HA for Enterprise Clouds:
Oracle Solaris Cluster & OpenStack

Eve Kleinknecht / Thorsten Frueauf
Oracle

Keywords:

OpenStack, High Availability, Solaris, Solaris Cluster, Oracle

Introduction:

More and more enterprises deploy their mission critical applications inside enterprise clouds. However, high availability requirements don't go away. As more and more consolidation occur, those requirements increase and become more diverse. Oracle Solaris Cluster provides high availability and disaster recovery to infrastructure and applications in Oracle Solaris environments. This paper describes how it can be used to secure and protect an Oracle Solaris-based OpenStack Cloud Controller to deliver a highly available enterprise grade cloud infrastructure deployment.

OpenStack and Oracle Openstack for Oracle Solaris

OpenStack

OpenStack is open source software that enables the easy creation and management of private and public cloud environments. OpenStack is a cloud solution that controls large pools of compute, storage, and networking resources throughout a datacenter, all managed through a dashboard that gives administrators control while empowering their users to provision resources through a web interface.

Oracle Openstack for Oracle Solaris

Oracle Solaris 11 includes a complete OpenStack distribution, allowing administrators to centrally share and manage data center resources through a single pane of management, including infrastructure and virtualization offerings provided by other vendors. Integrated into the core technology foundations such as Oracle Solaris Zones, the ZFS file system, Unified Archives and comprehensive software defined networking, Oracle OpenStack on Oracle Solaris provides self-service computing, allowing IT organizations to deliver services in minutes, with enterprise-grade reliability, security, and performance.

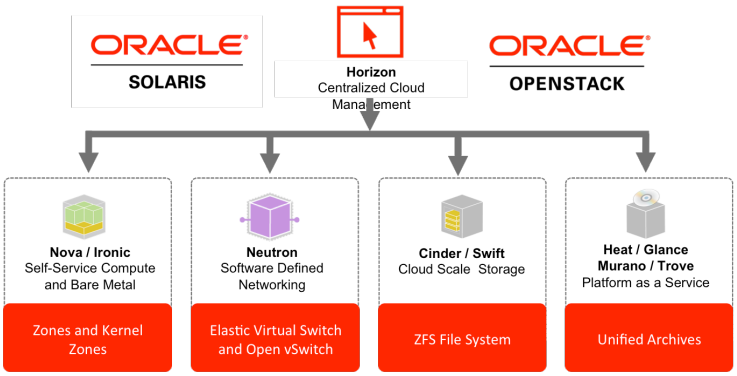

Illustration 1: Oracle OpenStack for Oracle Solaris components

Oracle Solaris Cluster

Oracle Solaris Cluster is a comprehensive high availability and disaster recovery solution for Oracle Solaris environments running on Oracle SPARC or x86 servers. It extends the Oracle Solaris operating system into a cluster operating system, delivering resiliency to outage for application and platform services by leveraging redundancy of components and systems.
Oracle Solaris Cluster offers:

- Efficient failure detection
- Automatic fault management
- Orchestrated recovery

Efficient failure detection: Oracle Solaris Cluster monitors all the components of the eco-system: server, network, storage, OS, VM and is fully integrated in the OS kernel for immediate server failure detection.

Automatic fault management: Once a failure is detected, component specific algorithms compute the correct reaction for isolating the failure and recovering from it at the lowest level possible, using integrated OS capabilities such as Oracle Solaris Fault Management Architecture (FMA) for low-level hardware outages, Oracle Solaris multi-pathing for storage failures or IPMP and DLMP for network level issues. At higher levels Oracle Service Management Framework (SMF) and specific cluster services manage the local restart of platform and application services.

Orchestrated recovery: If the failure is at the server level, then the whole ecosystem is restarted on an alternate healthy system: Oracle Solaris Cluster executes failover to an alternate server of all managed platform and applications services together with the resources the applications are dependent on, in the appropriate, pre-defined order. The recovery is orchestrated and optimized, enabling minimization of unplanned downtime.

