Campus Days
14.-16. januar 2010

Session UE-06 Exchange 2010
Flexible and Reliable
DAG and Transport Deep Dive
John Ferris

john.ferris@microsoft.com

http://blogs.technet.com/ferris/

Microsoft TechNet

John Ferris

Senior Messaging Consultant,

Microsoft Consulting Services,

Microsoft Denmark

Agenda

- Exchange 2007 High Availability Recap & Overview of Exchange 2010 HA

- Exchange 2010 High Availability Fundamentals

- High Availability Management

- Storage Improvements

- End to End Availability Improvements

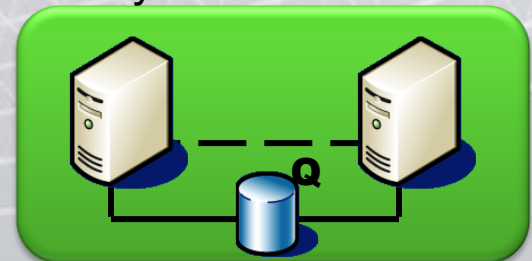- High Availability Design Examples

HA in Exchange Server 2007

- Two Main Technologies

  - Single Copy Clusters (SCC)

  - Continuous Replication (CCR, LCR, SCR)
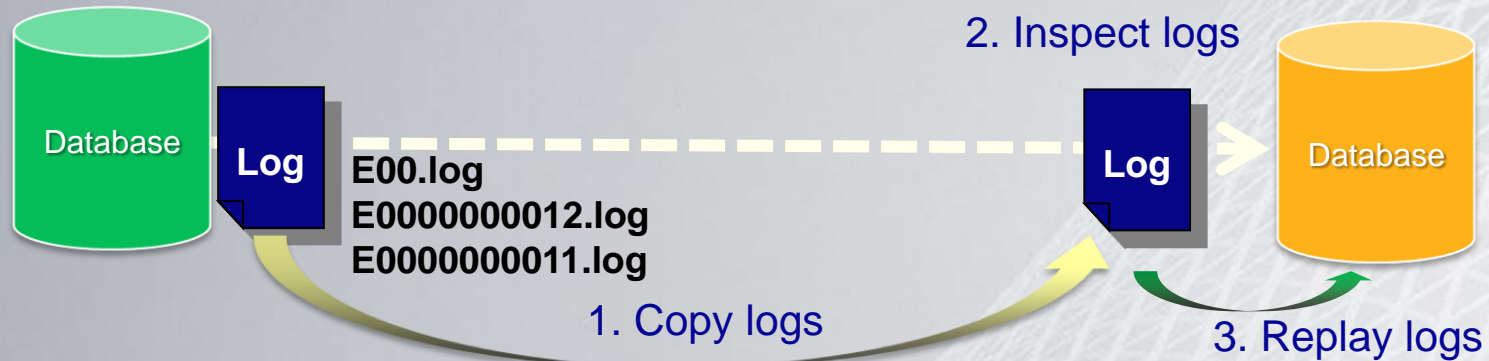
HA in Exchange Server 2007 Single Copy Clustering

- Single Copy Clustering (SCC) out of the box provides little High Availability (HA) value
  - On Store failure, SCC restarts store on the same machine; there is no actual cluster failover
  - SCC does not cover storage failures
  - SCC does not protect your data, your most valuable asset
  - SCC does not cover site failures
  - SCC redundant network support is only for Windows Failover Clustering's internal needs
- Conclusion
  - SCC only provides protection from server hardware failures and bluescreens, the relatively easy components to recover
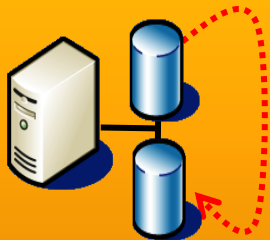 - Supports rolling upgrades without losing redundancy
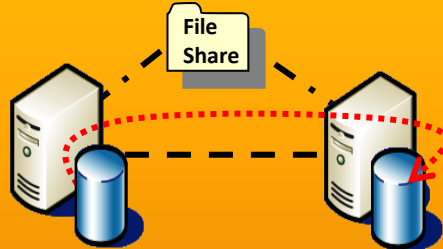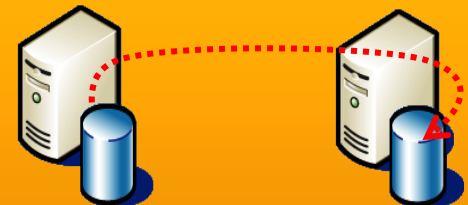
HA in Exchange Server 2007 Continuos Replication

Exchange Server 2007 Solution (CCR + SCR)

Outlook (MAPI) client

OWA, ActiveSync, or Outlook Anywhere

Manual "activation" of remote mailbox server

AD site: London

Client Access Server

Standby Server

DB4

DB5

DB6

AD site: Copenhagen

Client Access Server

Mailbox server can't co-exist with other roles

SCR

SCR managed separately; no GUI

CCR #1 Node A

CCR #1 Node B

Windows cluster

CCR #2 Node A

CCR #2 Node B

Windows cluster

Clustering knowledge required

DB1

DB2

DB3

DB1

DB2

DB3

DB4

DB5

DB6

DB4

DB5

DB6

Database failure requires server failover

High Availability Improvements

Improved mailbox uptime

- Improved failover granularity
- Simplified administration
- Incremental deployment
- Unification of CCR + SCR
- Easy stretching across sites
- Up to 16 replicated copies

