



High availability configurations

ONTAP Select

Barb Einarsen
November 19, 2019

Table of Contents

- High availability configurations 1
 - Two-node HA versus multi-node HA 2
 - Two-node HA versus two-node stretched HA (MetroCluster SDS) 3

High availability configurations

Discover high availability options to select the best HA configuration for your environment.

Although customers are starting to move application workloads from enterprise-class storage appliances to software-based solutions running on commodity hardware, the expectations and needs around resiliency and fault tolerance have not changed. An HA solution providing a zero recovery point objective (RPO) protects the customer from data loss due to a failure from any component in the infrastructure stack.

A large portion of the SDS market is built on the notion of shared-nothing storage, with software replication providing data resiliency by storing multiple copies of user data across different storage silos. ONTAP Select builds on this premise by using the synchronous replication features (RAID SyncMirror) provided by ONTAP to store an extra copy of user data within the cluster. This occurs within the context of an HA pair. Every HA pair stores two copies of user data: one on storage provided by the local node, and one on storage provided by the HA partner. Within an ONTAP Select cluster, HA and synchronous replication are tied together, and the functionality of the two cannot be decoupled or used independently. As a result, the synchronous replication functionality is only available in the multinode offering.

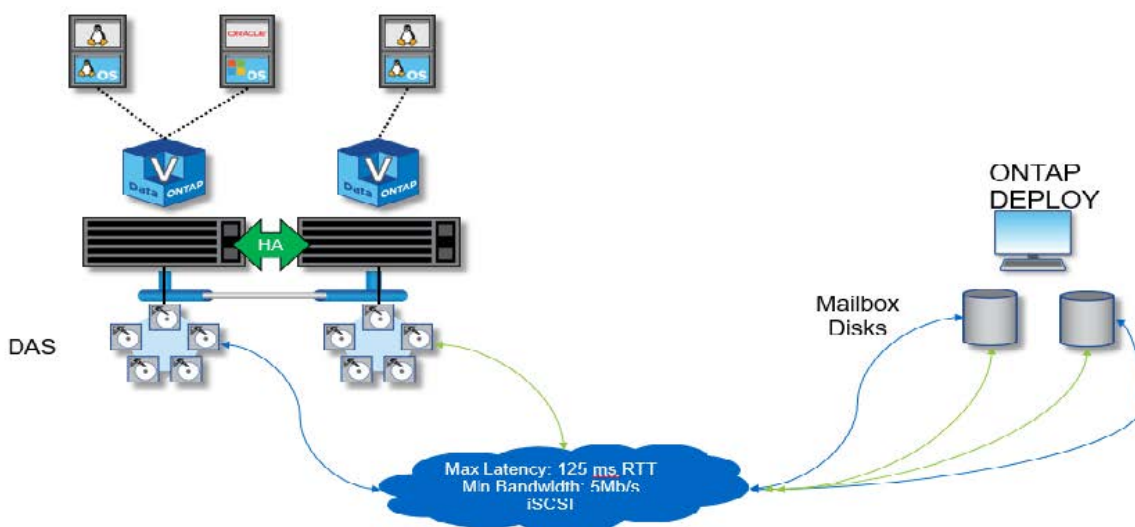


In an ONTAP Select cluster, synchronous replication functionality is a function of HA implementation, not a replacement for the asynchronous SnapMirror or SnapVault replication engines. Synchronous replication cannot be used independently from HA.

There are two ONTAP Select HA deployment models: the multinode clusters (four, six, or eight nodes) and the two-node clusters. The salient feature of a two-node ONTAP Select cluster is the use of an external mediator service to resolve split-brain scenarios. The ONTAP Deploy VM serves as the default mediator for all the two-node HA pairs that it configures.

The two architectures are represented in the following figures.

