



15-19 April 2013, Bologna Italy
HEPiX Spring 2013 Workshop

Wojciech Ozga

Faculty of Computer Science, Electronics and Telecommunication
AGH University of Science and Technology in Krakow, Poland
CERN, Geneva, Switzerland

**The CMS openstack, opportunistic, overlay,
online-cluster Cloud (CMSooooCloud)**

CMS HLT farm

Requirements

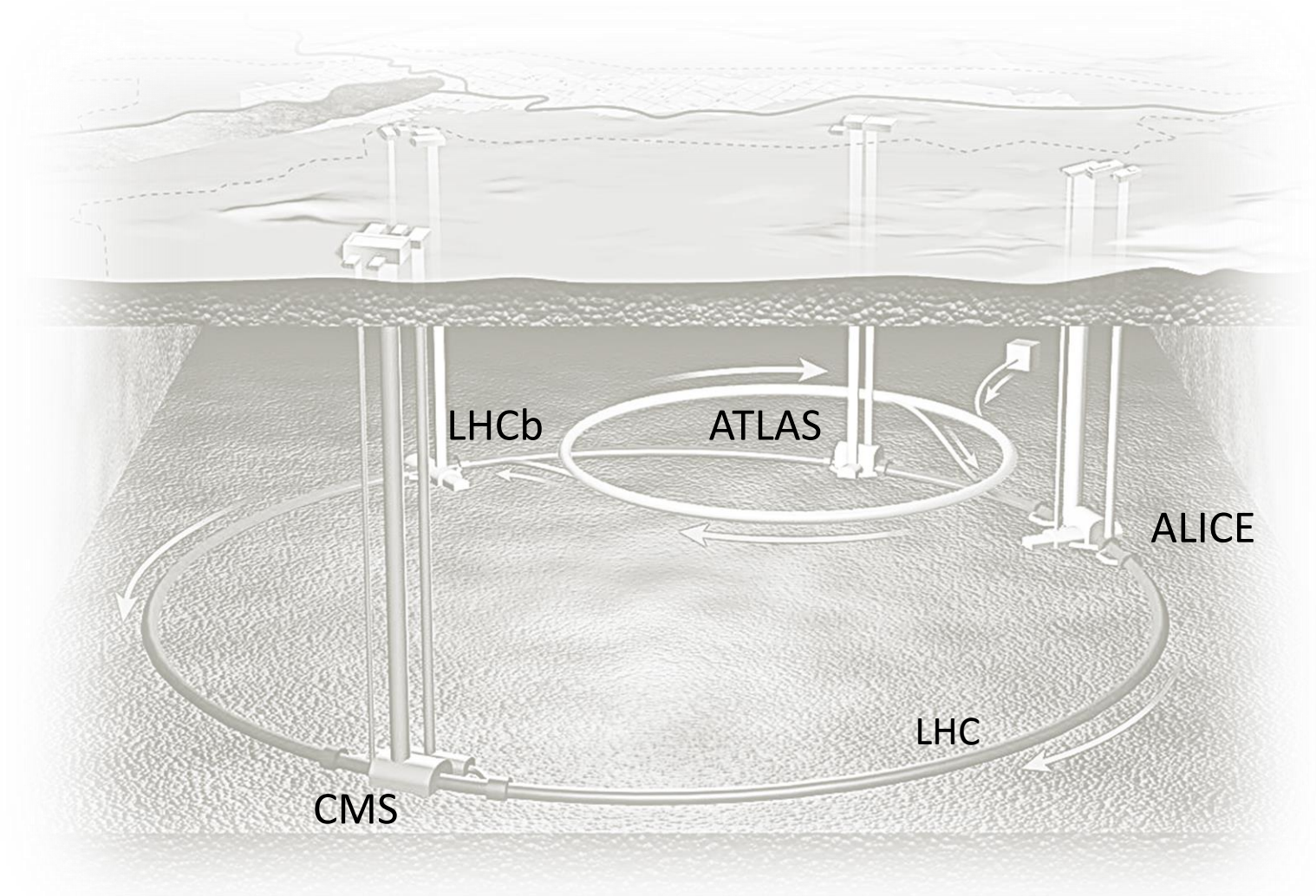
Network virtualization

Hardware virtualization

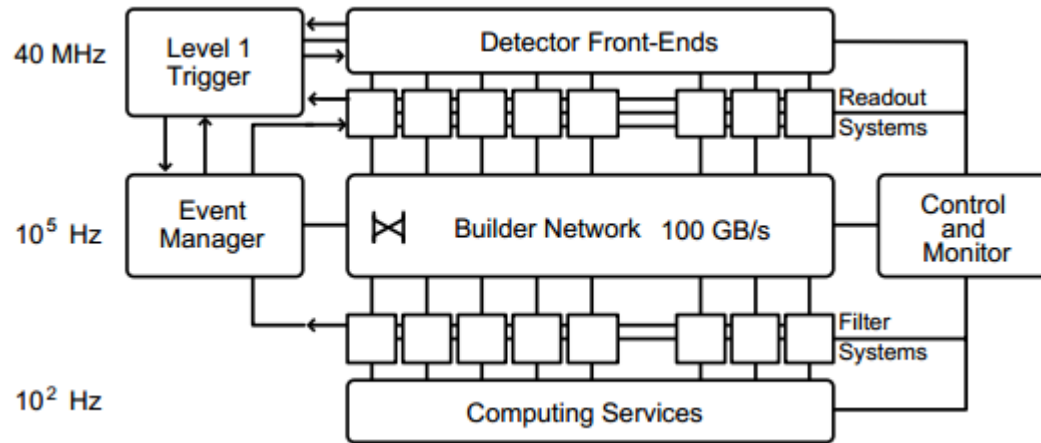
Conclusions

OUTLINE

CMS experiment @ CERN



DAQ system is designed for taking up to **20TBytes** of data per day



CMS HLT farm

Complex

varied hardware
different network schemas
hundreds of OS configurations

Large size

1264 nodes
50 switches

Powerful

13312 cores
26.6 Tbytes of RAM
232 Tbytes of storage

CMS HLT farm vs Tier_[0-2] CPU in HEP-SPEC06

	HLT farm	Tier0	Tier1	Tier2
HEP-SPEC06	195k	121k	150k	399k

<http://wlcg-rebus.cern.ch/apps/pledges/resources/>

Requirements

Minimal Changes

no special hardware reconfiguration
No network reconfiguration

Opportunistic Usage

Quick switching from data taking to other tasks
Flexibility in running different tasks