The Oracle Solaris Cluster infrastructure can also be used to limit and even suppress application downtime due to planned maintenance. Cold, warm or live migration (depending on technology and context) can be exercised by Oracle Solaris Cluster to move applications or virtual machines and their ecosystem to alternate server or locations, facilitating the maintenance of a server or a data center.

Oracle Solaris Cluster is capable of protecting a very wide range of configurations supporting bare-metal and virtualized environments, application- and virtual machine-level protection, local or multi-site infrastructure.

The main high availability enabling features provided by Oracle Solaris Cluster are:

Resource Group Manager
The Resource Group Manager (RGM) subsystem manages applications put under Oracle Solaris Cluster control and ensures that the resources that they require to function are available to them. RGM also manages dependencies between applications and resources or between applications as well as load distribution policies.

Failover and scalable IP addresses
To enable continuous client access to applications, the Oracle Solaris Cluster software provides two mechanisms: logical IP addresses which are used for failover (active/passive) applications and global IP addresses which are used for scalable (active/active) applications running on more than one node concurrently. The later can also be used with a built-in load-balancing feature.

Failover and scalable storage access
To ensure availability of data to applications, two types of file systems are offered: a failover file system type which is accessible by one node at a time (for active/passive configurations) or a proxy file system type which is accessible to multiple nodes (for active/active configurations). Oracle Solaris Cluster fencing and quorum algorithms protect both types of file systems from potential data corruption that can happen during split-brain or amnesia failure scenarios.

Data Services (aka agents)
Oracle Solaris Cluster provides a large set of built-in high availability data services for applications. The agents implement specific mechanism for orderly startup and shutdown, fault monitoring, and automatic failover of the applications enabling customized failure detection and fast recovery. In the context of OpenStack the MySQL and Apache data services are specifically of interest as well as the SMF proxy resources that can be used to extend Oracle Solaris SMF services into multi-node services. In addition Oracle Solaris Cluster includes the Generic Data Service, a mechanism which helps integrate custom applications without requiring cluster specific coding.

Virtualisation
Oracle Solaris Cluster offers different solutions when it comes to deploying workloads in virtualized environments such as Oracle Solaris Zones or Oracle VM for SPARC domains. The Oracle Solaris Zone Cluster offers a virtual cluster based on Zones which is isolated for resource and fault management. The ability to run multiple zone clusters within a global physical cluster, enables deployment of multi-tiered applications and securely separates application layers from each other, while dependencies between components are automatically managed. Zones can also be considered as complete workloads and can be configured as failover resources, where orderly booting, shutdown, fault monitoring and automated recovery as well as cold, warm and live migration are managed by the cluster framework.

We will see in the following paragraphs how those facilities can be used in the OpenStack context offering different possibilities of configuration depending on the use case and requirements.

High Availability for OpenStack Cloud Controller

Any enterprise-type OpenStack deployment requires a highly available OpenStack infrastructure that can sustain individual system failures. The OpenStack Cloud Controller provides the central management system for OpenStack deployments. The architecture for the Cloud Controller has therefore to be designed in such a way that it offers high availability and scalability. Other considerations also need to be taken into account such as life cycle management, flexibility and capacity for evolution. In the following paragraphs two different topologies will be proposed which can provide high availability to an OpenStack Cloud Controller but which can differ in term of service levels and flexibility.