More storage flexibility

- Further IO reductions
- RAID-less / JBOD support

Better end-to-end availability

- Improved transport resiliency
- Online mailbox moves

Key benefits

- ✓ Easier & cheaper to deploy
- ✓ Easier & cheaper to manage
- ✓ Better SLAs

- ✓ Reduced storage costs
- ✓ Larger mailboxes

- ✓ Easier & cheaper to manage
- ✓ Better SLAs

Unified Platform for High Availability and Disaster Recovery

Mailbox Server — DB1, DB2, DB3, DB4, DB5

Recover quickly from disk and database failures

Mailbox Server — DB1, DB2, DB3, DB4, DB5

Copenhagen

Replicate databases to remote datacenter

Mailbox Server — DB1, DB2, DB3, DB4, DB5

London
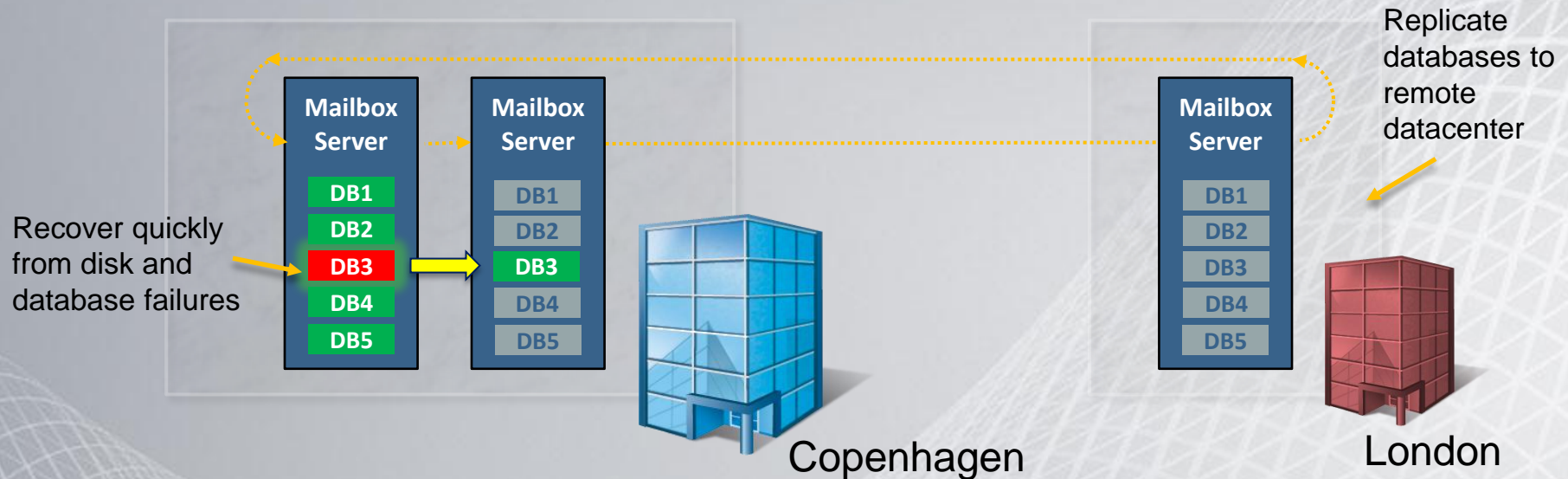
- Evolution of Continuous Replication technology
- Combines the capabilities of CCR and SCR into one platform
- Easier than traditional clustering to deploy and manage
- Allows each database to have up to 16 replicated copies
- Provides full redundancy of Exchange roles on two servers

