

SUSE® OpenStack Cloud Hypervisors

From KVM to Docker

Vincent Untz

Project Manager

vuntz@suse.com



OpenStack: “Ubiquitous”, “Open”

- Perceived as the “Linux of the cloud”
- Hypervisor-agnostic approach
 - Support as many hypervisors as possible
 - Not necessarily in-tree
 - Support running multiple hypervisors in the same cloud
- Impact on the non-compute components of the project

Hypervisor History in OpenStack

- Ancient times: libvirt (KVM, QEMU)
- Early days (2010/2011):
 - libvirt (KVM, QEMU, UML, Xen), XenAPI, first Hyper-V driver
 - LXC (through libvirt) and “VMware”
- 2012:
 - Hyper-V dropped... and coming back
 - Baremetal
 - PowerVM

Hypervisor History in OpenStack

- 2013:
 - “VMware” rewritten, using vCenter
 - Docker
 - Baremetal moving to Ironic
- 2014
 - Docker moving out of tree
 - Baremetal dropped
 - z/VM
- 2015
 - Improved integration for many drivers

Hypervisor Trends in OpenStack

- Hypervisor requests:
 - 2011: Xen was quite popular
 - 2012: KVM moved relatively fast as the “default”
 - 2013: People starting to ask for “VMware” and Hyper-V
 - 2014: Everyone starting to look at containers
- Multi-hypervisor support:
 - Most deployments used a single hypervisor
 - Multiple hypervisors in the same cloud can clearly benefit enterprise customers, hence becoming a need

Summary of Available Hypervisors

- libvirt: KVM, Xen, LXC, QEMU, UML
- XenAPI
- “VMware”
- Hyper-V
- Docker
- PowerVM
- Baremetal / Ironic
- z/VM

Feature	Status	Hyper-V	Ironic	Libvirt KVM (ppc64)	Libvirt KVM (s390x)	Libvirt KVM (x86)	Libvirt LXC	Libvirt QEMU (x86)	Libvirt Virtuooszo CT	Libvirt Virtuooszo VM	Libvirt Xen	VMware vCenter	XenServer
Attach block volume to instance	optional	✓	✗	✓	✓	✓	✗	✓	✗	✓	✓	✓	✓
Detach block volume from instance	optional	✓	✗	✓	✓	✓	✗	✓	✗	✓	✓	✓	✓
Set the host in a maintenance mode	optional	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓
Evacuate instances from a host	optional	?	?	?	✓	✓	?	?	✗	✗	?	?	?
Guest instance status	mandatory	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Guest host status	optional	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Live migrate instance across hosts	optional	✓	✗	✓	✓	✓	✗	✓	✗	✗	✓	✗	✓
Launch instance	mandatory	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Stop instance CPUs	optional	✓	✗	✓	✓	✓	✓	✓	✗	✓	✓	✗	✓
Reboot instance	optional	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Rescue instance	optional	✗	✗	✓	✓	✓	✗	✓	✗	✗	✓	✓	✓
Resize instance	optional	✓	✓	✓	✓	✓	✗	✓	✗	✗	✓	✓	✓
Restore instance	optional	✓	✗	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓
Service control	optional	✗	✗	✓	✓	✓	✗	✓	✗	✗	✗	✓	✓
Set instance admin password	optional	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓
Save snapshot of instance disk	optional	✓	✗	✓	✓	✓	✗	✓	✗	✗	✓	✓	✓
Suspend instance	optional	✓	✗	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓
Swap block volumes	optional	✗	✗	✓	✓	✓	✓	✓	✗	✓	✓	✗	✗
Shutdown instance	mandatory	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Resume instance CPUs	optional	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓
Auto configure disk	optional	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓



Ensuring High Quality

- Unit tests and functional tests executed for each code change
 - Obviously requires a huge amount of resources!
 - Difficult for some hypervisors; require external system
- Drivers not receiving enough attention are being deprecated
- Also explains why some drivers are out-of-tree

<i>Reviewer</i>	<i>Code-Review</i>	<i>Verified</i>	<i>Workflow</i>
Virtuozzo Storage CI			
Virtuozzo CI			
IBM PowerKVM CI			
VMware NSX CI			
Intel PCI CI			
Microsoft Hyper-V CI			
Jenkins		+1	
XenProject CI		+1	
Citrix XenServer CI		+1	

Why one hypervisor?
Why multiple hypervisors?