VIRTUALIZATION

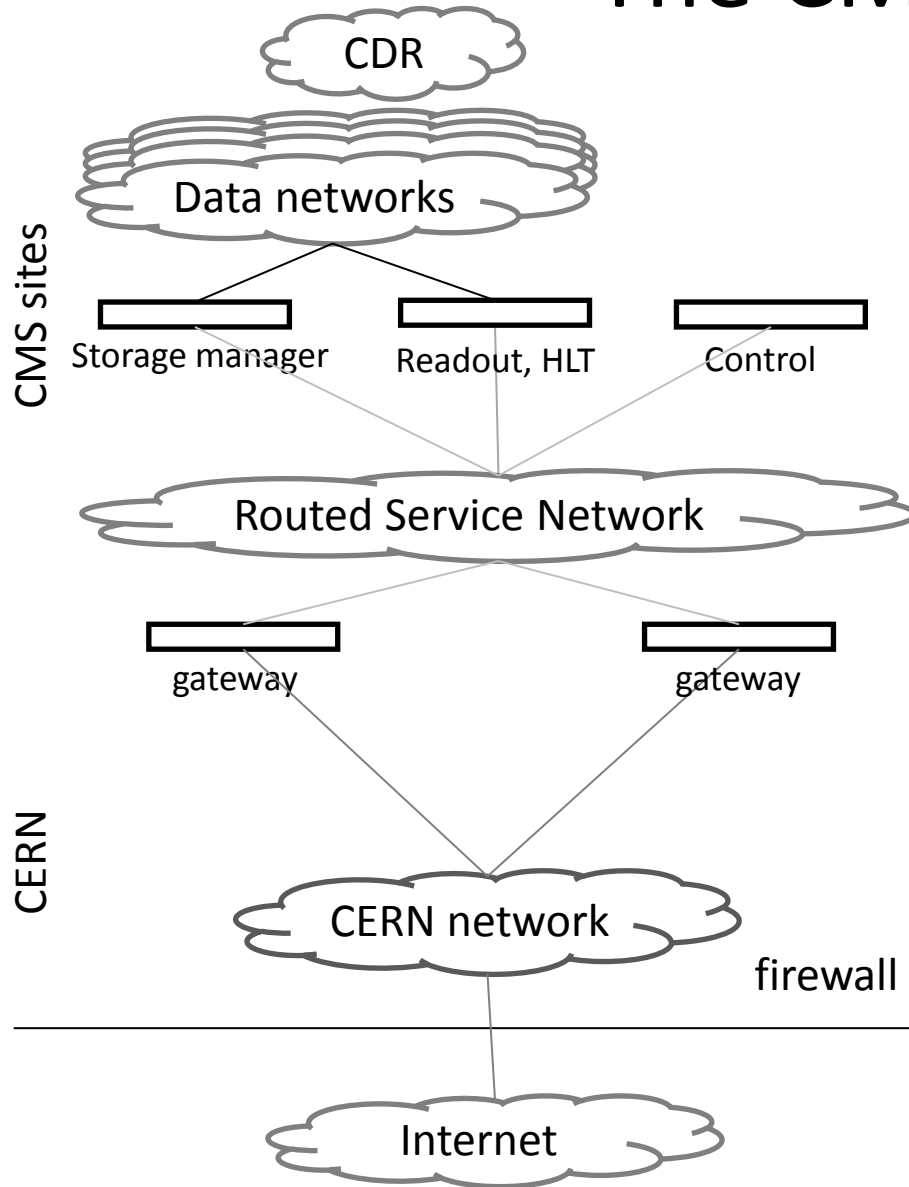
No impact on data taking

Nor during the setup phase
Neither during data taking on Heavy Ion runs

Full control

Full control on hardware and OS
no extra temporary sudoers

The CMS Online Cluster



Private CMS networks

- Service networks (per rack)
~3000 1Gbit ports
- Data networks (VLAN's)
~4000 1Gbit ports
- Central Data Recording (CDR)
Network to Tier0
- Oracle RAC networks
- Subdetectors networks

CERN network

Network virtualization

Integration of cloud network
with CMS network



network virtualization



+

No router/switch re-configuration.

