



# 45 Minutes of OpenStack Hate: A Reality Check

Kristian Köhntopp, Cloud Architect Old Fart  
SysEleven







***What is it that we want to do?***



***„Any VM, anywhere.“***

**–Operating POV**



# *„Infrastructure as Code.“*

–Developer POV

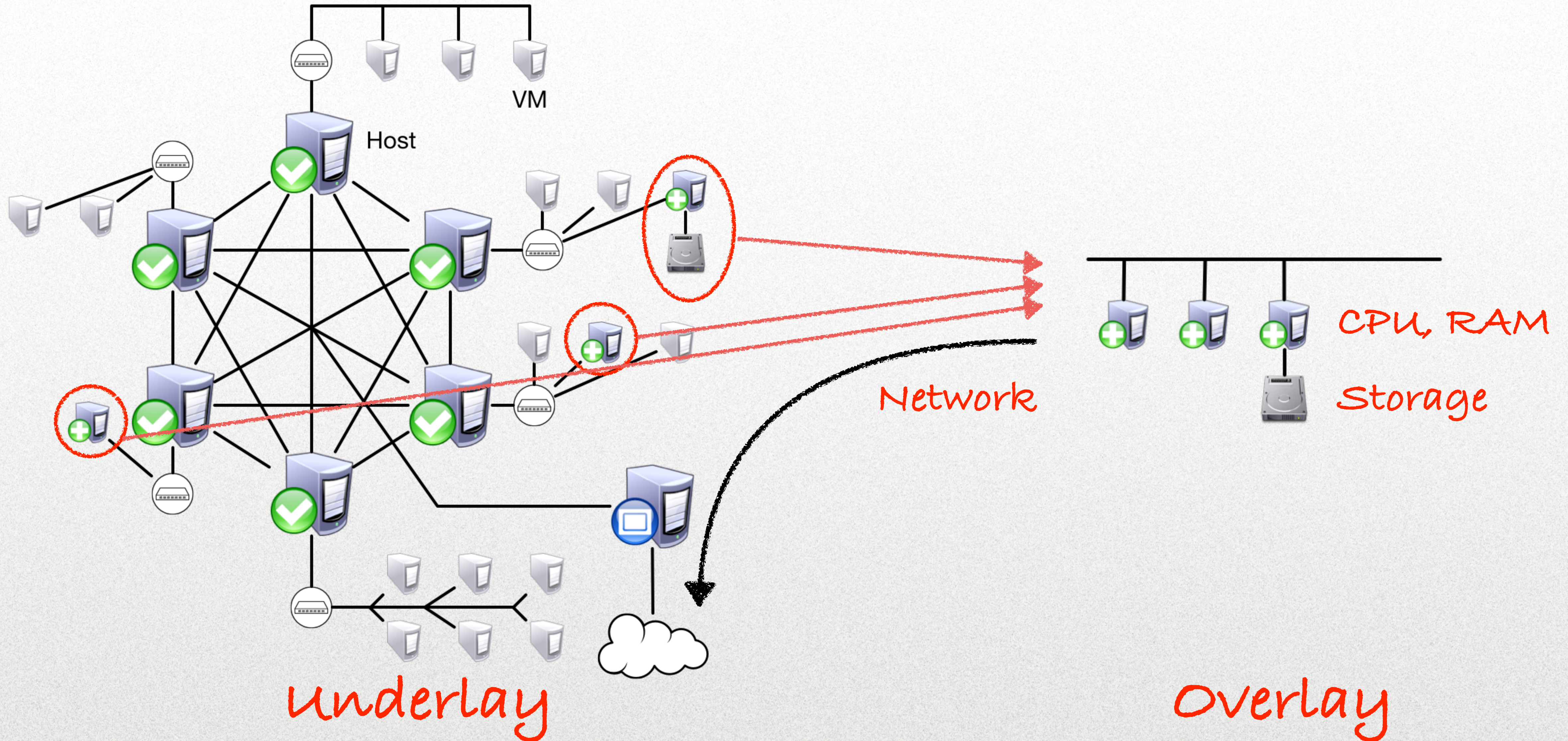


# Why would I need that?

- 24 Cores (48 Cores HT), 256G RAM, 2\* 10GBit und 12\* 3TB HDD including a solid BBU
- 10k EUR
- Applications that actually fill that box are rare. So we are cutting it up to sell off the parts.

