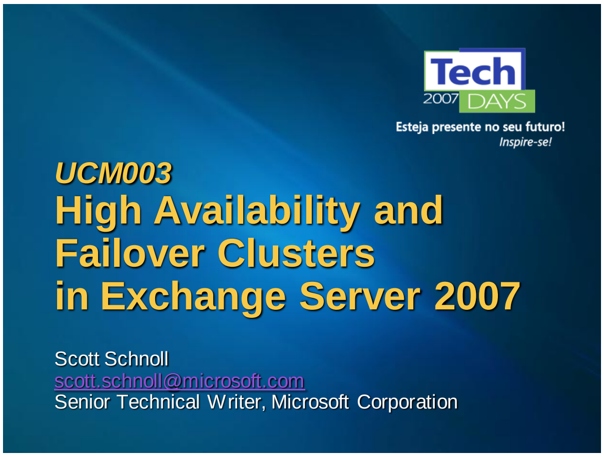
Tech
2007 DAYS
Esteja presente no seu futuro!
Inspire-se!


Tech
2007 DAYS
Esteja presente no seu futuro!
Inspire-se!

UCM003
High Availability and Failover Clusters in Exchange Server 2007

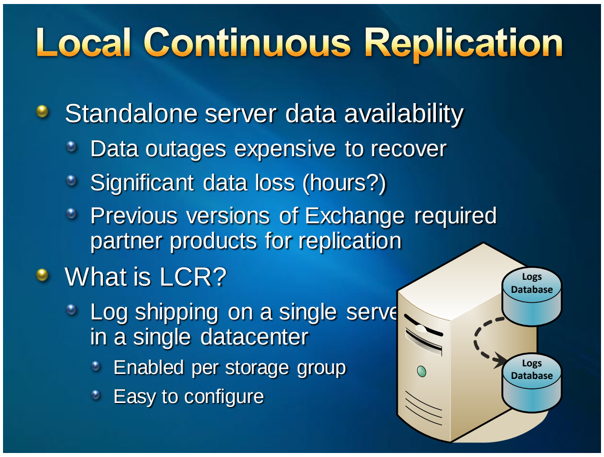
Scott Schnoll
scott.schnoll@microsoft.com
Senior Technical Writer, Microsoft Corporation

Patrocinadores

DELL
GFI
TSUNAMI COMPUTERS
UNISYS imagine it. done.
Actual Training
cpc Informática Sistemas
Novabase
QUEST SOFTWARE
Rumos Formação Profissional
tecnidata grupo
ca
compta
FCA
GALILEU
ParaRede
VANTYX

Agenda

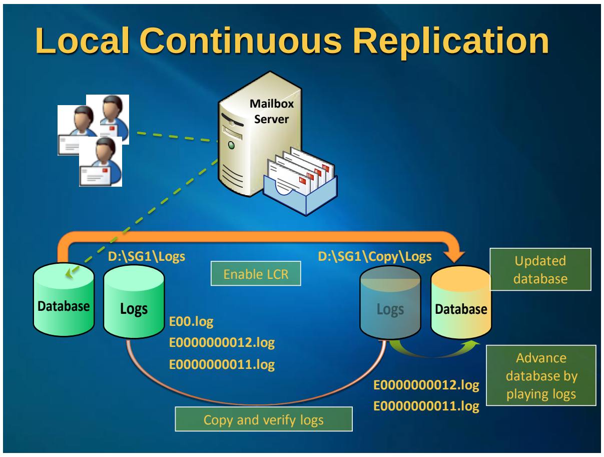
- Exchange 2007 High Availability Features
- Continuous Replication Internals



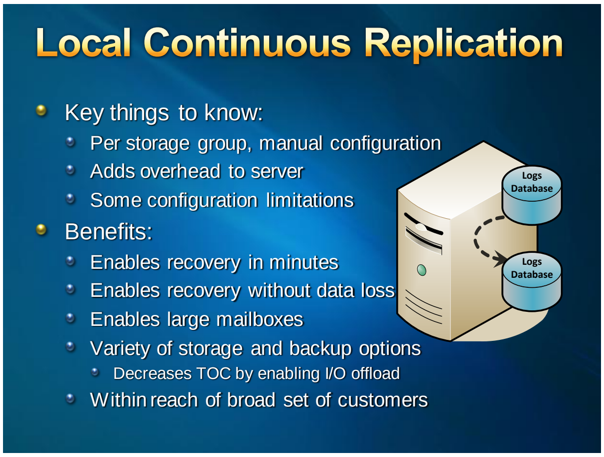
Local Continuous Replication

- Standalone server data availability
  - Data outages expensive to recover
  - Significant data loss (hours?)
  - Previous versions of Exchange required partner products for replication
- What is LCR?
  - Log shipping on a single serve in a single datacenter
    - Enabled per storage group
    - Easy to configure