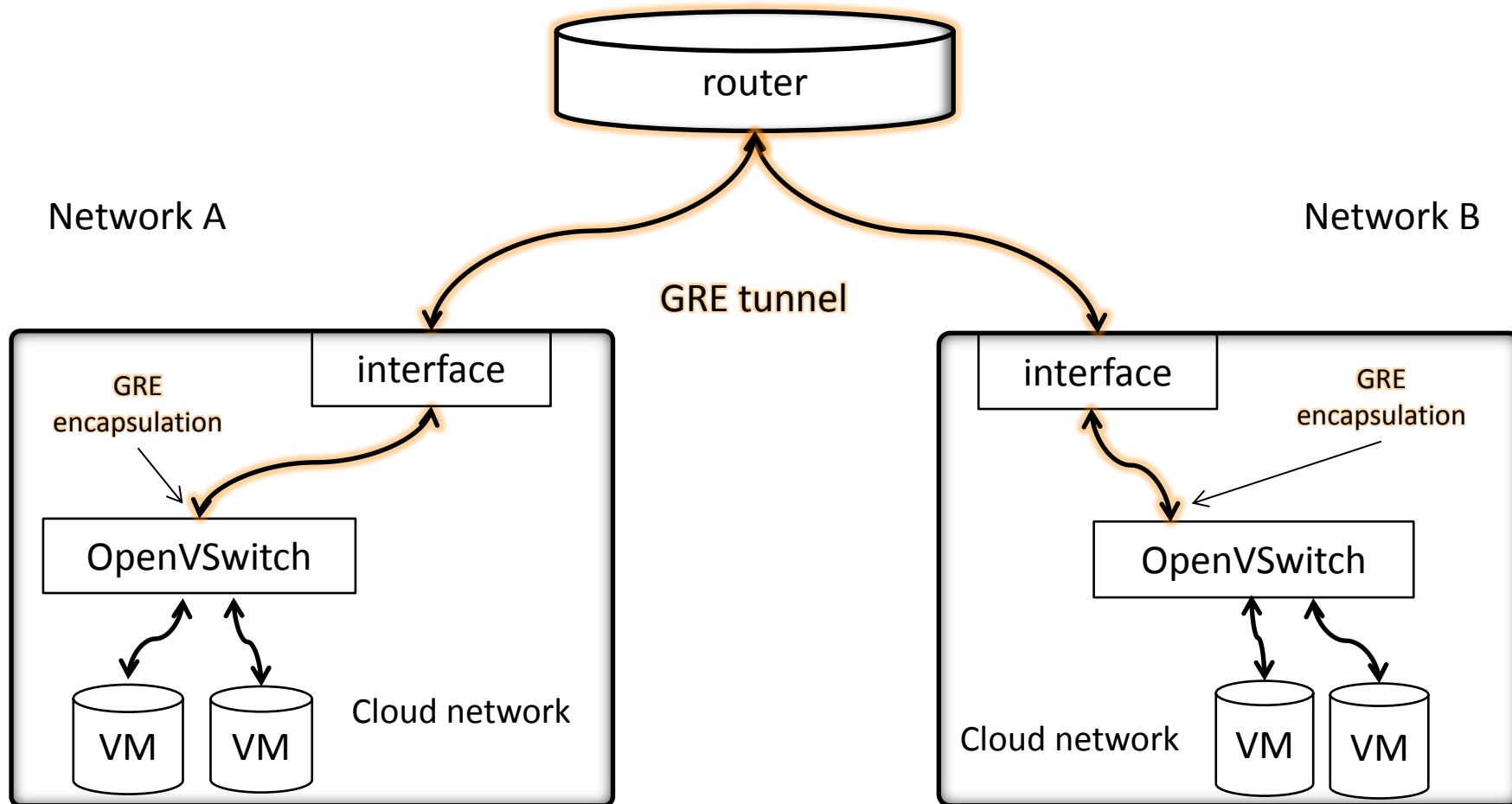
No dependencies between production and cloud network.