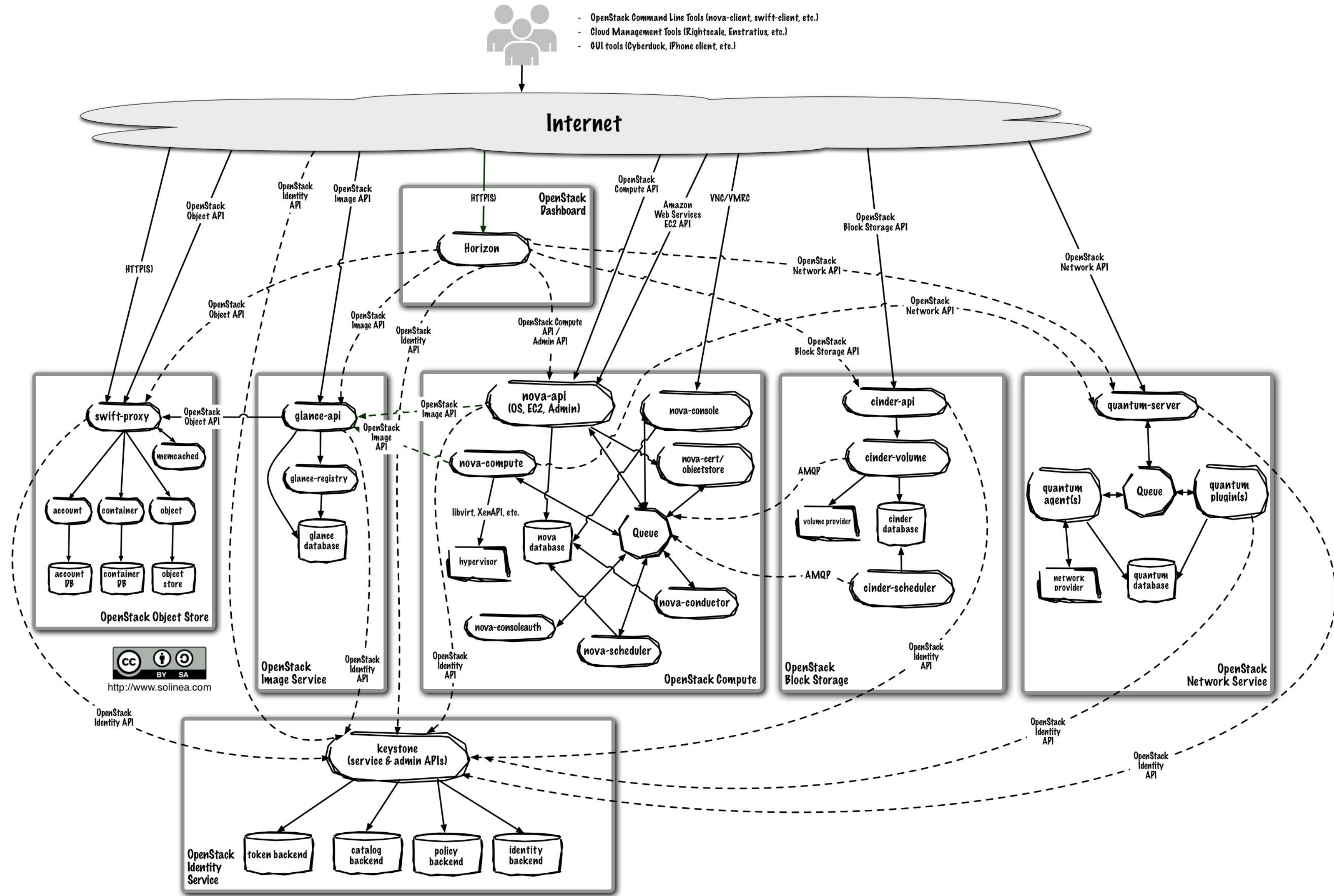


# "Any VM, anywhere"





# Openstack Overview (simplified)





**P A R E N T A L**

**A D V I S O R Y**

**E X P L I C I T C O N T E N T**

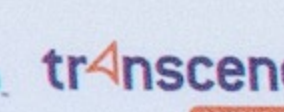
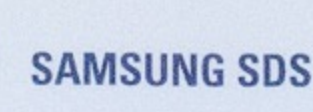
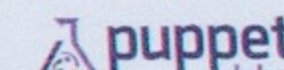
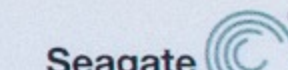
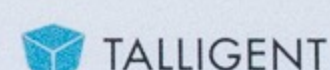
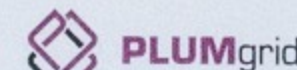
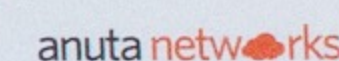
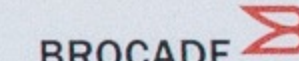
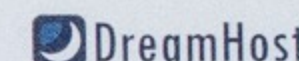
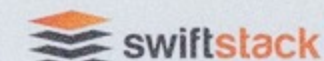
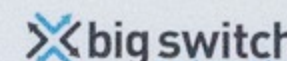
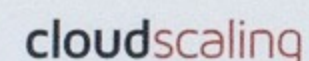
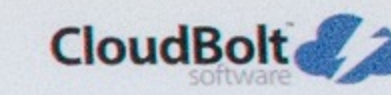
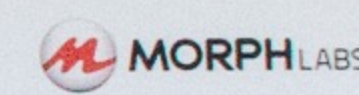
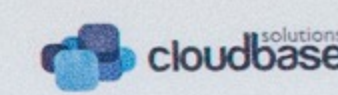
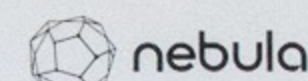
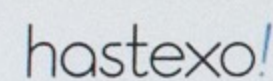
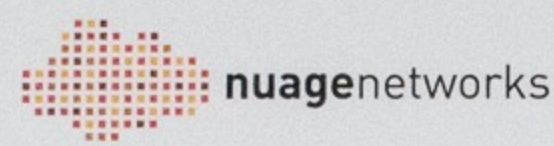
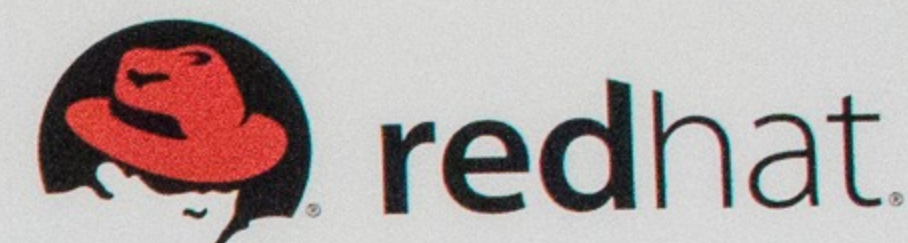
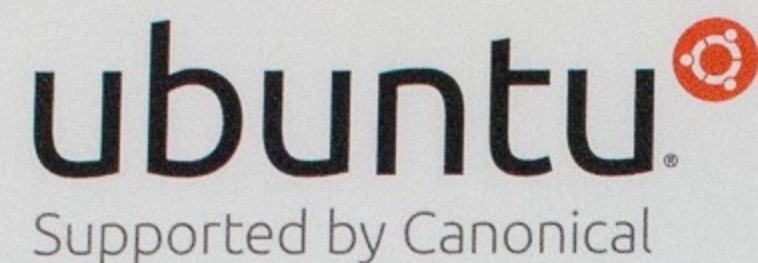




openstack™



users  
openstack  
summit  
devs  
PORTLAND // 2013



Widely supported by many vendors marketing departments



## Red Hat, Dell redouble OpenStack private cloud efforts

April 9, 2015 Written by [Business Cloud News](#)

Print Email



Red Hat and Dell have announced a series of co-engineered, high-density servers the companies claim are optimised for large-scale OpenStack deployments.

The co-engineered servers ship with Red Hat Enterprise Linux 7 and are based on Dell PowerEdge R630 and R730xd high-density rack servers, the latter ideal for compute and the latter optimised for storage utilisation.

"Enterprise customers are requiring robust and rapidly scalable cloud infrastructures that deliver business results," said Jim Ganther, vice president and general manager, Dell Engineered Solutions and Cloud.

"Dell and Red Hat continue to work together to make open source-based solutions more effective, open source-based solutions that provide greater agility to our customers, and leverage the best of breed technology from both."

Red Hat and Dell are co-developing OpenStack-based private cloud solutions

# Customers reporting interest in cloud, containers, Linux, OpenStack for 2015

**Summary:** 2014 has been a year of emerging technologies and great IT opportunities in Asia-Pacific. To provide some insights on the year so far and what lies ahead, we've asked a group of Red Hat customers where their priorities lie in 2015.



By Dirk-Peter van Leeuwen for Stacking up Open Clouds | December 18, 2014 -- 06:12 GMT (06:12 GMT)

Follow @ZDNET 227K followers

Get the ZDNet Must Read News Alerts - UK newsletter now

## CIO Journal

February 25, 2015, 8:08 AM ET

CIO Report | Consumerization | Big Data | Cloud | Talent & Management | Security

# The Morning Download: H-P Deal With Deutsche Bank Is Step Forward for OpenStack

Article Email Print

Comments



By STEVE ROSENBUSH  
Editor

CONNECT



Associated Press  
The HP logo is seen outside the headquarters in Palo Alto, Calif., Aug. 21, 2014. Credit: Associated Press

<http://on.wsj.com/TheMorningDownloadSignup>

The Morning Download comes from the editors of CIO Journal and cues up the most important news in business technology every weekday morning. Send us your tips, compliments and complaints. You can get The Morning Download emailed to you each weekday morning

<http://www.zdnet.com/article/customers-reporting-interest-in-cloud-containers-linux-openstack-for-2015>

<http://www.businesscloudnews.com/2015/04/09/red-hat-dell-redouble-openstack-private-cloud-efforts>

<http://blogs.wsj.com/cio/2015/02/25/the-morning-download-h-p-deal-with-deutsche-bank-is-step-forward-for-openstack>





# What Openstack Vendors promise...





**How Openstack Admins imagine their workplace...**





WELCOME TO CISCO UATH POP-6  
CISCO - LANDING STORE

**What is being delivered...**





**What the admin job actually looks like...**





# Storage

"Where the Internet lives", <http://www.google.com/about/datacenters/gallery/#/tech/12>



# Storage requirements

- VM with Ephemeral Storage
  - Not a problem, right? Because storage can be an un-RAID-ed local disk.
- But then you got no migration.
  - Maintenance sucks w/o Migration. For you and for your customers.



