



HNSciCloud PILOT PLATFORM

RHEA Consortium

29 November 2018

OUR VISION

Building an open, European platform for science data exploitation

- ✓ Solution delivers a **cost effective solution at scale** (10k, 1PB; 400k, 100 PB), with **minimal intrusion** into Buyers' sites
- ✓ Built on pure IaaS to ensure **no-lock-in** and **best market pricing**
- ✓ **Laying the foundations of the European Open Science Cloud (EOSC)**



CONSORTIUM

Established Partnership

RHEA: Prime contractor with proven management skills for large, institutional contracts and cyber-security expertise



Exoscale: Dynamic, innovative cloud provider that has worked with European scientific institutions like CERN



SixSq: Cloud experts provide Nuvla multi-cloud management platform and technical leadership



Cyfronet: Provides Onedata for data management and IT services to research teams and educational institutions



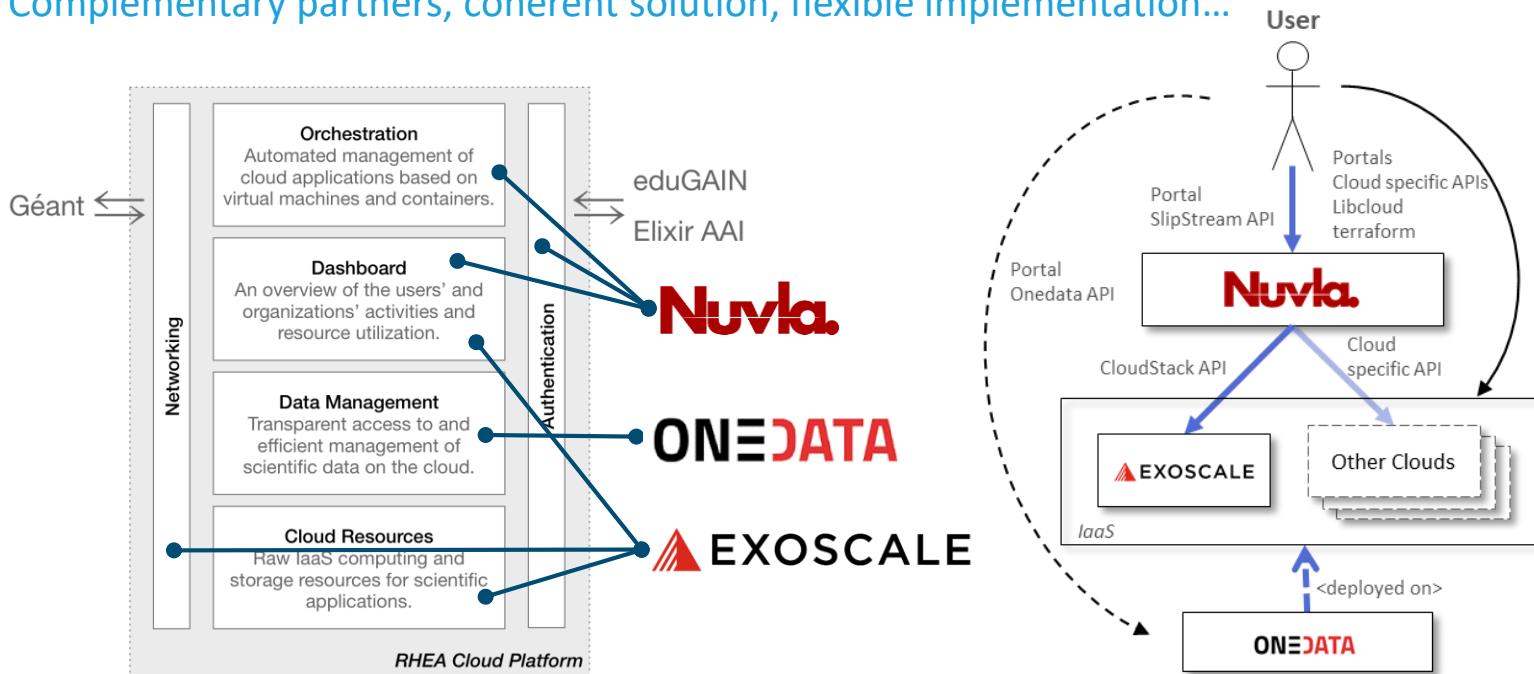
CHALLENGES

Deliver end-to-end optimised data access and analysis

- ↪ Provide seamless user access and security through **federated identities**
- ↪ Support a **wide range of applications**
- ↪ **High data throughput**: home institute ↔ cloud; data source ↔ application
- ↪ **Transparent data access** with **POSIX** support
- ↪ Support different **data layout and access patterns**
- ↪ Integrate **containers** and **container orchestration engines**
- ↪ ... **with minimum complexity!**

COMBINED SOLUTION

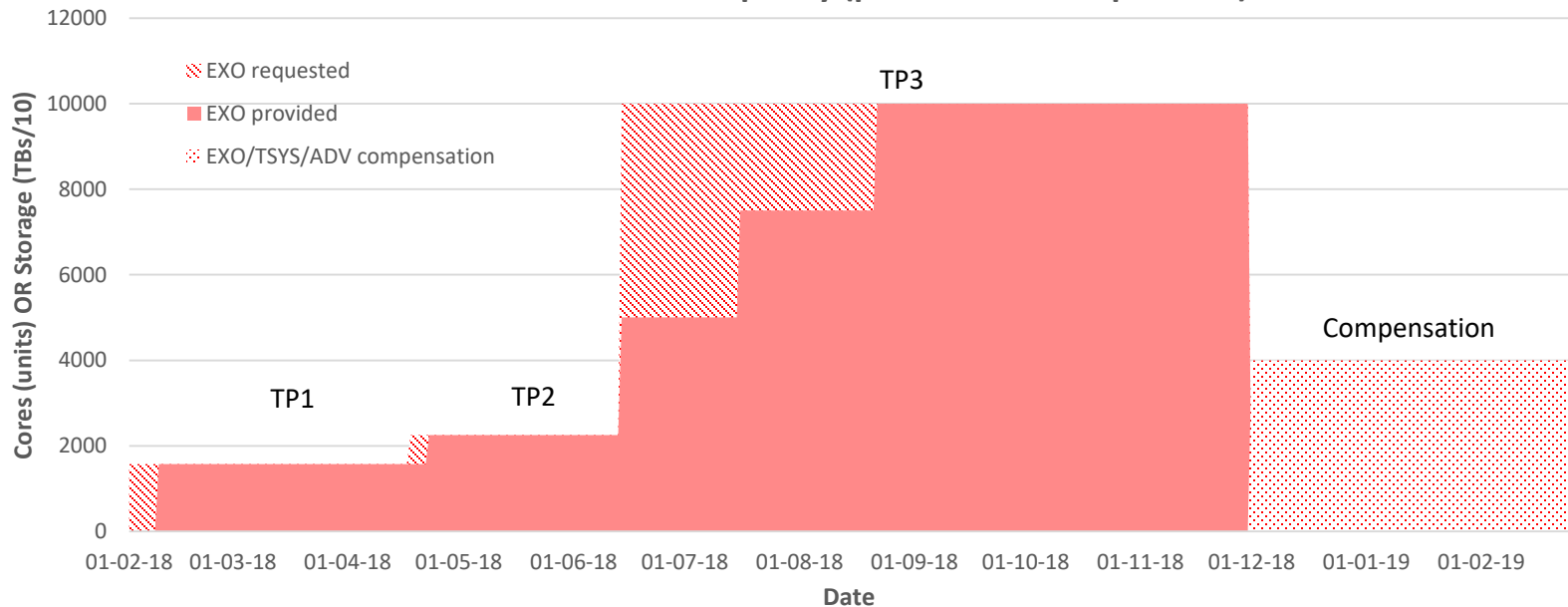
Complementary partners, coherent solution, flexible implementation...