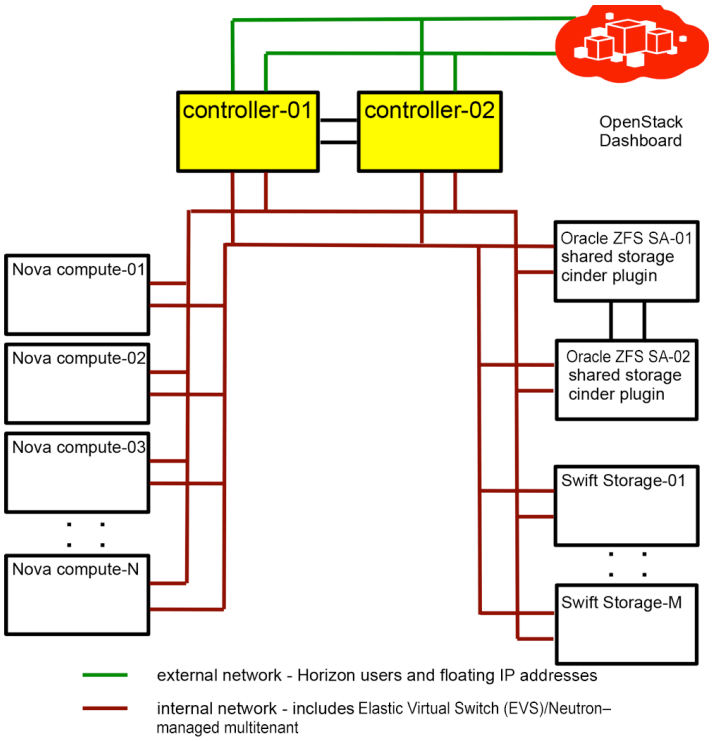

Illustration 2: Basic HA OpenStack deployment

Illustration 2 depicts just one example of a highly available OpenStack infrastructure deployment. The two cloud controller cluster nodes in yellow represent the highly available OpenStack Cloud Controller configuration running on 2 separate physical nodes. Of course, the OpenStack Cloud Controller components can be distributed across more nodes on a larger cluster or on virtualized environments that share the physical nodes with other functions such as compute. Details will be provided in the sections to follow.
Both architectures have in common :

- Clustered Cloud controller nodes managed by Oracle Solaris Cluster
- Clustered storage: a clustered Oracle ZFS Storage Appliance provides shared storage to the cluster nodes as well as the quorum device. It is also used as storage for the Nova compute nodes through the built-in iSCSI Cinder driver.
- Swift storage nodes

Topology 1

The objective of the first topology is to prioritize the flexibility and high availability level. This topology provides fine-grained control over the different OpenStack Cloud Controller services as well as fastest failure detection and recovery time. In this architecture the OpenStack services are under Oracle Solaris Cluster direct control. They can be separately started and stopped and monitored through specific probes. Resources such as virtual IP addresses and shared file systems are managed by the clustering software, offering the required high availability and data corruption protection as well as usage of the built-in cluster load balancer. Orchestration of the different services for start, shutdown and failover can also be managed through the use of inter-component dependencies, enabling error-free, automated bringup and recovery. In this architecture separate Zone clusters are used to separate three set of components: the MySQL database, RabbitMQ and the OpenStack services enabling fault and resource management isolation. Note: it is possible to further distribute the components into smaller sets and distribute them across multiple physical nodes with out loosing the advantages of fast failure detection and advanced orchestration as this is just a example of a possible configuration.

