Reference Architectures for Rackspace DataStax Cassandra Platform

Introduction

Apache Cassandra™ is a distributed storage system for managing very large amounts of structured data spread out across many commodity servers, while providing highly available service with no single point of failure. Datastax Enterprise (DSE), is one of the leading open Apache Cassandra™ distributions.

We have partnered with Datastax to bring Cassandra to our dedicated portfolio as well cloud. With this managed support offering, customers can:

· **Design an always-on distributed database configuration**: customizing to address multi-datacenter, replication, or no single point of failure.

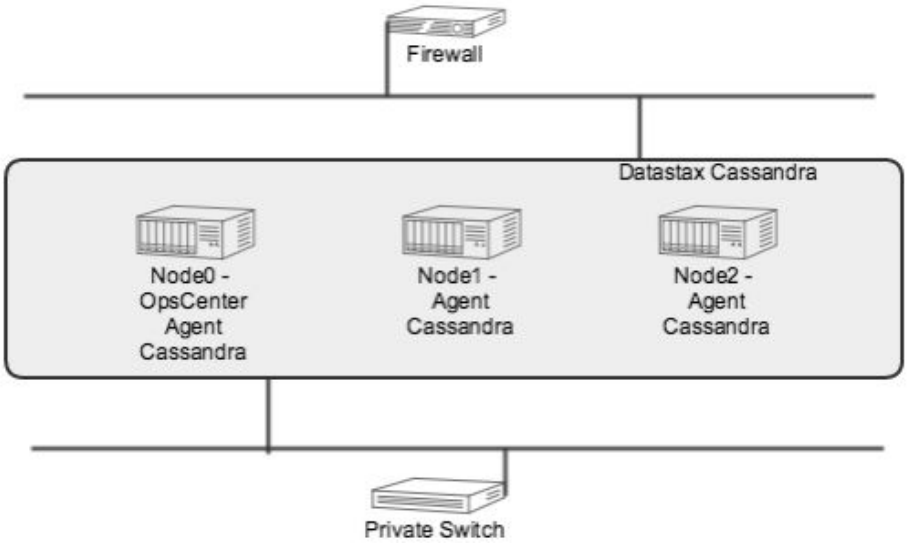
· **Reduce operational burden**: leveraging the deep expertise of Rackspace and Datastax, delivered fanatically 24x7, to deploy and maintain a Cassandra environment, which can be taxing to an IT department.

· **Analytics integration**: connecting to external Hadoop vendors such as HortonWorks allows low-latency and long-running analytics on Cassandra data.

We will start by outlining the components of a typical build and where those components normally reside.

We will then outline various reference architectures, which have come to suite our customers needs.

Small Reference Architecture (consolidated):



The above deployment is the baseline for our architectural reference. In this diagram we have the services associated with the given pieces of a Datastax cluster.

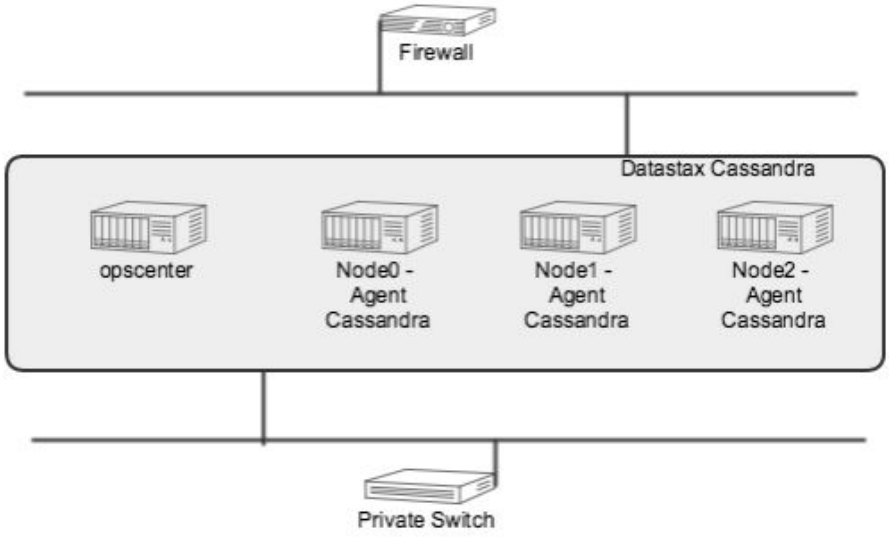
Private Switch: At the bottom we have a private switch. This switch is independent of any connectivity to the firewall. The purpose is to allow all cluster traffic for data processing to traverse this switch only. Intercluster traffic can tend to be heavy so it is recommended to include 10G switching in anything but the smallest of environments.

