# HNSciCloud
# Technical Aspects

Helge Meinhard / CERN
Chair, HNSciCloud Technical Sub-Committee
HNSciCloud Open Market Consultation 17-Mar-2016

*All information contained herein are for discussion purposes only and shall not be considered a commitment on the part of CERN or the Buyers group.*

# Outline

- Objective
- Use cases
- High-level requirements
- Highlights of implementation requirements
- Reference architecture
- Bidder’s qualifications
- Details on use cases

# Objective

- PCP project Helix Nebula Science Cloud (HNSciCloud)
  - Create a **common science cloud platform** for the European Research community across in-house resources, existing e-science infrastructure and commercial IaaS resources
  - Full **integration** with procurers' services in their **in-house data centres** and with other **e-science infrastructures**
    - Form a **common, hybrid infrastructure**
    - Aim: For the applications, it is entirely **transparent** where they run
  - Workloads more or less **data-intensive**
  - Focused on Infrastructure as a Service (IaaS), but just VMs are not sufficient

- Procurers
  - CERN, CNRS, DESY, EMBL, ESRF, IFAE, INFN, KIT, STFC, SurfSARA

- Experts
  - EGI.eu, Trust IT

# Use Cases

- High-Energy Physics
  - LHC Experiments (WLCG)
  - Belle II
  - COMPASS
- Life Sciences
  - ELIXIR
  - Euro-BioImaging
  - PanCancer
  - BBMRI/LSGA
  - HADDOCK
- Astronomy
  - CTA – Cherenkov Telescopic Array
  - MAGIC
  - Pierre Auger Observatory
- Photon/Neutron Science
  - Petra III, 3DIX, OCEAN, OSIRIS, European XFEL
- "Long tail of Science”

# High-Level Requirements (1)

- Supplies: Based on **“traditional” IaaS resources**
  - **Virtual machines** with local storage and network connectivity
  - **Persistent storage** shared between VMs at cloud instance level
    - Persistency refers to lifetime of project; long-term data preservation is addressed elsewhere
  - Performant and reliable **network connectivity** via GEANT
- Most applications can do with rather conventional/typical resources
  - Some need large memory, high core counts and fast (low-latency, high-bandwidth) inter-core connections for massively parallel applications, and/or larger/more performant local storage
- **Challenge** and **innovation** through demand for **integration** of services for **data-intensive applications** with procurers’ on-site services and other e-infrastructures, and across use cases
- Size of target community:
  - Typically less than 10 operational staff per procurer (less than 100 total) who interact directly with resource provider
  - Thousands of users of the services set up on IaaS resources, no direct interaction with resource provider

# High-Level Requirements (2)

- **Supplies** and/or **areas of common work: Integration** of
  - Orchestration, Multi-cloud management frameworks
  - Monitoring; dashboards; alerts; rapid reporting and accounting (APIs)
  - Storage at cloud instance level: persistent cache (?)
    - Service/software to transparently manage cache?
  - Service level agreements
  - Performance: Benchmarking, metrics for aggregate performance
  - AAI and Credential Translation schemes
  - Helpdesk; Computer Security Response Teams
  - Support for transparent deployment of containerised applications
  - Ordering, billing, invoicing
    - Address gap between cloud economic models and procurement procedures of public organisations