Why use a specific hypervisor?

- Better internal knowledge of technology
- Appliances requiring a specific platform
- ISV-provided application certified for specific hypervisor

Why use a specific hypervisor?

- KVM/Xen:
 - Very good support from Linux vendors
- XenAPI:
 - Desktop virtualization
- VMware:
 - Often an already existing infrastructure
 - Ability to use vCenter for advanced features
- Hyper-V:
 - Windows on Windows
 - Licensing model for Windows

Why use a specific hypervisor?

- LXC/Docker:
 - Lightweight
 - Benefits from Docker ecosystem
- Baremetal / Ironic
 - Full performance
- z/VM
 - Mainframe
 - Thousands of guests

Why use more than one hypervisor?

- Getting the best of each hypervisor
- Using OpenStack as a common interface for different technologies
 - Including self-service features
 - Make use of standard OpenStack features to enhance feature set of hypervisor
- Slowly migrate workloads from existing infrastructure to cloud
- Mixed cloud with (for instance) Linux on Linux, Windows on Windows, legacy application on VMware
- Optimized costs



Challenges of running multiple hypervisors

- Integration with other OpenStack components, such as:
 - Networking
 - Block Storage
 - Metering
- Unified interface, but not exact same feature set

Feature	Status	Hyper-V	Libvirt KVM	Libvirt KVM (ppc64)	Libvirt KVM (s390x)	Libvirt KVM (d86)	Libvirt QEMU (d86)	Libvirt QEMU (CT)	Libvirt VM	Libvirt Xen	Libvirt vCenter	Libvirt XenServer
Attach block volume to instance	optional	✓	✗	✓	✓	✓	✗	✓	✗	✓	✓	✓
Detach block volume from instance	optional	✓	✗	✓	✓	✓	✗	✓	✗	✓	✓	✓
Set the host in a maintenance mode	optional	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓
Evacuate instances from a host	optional	?	?	?	✓	✓	?	?	✗	✗	?	?
Guest instance status	mandatory	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Guest host status	optional	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓
Live migrate instance across hosts	optional	✓	✗	✓	✓	✓	✗	✓	✗	✓	✗	✓
Launch instance	mandatory	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Stop instance CPUs	optional	✓	✗	✓	✓	✓	✓	✗	✓	✓	✗	✓
Reboot instance	optional	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Rescue instance	optional	✗	✗	✓	✓	✓	✗	✓	✗	✗	✓	✓
Resize instance	optional	✓	✓	✓	✓	✓	✗	✗	✗	✓	✓	✓
Restore instance	optional	✓	✗	✓	✓	✓	✗	✓	✗	✓	✓	✓
Service control	optional	✗	✗	✓	✓	✓	✓	✗	✗	✗	✓	✓
Set instance admin password	optional	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓
Save snapshot of instance disk	optional	✓	✗	✓	✓	✓	✗	✓	✗	✗	✓	✓
Suspend instance	optional	✓	✗	✓	✓	✓	✗	✓	✓	✓	✓	✓
Swap block volumes	optional	✗	✗	✓	✓	✓	✓	✗	✓	✓	✗	✗
Shutdown instance	mandatory	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Resume instance CPUs	optional	✓	✗	✓	✓	✓	✓	✓	✓	✓	✗	✓
Auto configure disk	optional	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓
Instance disk I/O limits	optional	✗	✗	✓	✓	✓	✗	✓	✗	✗	✗	✗
Config drive support	choice	✓	✗	✗	✓	✓	✓	✗	✓	✓	✓	✓
Inject files into disk image	optional	✗	✗	✗	✗	✓	✗	✓	✗	✗	✗	✓
Inject guest networking config	optional	✗	✗	✗	✗	✓	✗	✓	✗	✗	✓	✓
Remote desktop over RDP	choice	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
View serial console logs	choice	✓	✗	✗	✓	✓	✗	✗	✗	✓	✓	✓
Remote interactive serial console	choice	✗	?	?	✓	✓	?	?	✗	✗	?	✗
Remote desktop over SPICE	choice	✗	✗	✗	✗	✓	✗	✓	✓	✓	✓	✓
Remote desktop over VNC	choice	✗	✗	✗	✗	✓	✗	✓	✓	✓	✓	✓
Block storage support	optional	✓	✗	✓	✓	✓	✓	✗	✓	✓	✓	✓
Block storage over Fibre channel	optional	✗	✗	✗	✓	✓	✓	✗	✓	✓	✗	✗
Block storage over iSCSI	condition	✓	✗	✓	✓	✓	✓	✗	✓	✓	✓	✓
CHAP authentication for iSCSI	optional	✓	✗	✓	✓	✓	✓	✗	✓	✓	✓	✓
Image storage support	mandatory	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Network firewall rules	optional	✗	✗	✓	✓	✓	✓	✓	✓	✓	✗	✓
Network routing	optional	✗	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓
Network security groups	optional	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓
Flat networking	choice	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VLAN networking	choice	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓



Challenges of running multiple hypervisors

- Integration with other OpenStack components:
 - Networking
 - Block Storage
 - Metering
- Unified interface, but not exact same feature set
- Scheduling images to the right compute node
- Full support from vendor (and its partners)?

Quick tour of some hypervisors in OpenStack

KVM / Xen

- KVM is the default
 - Heavily used in CI
 - “Just works”, no surprise
- Xen
 - Not supported by all distributors
 - Also “just works”
- In both cases:
 - Very simple setup: one node is a compute node
 - No special challenge in setting up the node

Hyper-V

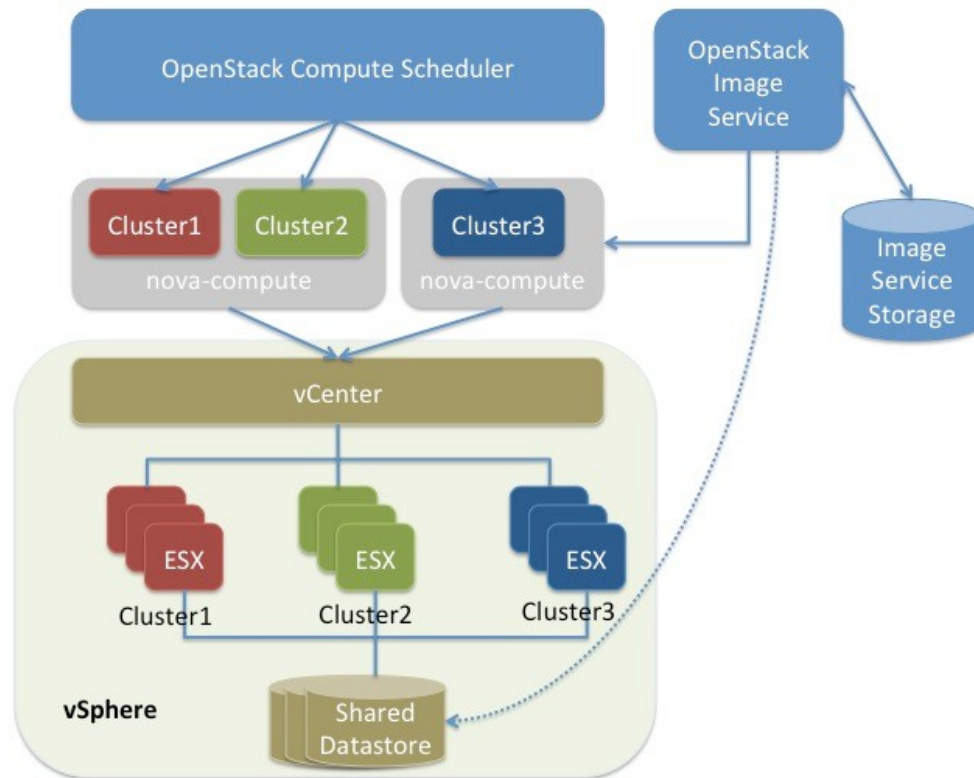
- Integration architecture with OpenStack is rather simple: a Windows node will be a compute node
- Manual setup of Windows as compute node is not trivial, though. Solutions:
 - Cloudbase provides an installer
 - SUSE OpenStack Cloud automates this step
- Windows Server 2012 R2 recommended
 - Hyper-V Server can be used for free

