



# Updating the CERN Computer Centre Infrastructure

Ian Bird, CERN  
GDB, 21<sup>st</sup> March 2012

# Outline

- Rationale
- Agile Infrastructure project @ CERN
  - Configuration management
  - Virtualisation/Openstack/etc.
  - Monitoring & Analysis
- Cloud initiatives
  - Helix Nebula initiative and project
- Remote extension of the Tier 0
  - Status & timescales

# Rationale

- When CERN developed existing tools, etc there were no other large-scale solutions
  - Now there are ...
- Why is CERN different?
  - Usually it is not
  - When it is, its often a self-imposed difference
- Long term: we need to reduce effort in developing and maintaining home-grown tools
  - Where there are proven solutions used in the wider world
- HEP is no longer large scale in terms of computer centres
- Continually ask why we need to do it that way (any longer)





# Some other reasons

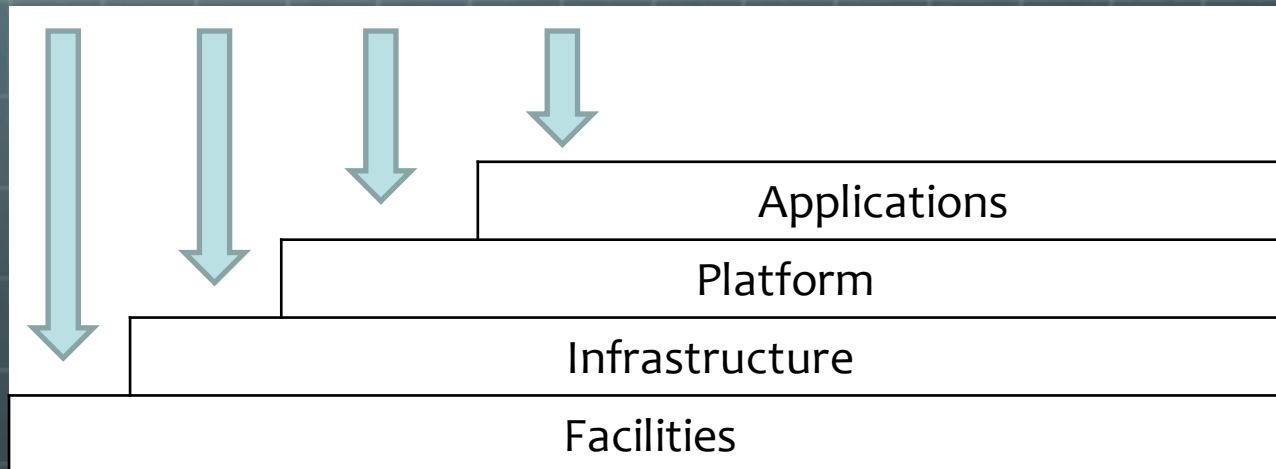
- Today: very hard to add services – time & effort
  - Time to provision a machine, define & set up a service
  - VOBBox a.k.a. PaaS a viable alternative for many services today
- More automation is essential; limited today by some of the tools
- Exploitation of remote Tier 0
- Business continuity
- Optimise resource allocation; live migration (if possible)
- Reduce hardware variations (as far as possible)
- Large community developing tools & enhancements & recipes
- Skills likely to be found in new hires; valuable for departing staff
  - Training and documentation available everywhere



# ?aaS

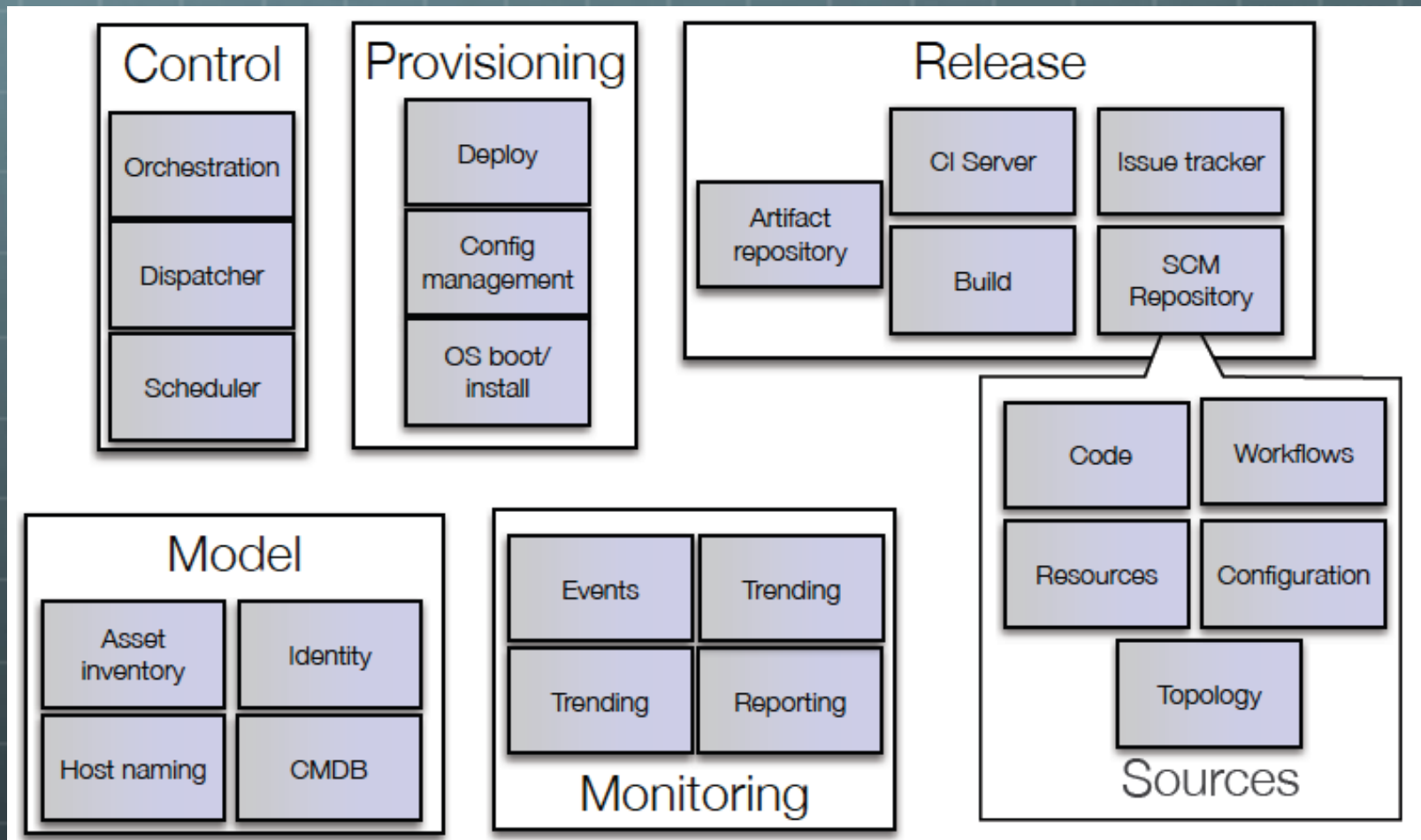
## We need to be able to provision at all levels