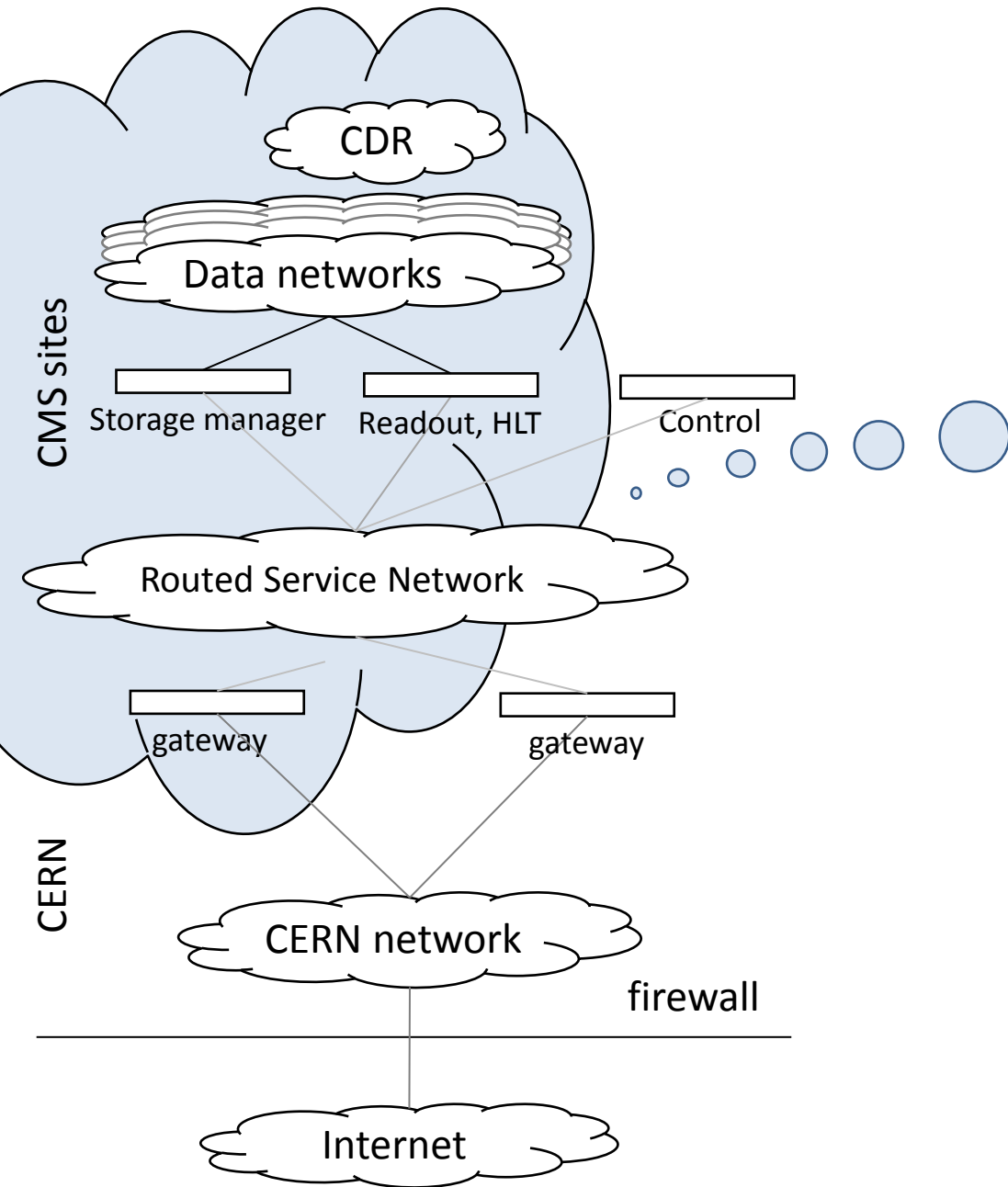
-

Encapsulation overhead.