Logs
Database
Logs
Database



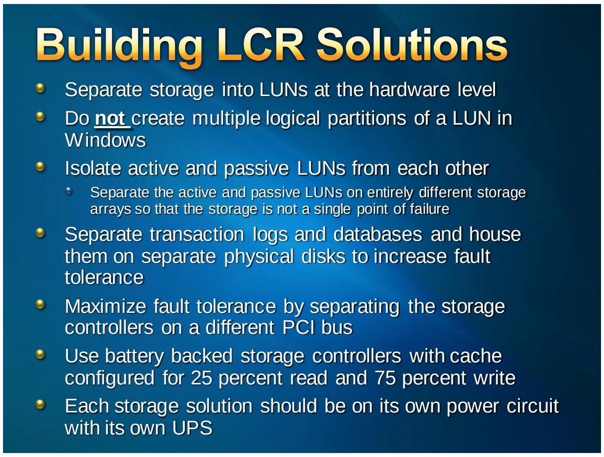
Local Continuous Replication



Local Continuous Replication

- Key things to know:
  - Per storage group, manual configuration
  - Adds overhead to server
  - Some configuration limitations
- Benefits:
  - Enables recovery in minutes
  - Enables recovery without data loss
  - Enables large mailboxes
  - Variety of storage and backup options
    - Decreases TOC by enabling I/O offload
  - Within reach of broad set of customers

Logs
Database
Logs
Database



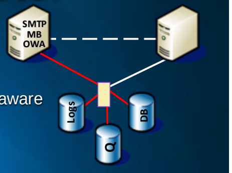
Building LCR Solutions

- Separate storage into LUNs at the hardware level
- Do **not** create multiple logical partitions of a LUN in Windows
- Isolate active and passive LUNs from each other
  - Separate the active and passive LUNs on entirely different storage arrays so that the storage is not a single point of failure
- Separate transaction logs and databases and house them on separate physical disks to increase fault tolerance
- Maximize fault tolerance by separating the storage controllers on a different PCI bus
- Use battery backed storage controllers with cache configured for 25 percent read and 75 percent write
- Each storage solution should be on its own power circuit with its own UPS

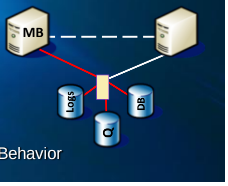
Building LCR Solutions

- Add overhead to Mailbox server design
  - Additional 20% CPU
  - Additional 1 GB memory
  - Passive LUNs require more disk I/O than active LUNs because log replay is a significant generator of both read and write I/O
- Proactive monitoring is required for high availability

Single Copy Cluster

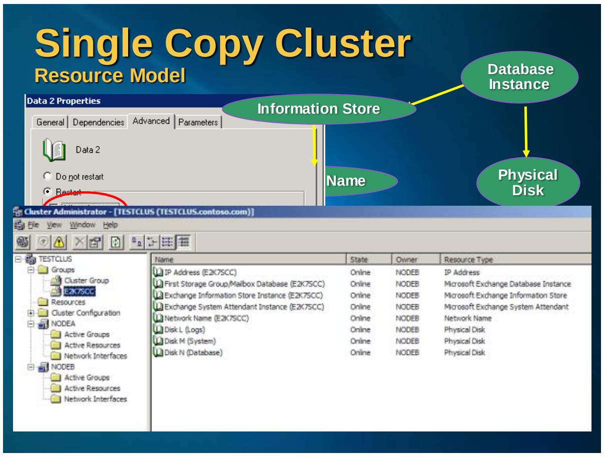


- Exchange Server 2003
  - Requires shared storage
  - SMTP, OWA, and Mailbox are cluster-aware
  - Single copy of mailbox data
  - Up to 8-node Active/Passive
  - 2-Node Active/Active
- Exchange Server 2007
  - Requires shared storage
  - Mailbox Only
    - Simple redundancy for other roles
  - Single copy of mailbox data
  - Up to 8-node Active/Passive
  - Active/Active cut
  - Improvements in Install, Management, Behavior



Single Copy Cluster

Resource Model



Data 2 Properties

General | Dependencies | Advanced | Parameters

Data 2

○ Do not restart
● Restart

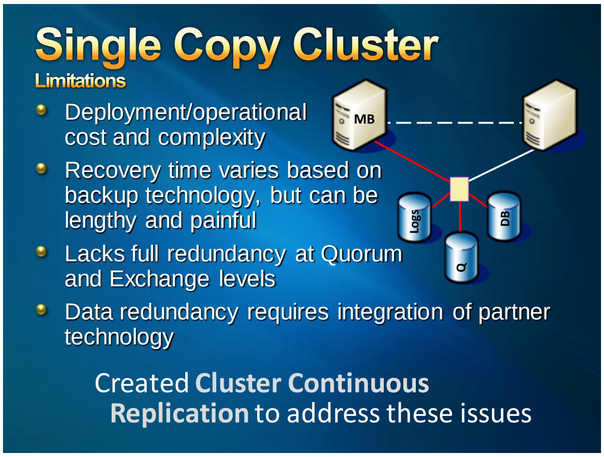
Cluster Administrator - [TESTCLUS (TESTCLUS.contoso.com)]

- TESTCLUS
  - Groups
    - Cluster Group
    - E2K7SCC
  - Resources
  - Cluster Configuration
  - NODEA
    - Active Groups
    - Active Resources
    - Network Interfaces
  - NODEB
    - Active Groups
    - Active Resources
    - Network Interfaces

Name	State	Owner	Resource Type
IP Address (E2K7SCC)	Online	NODEB	IP Address
First Storage Group/Mailbox Database (E2K7SCC)	Online	NODEB	Microsoft Exchange Database Instance
Exchange Information Store Instance (E2K7SCC)	Online	NODEB	Microsoft Exchange Information Store
Exchange System Attendant Instance (E2K7SCC)	Online	NODEB	Microsoft Exchange System Attendant
Network Name (E2K7SCC)	Online	NODEB	Network Name
Disk L (Logs)	Online	NODEB	Physical Disk
Disk M (System)	Online	NODEB	Physical Disk
Disk N (Database)	Online	NODEB	Physical Disk