***Lesson #1:***  
***Nope, local storage is not a good  
default.***



# Storage

- VM with Volume: All writes are always remote.
- And redundant.
- Relevant metrics:
  - Bandwidth, IOPS and Latency
  - MB/sec, multithreaded-fsync()/sec und sequential-fsync()/sec
- Where do the limits come from?



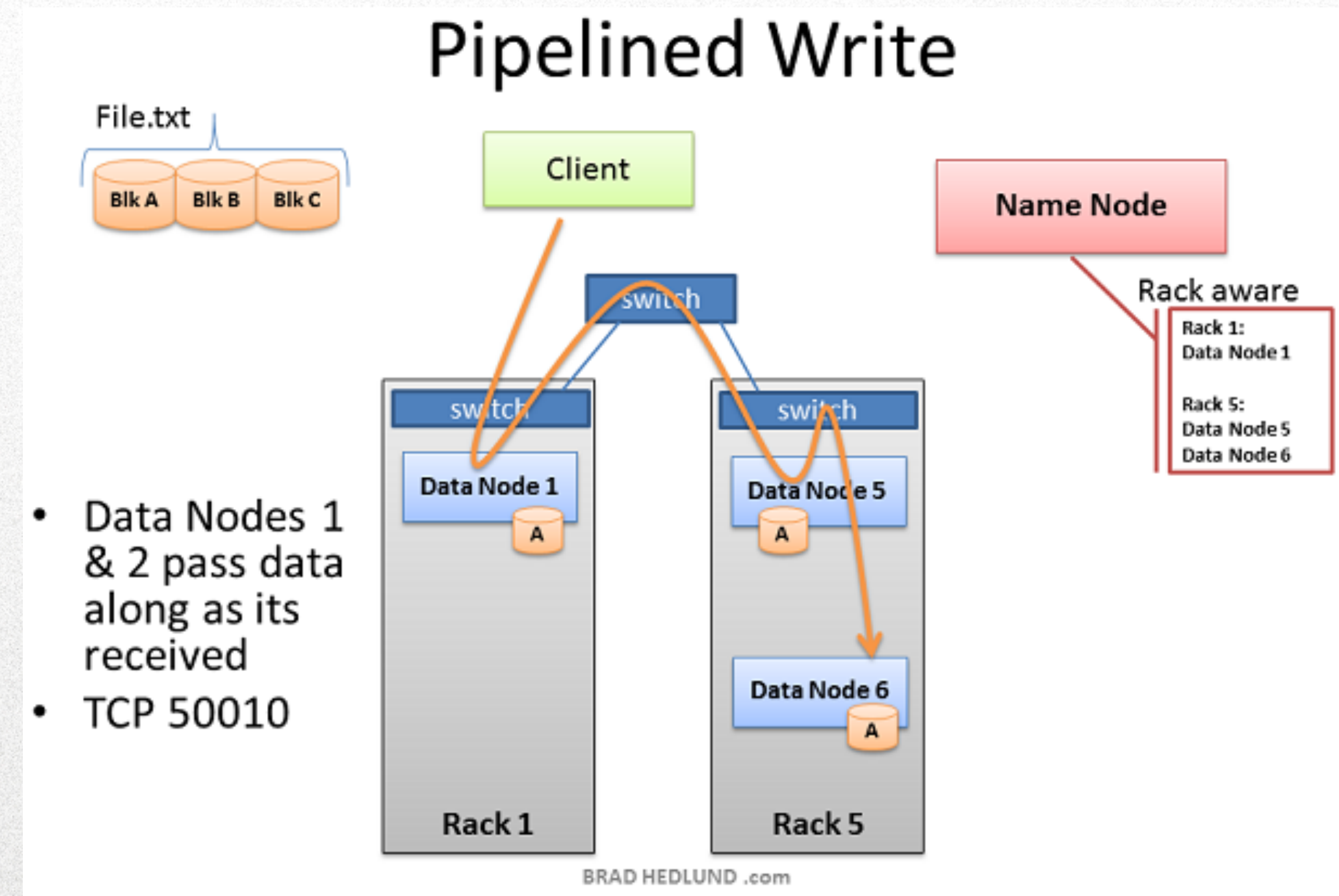
# Storage vs. requirements

- Scales easily:
  - Bandwidth, MT-IOPS
- Hard to scale:
  - Sequential-IOPS (b/c latency, target: 10k)
  - Default-Benchmark: "MySQL Slave on a volume"
  - "16 KB single-thread random-write on a datafile."



# Storage vs. requirements

- shared storage = at least one network write
- one write = 1200 MB/s
- two writes = half latency
- latency is at least 1/200000s
- at most 10k-20k commit/s





***Lesson #2:***  
***A Single-Threaded Random-Write  
Benchmark is a fine first evaluation.***



# Storage vs. actual Openstack delivery

- OpenStack Default is iSCSI und tgt
- »The problem is left as an exercise to the user.«
- Storage Vendors are totally in love with that approach.
- <https://wiki.openstack.org/wiki/CinderSupportMatrix>



# Storage: Ceph - the good

- Excellent bandwidth
- Good Multithreaded-IOPS
- Very robust, if you got spare memory and time for MTTR



# Storage: Ceph - the bad

- CRUSH
  - Layout follows topology
  - Change topology, move data needlessly
  - You do need a background network for storage to facilitate Rebalancing/Recovery/Cluster-Reorg



# Storage: Ceph - the ugly

- Logging
- IO is being serialised.
- Sequential IOPS fatally slow (200 IOPS).
- MySQL slave on our Ceph-Volumes @ 200 Commit/s
- Windows 8.1 boots from a Ceph-Volume in 15 Minutes



## ***Lesson #3:***

***Pure Play OpenStack will  
not work for production.***

***Corollary:***

***OpenStack is not a functioning  
Open Source Project.***



# Storage requirements vs. Multitenancy

- IOPS-Requirements need SSD to implement
- IOPS Quotas require SSD to implement
- SSD complicated - Price vs. guaranteed Performance
- Caches complicated

*Small IOPS = large performance variance.*

*You can't put a price on the US & the EU.*

*Single-Threaded Database Updates*

*"Magento Indexer"*

*Working Set > Cache means you are working the raw iron.*

*important than huge peak perf.*



## ***Lesson #4:***

***Caching is as much part of the problem  
as it is part of the solution.***





# Network



# Network requirements

- Hosting setup:
  - Topology: Wait-Free, Oversubscription-Free
  - Redundancy: SPOF-Free
  - Multi-Tenancy: Isolation and Quotas
- Also:
  - Capable of carrying the storage traffic