-  Facility: CERN CC; safehost; remote Tier 0
-  Infrastructure: {Ixccloud, CVI}
-  Platform: CERN Web services, “VOBox”
-  Software: Indico



-  Need to be more dynamic than now in order to support this

# Use standard tools



# Agile Infrastructure project

- Integrates all the activities in IT that are reworking all aspects of running and managing the CC (and the remote extensions)
- Started with 2 aspects: configuration management (full machine lifecycle); and virtualisation/internal cloud deployment
- In parallel we re-assessed the monitoring infrastructure IT-wide
  - To remove duplicities; Most important: to be able to data mine and analyse monitoring information (automation, trends, etc.)
- Integrated all 3 aspects together into one overall umbrella project
  - Coordinated by Bernd Panzer; with effort drawn from across most IT groups

# Aspects of the Agile Infrastructure

- Areas that deal with:
  - Overall architecture
  - Installation
  - Configuration management
  - Monitoring & accounting (billing?)
  - Orchestration/Scheduling and task/job placement
  - Storage infrastructure
  - Network infrastructure
  - Private and hybrid cloud infrastructure



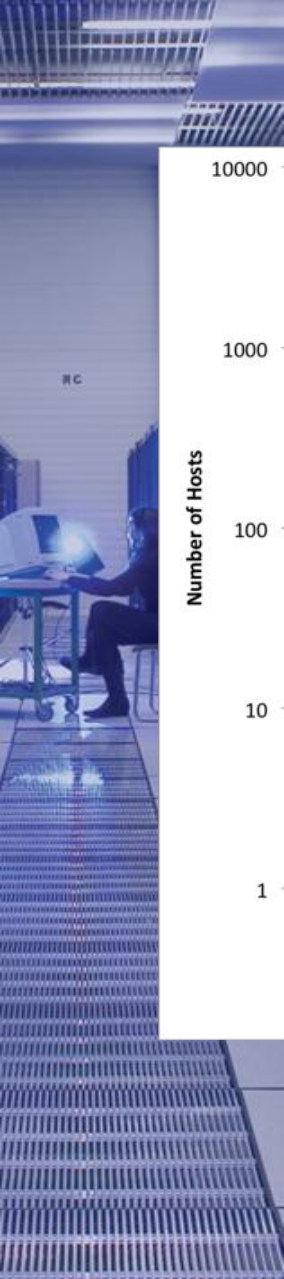
# Configuration management

- The project is reviewing the entire CERN computer-centre management toolset
  - What happens from the bare metal up
  - Asset management, inventory
  - Sysadmin tools and maintenance workflows
  - Service management and configuration tools
  - Dynamic configuration for ‘virtual’ hosts
  - Operations monitoring
  - Workflow automation and continuous deployment



