Exchange Server 2013
High Availability | Site Resilience

Scott Schnoll
Principal Technical Writer
Microsoft Corporation

Twitter: Schnoll
Blog: http://aka.ms/Schnoll
Agenda

• Storage
• High Availability
• Site Resilience
Storage Challenges

- Capacity is increasing, but IOPS are not
- Database sizes must be manageable
- Reseeds must be fast and reliable
- Passive copy IOPS are inefficient
- Lagged copies have asymmetric storage requirements
- Low agility from low disk space recovery
Storage Enhancements

- Multiple Databases Per Volume
- Automatic Reseed
- Automatic Recovery from Storage Failures
- Lagged Copy Enhancements
MULTIPLE DATABASE PER VOLUME
Multiple databases per volume

4-member DAG
4 databases
4 copies of each database
4 databases per volume

Symmetrical design with balanced activation preference

Number of copies per database = number of databases per volume

Active Passive Lagged
Multiple databases per volume

Single database copy/disk:
Reseed 2TB Database = \(~23\) hrs
Reseed 8TB Database = \(~93\) hrs
Multiple databases per volume

Single database copy/disk:
Reseed 2TB Database = ~23 hrs
Reseed 8TB Database = ~93 hrs

4 database copies/disk:
Reseed 2TB Disk = ~9.7 hrs
Reseed 8TB Disk = ~39 hrs
Multiple databases per volume

• **Requirements**
  – Single logical disk/partition per physical disk

• **Recommendations**
  – Databases per volume should equal the number of copies per database
  – Same neighbors on all servers
  – Balance activation preferences
AUTORESEED
Seeding Challenges

• Disk failure on active copy = database failover
• Failed disk and database corruption issues need to be addressed quickly
• Fast recovery to restore redundancy is needed
Seeding Enhancements

- Auto reseed - automatically restore redundancy after disk failure
Autoreseed

Periodically scan for failed and suspended copies → Check prerequisites: single copy, spare availability → Allocate and remap a spare → Start the seed → Verify that the new copy is healthy → Admin replaces failed disk
AutoReseed

Configure storage subsystem with spare disks

Create DAG, add servers with configured storage

Create directory and mount points

Configure DAG, including 3 new properties

Create mailbox databases and database copies

AutoDagDatabaseCopiesPerVolume = 1
Autoreseed

• **Requirements**
  – Single logical disk/partition per physical disk
  – Specific database and log folder structure must be used

• **Recommendations**
  – Same neighbors on all servers
  – Databases per volume should equal the number of copies per database
  – Balance activation preferences

• **Configuration instructions**
  – [http://aka.ms/autoreseed](http://aka.ms/autoreseed)
AUTOMATIC RECOVERY FROM STORAGE FAILURES
Recovery Challenges

• **Storage controllers are basically mini-PCs**
  – As such, they can crash, hang, etc., requiring administrative intervention

• **Other operator-recoverable conditions can occur**
  – Loss of vital system elements
  – Hung or highly latent IO
Recovery Enhancements

- Innovations added in Exchange 2010 carried forward
- New recovery behaviors added to Exchange 2013
  - Even more added to Exchange 2013 CU1

<table>
<thead>
<tr>
<th>Exchange Server 2010</th>
<th>Exchange Server 2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESE Database Hung IO (240s)</td>
<td>System Bad State (302s)</td>
</tr>
<tr>
<td>Failure Item Channel Heartbeat (30s)</td>
<td>Long I/O times (41s)</td>
</tr>
<tr>
<td>SystemDisk Heartbeat (120s)</td>
<td>MSExchangeRepl.exe memory threshold (4GB)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exchange Server 2013 CU1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bus reset (event 129)</td>
</tr>
<tr>
<td>Replication service endpoints not responding</td>
</tr>
</tbody>
</table>
LAGGED COPY ENHANCEMENTS
Lagged Copy Challenged

- Activation is difficult
- Lagged copies require manual care
- Lagged copies cannot be page patched
Lagged Copy Enhancements

• Automatic log file replay in a variety of situations
  – Low disk space (enable in registry)
  – Page patching (enabled by default)
  – Less than 3 other healthy copies (enable in AD; configure in registry)

• Integration with Safety Net
  – No need for log surgery or hunting for the point of corruption
HIGH AVAILABILITY
High Availability Challenges

• High availability focuses on database health
• Best copy selection insufficient for new architecture
• Management challenges around maintenance and DAG network configuration
High Availability Enhancements

- Managed Availability
- Best Copy and Server Selection
- DAG Network Autoconfig
MANAGED AVAILABILITY
Managed Availability

• Key tenet for Exchange 2013:
  – All access to a mailbox is provided by the protocol stack on the Mailbox server that hosts the active copy of the user’s mailbox

• If a protocol is down on a Mailbox server, all active databases lose access via that protocol

• Managed Availability was introduced to detect these kinds of failures and automatically correct them
  – For most protocols, quick recovery is achieved via a restart action
  – If the restart action fails, a failover can be triggered
    • Each protocol team designed their own recovery sequence, which is based on their experiences running Office 365 – service experience accrues to the on-premises admin!
Managed Availability

• An internal framework used by component teams
• Sequencing mechanism to control when recovery actions are taken versus alerting and escalation
• Enhances the best copy selection algorithm by taking into account server health
• Includes a mechanism for taking servers in/out of service (maintenance mode)
Managed Availability

• MA failovers are recovery action from failure
  – Detected via a synthetic operation or live data
  – Throttled in time and across the DAG