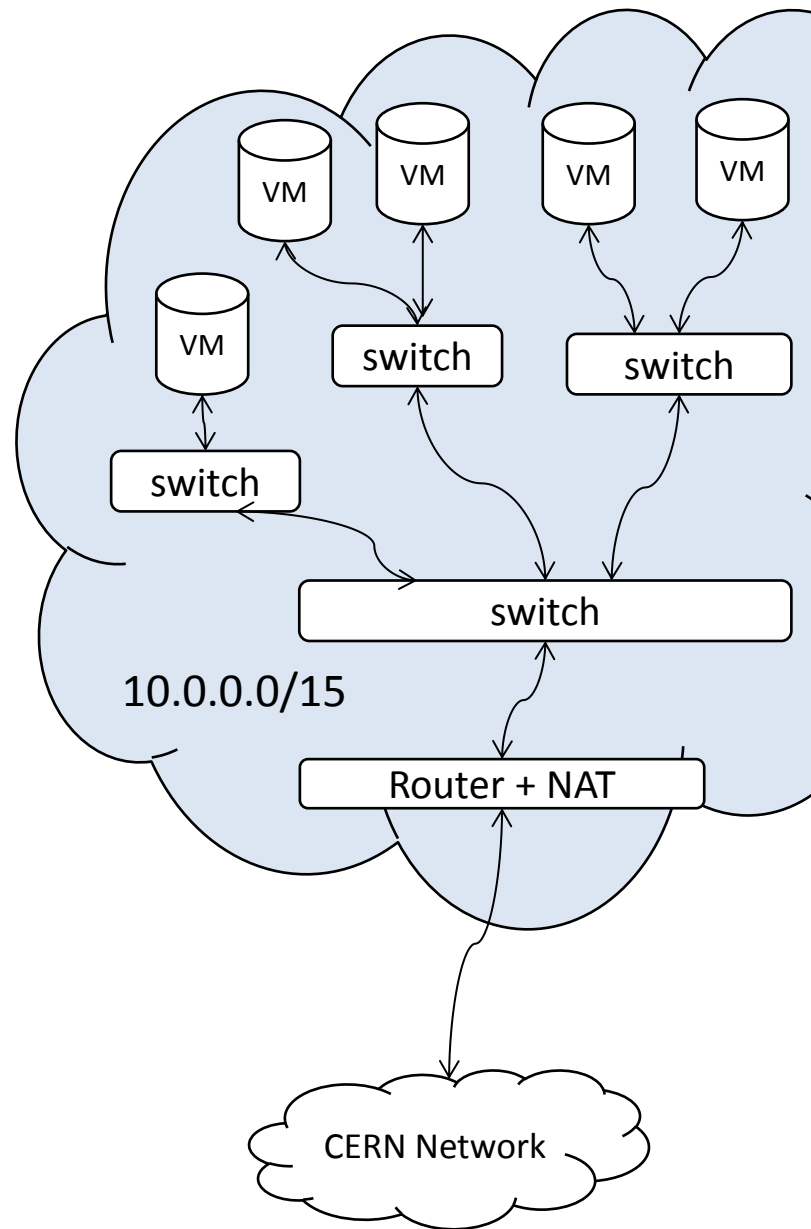
NATing to CERN network required.