RESOURCE CAPACITY

All the compute capacity (an more) has been used in HNSciCloud Phase 3

HNSC Exoscale capacity (provisioned and planned)

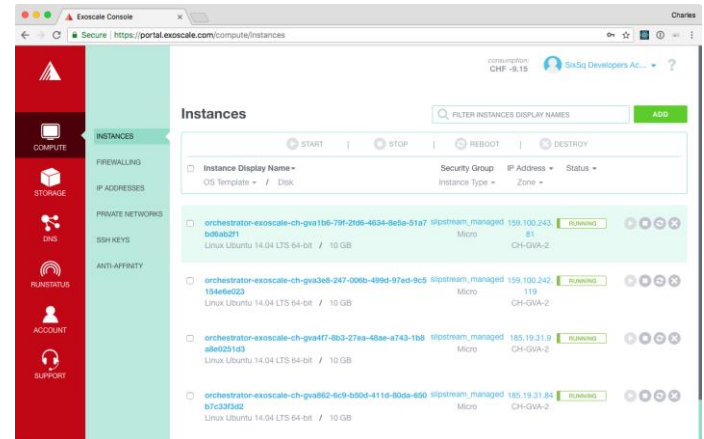


EXOSCALE: IAAS CLOUD SERVICES



Innovative cloud service provider optimised for cloud-native applications...

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



EXOSCALE COMPUTE & STORAGE

Simple to use IaaS with advanced features

- ✓ VMs start in less than 30s with
 - ✓ SSD, IPv4 and v6
 - ✓ Multiple networking options with public IP addresses for VMs
 - ✓ Scale from 1 to 225 GB of RAM
 - ✓ S3 compatible object storage
- ✓ Several GPU flavours supported
- ✓ Management features:
 - ✓ Optional **voucher** based registration to for new labs and users to try
 - ✓ Organization and sub-Organizations concept
 - ✓ Aggregated billing

Exoscale IaaS improvements (recent and upcoming)

✓ Achievements:

- ✓ IPv6
- ✓ Object storage in FRA1 & VIE1
- ✓ Private Connect
- ✓ Enforced 2FA
- ✓ Managed Private Networks (GVA2)

✓ Upcoming:

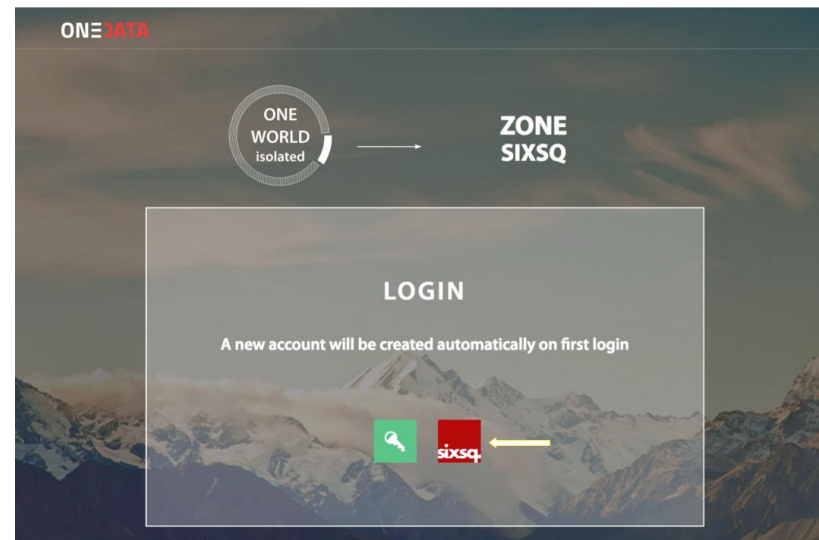
- ✓ IAM: Fine grain access
- ✓ Custom templates
- ✓ Load balancer
- ✓ Network block storage
- ✓ Managed Kubernetes
- ✓ IPSec gateway
- ✓ Pre-emptive VMs