Node0 (Primary): The primary node (Node0) acts like all your other nodes replicating data but if you are not want to have an extra server to run Ops Center its recommended to put it on Node0. Depending on your data size or transaction-rate it may require separation.

NodeN: These servers replication the data across the environment. By default each nodes in the SE cluster has a DataStax agent that is used by OpsCenter to manage and monitor the cluster. A replication factor of 3 is our default configuration, which means that you will have 3 sets of your data. Depending on your data size it may require additional Node servers.

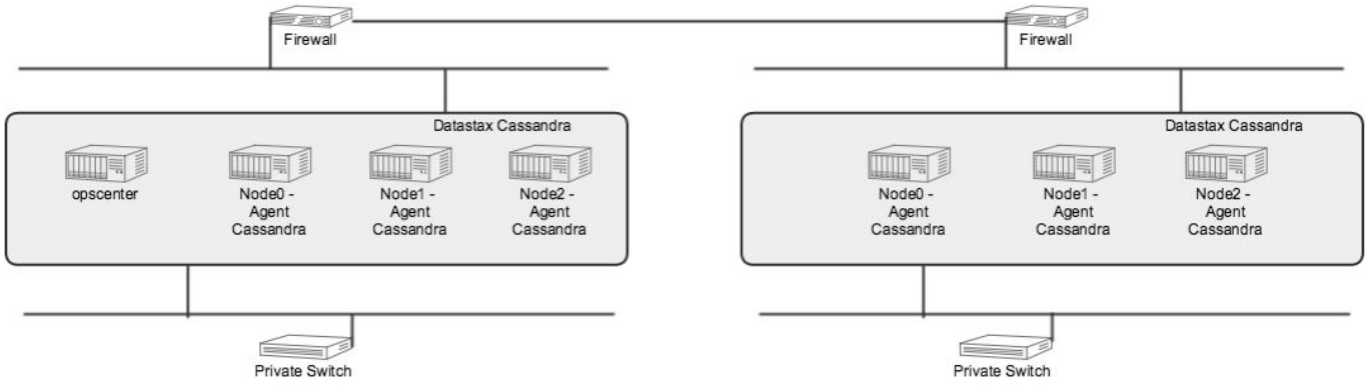
The traditional DSE architectural design allows for a small to medium sized distributed deployments that scale compute and capacity linearly.

Reference Architecture (Separate)



Our configuration represented here provides a scalable architecture by physically separating all the necessary components. In some budget conscious cases the edge node is optional but we as referenced in the typical build the customer should understand its use and what they lose.

Reference Architecture (Distributed)



This final configuration represents a distributed Cassandra cluster in multiple DCs. This is a high level view of the components and how they are connected. All Cassandra nodes would be configured to communicate as individuals to support the Cassandra Gossip Protocol.

In Conclusion

At Rackspace, you have choices about your Cassandra implementation. This fact sheet and reference architectures are only a sample of what we can do to make your DSE initiative a success. If you would like to discuss your business and technical needs in deeper detail, Cassandra experts will be glad to have this conversation with you.

Other questions just send an email to apollo@rackspace.com or visit www.rackspace.com

[END]