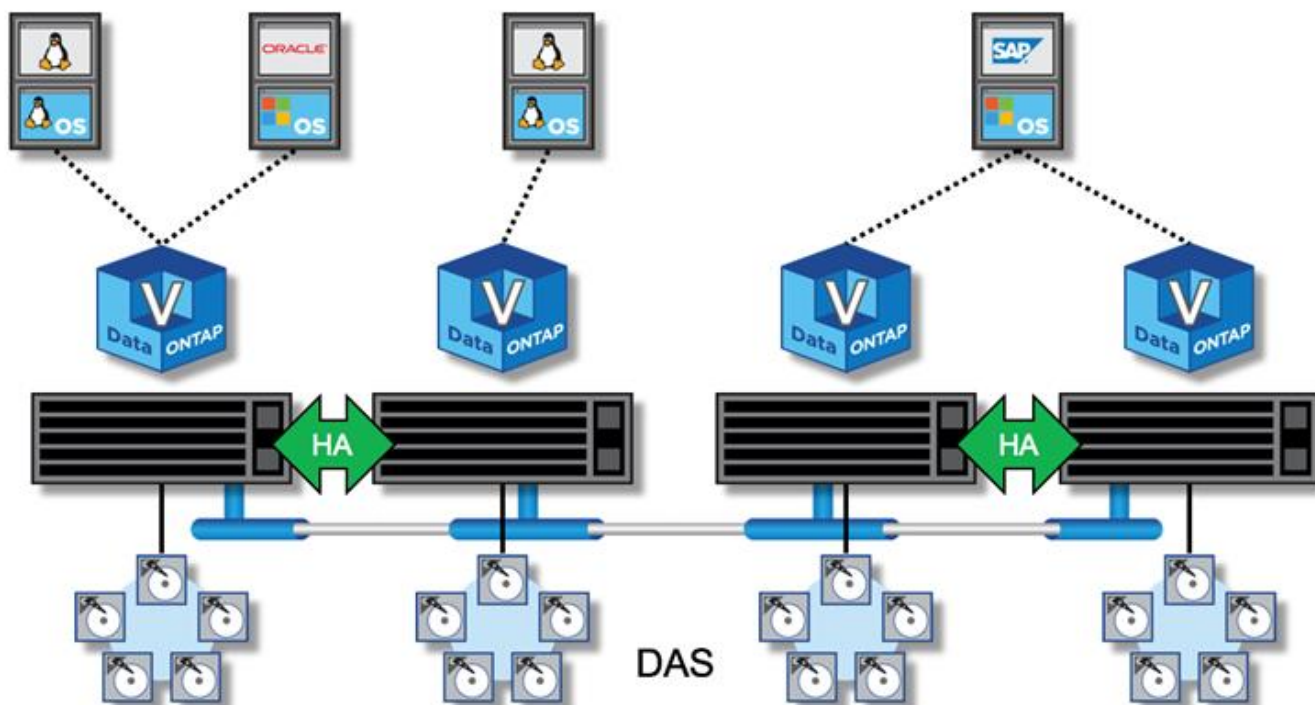
Two-node ONTAP Select cluster with remote mediator and using local-attached storage





The two-node ONTAP Select cluster is composed of one HA pair and a mediator. Within the HA pair, data aggregates on each cluster node are synchronously mirrored, and, in the event of a failover, there is no loss of data.

Four-node ONTAP Select cluster using local-attached storage



- The four-node ONTAP Select cluster is composed of two HA pairs. Six-node and eight-node clusters are composed of three and four HA pairs, respectively. Within each HA pair, data aggregates on each cluster node are synchronously mirrored, and, in the event of a failover, there is no loss of data.
- Only one ONTAP Select instance can be present on a physical server when using DAS storage. ONTAP Select requires unshared access to the local RAID controller of the system and is designed to manage the locally attached disks, which would be impossible without physical connectivity to the storage.

Two-node HA versus multi-node HA

Unlike FAS arrays, ONTAP Select nodes in an HA pair communicate exclusively over the IP network. That means that the IP network is a single point of failure (SPOF), and protecting against network partitions and split-brain scenarios becomes an important aspect of the design. The multinode cluster can sustain single-node failures because the cluster quorum can be established by the three or more surviving nodes. The two-node cluster relies on the mediator service hosted by the ONTAP Deploy VM to achieve the same result.

The heartbeat network traffic between the ONTAP Select nodes and the ONTAP Deploy mediator service is minimal and resilient so that the ONTAP Deploy VM can be hosted in a different data center than the ONTAP Select two-node cluster.



The ONTAP Deploy VM becomes an integral part of a two-node cluster when serving as the mediator for that cluster. If the mediator service is not available, the two-node cluster continues serving data, but the storage failover capabilities of the ONTAP Select cluster are disabled. Therefore, the ONTAP Deploy mediator service must maintain constant communication with each ONTAP Select node in the HA pair. A minimum bandwidth of 5Mbps and a maximum round-trip time (RTT) latency of 125ms are required to allow proper functioning of the cluster quorum.