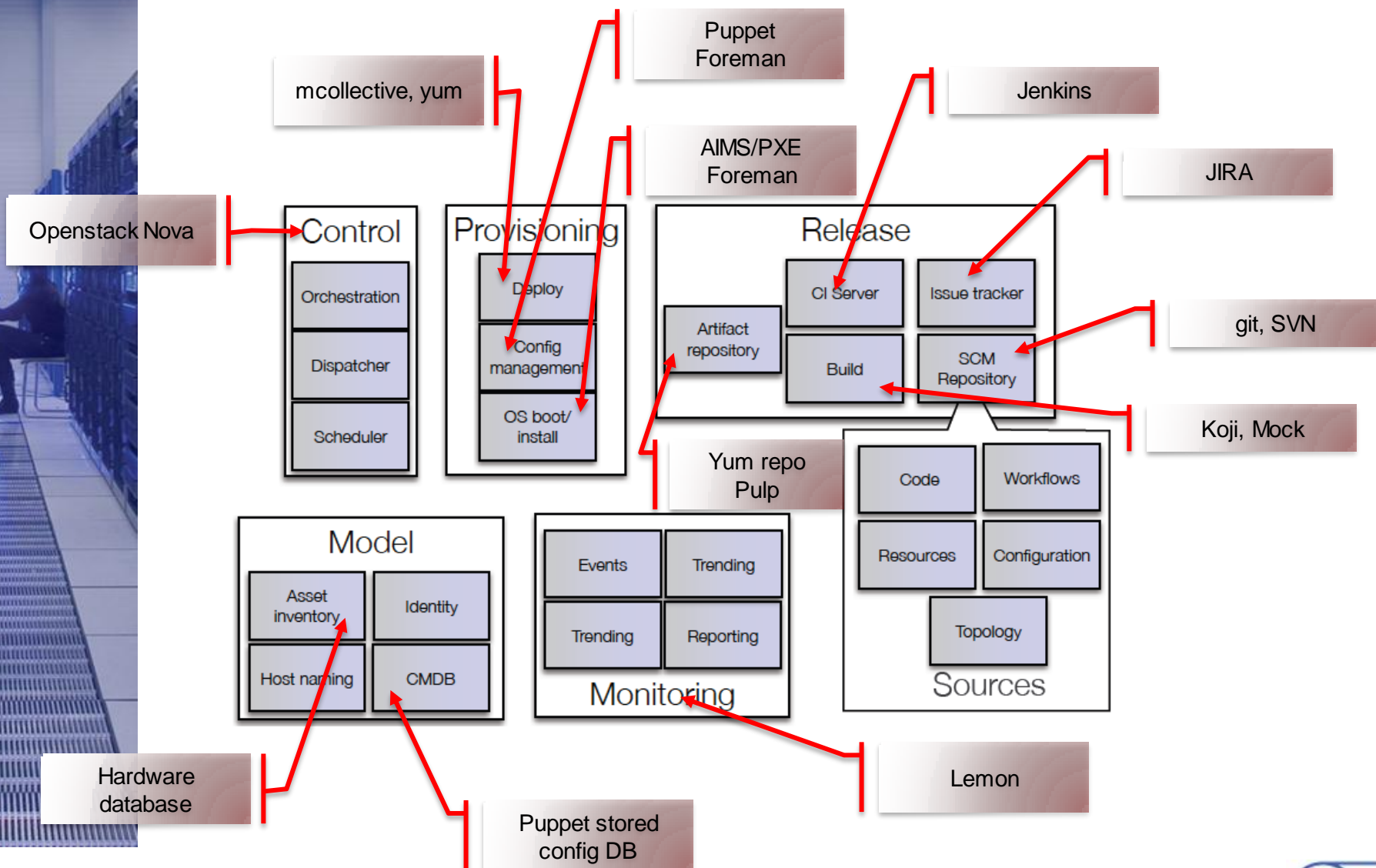
- Current production system built around the Quattor toolset is successfully managing 10k servers
  - (CERN) Quattor + many CERN components
- Currently we have  $O(10k)$  physical nodes
- IaaS approach:
  - Moving to virtual machines
  - More (smaller, load-balanced) service nodes
  - VMs for raw compute (batch or pilot jobs)
  - Homogeneous: compute + storage on the same node
- Add another computer centre, 24/48 SMT cores per node, you get **100k – 300k virtual nodes** to be managed
  - 99.6%<sup>(1)</sup> node update success-rate means **1200 manual interventions** to “fix it”

<sup>(1)</sup> in a recent intervention on lxbatch



- 

# Current tool snapshot (liable to change)





# VM management

# Today

- LXCloud pilot:
  - Based on OpenNebula, providing EC2 services for physics
- CVI (CERN Virtual Infrastructure) “kiosk”:
  - Uses Hyper-V providing long-lived server-like workloads (VOBoxes, etc)
- Neither is ideal for long-term support or scale
  - Different teams with different initial goals

# Some goals

- Support CERN needs and use-cases
  - Existing use cases of LXCloud and CVI
  - PaaS, bursting to commercial or academic clouds
  - Dynamic provisioning of services for users
- More efficient use of hardware
  - Reduce impact of failure
  - Can we reduce the amount of “specialized” hardware types?
- Improve usage accounting and monitoring
  - Understand usage and provisioning better
- Management & provisioning of remote data centre
  - Need a more dynamic way to provision services (how dynamic tbd!)
  - Will have no local admins
- Exploit work done in industry & elsewhere
  - Load balancing, schedulers, etc., etc.

# Openstack

- Open source software components to help run a cloud infrastructure
  - Is “Amazon-like” (i.e. EC2 and S3)
- Large community –
  - Significant industry support: 149 companies
  - Several large research organisations
- Active community:
  - Design summits, conferences, user groups
- Openstack Foundation:
  - Independent of a company (and thus associated risks)

# Openstack projects:

## The OpenStack Core Projects



Compute

OPENSTACK COMPUTE: open source software and standards for large-scale deployments of automatically provisioned virtual compute instances.



Object Storage

OPENSTACK OBJECT STORAGE: open source software and standards for large-scale, redundant storage of static objects.