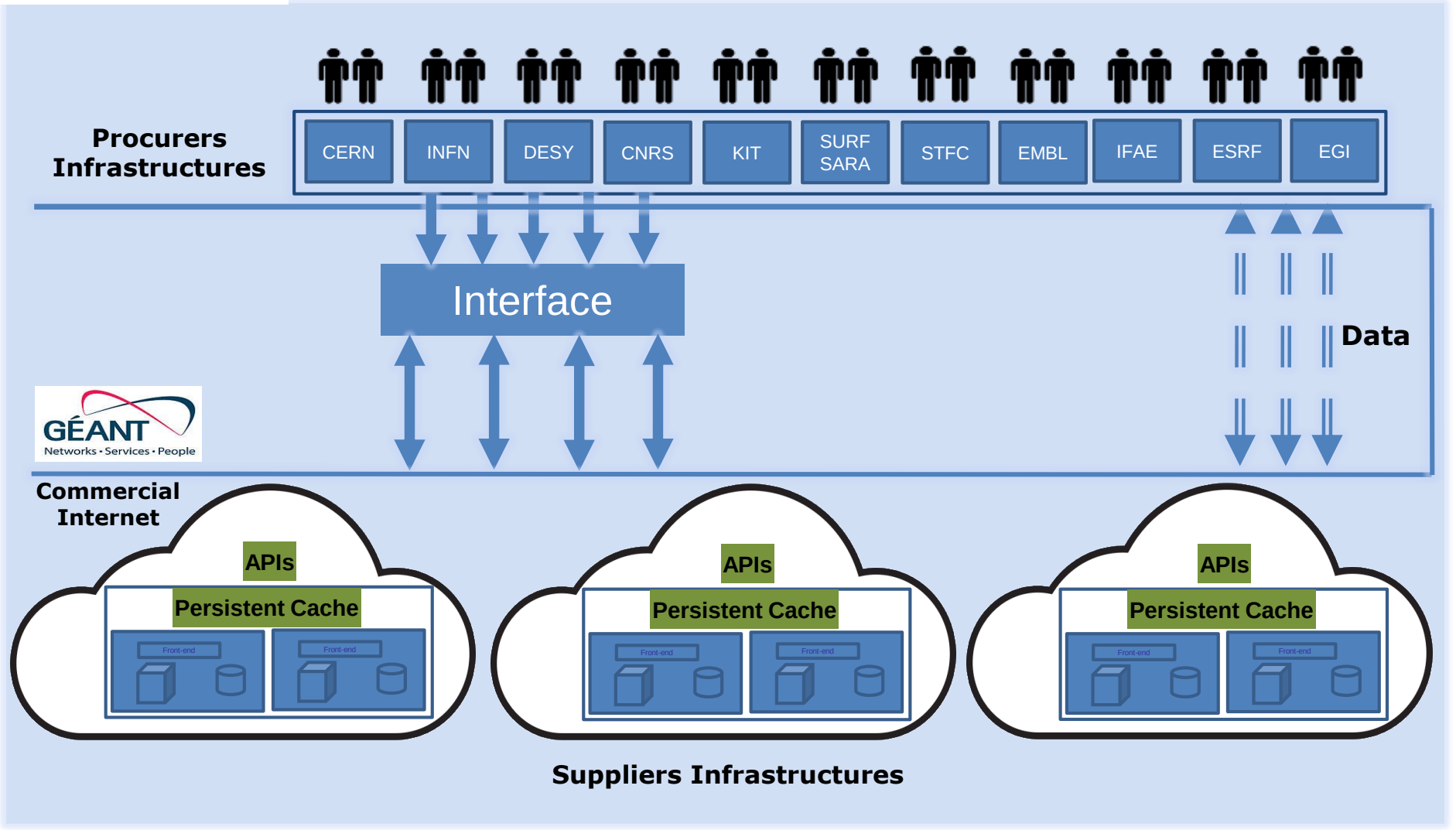
# Implementation Requirements: Highlights (1)

- Core count per VM:
  - Most use cases require 1…8 cores
  - Some use cases require 16 cores per VM
    - OCEAN, Euro BioImaging, PanCancer
  - Special requirement for massively parallel applications (128 cores or more) – high core count with fast interconnects (same box and/or Infiniband?)
- RAM per core: most use cases fine with 2…4 GB
- Local storage at VM: Some requirements for large capacity and high IOPS
- Images
  - Linux (CentOS 6+, Scientific Linux 6/7, Debian)
  - Docker containers
- VM life times: Some fraction of VMs expected to be stable over the whole project phase, others could be short-lived (e.g. for 12 hours every day); life times of much less than 12 hours probably not very useful

# Implementation Requirements: Highlights (2)

- Shared storage as block or clustered or NAS storage
- IPv4 addresses of VMs and storage endpoints (public or via VPN)
- Network bandwidth:
  - Between provider and GEANT: several  10 Gbps … 100 Gbps
  - Internal: tbd
- Workload bound by budget rather than demand
  - Minimum total capacity needed during implementation & sharing phases:
    - Prototype: 10’000 cores & 1 PB  during approx. 6 months in 2017
    - Pilot: 20’000 cores, 2 PB during approx. 12 months in 2017 – 2018
- Data privacy: Little sensitive data; some requirements at least at the level of user groups (e.g. WLCG, PanCancer, Euro BioImaging, …)

# HNSciCloud Reference Architecture



# Bidder's Qualifications

- Detailed requirements still to be defined
- Some indications:
  - Able to fulfil use cases with typical requirements
  - Addressing use cases with non-typical requirements as well would be an advantage
  - Able to conduct developments such as the ones listed above
  - Formal criteria such as ISO9001:2008, ISO/IEC 27017, ISO/IEC 27018, ISO/IEC-27036-4 being considered
  - Consortia and/or sub-contracting:
    - Likely to be limited for identical services
    - Possibly acceptable for covering different services (e.g. for use cases with non-typical requirements or dedicated developments)