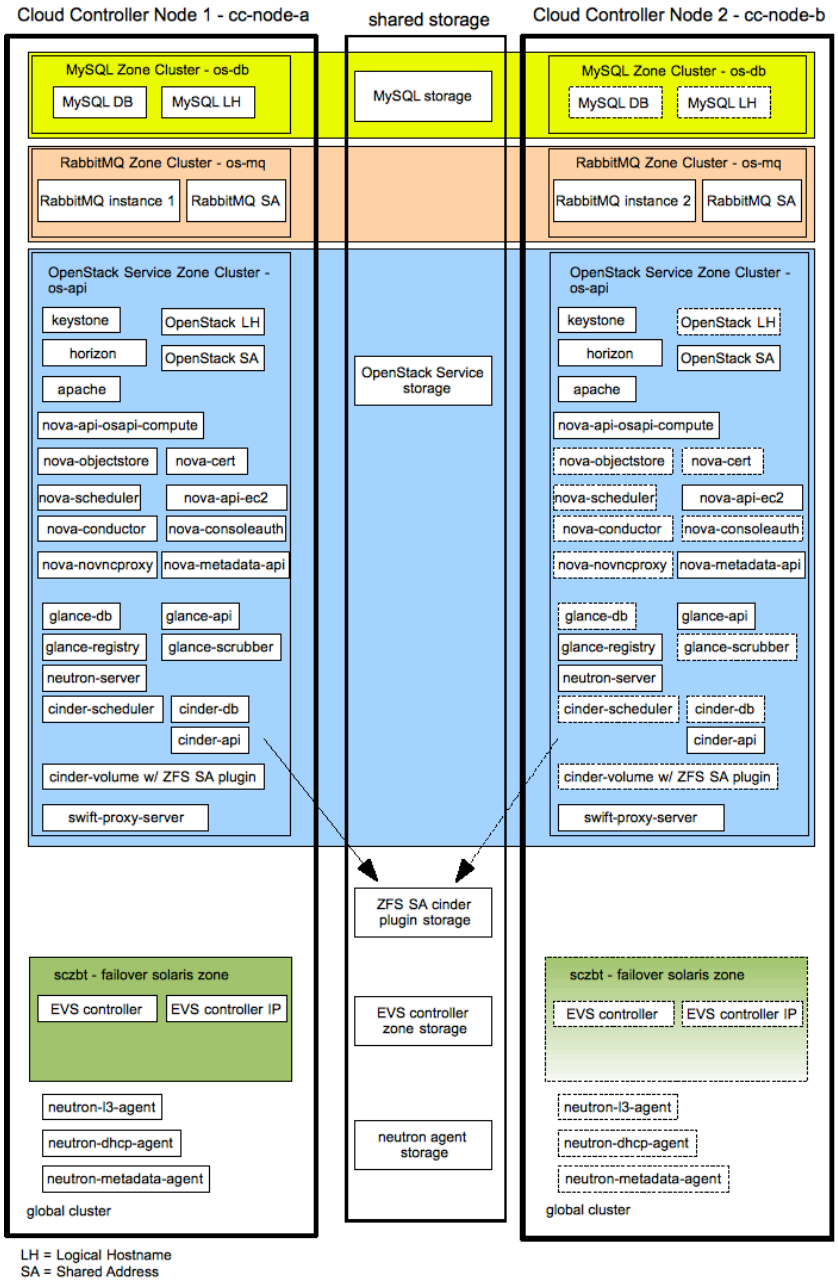

Illustration 3: HA OpenStack Cloud Controller – Topology 1

Topology 2

The objective of the second topology example is to prioritize simplicity of administration while retaining high availability characteristics. In this architecture the different services are is based on the usage for failover zones where the different services are configured inside the zones, but only the zones themselves are under Oracle Solaris Cluster control, which manages the start, stop and probing at the zone level. This means a simplification at the administrative level: the administration and upgrade of OpenStack services are nearly identical to non-HA setup. It is also possible to use kernel zones for further isolation and easier migration between physical nodes. In case of server failure, the different zones are restarted automatically on an alternate server delivering service resiliency.

However the overall failure to recovery time will be increased: the failure detection time might increase as the monitoring is done at the zone level and not as the service level and the recovery of the services will require first the booting of the corresponding zone.

