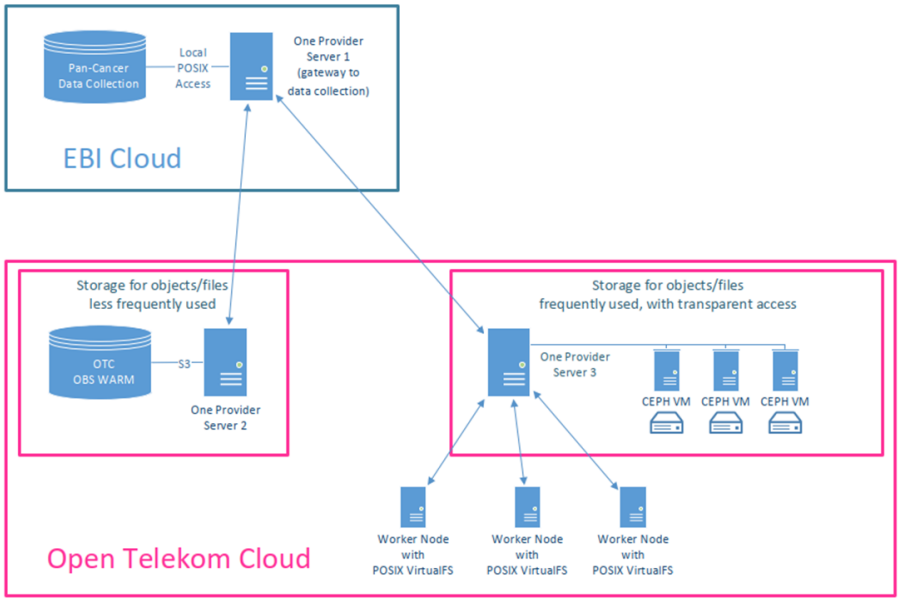


Figure 1: Pan-Cancer Data Management scenario

"The impact of flavour and VM size on TCO can be significant."

Table 5: ALICE job costs compared between CERN cloud and OTC public cloud

Job Type	TCO Benefit OTC Public Cloud
Monte Carlo	
Reconstruction	
Analysis	

Table 4: Pan-Cancer qualitative factors

Factor	Importance (H, M, L)	Effectiveness	Notes	Additional Information
Agility	H	+1	The solution provides easy and quick deployment and changes to the processing and storage functionality	
Contract Review and Negotiation	M	+1	Can be implemented with minimum review	
Elasticity and Scalability	H	+1	Solution facilitates easy and quick expansion of available processing and/or storage capacity	
Regulatory and Policy Requirements	H	+1	Solution adequately enables compliance with external regulations	User data can be stored encrypted in the public cloud. Through Onedata, data privacy can be maintained between public, restricted and confidential data. Confidential data would only be hosted on-premise.
Security	H	+1	Solution provides effective mechanisms by which constantly escalating security threats are prevented and security events or breaches are constantly monitored	Cloud service is certified e.g. for relevant ISO standards, EU GDPR and German national regulations.
Service Levels	M	+1	Service level targets were not provided. The solution would be able to match or improve the current SLA based on on-premise services. Especially with regards to performance, the cloud offers more granularity and diversity to tune performance to requirements when compared to on-premise resources.	Service Availability of 99,95% is achievable when using 2 availability zones and load balancing. Support is provided 24/7.

TCO results: RHEA System

Table 15: TCO Cost Summary for Each Use Case

Use-Case	Exoscale	AWS	CloudFerro
PANCANCER (no storage)	123,170	147,575	155,247
PANCANCER (with storage)	239,562	-	-
ALICE (all three use-cases) – 50000 jobs*	6,216,252	7,907,449	9,282,265
ALICE Monte Carlo – 1 job**	0.082	0.106	0.125
ALICE Raw Data Reconstruction – 1 job**	0.096	0.138	0.159
ALICE Analysis Trains – 1 job**	0.038	0.034	0.042

"The PANCANCER and ALICE use-cases each have their own specific requirements for cloud deployment. Both use-cases have more than one job type requiring different VM flavours and/or pricing options. Also, use-cases can be supported without having to storage large volumes of data in the cloud, minimising storage costs. Therefore, we have not included the cost data management solutions, since our assumptions for use-cases are that the data can be 'streamed' to the VMs. "

Lessons Learned

- *Framework agreements provide a convenient structure for service procurements in the scientific community*
- *Volume and requirement aggregation across a group of scientific organisations brings advantages*
- *Use a small number of cloud providers in parallel to avoid lock-in and ensure service continuity*
- *Need to repatriate data at the end of contracts*
- *Commercial clouds offer opportunities to rapidly scale cutting edge technology for R&D deployments*
- *Vouchers/Credits are a practical means to provide limited-scale access to commercial cloud services for end-users*
- *Commercial clouds providers offer services that are certified against international standards (e.g. ISO 27000 family) and consistent with legislation (e.g. GDPR)*

OCRE Open Clouds for Research Environments

http://www.ocre-project.eu

European Commission

- *Procurement of digital services for the European Open Science Cloud*
- *Procurement Budget: 9.5M€*
- *Starting Date: 1st January 2019*
- *Duration: 36 Months*
- *Coordinating Partner: GÉANT*

CERN GÉANT RHEA GROUP Trust IT

OCRE will become the procurement vehicle of the European Open Science Cloud (EOSC)

Procuring commodity services from multiple cloud providers across IaaS/PaaS/SaaS

Suppliers readiness will be technically validated with a test-suite

Procurement Framework: Oct 2019, Call-offs: Q1 2020 and Q1 2021

OCRE will provide opportunities for research communities to adopt commercial cloud services as part of their IT computing strategy

ARCHIVER

ARCHIVING AND PRESERVATION FOR RESEARCH ENVIRONMENTS

http://www.archiver-project.eu

European Commission

- *Pre-Commercial Procurement*
- *Focus: Archiving and Data Preservation Services in commercial clouds*
- *Procurement Budget: 3.4M€*
- *Starting Date: 1st January 2019*
- *Duration: 36 Months*
- *Coordinating Partner: CERN*

CERN DESY. EMBL-EBI PIC port d'informació científica addestino innovation delivered. Trust IT

ARCHIVER Pre-Commercial Procurement will run an open tender process for R&D in the next generation of digital archiving and Long-Term Data Preservation services

Experience gathered from HNSciCloud about the PCP process will be applied in ARCHIVER

Open Market Consultation Kick-off event: 8th of April 2019: https://goo.gl/rhWTmH