HELIX NEBULA
THESCIENCECLOUD

# Thank you
# Questions?

# Details on Use Cases

# WLCG use case

- CPU Requirements (hours)
  - 1000 VMs
- Peak Requirements (CPU, RAM, Disk Storage)
  - 4 vCPUs, 4Gb RAM/vCPU, 100 GB storage per VM
  - Server size homogeneous
- Data Requirements (Quantity of data)
  - 20 TB/day
- Single/Multiple Binaries
  - Multiple, Sequence (Simulation – Reconstruction – Analysis)
- Programming Model
  - Serial, may require multi-threading
- Interaction with user
  - Batch need to assess feasibility of interactive modes
- External connectivity
  - Needed to extensive number of sites
- Volumes of data
  - 20 TB/day
- Bandwidth
  - 10 Gb/s guaranteed BW from each supplier Data Centre to GEANT
- OSes
  - CENTOS
- APIs
  - Management API for VMs and Storage

# Belle II use case (INFN, KIT)

- CPU Requirements (hours)
  - 0.5M CPU hours
- Peak Requirements (CPU, RAM, Disk Storage)
  - 3GHz, 2GbRAM/Core, 1GB storage per core
  - Server size homogeneous
- Data Requirements (Quantity of data)
  - 1Gb/job
- Single/Multiple Binaries
  - Multiple
- Programming Model
  - Serial
- Interaction with user
  - Batch
- External connectivity
  - Need to copy to external site 1700 files (~30MB) after 100k CPU hours
- Volumes of data
  - High throughput for short periods (2TB copy in less than a day)
- Bandwidth
  - 10 Gb/s guaranteed BW from each supplier Data Centre to GEANT
- OSes
  - SL6
- APIs
  - Management API for VMs and Storage

# COMPASS use case (INFN, KIT)

- CPU Requirements (hours)
  - 6M CPU hours
- Peak Requirements (CPU, RAM, Disk Storage)
  - 4 vCPUs, 2Gb RAM/vCPU, 10 GB storage per VM
  - Server size homogeneous
- Data Requirements (Quantity of data)
  - 20 TB/day
- Single/Multiple Binaries
  - Single
- Programming Model
  - Serial
- Interaction with user
  - Batch
- External connectivity
  - Need to copy data when after cloud buffer full
- Volumes of data
  - 2 TB/day
- Bandwidth
  - 100 MB/s
- OSes
  - SL6
- APIs
  - Management API for VMs and Storage

# CTA use case (IFAE)

- CPU Requirements (hours)
  - 2M core hours
- Peak Requirements (CPU, RAM, Disk Storage)
  - 2.5 GHz, 8GB RAM/Core 500 GB storage per job. 2000 cores
  - Server size: 24 cores, 86 GB RAM
- Data Requirements (Quantity of data)
  - 20 TB 20M files
- Single/Multiple Binaries
  - Multiple, Sequence (Simulation – Reconstruction – Analysis)
- Programming Model
  - Serial, may require multi-threading
- Interaction with user
  - Batch need to assess feasibility of interactive modes
- External connectivity
  - 2 Gbps
- Volumes of data
  - 10 TB/day
- Bandwidth
  - 2 Gbps
- OSes
  - SL6
- APIs
  - Management API for VMs and Storage

# MAGIC use case (IFAE)

- CPU Requirements (hours)
  - 700 core hours/day during 6 months
- Peak Requirements (CPU, RAM, Disk Storage)
  - 20 servers 2 cores, 8GB RAM/vCPU, 2 TB shared storage
  - Server size: 2 cores 4 GB RAM per core
- Data Requirements (Quantity of data)
  - 2 TB shared storage
- Single/Multiple Binaries
  - Multiple
- Programming Model
  - Serial
- Interaction with user
  - Batch and interactive
- External connectivity
  - 10 Gbps
- Volumes of data
  - 100 GB/day
- Bandwidth
  - 1 Gb/s
- OSes
  - RH Linux
- APIs
  - Management API for VMs and Storage