Image Service

OPENSTACK IMAGE SERVICE: provides discovery, registration, and delivery services for virtual disk images.

- “Incubated” projects becoming part of the core now:
  - Keystone (AA across openstack projects and integration with existing authN systems)
  - Horizon (dashboard: portal for admins and users)
- Large number of community projects:
  - Network, load balancing, gateways, installation/maint., etc.

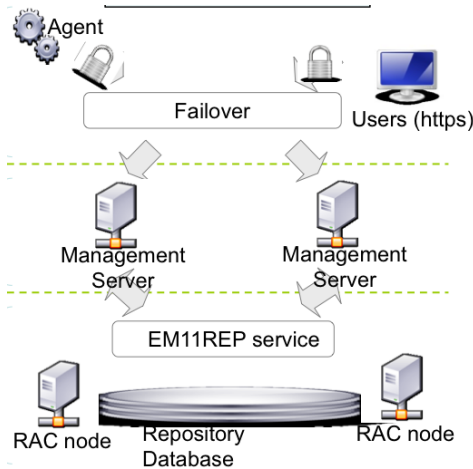
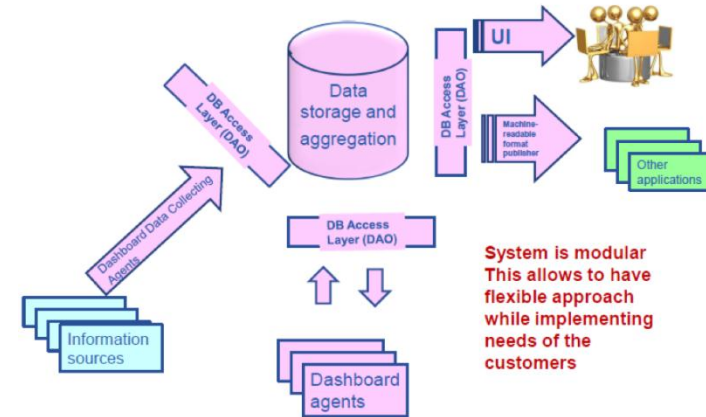
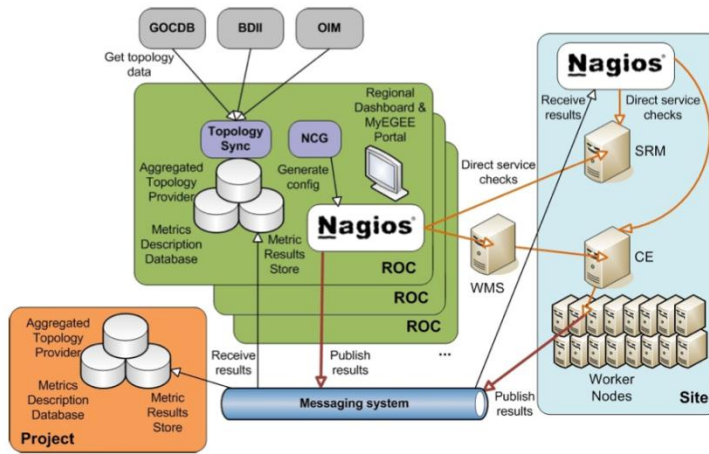
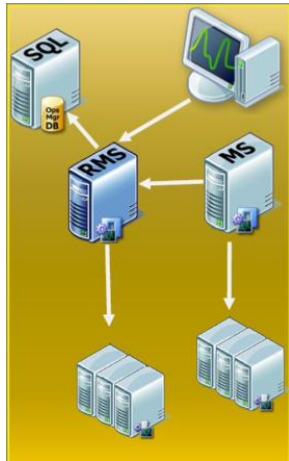


# Monitoring

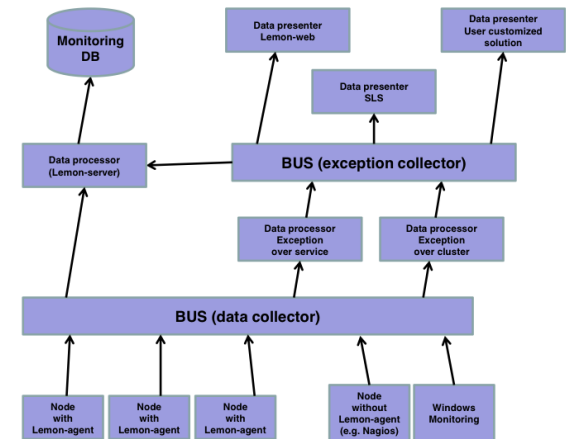
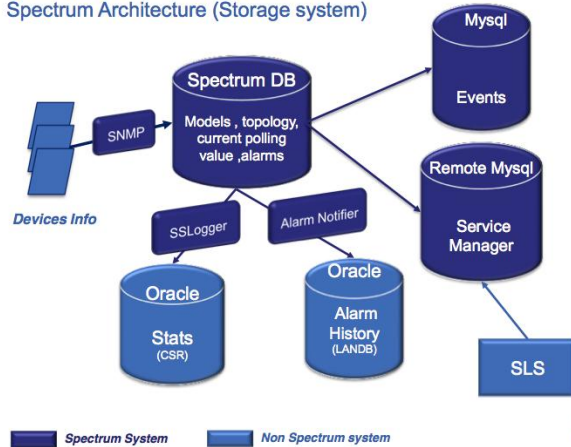
# Motives