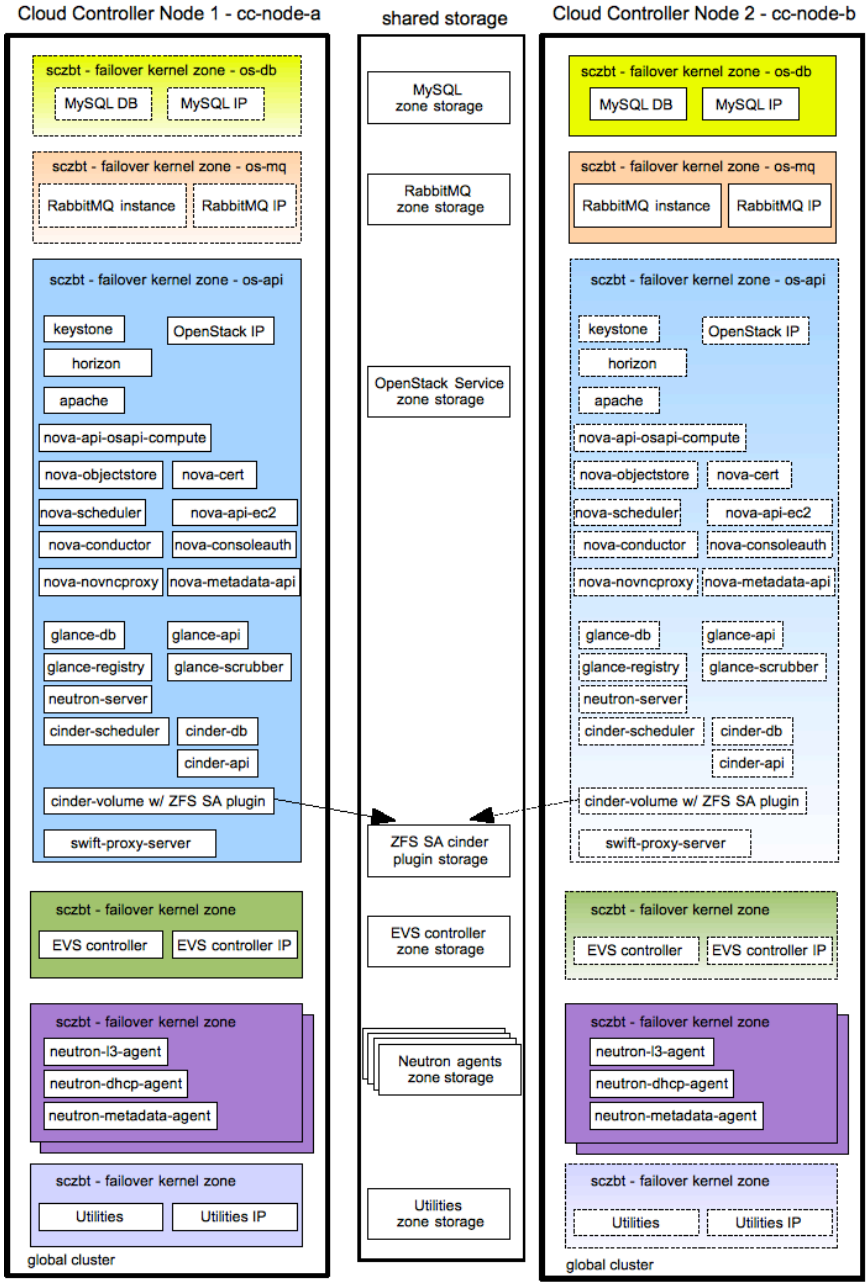

Illustration 4: HA OpenStack Cloud Controller – Topology 2

Conclusion

Many approaches are valid when it comes to designing the architecture for an highly available OpenStack Cloud Controller. To make the right choice it is important to not only look at the

requirements in terms of high availability but also to take into account scalability, flexibility and ease of operations. This will help to find the best approach for a successful deployment.

For more information

Oracle Openstack for Oracle Solaris
http://www.oracle.com/technetwork/server-storage/solaris11/technologies/openstack-2135773.html

Oracle Solaris Cluster
http://www.oracle.com/technetwork/server-storage/solaris-cluster/overview/index.html

White Paper: Providing High Availability to the OpenStack Cloud Controller on Oracle Solaris with Oracle Solaris Cluster
http://www.oracle.com/technetwork/server-storage/solaris-cluster/documentation/ha-for-openstack-cloud-2537455.pdf

Contact address:

Eve Kleinknecht
Oracle Corporation
180 Avenue de l'Europe
38330 Montbonnot Saint Martin
France
Email eve.kleinknecht@oracle.com

Thorsten Frueauf
Oracle Deutschland B.V. & Co. KG
Liebknechtstr. 35
70565 Stuttgart
Germany
Email thorsten.frueauf@oracle.com