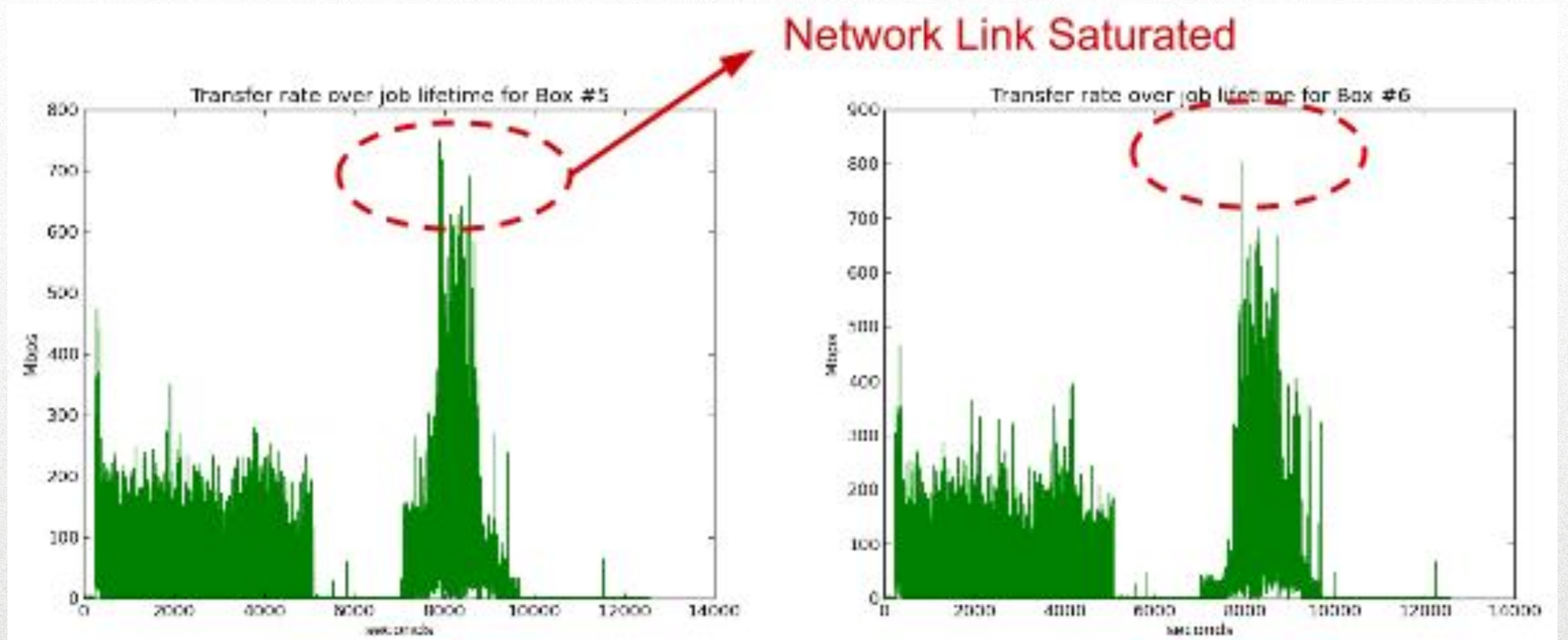


***„We just rolled out a few dozen VMs  
with Hadoop and played Terasort.“***

**–Joe Random Alphauser**



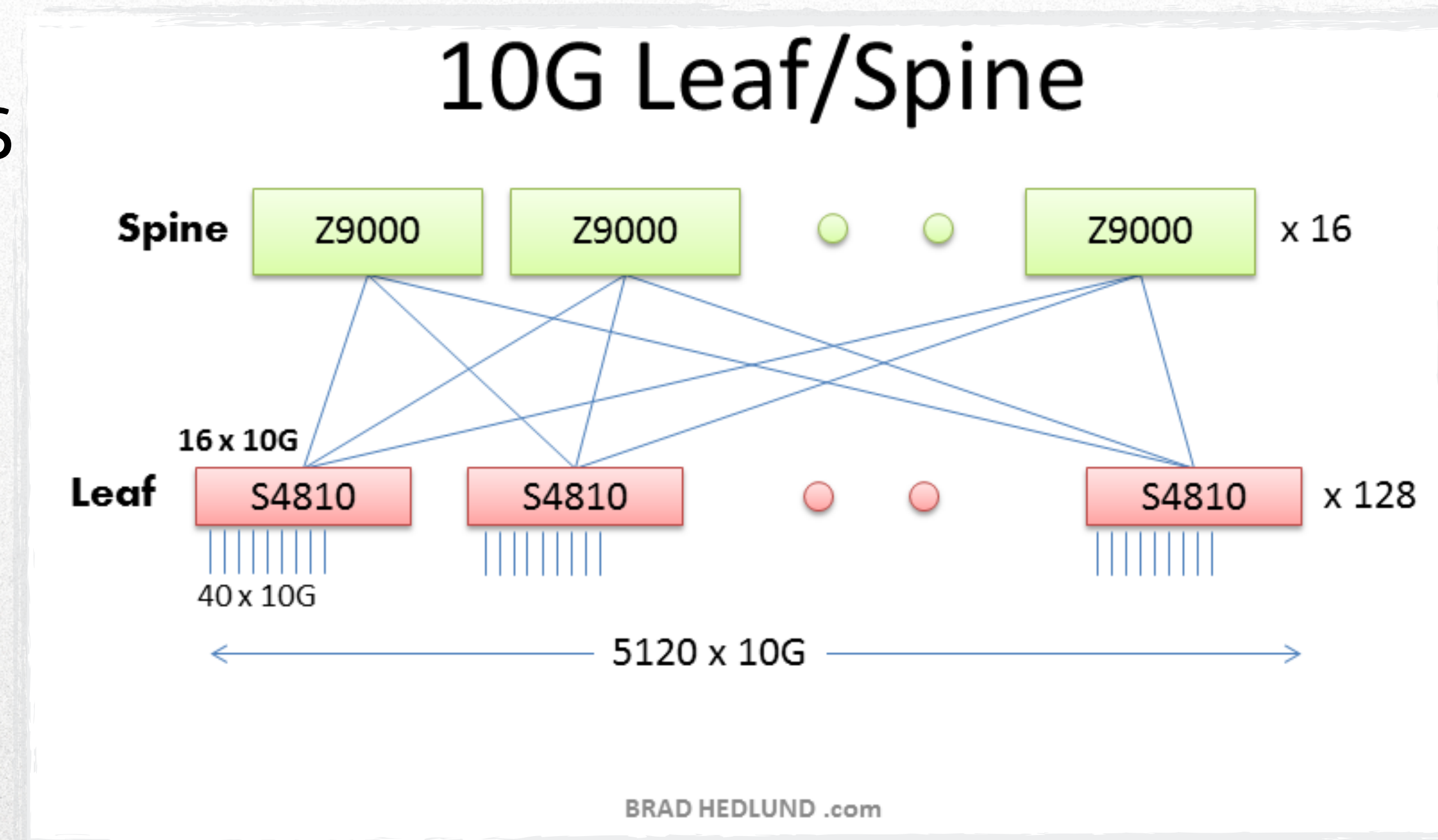
# Terasort to watch the world burn





# Oversubscription free networking

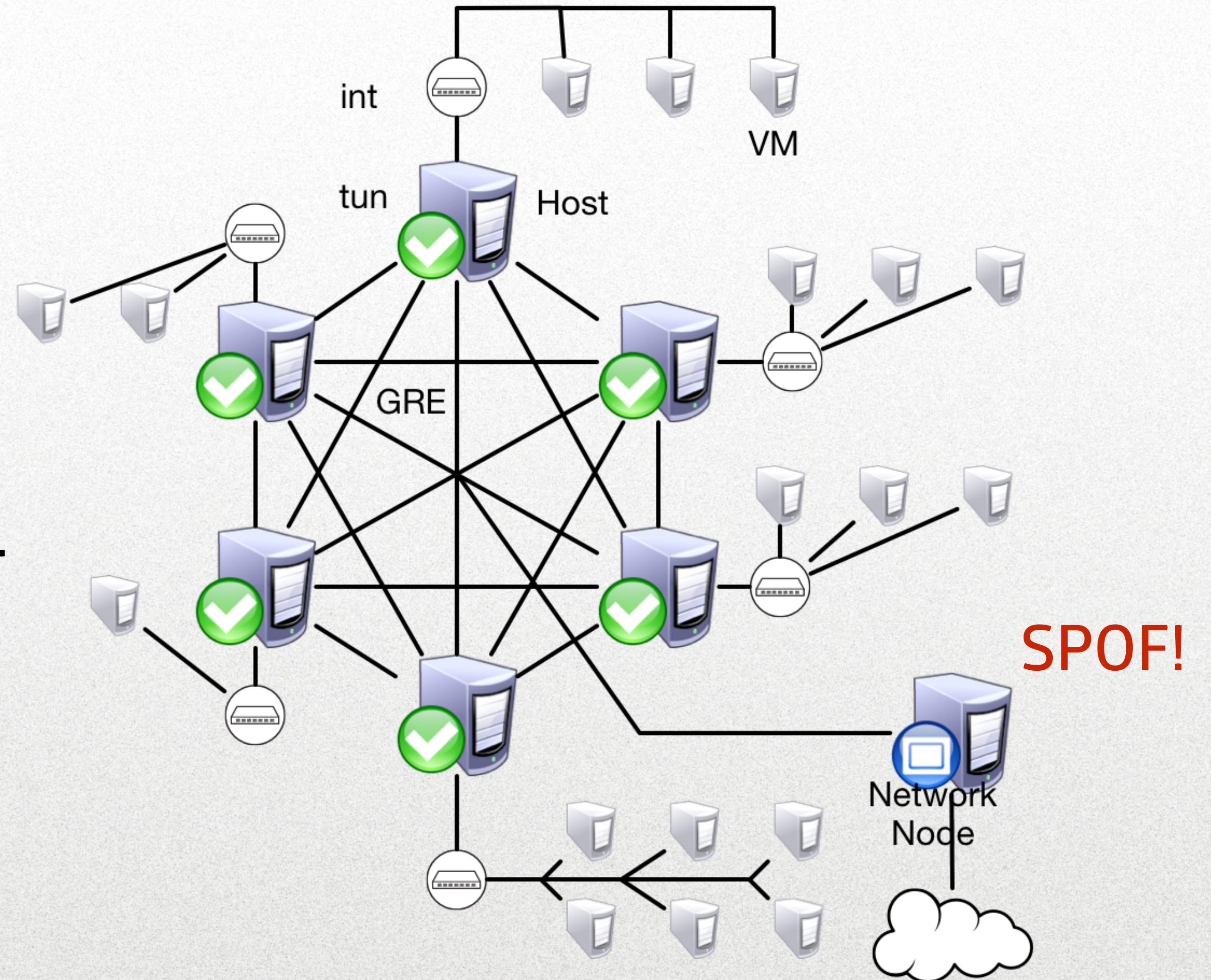
- Leaf and Spine Architecture
- More ports, more switches
- Price per port:
  - 200-300 EUR/10GBit





# Network: Openstack default offering

- Open vSwitch, GRE-Ball, Broadcast-Problem, Chokepoint, SPOF
- Current releases are only marginally less braindead.

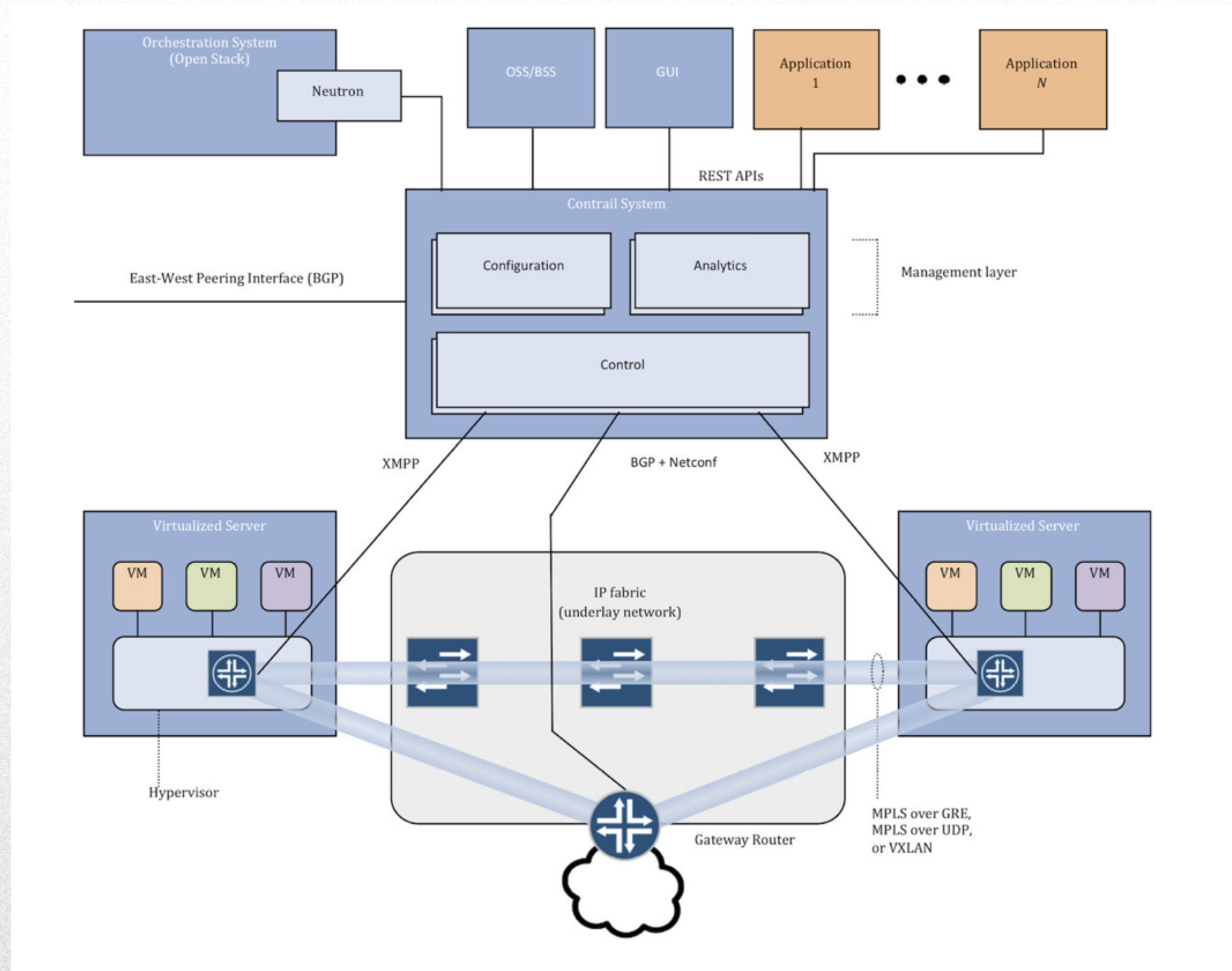




***Lesson #5:***  
***Ok, networking doesn't work, either.***



# Let's look around for a non-broken solution





# OpenContrail: the good, ...

- Open Source Project by Juniper
- Uses existing hardware routing infrastructure
  - scales, no SPOFs, no Chokepoints
- based on MPLS, BGP, and other well understood protocols
- actually delivers stuff (as opposed to e.g. OpenDaylite)



# OpenContrail: the bad, ...

- Juniper bought Contrail, doesn't understand how to market it
- little and outdated documentation, bad release management, very bad packaging
- funny support organisation  
(the support is awesome, the organisation is funny)



# OpenContrail: the bad, ...

- During the OpenContrail build process, the `scons`-based build-process will
  - download the full source of tools such as Bind and Curl
  - patch them wildly
  - put the resulting binaries into a `.deb` package
- which will of course conflict with half a dozen Ubuntu packages





**... and the ugly.**



# Technology-Jenga

- OpenContrail's "Stack"
  - vrouter (.ko), if-map, Sandesh (XML over Thrift!)
  - C++, Python, node.js, ironD (Java)
  - redis, Cassandra, Zookeeper
  - xmpp, BGP, MPLS
  - Up Next in OpenContrail 2.2: Kafka
  - Put that into a job profile, try to find a candidate. Is it Pedro?