- We have lots of monitoring – and a successful framework
  - Lemon and SLS widely used
- No coherent monitoring architecture
- Lots of teams starting to want additional facilities, including log mining and analysis
  - Potential for huge duplication of effort
- Recognise a need to be able to do across the board analysis
  - Trends, correlations, etc
  - Need to be able to automate responses to complex situations
  - Need to be able to react to changing usage patterns
- Had internal workshop, and working group to agree a common architecture and approach

# Monitoring Applications



Spectrum Architecture (Storage system)





- Producers
  - 40538
- Input Volume
  - 283 GB per day
- Input Rate
  - 697 M entries per min
  - 2,4 M entries per min without (batch) process accounting
- Query Rate
  - 52 M queries per day
  - 3,3 M entries per day without process accounting

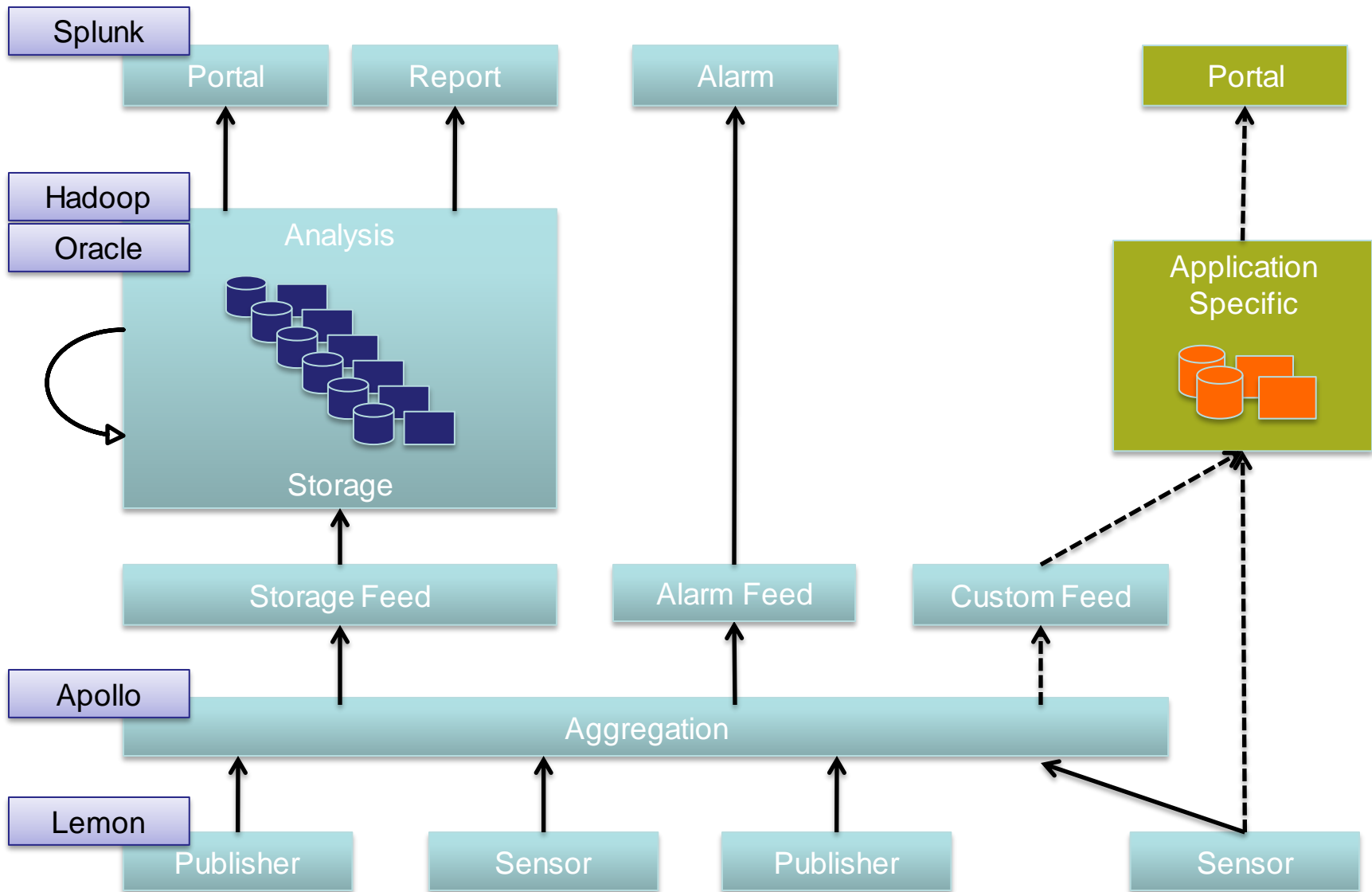
- Monitoring in IT covers a wide range of resources
  - Hardware, OS, applications, files, jobs, etc
- Many application-specific monitoring solutions
  - Some are commercial solutions
  - Based on different technologies
- Limited sharing of monitoring data
  - Maybe no sharing, simply duplication of monitoring data
- All monitoring applications have similar needs
  - Publish metric results, aggregate results, alarms, etc



- Focus on providing well established solutions for each layer of the monitoring architecture
  - Transport, storage, analysis
- Flexible architecture where a particular technology can be easily replaced by a better one
- Adopt whenever possible existing tools and avoid home grown solutions
- Follow a tool chain approach
- Allow a phased transition where existing applications are gradually integrated