Building SCC Solutions

- Entire solution must be listed in Cluster Solutions category of Windows Server Catalog
  - Geographically-dispersed solution must be listed in Geographically Dispersed Cluster Solution category of Windows Server Catalog
- Requires shared storage for SGs/DBs
  - Can use MNS or MNS w/FSW quorum
  - Storage must be properly configured before forming cluster
  - Disk resources must be configured for CMS after forming cluster
  - Disk resource dependencies must be configured after CMS is installed

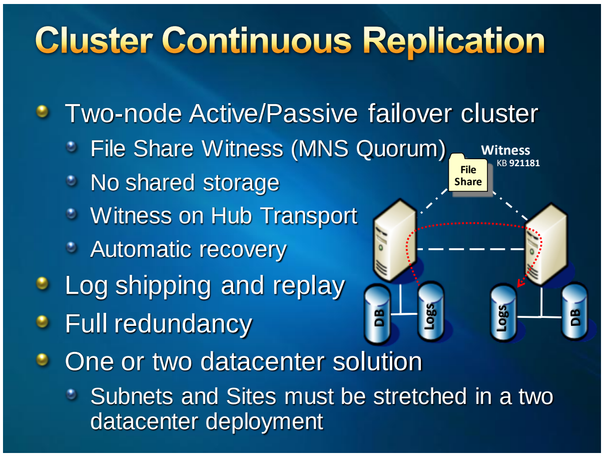


Single Copy Cluster

Limitations

- Deployment/operational cost and complexity
- Recovery time varies based on backup technology, but can be lengthy and painful
- Lacks full redundancy at Quorum and Exchange levels
- Data redundancy requires integration of partner technology

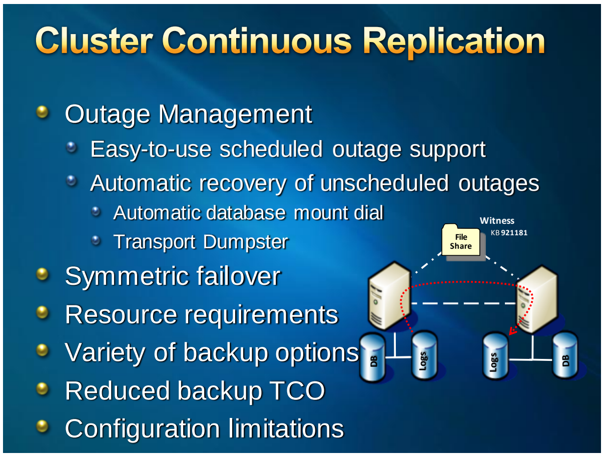
Created **Cluster Continuous Replication** to address these issues



Cluster Continuous Replication

- Two-node Active/Passive failover cluster
  - File Share Witness (MNS Quorum)
  - No shared storage
  - Witness on Hub Transport
  - Automatic recovery
- Log shipping and replay
- Full redundancy
- One or two datacenter solution
  - Subnets and Sites must be stretched in a two datacenter deployment

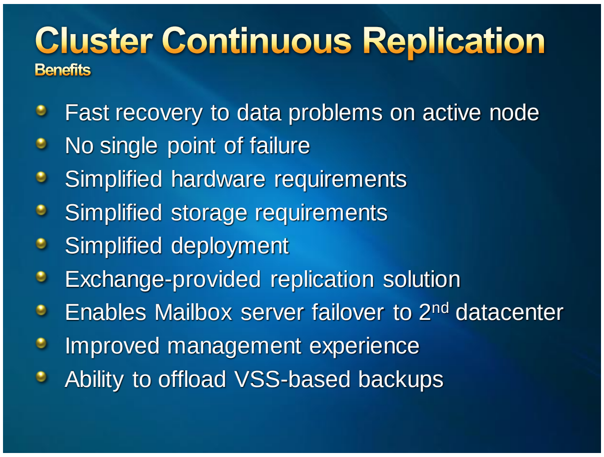
Witness
KB 921181
File Share
DB
Logs
Logs
DB



Cluster Continuous Replication

- Outage Management
  - Easy-to-use scheduled outage support
  - Automatic recovery of unscheduled outages
    - Automatic database mount dial
    - Transport Dumpster
- Symmetric failover
- Resource requirements
- Variety of backup options
- Reduced backup TCO
- Configuration limitations

Witness
KB 921181
File Share
DB
Logs
Logs
DB



Cluster Continuous Replication

Benefits

- Fast recovery to data problems on active node
- No single point of failure
- Simplified hardware requirements
- Simplified storage requirements
- Simplified deployment
- Exchange-provided replication solution
- Enables Mailbox server failover to 2nd datacenter
- Improved management experience
- Ability to offload VSS-based backups

Cluster Continuous Replication

CCR Failover Behavior

- Cluster service monitors the resources
  - Failure detection is not instantaneous
- IP Address or Network Name resource failures cause failover
  - A machine, or network access to it, has failed completely
- Exchange service failure or timeout doesn't cause failover
  - The service is restarted on the same node
- Database failure doesn't cause failover
  - Don't want to move 49 databases because 1 failed

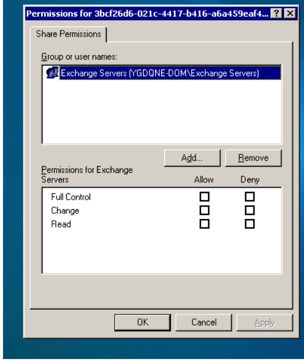
Cluster Continuous Replication

Log shipping file share

- Replication service runs remotely but needs access to log files
- Share created on the active node
- Readable by 'Exchange Servers' universal security group
  - Machine accounts of all Exchange servers
  - Run as LocalSystem to access the share
- 'Exchange Servers' group granted R/O access to files
  - CCR servers only

Cluster Continuous Replication

File Share Permissions



Permissions for 3bcf26d6-021c-4417-b416-a6a459eaf4...