ONEDATA: DATA MANAGEMENT

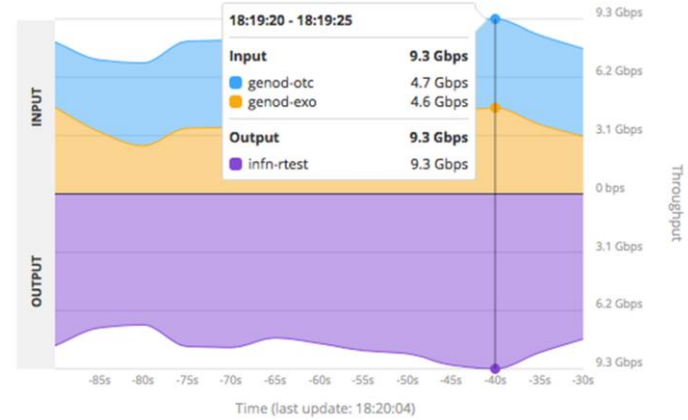
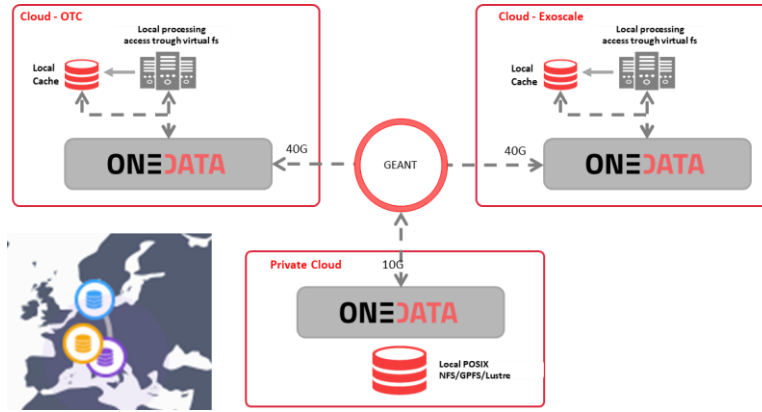


Flexible, effective data management solution built around shared “spaces”...

- ✓ **Uniform view** of all shared data sets, no matter where they are stored
- ✓ **Transparent access** to all data sets with **dynamic migration**
- ✓ Support for **POSIX** access
- ✓ Integrated with the **eduGAIN** and **Elixir AAI** identity federations



DATA PROCESSING IN MULTI CLOUD ENVIRONMENT



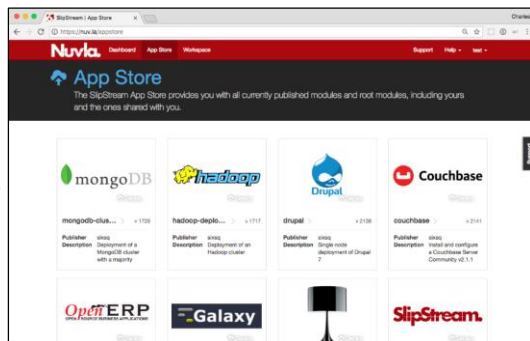
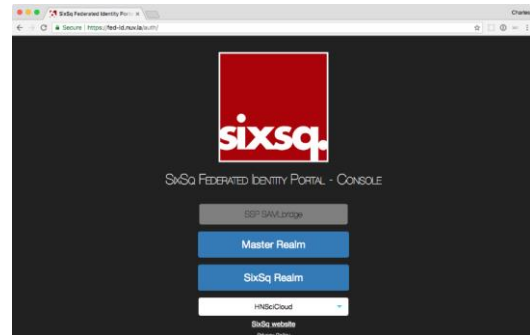
- ✓ Locked-in data collections on **NFS or GPFS** used for remote data processing in multi clouds
- ✓ **Transparent** access to data in many public clouds without changing application
- ✓ Data accessible as a **POSIX** file system
- ✓ On-the-fly / job replications

AUTHENTICATION & ORCHESTRATION



Simplifies and automates management of cloud applications...