If the ONTAP Deploy VM acting as a mediator is temporarily or potentially permanently unavailable, a secondary ONTAP Deploy VM can be used to restore the two-node cluster quorum. This results in a configuration in which the new ONTAP Deploy VM is unable to manage the ONTAP Select nodes, but it successfully participates in the cluster quorum algorithm. The communication between the ONTAP Select nodes and the ONTAP Deploy VM is done by using the iSCSI protocol over IPv4. The ONTAP Select node management IP address is the initiator, and the ONTAP Deploy VM IP address is the target. Therefore, it is not possible to support IPv6 addresses for the node management IP addresses when creating a two-node cluster. The ONTAP Deploy hosted mailbox disks are automatically created and masked to the proper ONTAP Select node management IP addresses at the time of two-node cluster creation. The entire configuration is automatically performed during setup, and no further administrative action is required. The ONTAP Deploy instance creating the cluster is the default mediator for that cluster.

An administrative action is required if the original mediator location must be changed. It is possible to recover a cluster quorum even if the original ONTAP Deploy VM is lost. However, NetApp recommends that you back up the ONTAP Deploy database after every two-node cluster is instantiated.

Two-node HA versus two-node stretched HA (MetroCluster SDS)

It is possible to stretch a two-node, active/active HA cluster across larger distances and potentially place each node in a different data center. The only distinction between a two-node cluster and a two-node stretched cluster (also referred to as MetroCluster SDS) is the network connectivity distance between nodes.

The two-node cluster is defined as a cluster for which both nodes are located in the same data center within a distance of 300m. In general, both nodes have uplinks to the same network switch or set of interswitch link (ISL) network switches.

Two-node MetroCluster SDS is defined as a cluster for which nodes are physically separated (different rooms, different buildings, and different data centers) by more than 300m. In addition, each node's uplink connections are connected to separate network switches. The MetroCluster SDS does not require dedicated hardware. However, the environment should adhere to requirements for latency (a maximum of 5ms for RTT and 5ms for jitter, for a total of 10ms) and physical distance (a maximum of 10km).

MetroCluster SDS is a premium feature and requires a Premium license. The Premium license supports the creation of both small and medium VMs, as well as HDD and SSD media.



MetroCluster SDS is supported with both local attached storage (DAS) and shared storage (vNAS). Note that vNAS configurations usually have a higher innate latency because of the network between the ONTAP Select VM and shared storage. MetroCluster SDS configurations must provide a maximum of 10ms of latency between the nodes, including the shared storage latency. In other words, only measuring the latency between the Select VMs is not adequate because shared storage latency is not negligible for these configurations.

Copyright Information

Copyright © 2020 NetApp, Inc. All rights reserved. Printed in the U.S. No part of this document covered by copyright may be reproduced in any form or by any means-graphic, electronic, or mechanical, including photocopying, recording, taping, or storage in an electronic retrieval system-without prior written permission of the copyright owner.

Software derived from copyrighted NetApp material is subject to the following license and disclaimer:

THIS SOFTWARE IS PROVIDED BY NETAPP "AS IS" AND WITHOUT ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE, WHICH ARE HEREBY DISCLAIMED. IN NO EVENT SHALL NETAPP BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

NetApp reserves the right to change any products described herein at any time, and without notice. NetApp assumes no responsibility or liability arising from the use of products described herein, except as expressly agreed to in writing by NetApp. The use or purchase of this product does not convey a license under any patent rights, trademark rights, or any other intellectual property rights of NetApp.

The product described in this manual may be protected by one or more U.S. patents, foreign patents, or pending applications.

RESTRICTED RIGHTS LEGEND: Use, duplication, or disclosure by the government is subject to restrictions as set forth in subparagraph (c)(1)(ii) of the Rights in Technical Data and Computer Software clause at DFARS 252.277-7103 (October 1988) and FAR 52-227-19 (June 1987).

Trademark Information

NETAPP, the NETAPP logo, and the marks listed at <http://www.netapp.com/TM> are trademarks of NetApp, Inc. Other company and product names may be trademarks of their respective owners.