Share Permissions

Group or user names:

Exchange Servers (YGDQNE-DOM\Exchange Servers)

Permissions for Exchange Servers

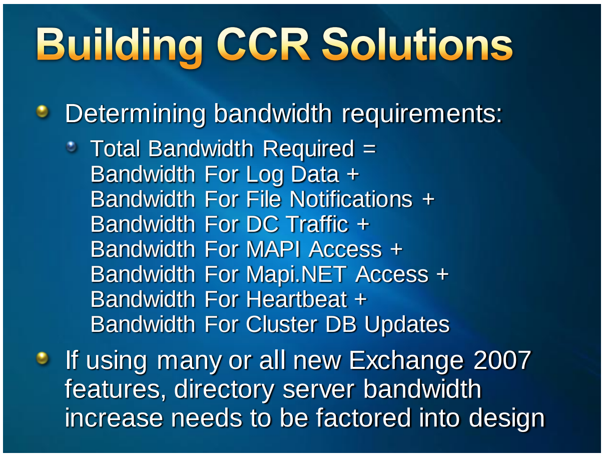
	Allow	Deny
Full Control	□	□
Change	□	□
Read	□	□

This is normal! (Permissions are very restrictive)

Building CCR Solutions

- File share for FSW should be on Hub Transport server located in primary datacenter
- Replacement FSW can be provisioned ahead of time using fake DNS record
- After site failure, you provision a new FSW in the surviving datacenter on a Hub Transport
- Tolerance for missed heartbeats must be properly configured

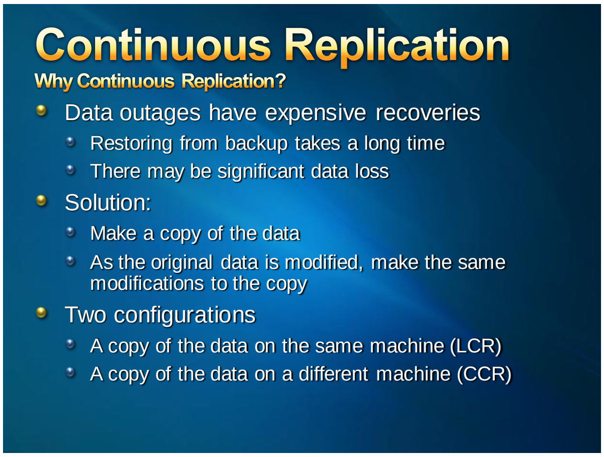
cluster *name* /priv HeartBeatLostInterfaceTicks=10:DWORD
cluster *name* /priv HeartBeatLostNodeTicks=10:DWORD



Building CCR Solutions

- Determining bandwidth requirements:
  - Total Bandwidth Required =
    Bandwidth For Log Data +
    Bandwidth For File Notifications +
    Bandwidth For DC Traffic +
    Bandwidth For MAPI Access +
    Bandwidth For Mapi.NET Access +
    Bandwidth For Heartbeat +
    Bandwidth For Cluster DB Updates
- If using many or all new Exchange 2007 features, directory server bandwidth increase needs to be factored into design

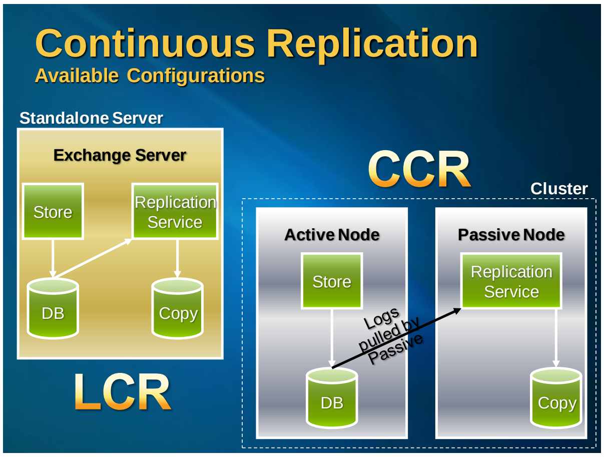
Continuous Replication Internals



Continuous Replication

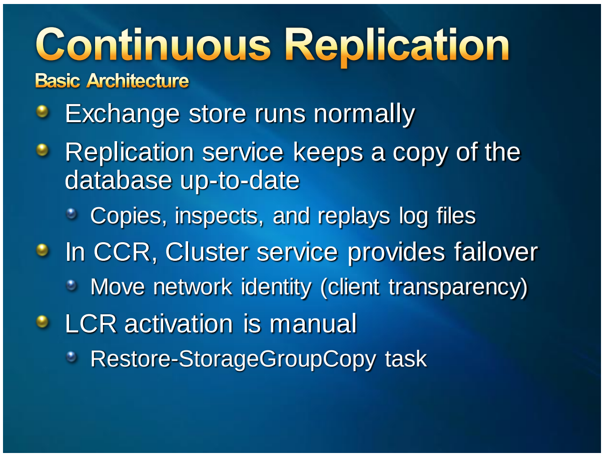
Why Continuous Replication?