***Lesson #6:***  
***The Jenga problem is actually  
not limited to Contrail scope.***





**Let's look what else is in the box...**



# General requirements as of 2015


- Distributed Anything:
  - NTP, centralized Logging, centralized Monitoring
  - functional validation before components before join, clean cluster startup
  - cluster comms are Kyle-Kingsbury-proof
  - CA, encrypted data in flight, optionally encrypted data at rest
  - User-Story regarding Live-Upgrades, with Canaries





**OpenStack delivers this...**





**General  
Pain**

**A wild hack session appears...**



# Not an isolated problem...

- Start a HEAT stack with a few networks and 20 VMs
- ... and the cluster switches off.
- Nova does not communicate with Cinder, but has a hard timeout.
- "Are we stupid or is OpenStack stupid? Let's test a few public clouds..."



# Result...

- Phone ringing...
- "Whatever you are doing, could you please do something else?"





# Not an isolated problem...

os-274 puppet-keystone openstack role/user ensuring braucht jahrzehnte

Relates to: [x os-246](#)

Is duplicated by: [x os-276](#)

Iteriert über alle user und tenants -> nicht flott -> nicht fröhlich

Upstream issue: <https://review.openstack.org/#/c/150200/>

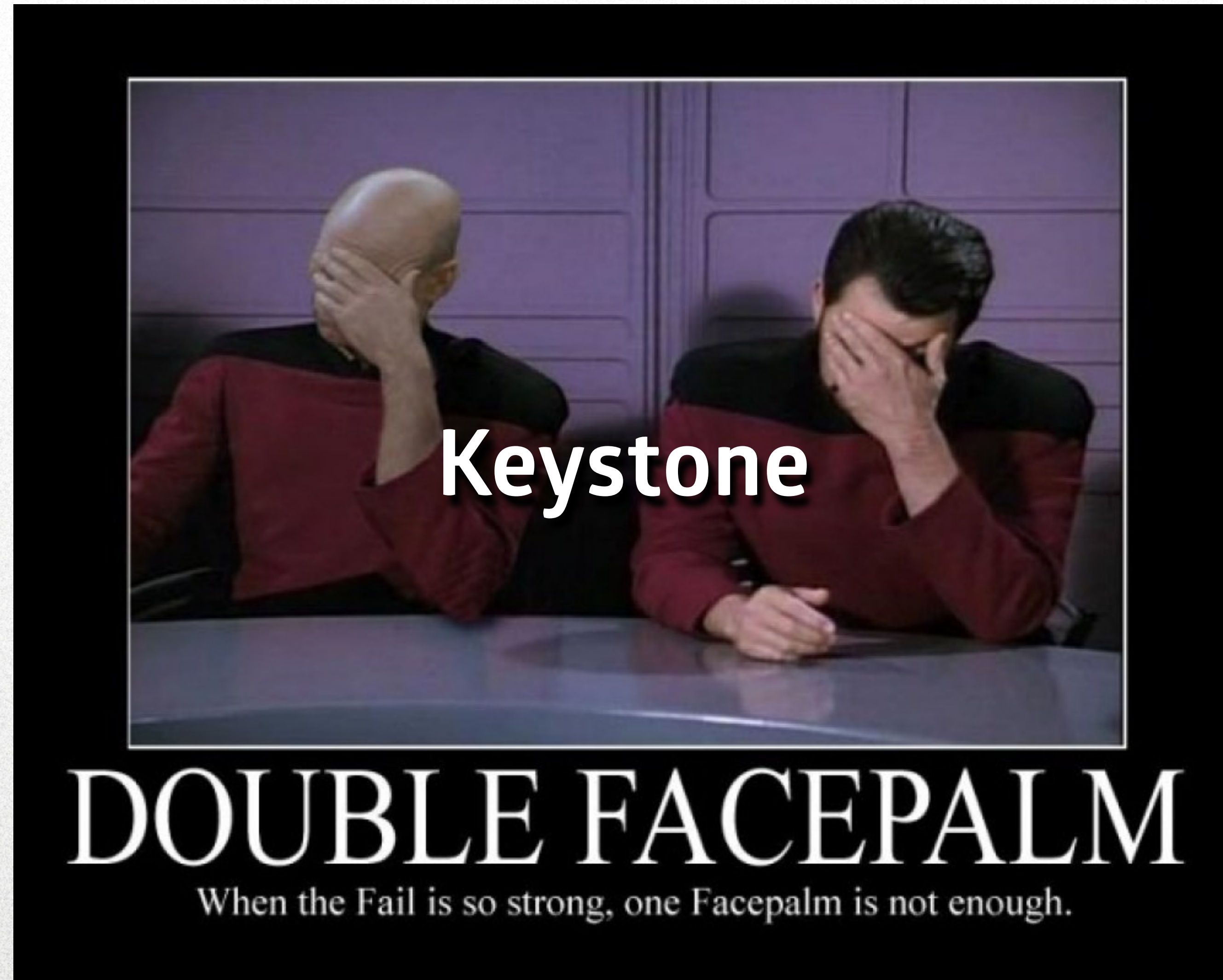
## Implement caching lookup for keystone\_user\_role

The new implementation of keystone\_user\_role failed to implement a previously existing method, user\_role\_hash, which helped cache the user role matrix. This helped prevent subsequent runs of exists? from re-requesting full lists of projects, users, and roles for every found instance of keystone\_user\_role. The failure to implement this caching causes extremely slow puppet runs even when no changes were needed since the provider would rebuild the hash for every call of exists?, which even for a small number of users and projects is very time consuming.

*Implementation tested  
on Devstack in VMware Fusion  
on a MacBook Air in St. Oberholz?*



# Not an isolated problem...





# Nova ~~can't count~~ has no clue whatsoever

## Nova quota usage - synchronization

Nova quota usage gets frequently out of sync with the real usage consumption. We are hitting this problem since a couple of releases and it's increasing with the number of users/tenants in the CERN Cloud Infrastructure.

```
if not CONF.workarounds.destroy_after_evacuate:
    LOG.warning(_LW('Instance %(uuid)s appears to have been '
                    'evacuated from this host to %(host)s. '
                    'Not destroying it locally due to '
                    'config setting '
                    '"workarounds.destroy_after_evacuate". '
                    'If this is not correct, enable that '
                    'option and restart nova-compute.'),
                {'uuid': instance.uuid,
                 'host': instance.host})
    continue
```

```
[workarounds] destroy_after_evacuate = True
```

(BoolOpt) Whether to destroy instances on startup when we suspect they have previously been evacuated. This can result in data loss if undesired. See <https://launchpad.net/bugs/1419785>



# Not an isolated problem...







**What is the real problem?**



***„Any VM, anywhere.“***



***„As a hoster/enterprise/department,  
I want a virtualization platform  
which does...“***



# Problemspec vs. Produktspec

- Requirements regarding
  - Tenant Isolation,
  - Billing Model,
  - Operations Model,
  - Development Model,
  - Scalability





# Prototypes vs. Product

"Prototype in the round file", <https://www.flickr.com/photos/generated/3313311558> Jared Tarbell (CC-BY 2.0)



***Stabilize the core with actual engineering***

***VS.***

***Moar components***





# “The Big Tent”





**"The problem with the Big Tent is that it is full of clowns."**



***„Instead of having a functional solution I can now choose between 13 differently deficient components.“***

**-Maik Zumstrull**



# Democracy as a software architecture model

- Product definition == target specification
- Who is our customer and what do they need?
- What are the mandatory/desireable/optional properties of the product to become part of the solution space instead of the problem space?