# Pierre Auger use case (KIT)

- CPU Requirements (hours)
  - 1 000 000
- Peak Requirements (CPU, RAM, Disk Storage)
  - 2000 CPUs, 2GB RAM/core
  - Server size: 1CPU, 2GB RAM
- Data Requirements (Quantity of data)
  - 10 TB local
- Single/Multiple Binaries
  - Multiple
- Programming Model
  - Serial
- Interaction with user
  - Batch
- External connectivity
  - 500MB output data
- Volumes of data
  - 2 TB/day
- Bandwidth
  - 250 Mbps
- OSes
  - SL6
- APIs
  - Management API for VMs and Storage

# ELIXIR use case (EMBL)

- CPU Requirements (hours)
  - >4M hours
- Peak Requirements (CPU, RAM, Disk Storage)
  - 1000 cores, 4.5TB RAM total, 1PB storage total
  - Server size: 16 cores, 64GB RAM, 1TB SSD storage
- Single/Multiple Binaries
  - Multiple
- Programming Model
  - Multiple
- Interaction with user
  - Batch and Interactive
- External connectivity
  - 1PB ingress, 0.25 PB egress
- Volumes of data
  - Driven by dataset import, otherwise minimal
- Bandwidth
  - 10 Gb/s
- OSes
  - CENTOS
- APIs
  - Management API for VMs and Storage

# Euro-BioImaging use case (EMBL)

- CPU Requirements (hours)
  - 700k/840k/980k hours/year (Y1/Y2/Y3)
- Peak Requirements (CPU, RAM, Disk Storage)
  - 160 cores, 480 GB RAM, 4230 GB storage (15 VMs, 250 GB local storage + 32 GB image)
  - Server size: 16 cores, 32GB RAM (i.e. 5 VMs each 16 cores, 32 GB RAM; 10 VMs each 8 cores, 32 GB RAM, 32 GB image, 250 GB local data storage)
- Data Requirements (Quantity of data)
  - 300 TB + 10 TB scratch
- Single/Multiple Binaries
  - Multiple
- Programming Model
  - Multi-threading
- Interaction with user
  - Batch and interactive
- External connectivity
  - Initial import of 40 TB, minimal for the rest of the time
- Aggregate transfer volume
  - 20 TB/day
- Bandwidth
  - 10 Gb/s guaranteed BW from each supplier Data Centre to GEANT)
- OSes
  - Linux RH
- APIs
  - Management API for VMs and Storage

# PanCancer use case (EMBL)

- CPU Requirements (hours)
  - 4M hours
- Peak Requirements (CPU, RAM, Disk Storage)
  - 1000 Cores, 4.3 TB RAM, 1 PB
  - Server size: 16  cores, 64GB RAM, 1TB SSD storage
- Data Requirements (Quantity of data)
  - 1 PB + 25 TB for scratch
- Single/Multiple Binaries
  - Multiple
- Programming Model
  - Parallel
- Interaction with user
  - Batch
- External connectivity
  - 1 PB ingress 0.25 egress
- Volumes of data
  - Minimal
- Bandwidth
  - 10 Gb/s
- OSes
  - CENTOS
- APIs
  - Management API for VMs and Storage

# PETRA III use case (DESY)

- CPU Requirements (hours)
  - 100 jobs/day
- Peak Requirements (CPU, RAM, Disk Storage)
  - 1-128 cores on single node, 300MB-2GB RAM/socket, 300-1GB disk
  - Server size: 64 cores CPU, 32GB RAM, 200 GB disk
- Data Requirements (Quantity of data)
  - 20-100 GB
- Single/Multiple Binaries
  - Multiple
- Programming Model
  - Serial and multi-threading
- Interaction with user
  - Batch
- External connectivity
  - 1GBps
- Volumes of data

- Bandwidth
  - 10 Gbps
- OSes
  - Linux 64bit
- APIs
  - Management API for VMs and Storage

# NAF use case (DESY)

- CPU Requirements (hours)
  - 4h-2d VM allocation: week
- Peak Requirements (CPU, RAM, Disk Storage)
  - 5000 cores, 3GB RAM/core, 20GB/core local storage
  - Server size: 8 cores CPU, 16GB RAM, 160 GB local storage
- Data Requirements (Quantity of data)
  - No Cloud storage needed
- Single/Multiple Binaries
  - Multiple
- Programming Model
  - Serial and multi-threading
- Interaction with user
  - Batch
- External connectivity
  - 1Mbps min, 5Mbps max
- Volumes of data

- Bandwidth
  - 1Mbps
- OSes
  - SL6
- APIs
  - Management API for VMs and Storage