- Data outages have expensive recoveries
  - Restoring from backup takes a long time
  - There may be significant data loss
- Solution:
  - Make a copy of the data
  - As the original data is modified, make the same modifications to the copy
- Two configurations
  - A copy of the data on the same machine (LCR)
  - A copy of the data on a different machine (CCR)



Continuous Replication

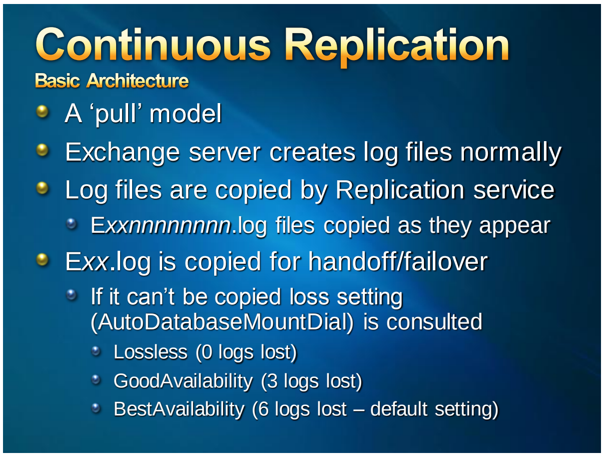
Available Configurations



Continuous Replication

Basic Architecture

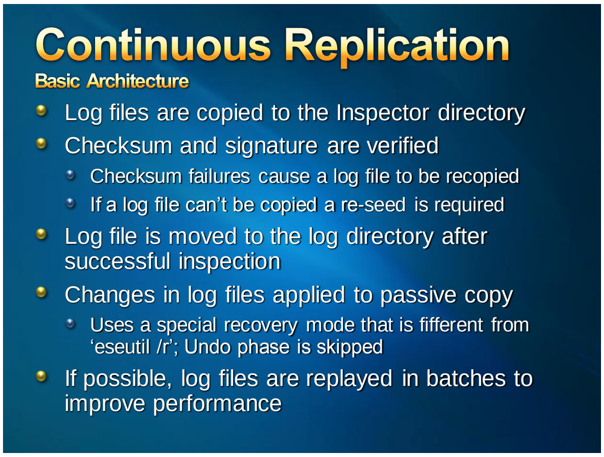
- Exchange store runs normally
- Replication service keeps a copy of the database up-to-date
  - Copies, inspects, and replays log files
- In CCR, Cluster service provides failover
  - Move network identity (client transparency)
- LCR activation is manual
  - Restore-StorageGroupCopy task



Continuous Replication

Basic Architecture

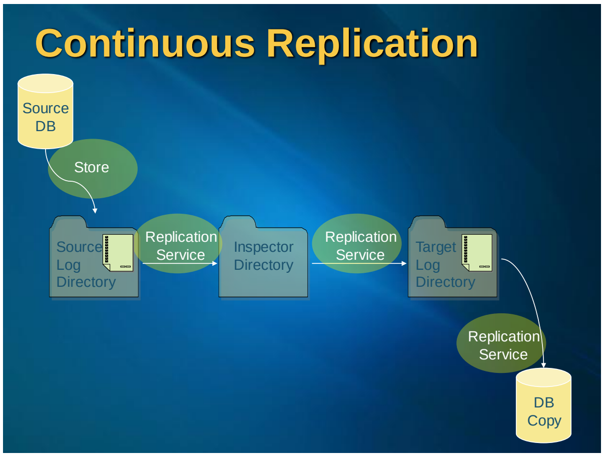
- A 'pull' model
- Exchange server creates log files normally
- Log files are copied by Replication service
  - E*xxnnnnnnnn*.log files copied as they appear
- E*xx*.log is copied for handoff/failover
  - If it can't be copied loss setting (AutoDatabaseMountDial) is consulted
    - Lossless (0 logs lost)
    - GoodAvailability (3 logs lost)
    - BestAvailability (6 logs lost – default setting)



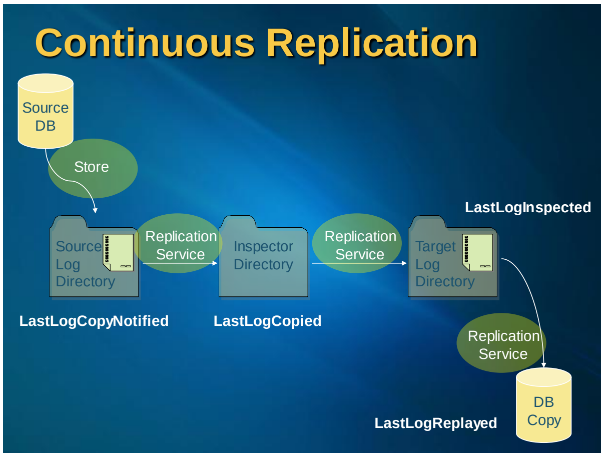
Continuous Replication

Basic Architecture

- Log files are copied to the Inspector directory
- Checksum and signature are verified
  - Checksum failures cause a log file to be recopied
  - If a log file can't be copied a re-seed is required
- Log file is moved to the log directory after successful inspection
- Changes in log files applied to passive copy
  - Uses a special recovery mode that is fifferent from 'eseutil /r'; Undo phase is skipped
- If possible, log files are replayed in batches to improve performance