Exchange 2010 High Availability Overview
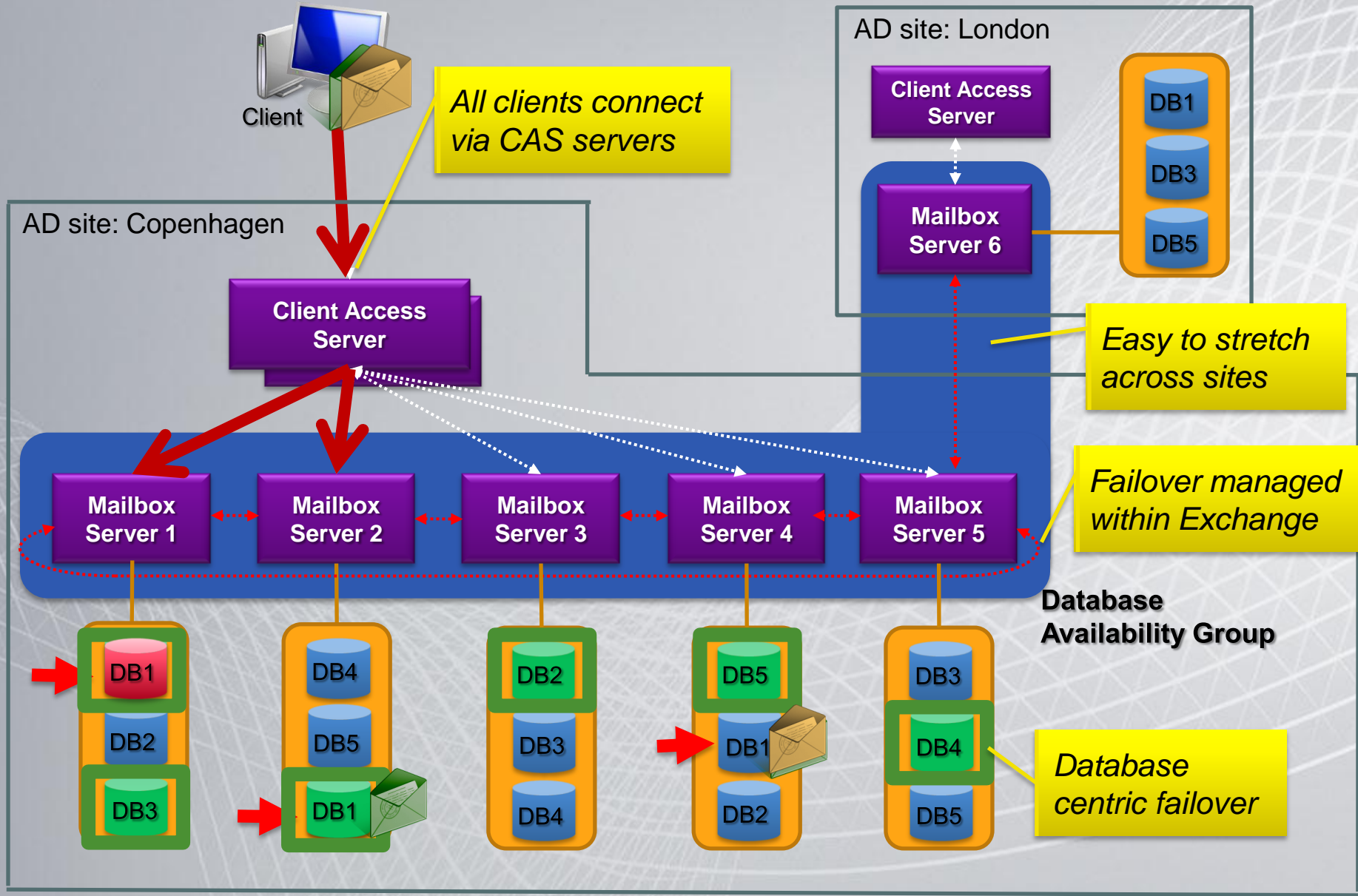
Agenda

- Exchange 2010 High Availability Fundamentals

- High Availability Management

- Storage Improvements

- End to End Availability Improvements

- High Availability Design Examples

Exchange 2010 HA Fundamentals
Database Availability Group (DAG)

DAG Construct….

- Group of up to 16 servers
- Wraps around Windows Failover Cluster
- Defines the boundary of replication and failover/switchover

Mailbox Servers ….

- Host the active and passive copies of multiple mailbox databases
- Support up to 100 Databases per server

Exchange 2010 HA Fundamentals
Mailbox Databases and Copies

- Mailbox Database
  - Unit of Failover/Switchover
  - 30 second Database Failover/Switchover
  - Database names are unique across a forest

- Mailbox Database Copy
  - A database has 1 Active copy in a DAG
  - A server may not host more than 1 copy of a given database
  - Replication of copies using Log Shipping
  - System tracks health of each copy

Exchange 2010 HA Fundamentals
Mailbox Database Copy Status

- Healthy
- Initializing
- Failed
- Suspended

- Mounted
- Dismounted
- ServiceDown
- FailedandSuspended

- Resynchronizing
- Seeding
- ActivationSuspended

Exchange Server 2010 HA Fundamentals
Log Shipping

- Log shipping in Exchange Server 2010 leverages TCP sockets
  - Supports encryption and compression
- Target Replication service notifies the active instance the next log file it expects
- Source Replication service responds by sending the required log file(s)
- Copied log files are placed in the target's Inspector directory
- Validation tests are performed prior to log replay