- Full support for **eduGAIN** and **Elixir AAI** identity federations
 - Keycloak** based implementation allows management of **users**, **groups**, and their **attributes**
- Nuvla
 - Automated deployment** and dynamic management of applications to **minimise costs** and **maximise performance**
 - Predefined applications: **SLURM**, **Docker Swarm**, **Kubernetes**, etc.
 - Dashboard** showing cloud resource utilisation, quotas...



DASHBOARDS

Nuvla – latest development (WebUI)

- Usage monitoring and metering
- Quota management
- Views on multi-tenant user orgs

The image displays two screenshots of the CUBIC web interface. The top screenshot shows the 'usage' dashboard, and the bottom screenshot shows the 'quota' dashboard.

Usage Dashboard:

- Navigation: deployment, application, usage (selected), quota, cimi, en
- Filters: last 30 days, from Wed, 8 August 2018, to Thu, 6 September 2018
- Options: billable only (checked), filter by users or roles (e.g. SixSq:can_deploy), all credentials
- Summary: 29/29 CREDENTIALS, 4'156 VMS [H], 10'154 CPUS [H], 30'741 RAM [GB-H], 42'867 DISK [GB-H], 94 € PRICE [€]
- Table:

credential	cloud	VMs [h]	CPUs [h]	RAM [GB-h]	DISK [GB-h]	PRICE [€]
		4'174.63	7'034.15	16'341.33	73.95	
		343.03	595.42	1'540.33	6.29	
		320.70	320.70	0.00	5.77	
		331.73	331.73	0.00	3.69	
		166.75	83.38	1'642.00	1.27	
		74.92	74.92	0.00	1.08	
		31.35	31.35	0.00	0.41	
		25.12	25.12	0.00	0.32	
		34.45	17.23	344.50	0.26	
		14.40	14.40	0.00	0.20	
		15.22	15.22	0.00	0.17	
		19.78	19.78	0.00	0.13	
		8.92	8.92	0.00	0.12	

Quota Dashboard:

- Navigation: deployment, application, usage, quota (selected), cimi, en
- Refresh button
- Summary Cards:

- exoscale-ch-dk** (Exoscale in Zurich, Switzerland): VMs [0/1000]
- exoscale-at-vie** (Exoscale in Vienna, Austria): VMs [1/1000]
- exoscale-ch-gva** (Exoscale in Geneva, Switzerland): VMs [10/1000]
- exoscale-de-fra** (Exoscale in Frankfurt, Germany): VMs [0/1000]

... enabling users and community managers to control and understand their cloud usage

HPCAAS

Dynamically deployed batch clusters using high-performance hardware...

- Automated **SLURM** deployment supporting normal and MPI jobs:
 - CERN: ML experiments
 - ESRF/FDMES
 - SURFSara
- Range of **Titan VMs** with up to 16 vCPUs and 128 GB of RAM.
- Two profiles with available Small and Large **GPUs**

GPU Profiles		
vCPU	12	48
RAM	56 GB	225 GB
SSD (max.)	2 TB	8 TB
GPU Cores	3584	14336
GPU NVIDIA Tesla P100	1	4

TOTAL COST OF OWNERSHIP

Use Cases

UseCase	VMs/Jobs	Utilisation	Data Rate	Bulk Storage	Recommended VM	
PANCANCER						Assumes use of OneData/GlusterFS/FUSE to provide POSIX on object storage. Streaming alternative instead of storage
Continuous	7	100%	Minimal	500TB	Huge 50GB Reserved	
Burst	250	13%	4Gbps/egress	500TB	Huge 50GB PAYG	
ALICE	50000 jobs at any one time					73 million jobs per year 2 or 8 cores per VM?
Montecarlo	35000 (70%)	100%	0.3Mbps per job 10.25Gbps	None	Medium 50GB Reserved	
Raw Data Reconstruction	10000 (20%)	100%	1.0Mbps per job 9.86Gbps	None	Medium 50GB Reserved	
Analysis Trains	5000 (10%)	100%	50.0Mbps per job 244Gbps	None	Medium 50GB Reserved	Network costs potentially high