Hyper-V

- No specific issue with Cinder nor Glance
 - iSCSI or SMB for volumes
- Integration with Ceilometer
- Limitations:
 - Images must use the VHD/VHDX format
 - GRE/VXLAN support requires experimental Open vSwitch on Windows
- Development in the community is quite active, closing the gap in terms of features
 - RDP, soft reboot/shutdown, console log, ephemeral storage, etc.
 - Hyper-V generation 2 VM

vCenter (“VMware”)

- Compute host communicates with vCenter
 - Enables the use of features such as vMotion, High Availability, Dynamic Resource Scheduling (DRS)
- A cluster of ESX hosts is seen as one hypervisor entity
 - vCenter uses DRS to schedule inside the cluster
- One compute host can manage several clusters



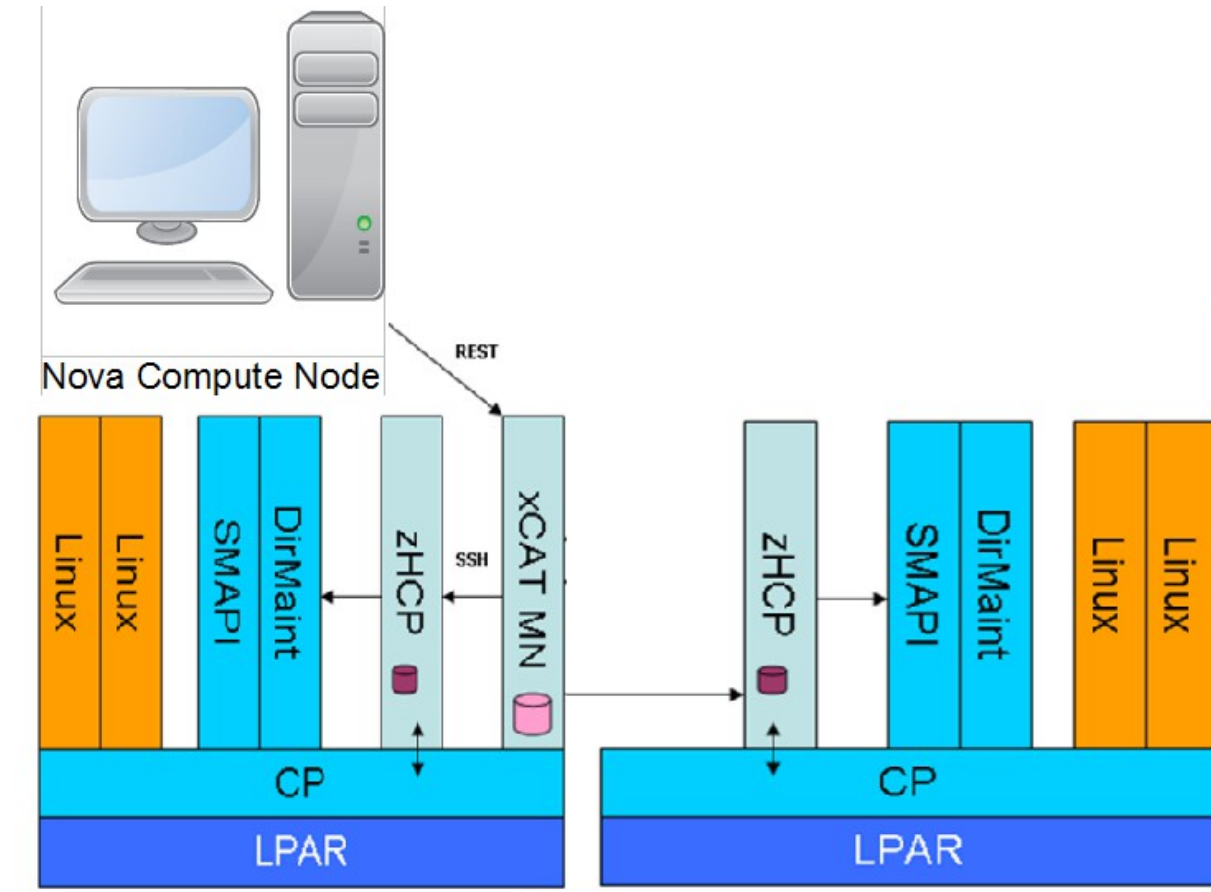
vCenter (“VMware”)