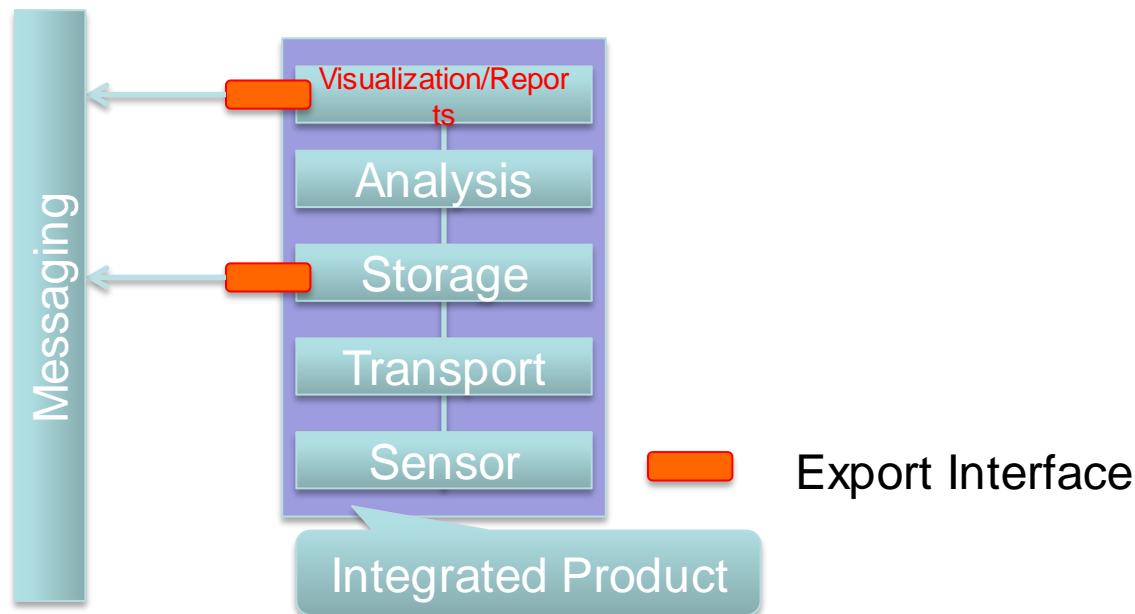
- User stories were collected from all IT groups and commonalities between them were identified
- To guarantee that different types of user stories were provided three categories were established:
  - Fast and Furious (FF)
    - Get metrics values for hardware and selected services
    - Raise alarms according to appropriate thresholds
  - Digging Deep (DD)
    - Curation of hardware and network historical data
    - Analysis and statistics on batch job and network data
  - Correlate and Combine (CC)
    - Correlation between usage, hardware, and services
    - Correlation between job status and grid status

# Architecture Overview



- Hadoop is a good candidate to start with
  - Prior positive experience in IT and the experiments
  - Map-reduce paradigm is a good match for the use cases
  - Has been used successfully at scale
  - Many different NoSQL solutions use Hadoop as backend
  - Many tools provide export and import interfaces
  - Several related modules available (Hive, HBase)
- Document based store also considered
  - CouchDB/MongoDB are good candidates
- For some use cases a parallel relational database solution (based on Oracle) could be considered

- External (commercial) monitoring
  - Windows SCOM, Oracle EM Grid Control, Spectrum CA
- These data sources must be integrated
  - Injecting final results into the messaging layer
  - Exporting relevant data at an intermediate stage





Monitoring.v1  
Q1 2012

- AI nodes monitored with Lemon (dependency on Quattor)
- Deployment of Messaging Broker and Hadoop cluster
- Testing of other technologies (Splunk)

Monitoring.v2  
Q2 2012

- AI nodes monitored with Lemon (no dependency on Quattor)
- Lemon data starts to be published via messaging

Monitoring.v3  
Q4 2012

- Several clients exploiting the messaging infrastructure
- Messaging consumers for real time alarms and notifications
- Initial data store/analysis for select use cases

Monitoring.v4  
Q4 2013

- Monitoring data published to the messaging infrastructure
- Large scale data store/analysis on Hadoop cluster

# Monitoring: note

- This architecture is (surprise!) essentially that of the WLCG monitoring infrastructure
  - Although scope and scales are different
  - Anticipate significant re-use of experience & tools – in both directions
    - E.g. we start to prototype an analysis service based on Hadoop and ...

# Agile Infrastructure project: summary

Year	What	Actions
2011		Agree overall principles
2012		Prepare formal project plan Establish IaaS in CERN CC Production Agile Infrastructure Monitoring Implementation as per WG Migrate Ixcloud Early adopters to Agile Infrastructure
2013	LS 1 New Data Centre	Extend IaaS to remote CC Business Continuity Support Experiment App re-work Migrate CVI General migration to Agile with SLC6 and Windows 8
2014	LS 1 (to November)	Phase out Quattor/CDB/...

# Clouds & virtualisation

- Several prongs:
- Use of virtualisation in the CERN CC:
  - Lxcloud pilot + CVI → dynamic virtualised infrastructure (which may include “bare-metal” provisioning)
  - No change to any grid or service interfaces (but new possibilities)
  - Likely based on Openstack – see above
  - Other WLCG sites also virtualising their infrastructures
- Use of commercial clouds – “bursting”
  - Additional resources;
  - potential of outsourcing some services?
  - Helix Nebula project – see next slides
- Not discussed here:
  - Can cloud technologies supplement/replace some of the grid services?  
On what timescale?
  - Was hoping TEGs would comment here...