Network virtualization





Physical network schema



Logical cloud network schema

Hardware virtualization

Standard approach

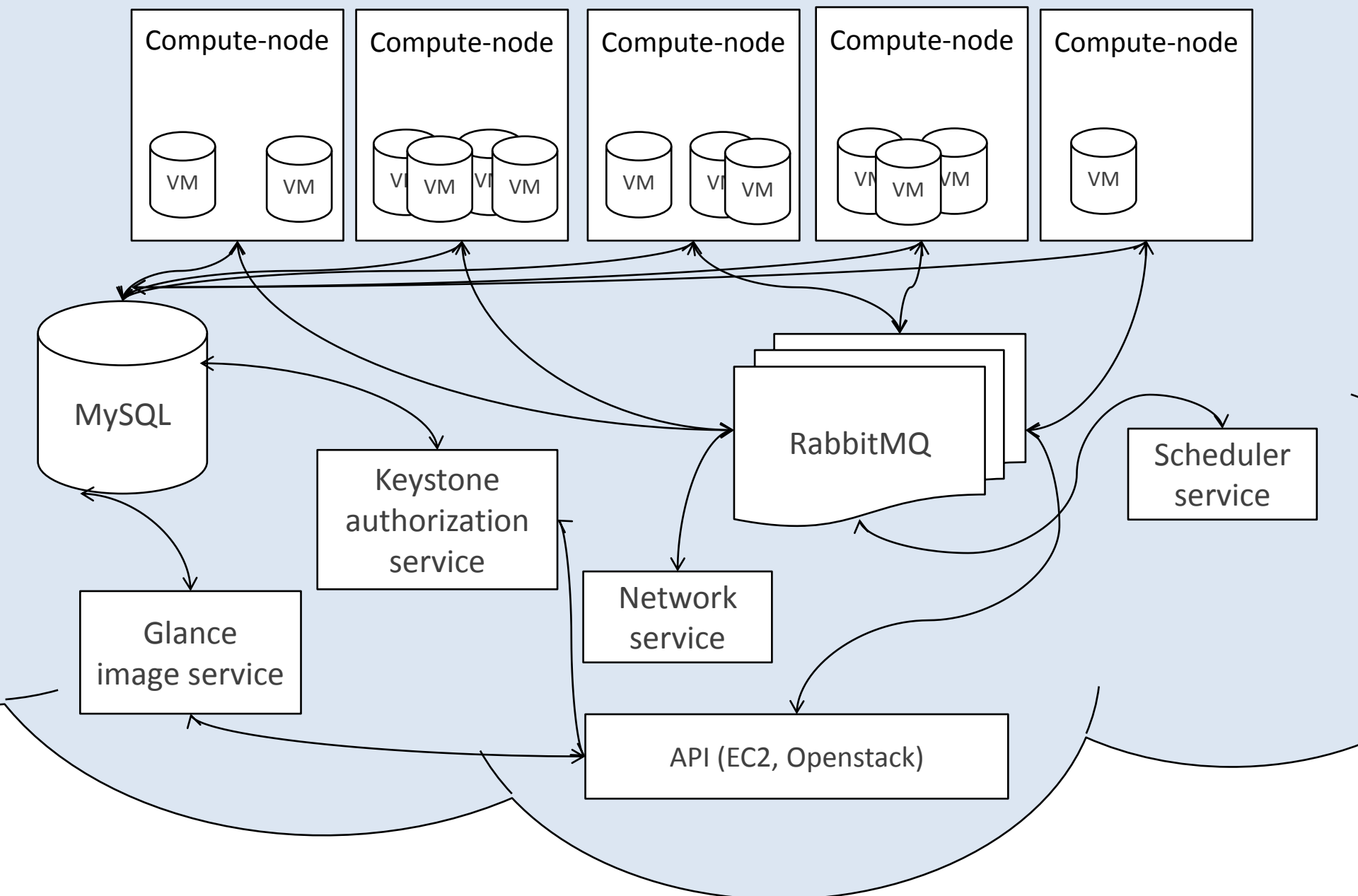
Configure physical machine to run specific task

$$\text{Number of configurations} = \text{Number of tasks} \times \text{number of different hardware configurations}$$

Cloud approach

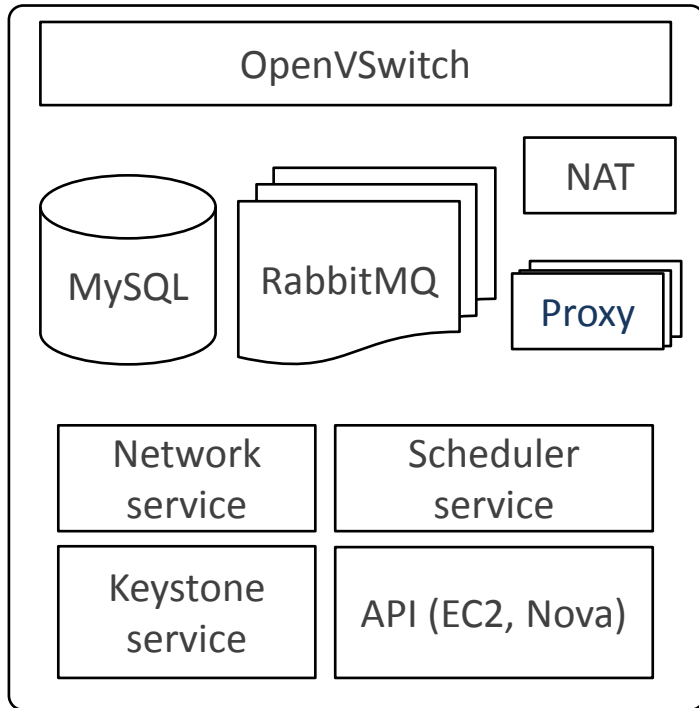
Launch task specific **Virtual Machine** over hypervisor

$$\text{Number of configurations} = \text{Number of tasks} + \text{number of different hardware configurations}$$