Exchange 2010 HA Fundamentals
Active Manager

Active Manager

- High Availability's Brain
- Manages which database copies should be active and passive
- Source of definitive information on where a database is active and mounted
  - Active Directory is primary source for configuration information
  - Active Manager is primary source for changeable state information such as active and mounted
- A process that runs on every server in DAG
  - Primary Active Manager (PAM)
  - Secondary Active Manager(s) (SAMs)
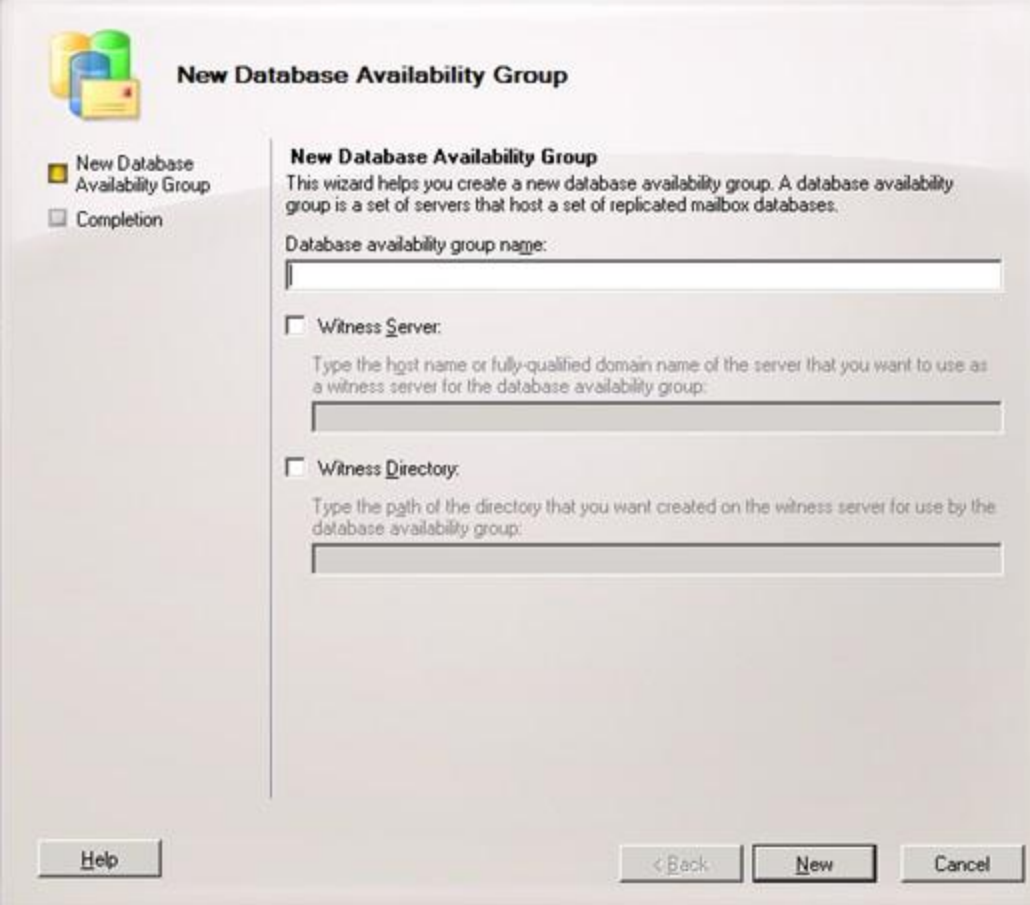
Exchange 2010 HA Fundamentals
Active Manager Selection of Active Database Copy

- Active Manager selects the "best" copy to become when the active fails;

  1. Ignores servers that are unreachable or activation is activation blocked

  2. Sorts copies by currency (last log replayed)

  3. If more than one current copy, the Activation Preference number is used to make decision

  4. Other criteria such as the copy state, and status of Content Indexing etc. are used as final tie breakers

Agenda

- Exchange 2010 High Availability Fundamentals
- High Availability Management
- Storage Improvements
- End-to-End Availability Improvements
- High Availability Design Examples

Creating a Database Availability Group
Exchange Management Console

Creating a Database Availability Group
Exchange Management Console

Creating a Database Availability Group
Exchange Management Console

Creating a Database Availability Group
Exchange Management Shell

- Create DAG

  `New-DatabaseAvailabilityGroup`

- Add servers to a DAG

  `Add-DatabaseAvailabilityGroupServer`

- Add database copies to a server in a DAG

  `Add-MailboxDatabaseCopy`

High Availability Management

demo

Creating the DAG

What's a Switchover?

- A controlled move of a single DB instance is known as a DB Switchover.

- A controlled move of all DB instances on a server is known as a Server Switchover.

- Planned, Expected, No surprises!

- Patching, Upgrades, etc

High Availability Management

demo

Database Switchover and Database Copy Status

What's a Failover?

- An unplanned move of a single Database instance is known as a Database Failover.

- A unplanned move of all DB instances on a server is known as a Server Failover over.

- Unplanned, Unexpected,

- Hardware Failures, Site Failure, Bad Stuff…

High Availability Management

demo

Database Failover….