Continuous Replication



Continuous Replication

Source DB

Store

Source Log Directory

Replication Service

Inspector Directory

Replication Service

Target Log Directory

LastLogInspected

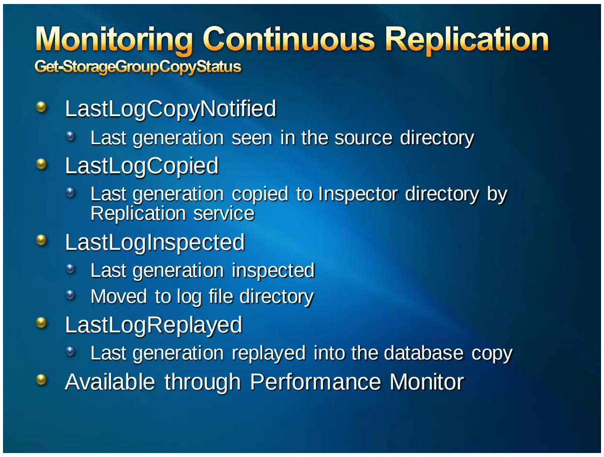
LastLogCopyNotified

LastLogCopied

Replication Service

DB Copy

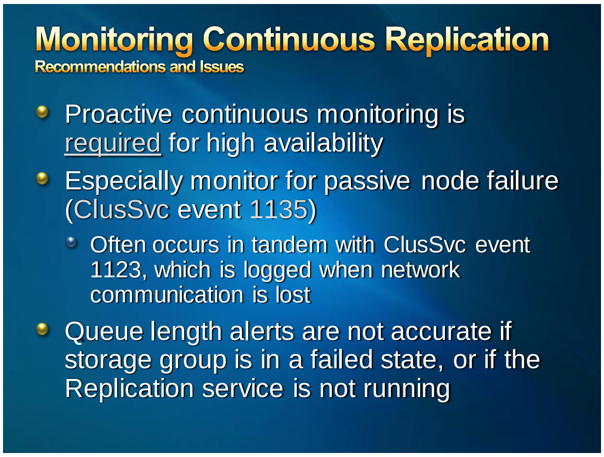
LastLogReplayed



Monitoring Continuous Replication

Get-StorageGroupCopyStatus

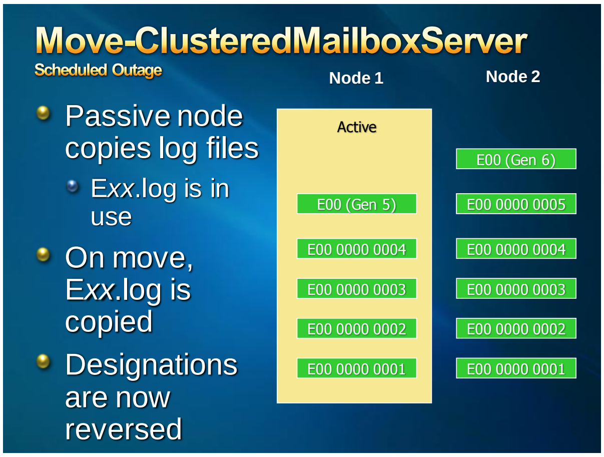
- LastLogCopyNotified
  - Last generation seen in the source directory
- LastLogCopied
  - Last generation copied to Inspector directory by Replication service
- LastLogInspected
  - Last generation inspected
  - Moved to log file directory
- LastLogReplayed
  - Last generation replayed into the database copy
- Available through Performance Monitor



Monitoring Continuous Replication

Recommendations and Issues

- Proactive continuous monitoring is required for high availability
- Especially monitor for passive node failure (ClusSvc event 1135)
  - Often occurs in tandem with ClusSvc event 1123, which is logged when network communication is lost
- Queue length alerts are not accurate if storage group is in a failed state, or if the Replication service is not running

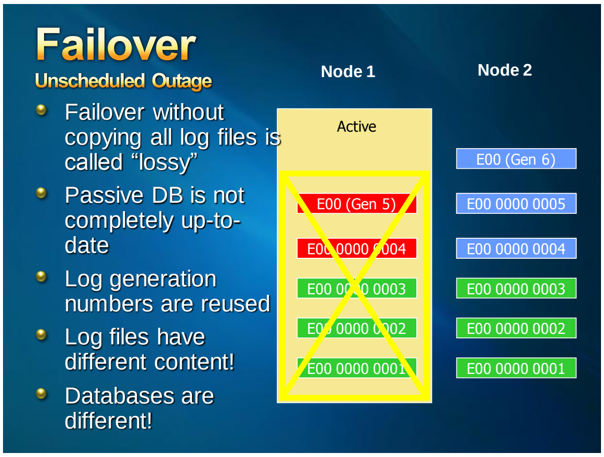


Move-ClusteredMailboxServer

Scheduled Outage

- Passive node copies log files
  - Exx.log is in use
- On move, Exx.log is copied
- Designations are now reversed

Node 1	Node 2
Active	
	E00 (Gen 6)
E00 (Gen 5)	E00 0000 0005
E00 0000 0004	E00 0000 0004
E00 0000 0003	E00 0000 0003
E00 0000 0002	E00 0000 0002
E00 0000 0001	E00 0000 0001