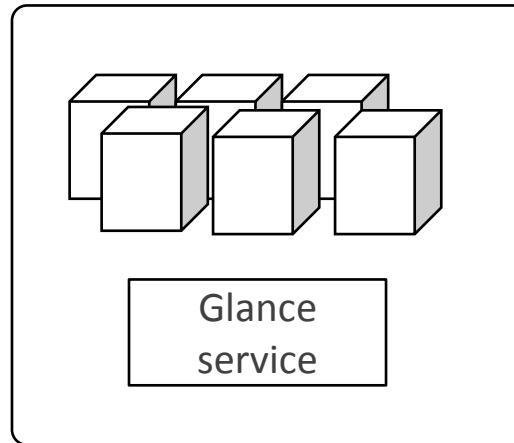
Openstack Cloud Manager

Cloud Architecture



1 x Fat Controller Node

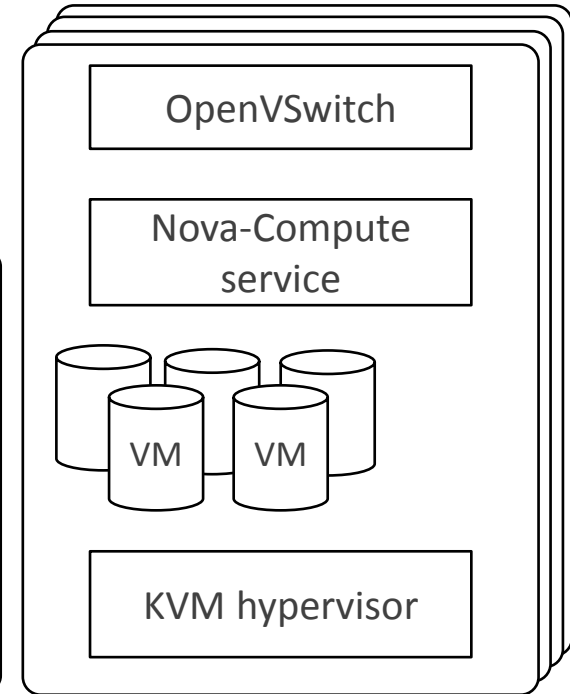
48Gbytes, 8 CPU, 8x1Gbit Ethernet



1 x VM Image Store

High bandwidth

Cache mechanisms



1300 x Compute node

KVM virtualization

OVS network virtualization

"The best advice I know is 'Don't distribute.'"

Adam Bien

CMS Cloud usage

8th – 12th of October 2012

First big scale test: **Folding@Home**

Cloud controls 1300 hypervisors

768 completed Work Units in 12h

21st of December 2012 – 11th of January 2013

Test of stability, running **Folding@Home**

Stable run of 1000 VMs during 3 weeks

Deployment of 250VMs in ~5min

54625 completed Work Units

Since January 2013

Finalizing the integration as a CERN's GRID resource

Outlook

Migration to the new version of OpenStack – Grizzly release
emphasize on **high availability** features

Increase the network connectivity to CERN Tier0
bandwidth connection up to 40Gbit/s

Cloud network reconfiguration while we are in the long shutdown period
benefit from **VLANS** at the expense of Openvswitch

Conclusions

An **Overlay Cloud layer** deployed on CMS HLT cluster
with **Zero impact** on data taking.

Man power dedicated to cloudify the CMS HLT cluster
was about **~1.5 FTE** for **~6months**

CMS online **share knowledge** about cloud installation

CMS HLT cluster computing power is being used in an **opportunistic way**
leaving **full control** over cloud availability to DAQ CMS

Thank you. Questions?