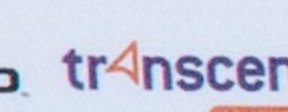
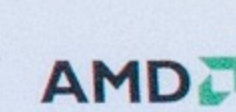
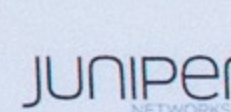
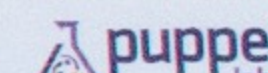
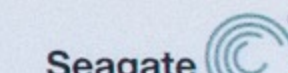
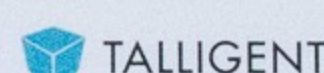
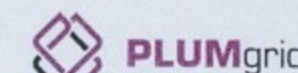
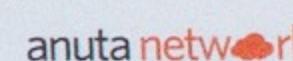
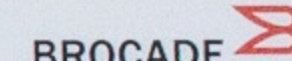
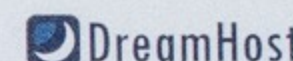
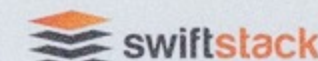
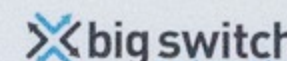
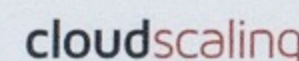
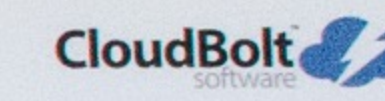
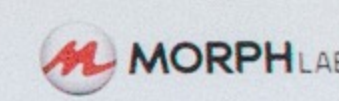
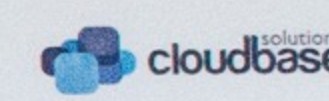
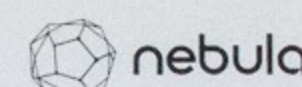
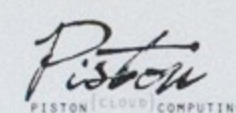
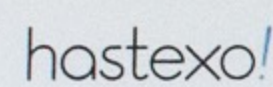
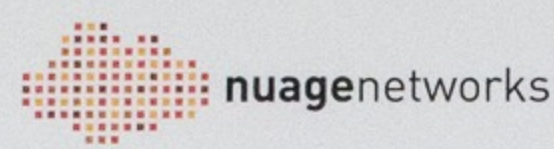
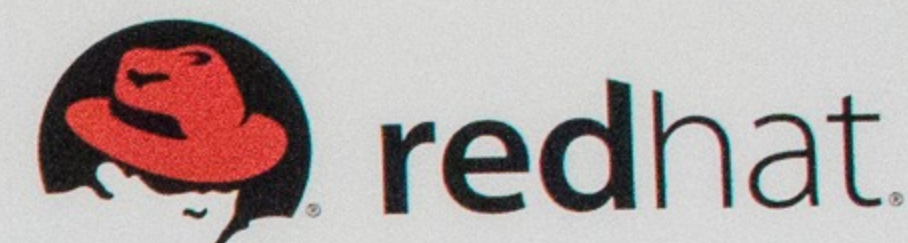
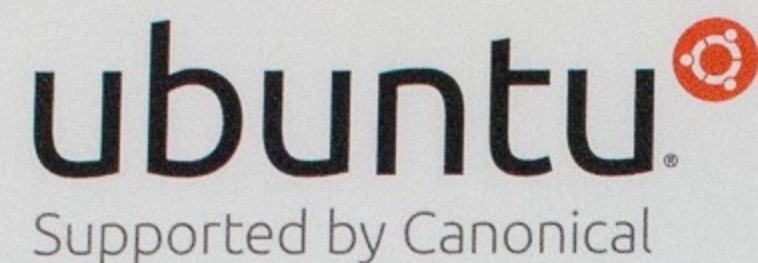


# Democracy as a software architecture model

- Quality Assurance
  - "But OpenStack does have Blueprints, CI and code review"
  - Without a target specification you have no idea about requirements on scale, workload types, security, ...



users  
openstack  
summit  
devs  
PORTLAND // 2013



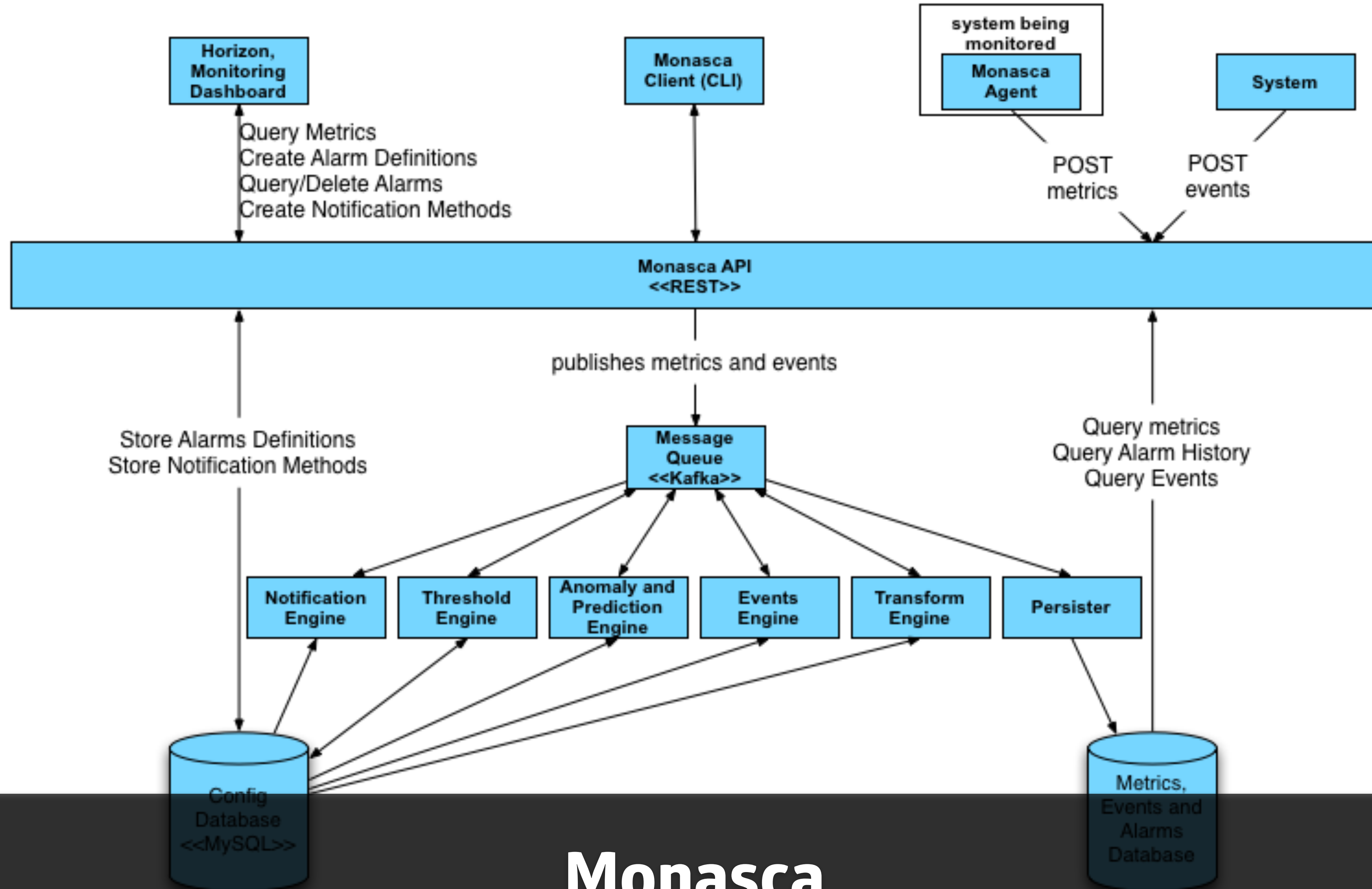
Who actually votes on your commit...