Multiple Database Copies Enable Backupless Configurations

- Site/Server/Disk failure ➡ Exchange 2010 HA

- Archiving/Compliance ➡ E-mail Archive

- Recover deleted items ➡ Extended / Protected Dumpster Retention

Database Availability Group

| Mailbox Server 1 | Mailbox Server 2 | Mailbox Server 3 |

DB1 DB1 DB1

7-14 day lag copy

DB2 DB2 DB2

DB3 DB3 DB3
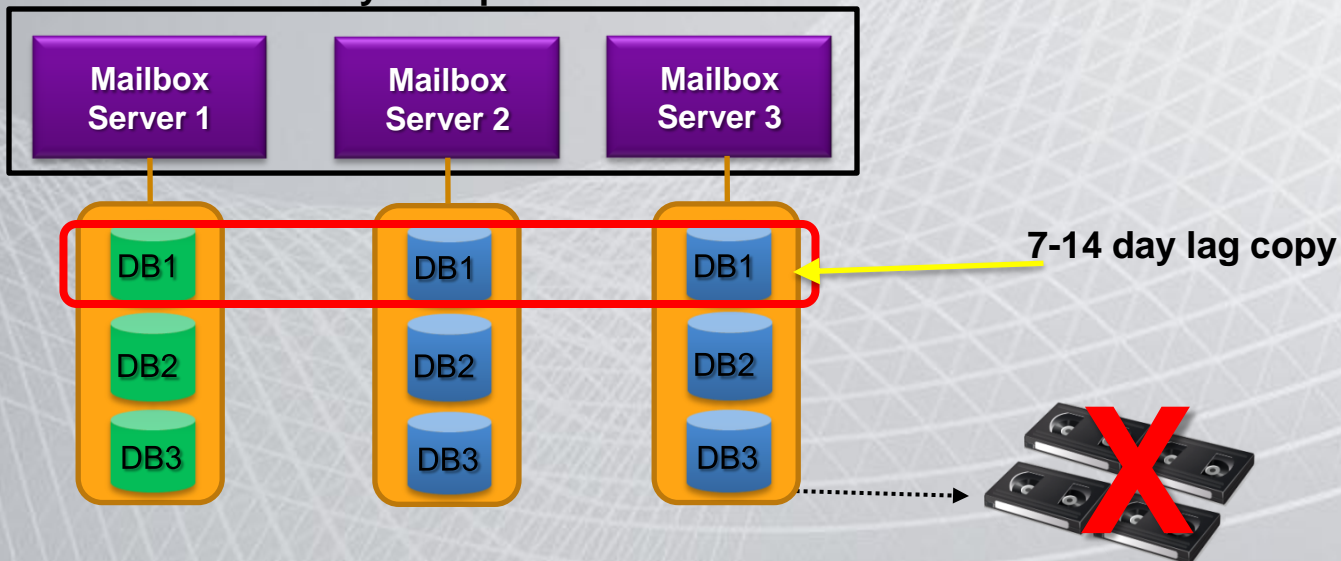
Exchange Server 2010 Backups

- Use a VSS backup solution
  - Backup from any copy of the database/logs
  - Always choose Passive (or Active) copy
  - Backup an entire server
  - Designate a dedicated backup server for a given database
- Restore from any of these backups
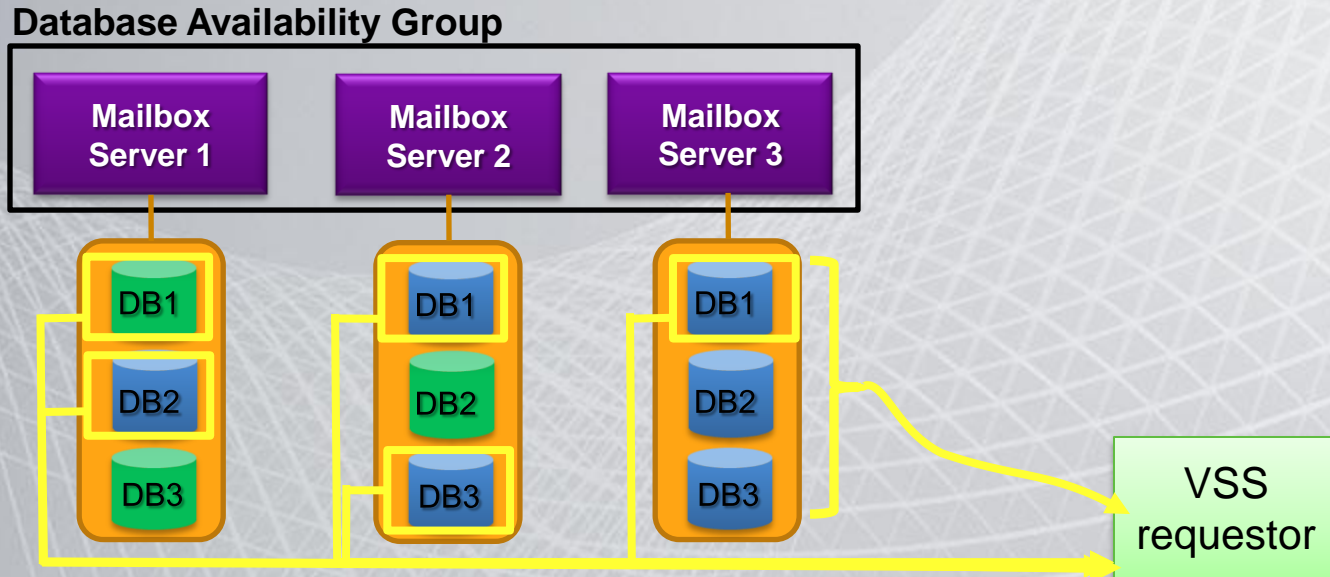
Database Availability Group