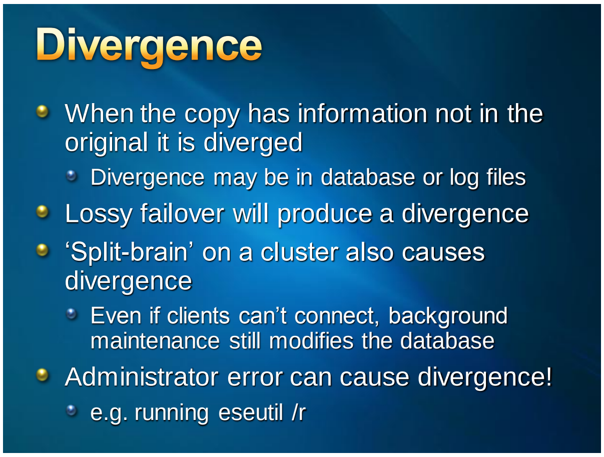


Failover

Unscheduled Outage

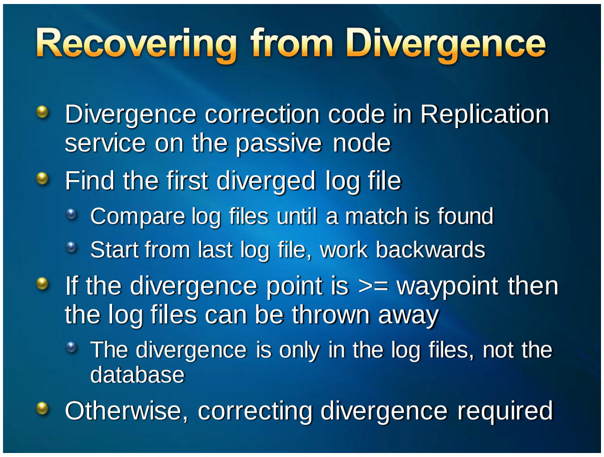
- Failover without copying all log files is called "lossy"
- Passive DB is not completely up-to-date
- Log generation numbers are reused
- Log files have different content!
- Databases are different!

Node 1	Node 2
Active	
	E00 (Gen 6)
E00 (Gen 5)	E00 0000 0005
E00 0000 0004	E00 0000 0004
E00 0000 0003	E00 0000 0003
E00 0000 0002	E00 0000 0002
E00 0000 0001	E00 0000 0001



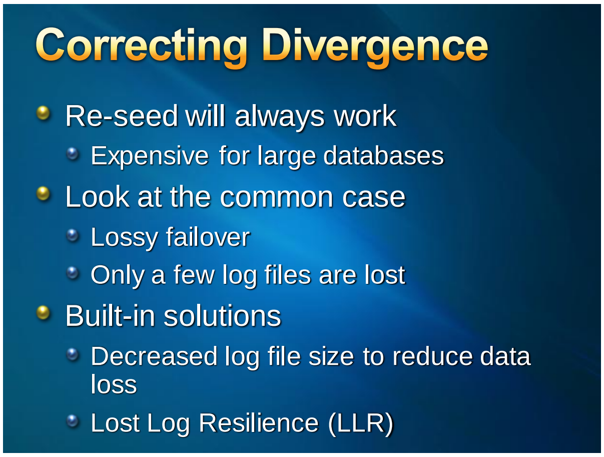
Divergence

- When the copy has information not in the original it is diverged
  - Divergence may be in database or log files
- Lossy failover will produce a divergence
- 'Split-brain' on a cluster also causes divergence
  - Even if clients can't connect, background maintenance still modifies the database
- Administrator error can cause divergence!
  - e.g. running eseutil /r



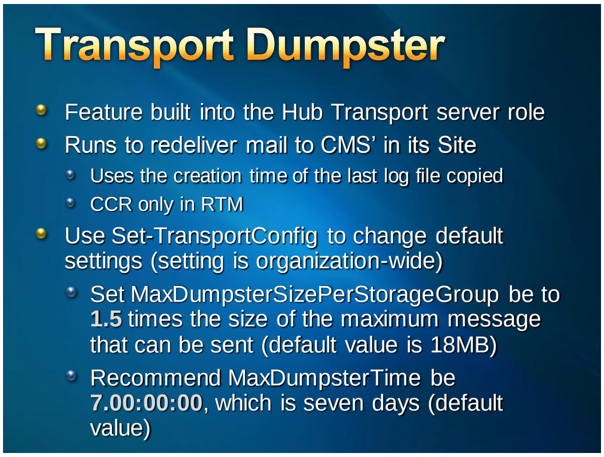
Recovering from Divergence

- Divergence correction code in Replication service on the passive node
- Find the first diverged log file
  - Compare log files until a match is found
  - Start from last log file, work backwards
- If the divergence point is >= waypoint then the log files can be thrown away
  - The divergence is only in the log files, not the database
- Otherwise, correcting divergence required



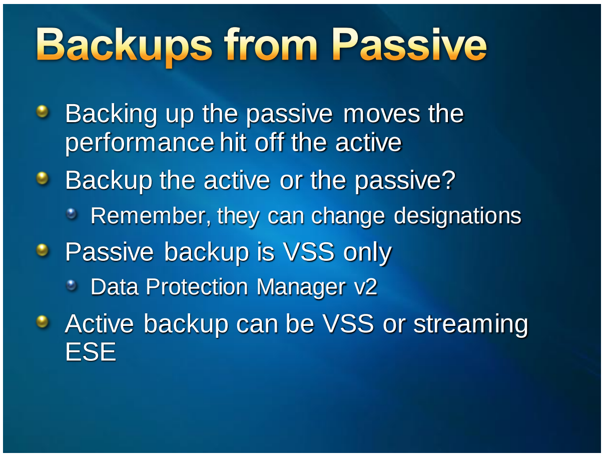
Correcting Divergence

- Re-seed will always work
  - Expensive for large databases
- Look at the common case
  - Lossy failover
  - Only a few log files are lost
- Built-in solutions
  - Decreased log file size to reduce data loss
  - Lost Log Resilience (LLR)