# Democracy as a software architecture model

- Limited, learnable technology stack
- One recommended and tested solution with a documented best practice for deployment instead of a plugin architecture.
- Reuseable solutions for comparable problems in neighboring subprojects (“Architecture Refactoring”).





# Monasca



# How to win at Hipsterbingo...

- »Monasca is an open-source multi-tenant, highly scalable, performant, fault-tolerant monitoring-as-a-service solution that integrates with OpenStack.

It uses a REST API for high-speed metrics processing and querying and has a streaming alarm engine and notification engine.«

**Blingo!**



# How to win at Hipsterbingo...

- »Uses a number of underlying technologies:

Apache Kafka, Apache Storm, Zookeeper, MySQL, Vagrant, Dropwizard, InfluxDB, Vertica.«



# Democracy as a software architecture model

- Useful vs. Hyped
  - “Docker is like OpenStack, but from a Dev instead of an Ops perspective.


Wouldn't it be great to unify the projects?”



<https://twitter.com/martinisoft/status/527191803603468288>



# And finally...




**Florian Haas**  
Shared publicly - Oct 31, 2014

One thing about OpenStack Summits is that the parties have become a noisy, alcohol-infused one-upping of vendors blowing obscene marketing budgets, which I am getting increasingly annoyed by. So for those of you interested in skipping out, having dinner and actually having a **conversation** (I know, shocking) about things OpenStack or non-OpenStack, let me know. Rumor has it that I travel with a restaurant shortlist everywhere I go.

#I


---




OpenStack Summit November 2014 Paris:  
Schedule For Evening Events @ Le Palais des  
Congrès

[openstacksummitnovember2014paris.sched.org](http://openstacksummitnovember2014paris.sched.org)

+12







***TL;DR please?***



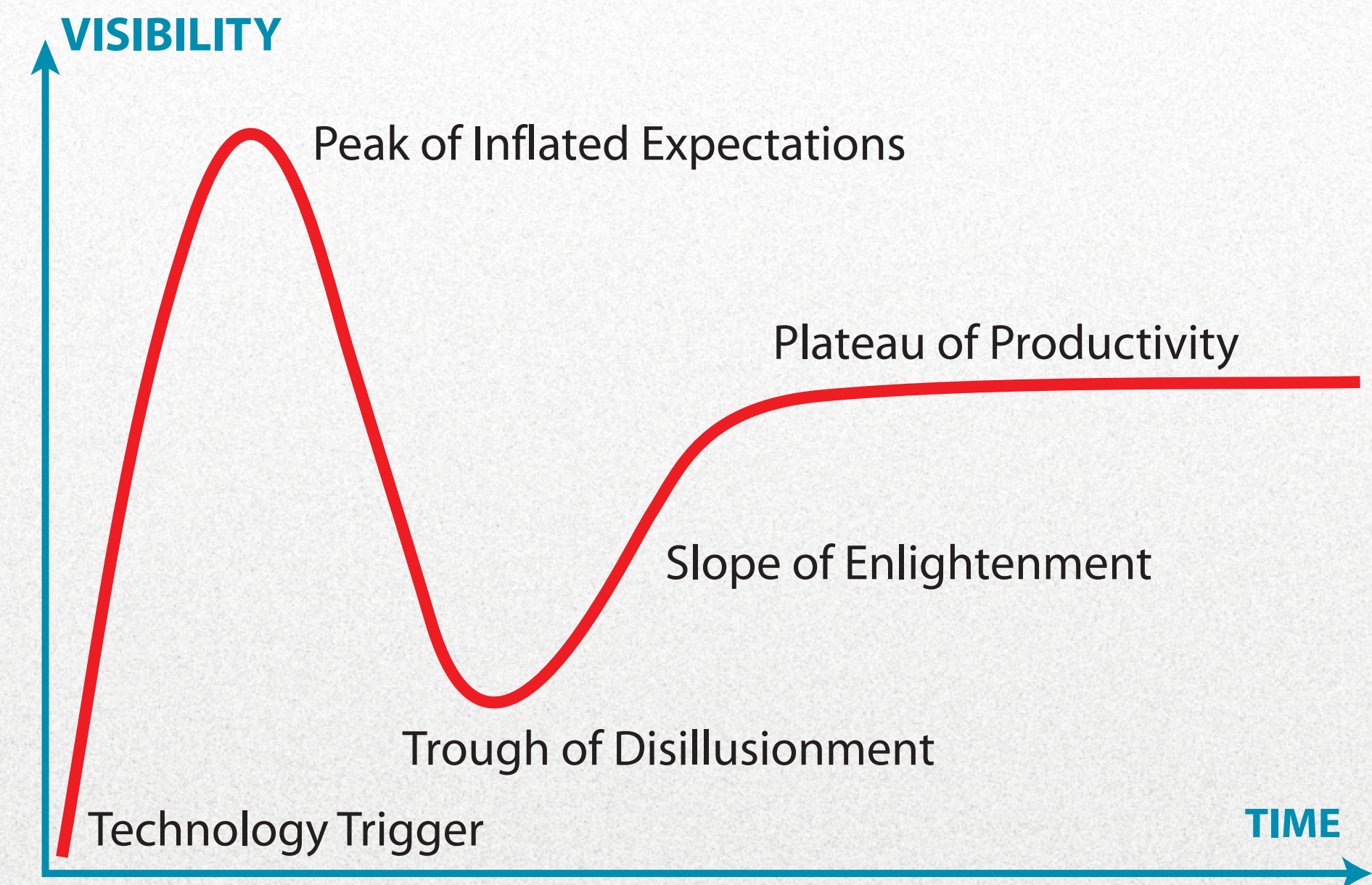
# TL;DR

- OpenStack right now: not a product, but a box of nuts and bolts. Significant assembly required.
- Not all parts are actually bad, but the quality varies. A lot.
- No other project has the momentum.
- Most other projects do not have multi-tenancy, multi-node.
- Many haven't even grasped the actual problem. Including Docker.



# TL;DR

- Deep in the Gartner hype cycle phase 2:
  - Users and hence real-world requirements missing.
  - No unified view on acceptable solutions.
  - Vendors think they can "win" and "own" the market or products, pushing their embrace and extend agendas.











**SysEleven**

Hosting. Skaliert.

Freitag / Friday

**Save  
the  
date**

**25.9.**  
**2015**

**Konferenz**  
10<sup>00</sup> bis 17<sup>30</sup>

**Sommerfest**  
ab 18<sup>00</sup>

**Kalkscheune**

Johannisstraße 2, 10117 Berlin

Tickets ab sofort unter:

[de.amiando.com/sys11Tag](http://de.amiando.com/sys11Tag)