Mailbox Server 1 · Mailbox Server 2 · Mailbox Server 3

DB1 · DB2 · DB3

VSS requestor

Agenda

- Exchange 2010 High Availability Fundamentals

- High Availability Management

- Storage Improvements

- End-to-End Availability Improvements

- High Availability Design Examples

Storage Improvements
Performance Enhancements Enable New Options
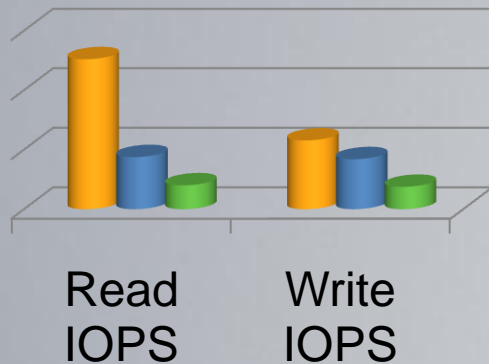
- Ex 2003
- Ex 2007
- Ex 2010

Read IOPS   Write IOPS

Exchange 2010 Storage Enhancements

- 70% reduction in IOPS
- Smoother IO patterns
- Resilience against corruption

Choose from a wide range of storage technologies without sacrificing system availability:

Storage Area Network (SAN)

Direct Attached w/ SAS Disks

Direct Attached w/ SATA Disks

JBOD SATA (RAID-less)

Microsoft® Exchange Server 2003

Microsoft® Exchange Server 2007

Microsoft® Exchange Server 2010

Lowering Exchange 2010 Storage Costs

- Optimized for DAS storage

- Use larger, slower, cheaper disks

  - Support larger mailboxes at lower cost

- HA provides resilience from disk failures

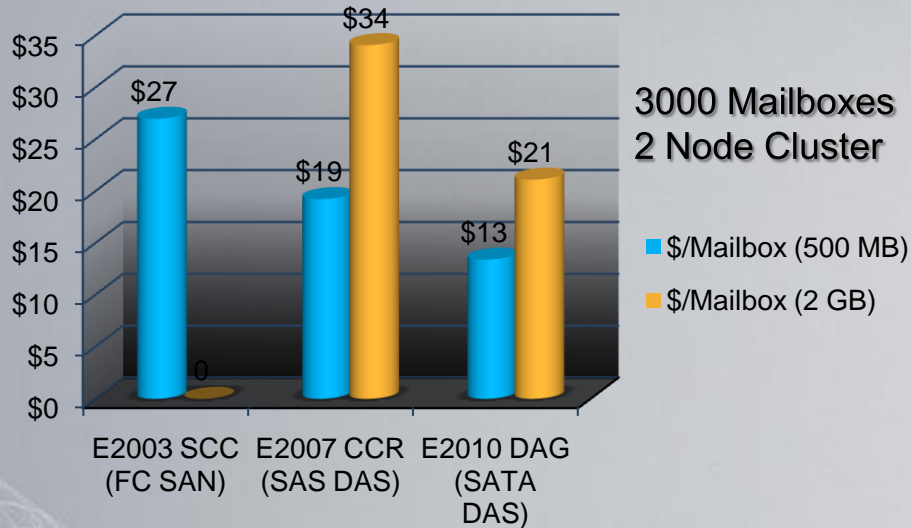 - HA Solution remains unchanged regardless of data volume size

- JBOD/RAID-less storage now an option

  - Requires 3+ DB Copies

Exchange 2010 Cost Savings

Server/Storage Capex $/Mailbox

3000 Mailboxes
2 Node Cluster

- $/Mailbox (500 MB)
- $/Mailbox (2 GB)

Values: E2003 SCC (FC SAN): $27, 0; E2007 CCR (SAS DAS): $19, $34; E2010 DAG (SATA DAS): $13, $21

Hardware Capex $/Mailbox

Double Server/Disk Failure Resiliency

24,000 Mailboxes

- $/Mailbox (2GB)

Values: 4 x 2 Node CCR 2 copies (RAID) / E2007 CCR (SAS DAS): $32; E2010 DAG (SATA DAS) 6 Node DAG 3 copies (JBOD): $8
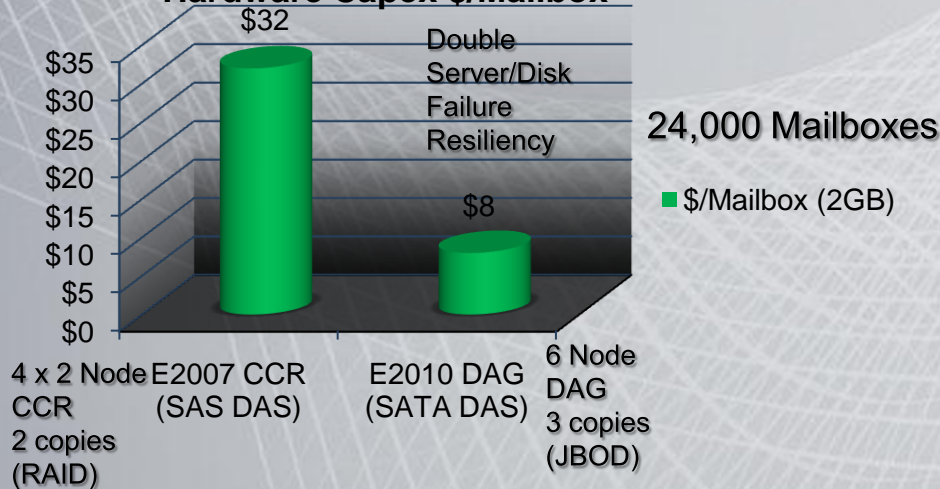
Storage Cost savings examples