TOTAL COST OF OWNERSHIP

Conclusions

✓ PANCANCER

- ✓ 1PB only available from Exoscale as Object Storage, so need GlusterFS, FUSE, OneData or similar
- ✓ Streaming of data from institute is feasible & ~50% the price

✓ ALICE

- ✓ 90% of use case can be covered and affordable
- ✓ Data rates mean networking needed for Analysis Trains is costly
- ✓ Exoscale prices are competitive against other suppliers but multi-cloud could provide significant cost benefits
- ✓ For large scale jobs, VMs need to be selected to closely match the job needs to minimise costs

COMMERCIALISATION PLAN

The Market

- ✓ Potentially a huge demand for commercial cloud services in the science community but...
 - ✓ Is the total cost of ownership compelling?
 - ✓ Can it provide a better services?
 - ✓ Is the science community willing to embrace?
 - ✓ How can it co-exist with existing research infrastructures?
 - ✓ Are scientists and their applications cloud ready?
 - ✓ Can existing procurement practices support it?
 - ✓ Are there strong enough incentives from funders (e.g. Local/Regional/National governments, EU etc.)?

COMMERCIALISATION PLAN

NuvlaScience

- ✓ Multi-cloud Access
 - ✓ Single contract – HNSciCloud Cloud Service Agreement
 - ✓ Self-service, brokerage, managed service
 - ✓ Resource pricing, usage monitoring & billing
 - ✓ Unified authentication
 - ✓ BYOL
- ✓ Simplifies cloud access and management for non-expert users
- ✓ Expert users can access underlying cloud(s) directly
- ✓ Process for integrating a new cloud provider clear (subject to GEANT access)



COMMERCIALISATION PLAN

Distributed Data Management

OneData

- Optional service provided by Onedata.io S.A.
- Monthly pricing per PB per Oneprovider
- Silver, Gold and Platinum service levels

Nuvla

- In collaboration with the European Space Agency (ESA), SixSq's data management solution, initially proposed to HNSciCloud, is finding a natural fit
- Working at 20 TB daily, the solution could be tested at peta scale in the near future
- Supports POSIX for small scale and Object Store for large scale
- Solution fully integrated in Nuvla as a standard feature

The logo for ONE DATA, with "ONE" in black and "DATA" in red.The logo for Nuvla, in red.

COMMERCIALISATION PLAN

Pricing

- ✓ Nuvla fees within prices – help desk and support
- ✓ Various payment models supported
- ✓ Benchmarking shows Exoscale prices are competitive
- ✓ Discounts for
 - ✓ Reserved instances (1 year, 3 years) & Volume (>100K, >1M per annum)
 - ✓ Pre-emptive VMs in 2019
- ✓ Voucher Scheme
 - ✓ Used with Exoscale during Phase 3, will roll-out more widely in H2020 OCRE
- ✓ Consulting & Support
 - ✓ Based on a daily rate for support



COMMERCIALISATION PLAN

Market Scorecard

	Business	Research	Education	Healthcare	Government	Other
Business						X
Research						X
Education						X
Healthcare						X
Government						X
Other						N/A
Overall						High

	Comment	Are we there yet?
Is the total cost of ownership compelling?	We think so, but only the BG can confirm.	Green
Can it provide a better services?	Together with the institutes, potentially yes.	Yellow
Is the science community willing to embrace?	Unclear. Needs a cultural shift. Without external factors could take a long time.	Yellow
How can it co-exist with existing research infrastructures?	Hybrid cloud model will only work if the interaction is near seamless. Needs work on institutional and supplier sides.	Green
Are scientists and their applications cloud ready?	Partially. Some legacy code may not be able to fully benefit.	Yellow
Can existing procurement practices support it?	Need innovative approaches, aggregating demand, shift from CAPEX to OPEX	Yellow
Are there strong enough incentives from funders (e.g. Local/Regional/National governments, EU etc.)?	Varies from country to country. EOSC is major EU initiative.	Green