# 3DIX use case (ESRF)

- CPU Requirements (hours)
  - 1 Month
- Peak Requirements (CPU, RAM, Disk Storage)
  - 32 cores, 64 GB RAM, 1 TB
  - Server size homogeneous
- Data Requirements (Quantity of data)
  - 1 TB
- Single/Multiple Binaries
  - Multipe
- Programming Model
  - Serial, may reqmulti-threading
- Interaction with user
  - Batch and interactive
- External connectivity
  - Input 100 GB, output 1 GB
- Volumes of data
  - 100 GB/hour
- Bandwidth
  - 1 Gb/s
- OSes
  - Debian based
- APIs
  - Management API for VMs and Storage

# OCEAN use case (ESRF)

- CPU Requirements (hours)
  - 4400
- Peak Requirements (CPU, RAM, Disk Storage)
  - 128 Cores, 64 GB RAM, 2 TB
  - Server size: 32 Cores
- Data Requirements (Quantity of data)
  - 2TB
- Single/Multiple Binaries
  - Multiple (Infiniband)
- Programming Model
  - Multiple (MPI on Infiniband)
- Interaction with user
  - Batch and Interactive
- External connectivity
  - 10 MB in/out
- Volumes of data
  - 8Mb/s
- Bandwidth
  - 8Mbps
- OSes
  - Debian Linux
- APIs
  - Management API for VMs and Storage

# OSIRIS use case (DESY)

- CPU Requirements (hours)
  - 100k for 1M core hours
- Peak Requirements (CPU, RAM, Disk Storage)
  - 20 servers with a shared storage (flexible allocation depending on user demand): 2 core and 8G each one, 2TB shared storage Server size: 10 000 cores
  - Server Size: 100-10000 core, 1-2 GB per core
- Data Requirements (Quantity of data)
  - 10 TB
- Single/Multiple Binaries
- Programming Model
  - Serial, may require multi-threading
- Interaction with user
  - Batch and  intractive
  - External connectivity
  - 1 Gbps
  - Needed to extensive number of sites
- Volumes of data
  - 20 TB/day
- Bandwidth
  - 10 Gb/s guaranteed BW from each supplier Data Centre to GEANT
- Oses
  - Linux
- APIs
  - Management API for VMs and Storage

# BBMRI/LSGA use case (SurfSARA)

- CPU Requirements (hours)
  - 100 CPU hours/sample 50 000 CPU hours total
- Peak Requirements (CPU, RAM, Disk Storage)
  - 12 GB RAM, 250 GB/sample of scratch space
  - Server Size: 8-16 cores, 12-16 GB RAM
- Data Requirements (Quantity of data)
  - 150 TB
- Single/Multiple Binaries
  - Multiple
- Programming Model
  - Parallel using shared memory (OpenMP or similar)
- Interaction with user
  - Batch and interactive
- External connectivity
  - 100 GB per sample (in/out)
- Volumes of data
  - 10 TB/day
- Bandwidth
  - 10 Gb/s
- OS
  - Linux
- APIs
  - Management API for VMs and Storage

# WENMR/HADDOCK use case (SurfSARA)

- CPU Requirements (hours)
  - 2000000 core hours
- Peak Requirements (CPU, RAM, Disk Storage)
  - 2GHz, 4Gb RAM, 500GB disk
  - Server Size: 4vCPUs, 8 GB RAM, 2GB local scratch, up to 500 GB in /home
- Data Requirements (Quantity of data)
  - 100 TB
- Single/Multiple Binaries
- Programming Model
  - Serial, and embarrassingly parallel ☺ (high number of concurrent jobs)
- Interaction with user
  - Batch and interactive
- External connectivity
  - 10MB flat input/output
- Volumes of data
  - 100 TB/year
- Bandwidth
  - 1 Gbps
- OS
  - Scientific Linux
- APIs
  - Management API for VMs and Storage

# Long tail of Science use case (CERN, EGI)

- CPU Requirements (hours)
  - <10 CPU hours
- Peak Requirements (CPU, RAM, Disk Storage)
  - 2 vCPUs, 2GB RAM/vCPU, 50GB storage per VM
  - Server size in function of number of users
- Data Requirements (Quantity of data)
  - 10 GB files
- Single/Multiple Binaries
  - N/A
- Programming Model
  - Serial
- Interaction with user
  - Batch
- External connectivity volumes
  - 100Mbps
- Volumes of data
  - 1Gbps
- Bandwidth
  - 10 Gb/s (between external sites and block storage)
- OSes
  - CENTOS
- APIs
  - Management API for VMs and Storage