The background of the slide is a deep space image featuring the Helix Nebula. The nebula is a ring-shaped planetary nebula, appearing as a glowing, reddish-brown and blue structure against a dark, star-filled sky. The text "Helix Nebula project" is centered over the nebula in a white, sans-serif font.

# Helix Nebula project

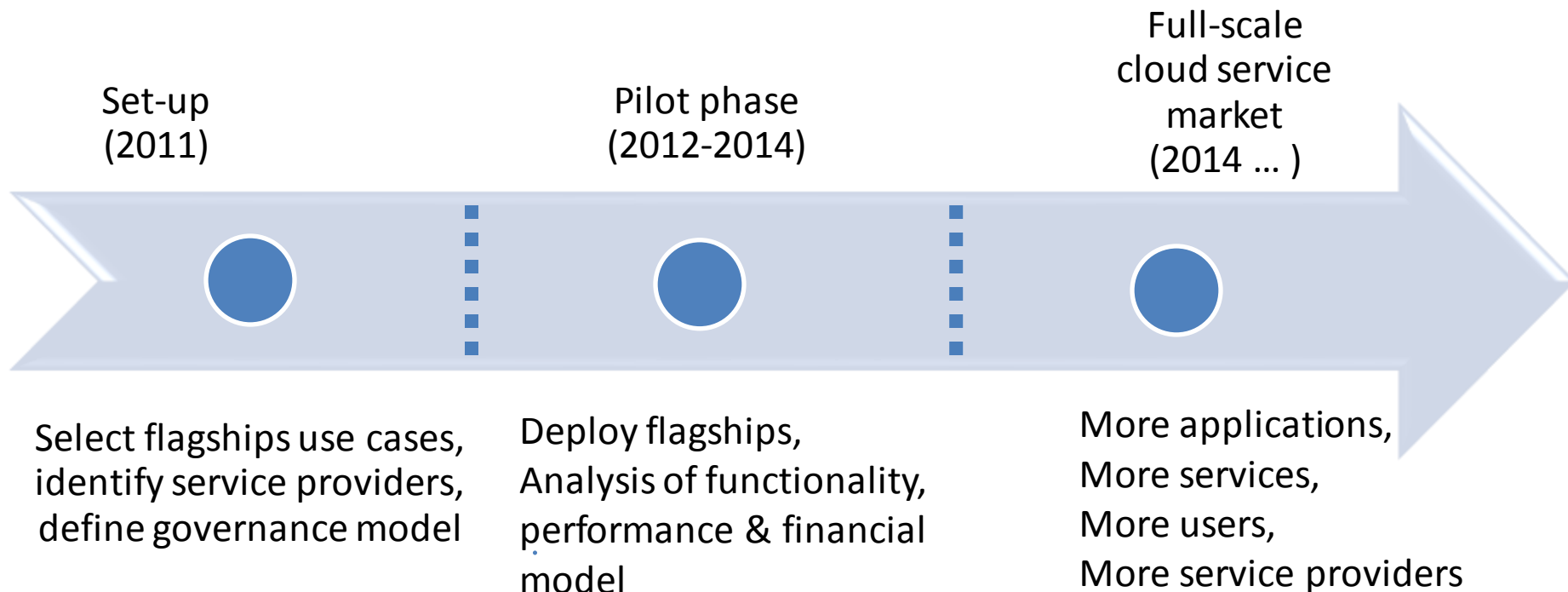


# Origin of the initiative

- Conceived by ESA as a prospective for providing cloud services to space sector in Europe
- Presented to the IT working group of the EIROforum where other members (CERN, EMBL) joined
- Two workshops held during 2011
  - June: hosted by ESA in Frascati
  - October: hosted by EMBL in Heidelberg

EIROforum: CERN, EFDA-JET, EMBL, ESA, ESO, ESRF, European XFEL, ILL

# Timeline



# Role of Helix Nebula: The Science Cloud

Vision of a unified cloud-based infrastructure for the ERA based on Public/Private Partnership, 4 goals building on the collective experience of all involved.

- **Goal One** : Establish HELIX NEBULA – *the Science Cloud* – as a cloud computing infrastructure addressing the needs of the ERA and capable of serving as a platform for innovation and evolution of the overall e-infrastructure.
- **Goal Two**: Identify and adopt suitable policies for trust, security and privacy on a European-level
- **Goal Three**: Create a light-weight governance structure that involves all the stakeholders - and which can evolve over time as the infrastructure, services and user-base grows.
- **Goal Four**: Define a funding scheme involving all the stake-holder groups (service suppliers, users, EC and national funding agencies) for PPP to implement a Cloud Computing Infrastructure that delivers a sustainable and profitable business environment adhering to European- level policies.