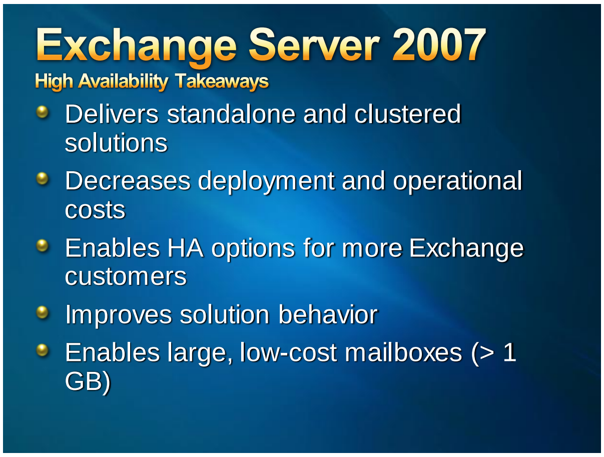
Transport Dumpster

- Feature built into the Hub Transport server role
- Runs to redeliver mail to CMS' in its Site
  - Uses the creation time of the last log file copied
  - CCR only in RTM
- Use Set-TransportConfig to change default settings (setting is organization-wide)
  - Set MaxDumpsterSizePerStorageGroup be to **1.5** times the size of the maximum message that can be sent (default value is 18MB)
  - Recommend MaxDumpsterTime be **7.00:00:00**, which is seven days (default value)



Backups from Passive

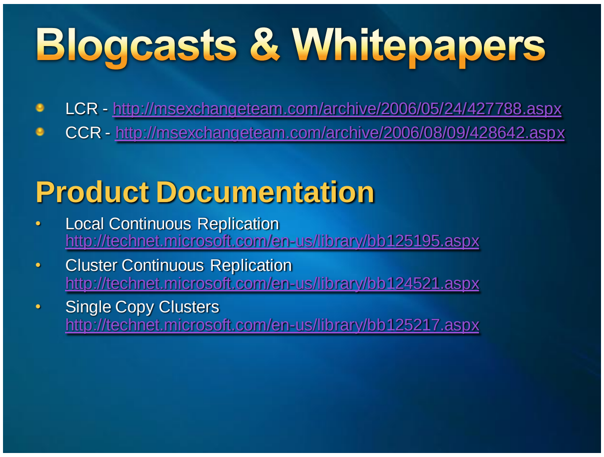
- Backing up the passive moves the performance hit off the active
- Backup the active or the passive?
  - Remember, they can change designations
- Passive backup is VSS only
  - Data Protection Manager v2
- Active backup can be VSS or streaming ESE



Exchange Server 2007

High Availability Takeaways

- Delivers standalone and clustered solutions
- Decreases deployment and operational costs
- Enables HA options for more Exchange customers
- Improves solution behavior
- Enables large, low-cost mailboxes (> 1 GB)



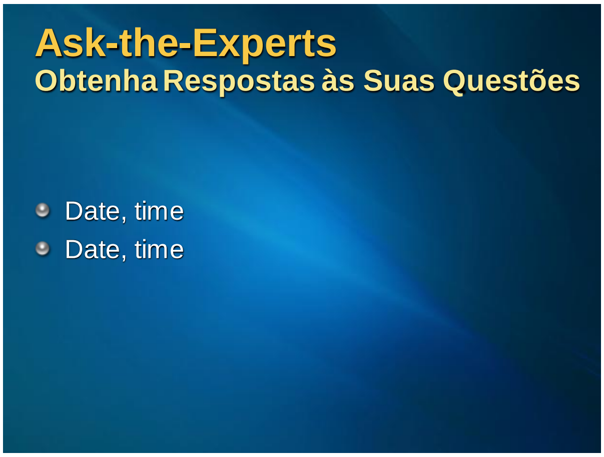
Blogcasts & Whitepapers

- LCR - http://msexchangeteam.com/archive/2006/05/24/427788.aspx
- CCR - http://msexchangeteam.com/archive/2006/08/09/428642.aspx

Product Documentation

- Local Continuous Replication
  http://technet.microsoft.com/en-us/library/bb125195.aspx
- Cluster Continuous Replication
  http://technet.microsoft.com/en-us/library/bb124521.aspx
- Single Copy Clusters
  http://technet.microsoft.com/en-us/library/bb125217.aspx

Q&A



Ask-the-Experts
Obtenha Respostas às Suas Questões

- Date, time
- Date, time

Outros Recursos
Para Profissionais de TI

- TechNet Plus
  - 2 incidentes de suporte gratuito profissional
  - software exclusivo: Capacity Planner
  - software Microsoft para avaliação
  - actualizações de segurança e service packs
  - acesso privilegiado à knowledge base
  - formação gratuita
  - e muito mais.

www.microsoft.com/portugal/technet/subscricoes

Tenha acesso directo a ferramentas e recursos técnicos essenciais
Microsoft TechNet Plus Subscription

Questionário de Avaliação
Passatempo!

- Complete o questionário de avaliação e devolva-o no balcão da recepção.
- Habilite-se a ganhar uma Xbox 360 por dia!

UCM003

High Availability and Failover Clusters in Exchange Server 2007

Microsoft®

Your potential. Our passion.™

This presentation is for informational purposes only. Microsoft makes no warranties, express or implied, in this summary.