• MA failovers come in two forms
  – Server: Protocol failure can trigger server failover
  – Database: Store-detected database failure can trigger database failover

• MA includes Single Copy Alert
  – Alert is per-server to reduce flow
  – Still triggered across all machines with copies
  – Monitoring triggered through a notification
  – Logs 4138 (red) and 4139 (green) events
BEST COPY AND SERVER SELECTION
Best Copy Selection Challenges

• Exchange 2010 used several criteria
  – Copy queue length
  – Replay queue length
  – Database copy status – including activation blocked
  – Content index status

• Using just this criteria is not good enough for Exchange 2013, because protocol health is not considered
Best Copy and Server Selection

• Still an Active Manager algorithm performed at *over time based on extracted health of the system
  – Replication health still determined by same criteria and phases
  – Criteria now includes health of the entire protocol stack

• Considers a prioritized protocol health set in the selection
  – Four priorities – critical, high, medium, low (all health sets have a priority)
  – Failover responders trigger added checks to select a “protocol not worse” target
Best Copy and Server Selection

1. All Healthy
   Checks for a server hosting a copy that has all health sets in a healthy state.

2. Up to Normal Healthy
   Checks for a server hosting a copy that has all health sets Medium and above in a healthy state.

3. All Better than Source
   Checks for a server hosting a copy that has health sets in a state that is better than the current server hosting the affected copy.

4. Same as Source
   Checks for a server hosting a copy of the affected database that has health sets in a state that is the same as the current server hosting the affected copy.
DAG NETWORK AUTOCONFIG
DAG Network Challenges

• DAG networks must be manually collapsed in a multi-subnet deployment
• Continuing to reduce administrative burden for deployment and initial configuration
DAG Network Enhancements

• DAGs now default to automatic configuration
  – Still requires specific configuration settings on NICs
  – Manual edits and EAC controls blocked when automatic networking is enabled
  – Set DAG to manual network setup to edit or change DAG networks

• Multi-subnet DAG networks automatically collapsed
DAG Network Enhancements

```
<table>
<thead>
<tr>
<th>SUBNET</th>
<th>STATUS</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.0.0.1</td>
<td>Up</td>
</tr>
<tr>
<td>10.0.0.2</td>
<td>Up</td>
</tr>
</tbody>
</table>
```

Enable replication
SITE RESILIENCE
Site Resilience Challenges

• Operationally complex
• Mailbox and Client Access recovery connected
• Namespace is a SPOF
Site Resilience Enhancements

- Operationally simplified
- Mailbox and Client Access recovery independent
- Namespace provides redundancy
Site Resilience – Operationally Simplified

• Previously loss of CAS, CAS array, VIP, LB, some portion of the DAG required admin to perform a datacenter switchover

• In Exchange Server 2013, recovery happens automatically
  – The admin focuses on fixing the issue, instead of restoring service
Site Resilience – Recovery Independent

• Previously, CAS and Mailbox server recovery were tied together in site recoveries

• In Exchange Server 2013, recovery is independent, and may come automatically in the form of failover
Site Resilience – Namespace Redundancy

- DNS resolves to multiple IP addresses
- Almost all protocol access in Exchange 2013 is HTTP
- HTTP clients have built-in IP failover capabilities
- Clients skip past IPs that produce hard TCP failures
- Admins can switchover by removing VIP from DNS
- Namespace no longer a SPOF
- No dealing with DNS latency
Site Resilience – Three Locations

• With the namespace simplification, consolidation of server roles, separation of CAS array and DAG recovery, and load balancing changes, three locations can simplify mailbox recovery and provide datacenter failovers

• You must have at least three locations
  – Two locations with Exchange; one with witness server
  – Exchange sites must be well-connected
  – Witness server site must be isolated from network failures affecting Exchange sites
Site Resilience

With multiple active load balancing DNS servers, one can set up failover automatically to an alternate VIP and just work!

VIP: 192.168.1.50

VIP: 10.0.1.50

primary datacenter: Redmond
alternate datacenter: Portland
Site Resilience

Assuming MBX3 and MBX4 are operating and one of them can lock the witness.log file, *automatic failover should occur*.
Site Resilience

Primary datacenter: Redmond

Alternate datacenter: Portland
1. Mark the failed servers/site as down: Stop-DatabaseAvailabilityGroup DAG1 –ActiveDirectorySite:Redmond
2. Stop the Cluster Service on Remaining DAG members: Stop-Clussvc
3. Activate DAG members in 2nd datacenter: Restore-DatabaseAvailabilityGroup DAG1 –ActiveDirectorySite:Portland
SUMMARY
Summary

• Many storage enhancements targeted towards JBOD environments
• Numerous high availability improvements
• Site resilience operationally simplified
Questions?

Scott Schnoll
Principal Technical Writer
scott.schnoll@microsoft.com
http://aka.ms/schnoll
/twitter schnoll
Formez-vous en ligne
Retrouvez nos événements
Faites-vous accompagner gratuitement
Essayer gratuitement nos solutions IT
Retrouver nos experts Microsoft

Pros de l’IT

www.microsoftvirtualacademy.com
http://aka.ms/itcamps-france
http://aka.ms/telechargements

Les accélérateurs Windows Azure, Windows Phone, Windows 8

L’IT Team sur TechNet
http://aka.ms/itteam

Développeurs

http://aka.ms/generation-app
http://aka.ms/evenements-developpeurs

La Dev’Team sur MSDN
http://aka.ms/devteam