# Specific outcomes

- Develop **strategies for extremely large or highly distributed and heterogeneous scientific data** (including service architectures, applications and standardisation) in order to manage the upcoming data deluge
- Analyse and promote trust building towards open scientific data e-Infrastructures covering **organisational, operational, legal and technological aspects**, including authentication, authorisation and accounting (AAA)
- Develop strategies and establish structures aiming at **co-ordination between e-infrastructure operators**
- Create frameworks, including **business models for supporting Open Science and cloud infrastructures based on PPP**, useful for procurement of computing services suitable for e- Science

# Scientific Flagships

- CERN LHC (ATLAS):
  - High Throughput Computing and large scale data movement
- EMBL:
  - Novel *de novo* genomic assembly techniques
- ESA:
  - Integrated access to data held in existing Earth Observation “Super Sites”
- Each flagship brings out very different features and requirements and exercises different aspects of a cloud offering



# ATLAS use case

- Simulations (~no input) with stage out to:
  - Traditional grid storage vs
  - Long term cloud storage
- Data processing (== “Tier 1”)
  - This implies large scale data import and export to/from the cloud resource
- Distributed analysis (== “Tier 2”)
  - Data accessed remotely (located at grid sites), or
  - Data located at the cloud resource (or another?)
- Bursting for urgent tasks
  - Centrally managed: urgent processing
  - Regionally managed: urgent local analysis needs
- All experiences immediately transferable to other LHC (& HEP) experiments



# Immediate and longer term goals

- Determine costs of commercial cloud resources from various sources
  - Compute resources
  - Network transfers into and out of cloud
  - Short and long term data storage in the cloud
- Develop understanding of appropriate SLA's
  - How can they be broadly applicable to LHC or HEP
- Understand policy and legal constraints; e.g. in moving scientific data to commercial resources
- Performance and reliability – compared to WLCG baseline
- Use of standards (interfaces, etc.) & interoperability between providers
- Can CERN transparently offload work to a cloud resource
  - Which type of work makes sense?
- Long term: Can we use commercial services as a significant fraction of overall resources available to CERN experiments?
  - At which point is it economic/practical to rely on 3<sup>rd</sup> party providers?

# Service Procurement

- Assuming pilot phase proves successful, the provision of commercial Cloud services would need to be integrated into the ICT procurement process of the demand-side organisations
- For the initial flagships this implies:
  - Inter-governmental organisations
    - Jurisdiction (governing laws & arbitration), tax-free status, etc.
    - Return on Investment: preference for procurement from each organisation's member-states
  - Pool of commercial service providers that can respond to calls for tender
  - Cannot integrate procurement processes of all demand-side organisations but can converge:
    - Technical specifications & standards
    - Terms and conditions
- EC published Guide for the procurement of standards based ICT Elements of Good Practice (21 Dec 2011)

# Summary

- The objective of this initiative is to establish a sustainable cloud computing infrastructure for the European Research Area based on commercially provided services
- It is a collaborative initiative bringing together all the stakeholders to establish a public-private partnership
- Interoperability with existing e-infrastructures is a goal of the initiative
- It has commitments from the IT industry and user organisations – flagship deployment started Jan 2012
- 3 initial flagship use cases identified
- Framework collaboration EC project to start summer 2012

# Summary

- **“Agile Infrastructure” Project:** CERN IT is re-working the tools used to manage & monitor the CC (and extensions!)
  - To become more industry-standard
  - To address issues of scalability and management effort
  - To be able to more flexibly and dynamically provide services
  - To be able to better monitor and analyse the use of resources
  - Increase automation
- **Cloud investigations:**
  - Internally as part of the above
  - Use of commercial clouds: issues of policy, overall feasibility, data management, cloud models, cost
    - Helix Nebula project is one step in this direction
    - Will consider other direct investigations with other providers
  - **CERN openlab:**
    - New partner Huawei: prototype cloud storage appliance (S3, and others as services)



# Extension of the Tier 0

- In 2006 we first recognised the need for an extension of the Tier 0
- Eventually it was decided that this should not be at CERN
- Several years of investigations and proposals
- Last year – call for proposals
  - Followed by detailed technical follow ups
  - Led to call for tender
- Results adjudicated last week in the Finance Committee (of Council)
- Winner is Budapest, Hungary (Wigner Inst.)

# Remote Tier 0

- Timescales
  - Prototyping in 2012 hopefully
  - Testing in 2013
  - Production in 2014
- This will be a “hands-off” facility for CERN
  - They manage the infrastructure and hands-on work
  - We do everything else remotely
- Need to be reasonably dynamic in our usage
  - Will pay power bill
- Various possible models of usage – sure to evolve
  - The less specific the hardware installed there, the easier to change function
- This facility is another reason for modernising the tools and the way we deliver services
  - (More) Automation is essential to be able to use this effectively



# Summary

- CERN is re-working how we manage the CC infrastructure
  - Address sustainability of solutions
  - Better ways to deliver services
  - More dynamic response to changing environment
  - Need to be able to integrate the remote Tier 0
- Agile Infrastructure project has 3 aspects
  - Management tool chain
  - Use of virtualisation
  - Monitoring
- Also investigate feasibility of using cloud providers

# Milestones

- July 2012
  - 10% resources in CC managed by IaaS system (1000 nodes) including installation, configuration, etc.
- January 2013
  - 50 % resources managed (over 3 sites)
- July 2013
  - 95% resources managed
- Initial focus on service/server consolidation use cases

# Technical details

- IT Technical Forum (Internal IT Dept. seminars):
  - Configuration management:  
<http://indico.cern.ch/conferenceDisplay.py?confId=172002>
  - Openstack:  
<http://indico.cern.ch/conferenceDisplay.py?confId=175809>
  - Monitoring:  
<http://indico.cern.ch/conferenceDisplay.py?confId=175813>
- Will be presentations at HEPiX, CHEP, ...