- Drivers also exist for Cinder and Glance
 - Volumes and images on vCenter managed data stores
- Integration with Ceilometer
- Limitations:
 - Shared storage required for ESX hosts
 - Images must use the VMDK format
 - Security groups only work with the NSX neutron plug-in
 - No GRE support with just Open vSwitch
 - vSphere 5.0 and earlier requires some manual configuration
- VMware (and others) working in the community to keep improving support for this driver



z/VM

- Out-of-tree driver
- Using xCAT to “translate” the requests to z/VM language
 - Also for networking parts
- 1:1 relationship between compute node and z/VM hypervisor

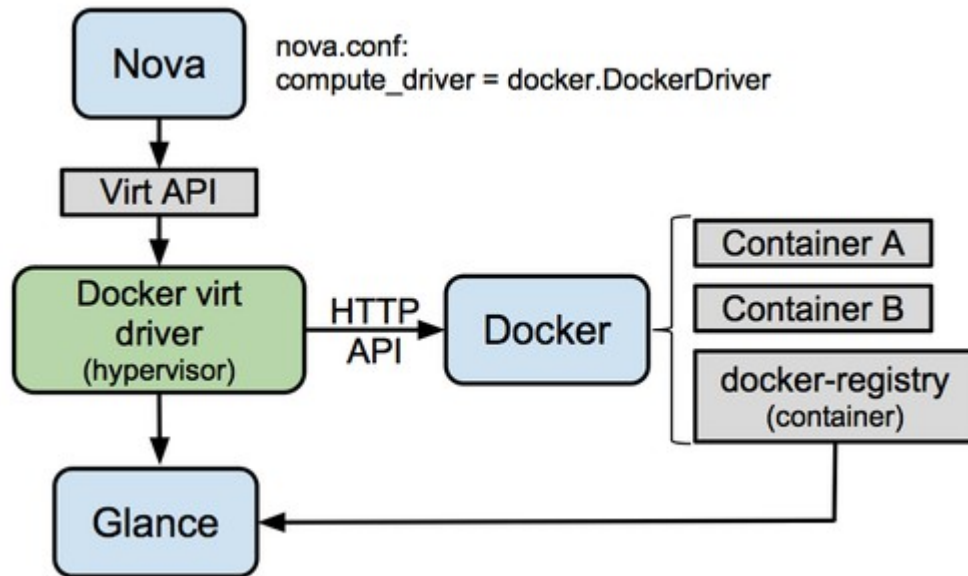


z/VM

- No specific issue with Cinder or Glance
 - iSCSI for Cinder
- Limitations:
 - No integration with Ceilometer
 - Requires flat or VLAN networking

Docker

- Out-of-tree driver
- Using Docker to simplify handling of containers, not necessarily for Docker features
 - Note that there's also a LXC driver!
- An instance is a container directly running on the compute node
 - “Containers inside a VM” scenario can however be achieved through a Heat plugin
- Standard features of OpenStack can be used with containers
 - For instance the orchestration features, including load balancing, scale up/scale down, etc.