- Storage flexibility
- Simplified management
- Simplified site resilience
- All server roles on one server (Small deployments)

Agenda

- Exchange 2010 High Availability Fundamentals

- High Availability Management

- Storage Improvements

- Transport Improvements

- High Availability Design Examples

Exchange Server 2010 Transport
Key Design Goals

- Increased availability
- Better administrative control
- Operational Excellence
- Lowering costs
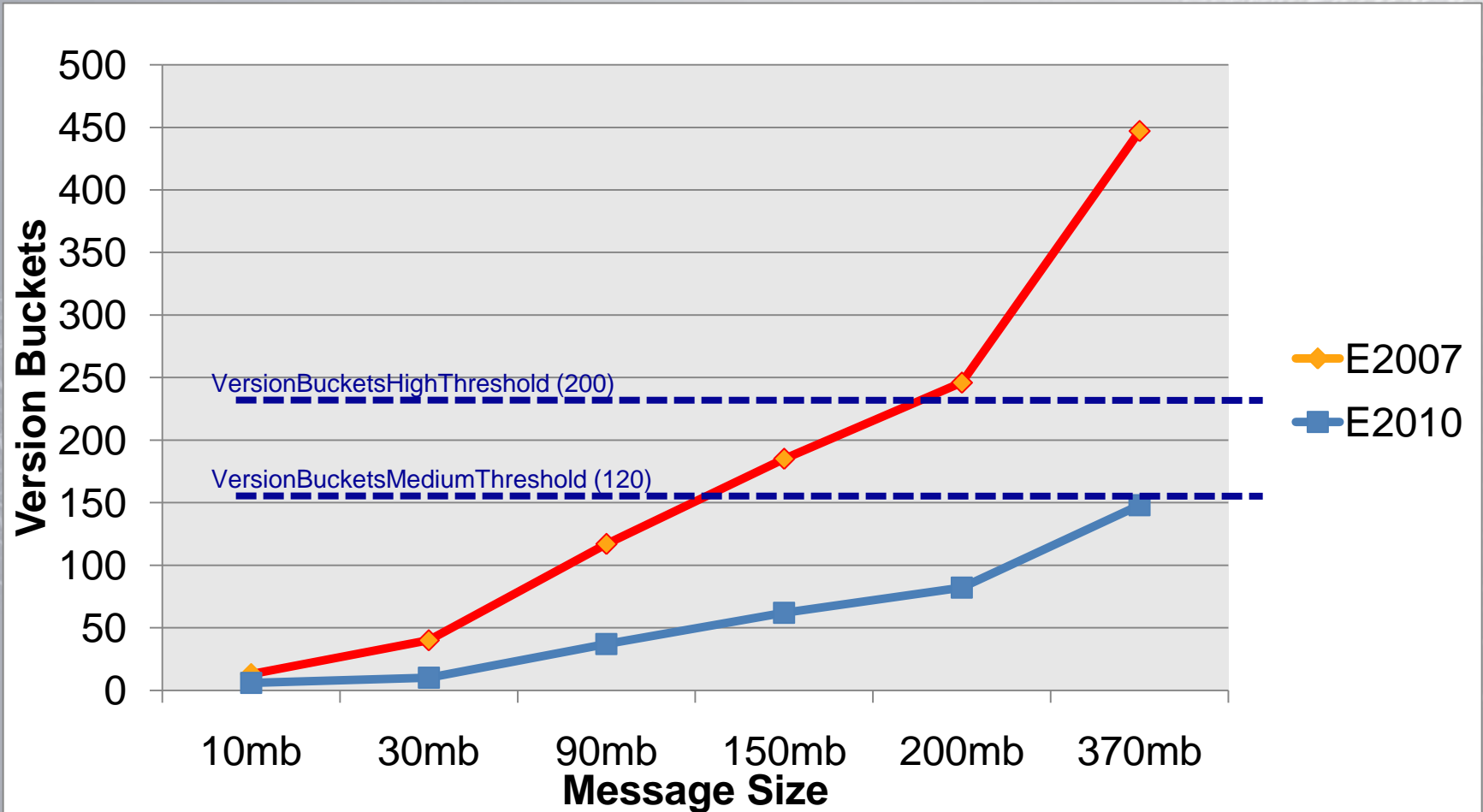
Increased Availability
Transport Performance - Improvements

- mail.que database improvements

    - Increased Extensible Storage Engine (ESE) page size to 32 KB
    - ESE Database (DB) page compression
    - ESE version store maintenance
    - Increase DB cache size and checkpoint depth