Docker

- Limitations

- No support for volumes
- No integration with Ceilometer
- Only containers, no full support for all docker features
 - No registry (portus)
 - No layers in images

OpenStack Magnum

- Magnum = Container-as-a-Service
 - Docker specific... for now!
- Purpose is to provide a service specifically tailored at containers
 - Unlike nova-docker integration, API is not centered about generic instance concept, but about containers
 - Can benefit from container-specific features
- Make use of existing container orchestration engines
 - Kubernetes
 - Docker Swarm
- Leverage OpenStack for orchestration, multi-tenancy, identity management, etc.

How do I choose my hypervisor
in a multi-hypervisor cloud?

The end user point of view

- No easy way to do all this from the OpenStack Dashboard by default
- Choosing hypervisor while uploading an image:
 - hypervisor_type property:
`glance image-create --property hypervisor_type=kvm [...]`
 - vm_mode property. For Xen:
`vm_mode=xen, vm_mode=hvm`
- Choosing hypervisor while booting an instance:
 - Picking a specific host for the instance:
`nova boot --availability-zone zone:host [...]`

The end user point of view

- Alternative: using host aggregates
 - Group compute nodes in a host aggregate with a specific property
 - Create flavors using this property
 - Downside: only doable by administrators
- Alternative: defining availability zones
 - One availability zone per hypervisor type
 - Downside: only doable by administrators

How to deploy multiple hypervisors?

What about deployment?

- “Simply” a matter of configuring all compute nodes the way you want
- Need to be careful about networking technology used
 - Generally, using VLAN is a safe bet (even though not as convenient as GRE or VXLAN)
- Most deployment tools are targeting the one hypervisor use case, though
 - Not SUSE OpenStack Cloud!

Concrete example: using KVM

- In nova.conf of compute node:

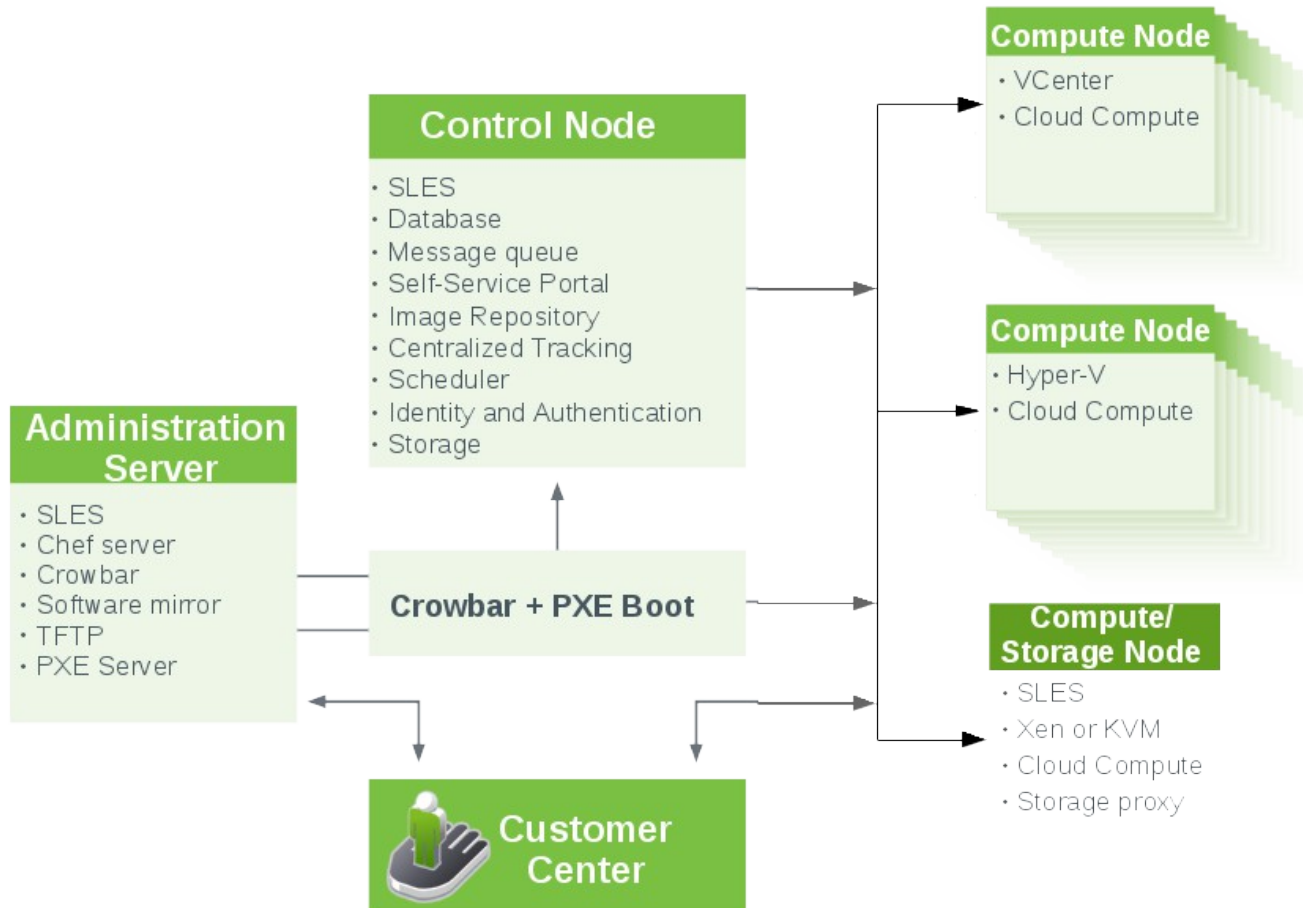
```
compute_driver=libvirt.LibvirtDriver  
libvirt_type=kvm
```

- Of course, there are many other options to tweak the behavior of each hypervisor
- Glance, Cinder and Neutron configurations need adjustments too

Multiple Hypervisors with SUSE OpenStack Cloud

- Support for:
 - KVM
 - Xen
 - Hyper-V
 - VMware vCenter
- Tech preview for:
 - Docker
- Coming in SUSE OpenStack Cloud 6:
 - z/VM
- All in the same cloud!
- Install from baremetal for KVM, Xen and Hyper-V compute nodes





Which hypervisor(s) are you interested in? Why?

Questions?

Thank you.







Corporate Headquarters
Maxfeldstrasse 5
90409 Nuremberg
Germany

+49 911 740 53 0 (Worldwide)
www.suse.com

Join us on:
www.opensuse.org

Unpublished Work of SUSE LLC. All Rights Reserved.

This work is an unpublished work and contains confidential, proprietary and trade secret information of SUSE LLC. Access to this work is restricted to SUSE employees who have a need to know to perform tasks within the scope of their assignments. No part of this work may be practiced, performed, copied, distributed, revised, modified, translated, abridged, condensed, expanded, collected, or adapted without the prior written consent of SUSE. Any use or exploitation of this work without authorization could subject the perpetrator to criminal and civil liability.

General Disclaimer

This document is not to be construed as a promise by any participating company to develop, deliver, or market a product. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. SUSE makes no representations or warranties with respect to the contents of this document, and specifically disclaims any express or implied warranties of merchantability or fitness for any particular purpose. The development, release, and timing of features or functionality described for SUSE products remains at the sole discretion of SUSE. Further, SUSE reserves the right to revise this document and to make changes to its content, at any time, without obligation to notify any person or entity of such revisions or changes. All SUSE marks referenced in this presentation are trademarks or registered trademarks of Novell, Inc. in the United States and other countries. All third-party trademarks are the property of their respective owners.