Increased Availability
Transport Performance

Reducing Version Bucket Resource Pressure
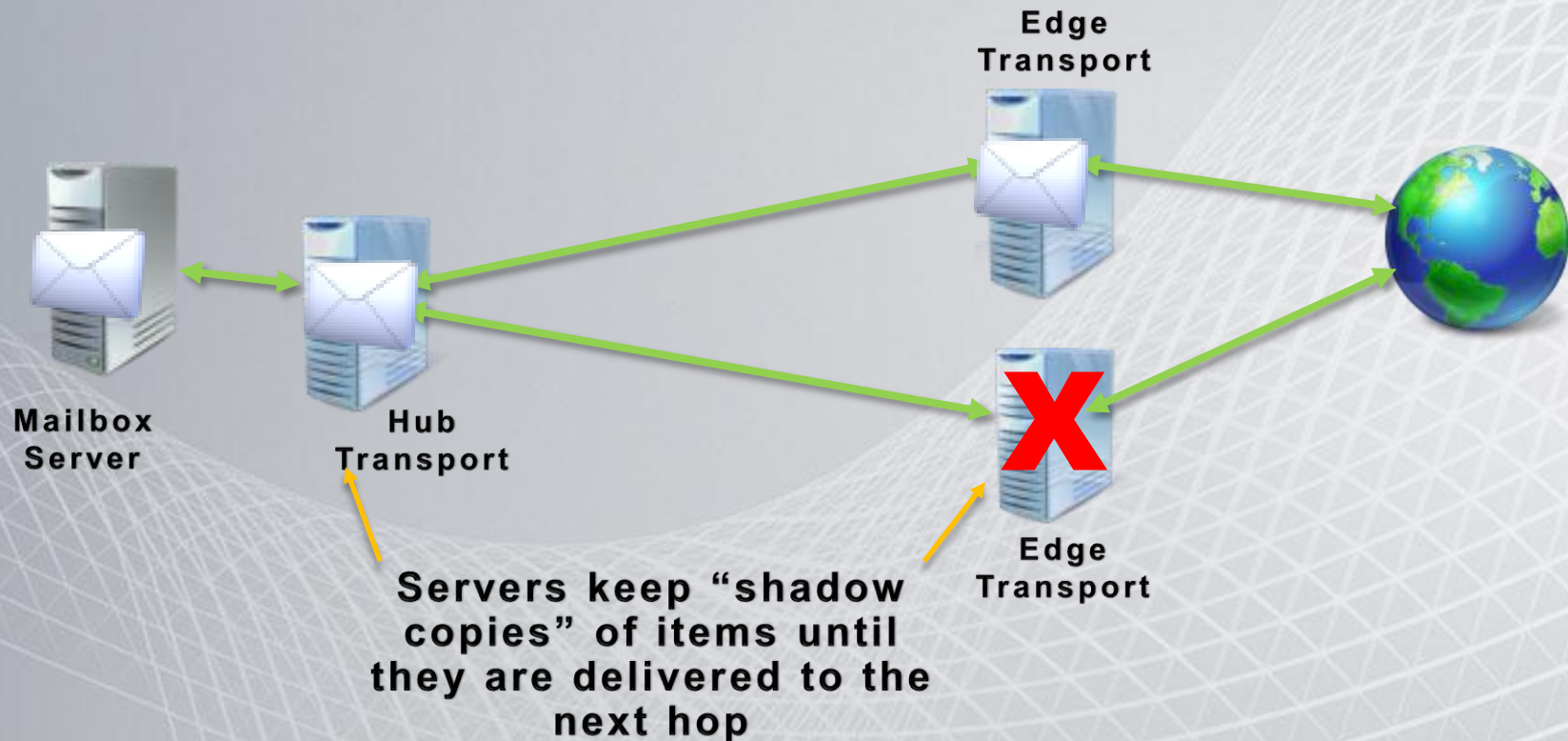
Increased Availability
Shadow Redundancy

- Goals
  - Increased reliability without increased hardware costs
  - Enabled by default
  - Shadow redundancy similar to transport dumpster
  - Data retained on previous hop until delivered
  - When failure in next hop detected, previous hop resubmits
  - SMTP extensions used (create little overhead)
  - Ellimination of RAID overhead

Improved Transport Resiliency
Automatic Protection Against Loss of Queued Emails Due to Hardware Failure

Edge Transport

Mailbox Server

Hub Transport

Servers keep "shadow copies" of items until they are delivered to the next hop

Edge Transport

- Simplifies Hub and Edge Transport Server upgrades and maintenance

Increased Availability
How does it work?

Hub

1

Edge1   **Edge2**

2

**Foreign
MTA**
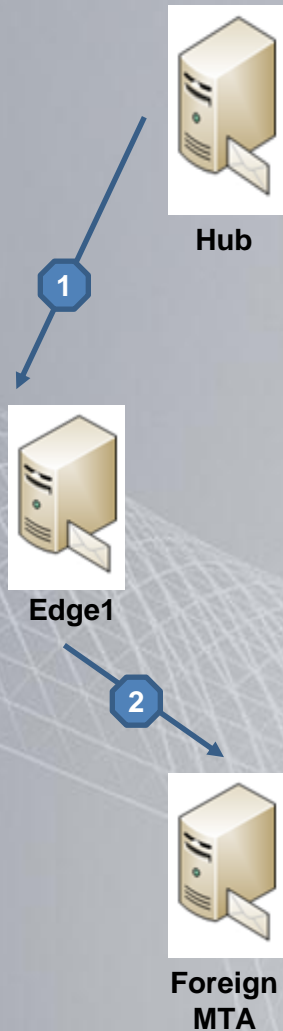
1. **Hub delivers message to Edge1**
   Detects that Edge1 supports Transport
   redundancy through XSHADOW verb
   *Hub moves message to shadow queue and stamps
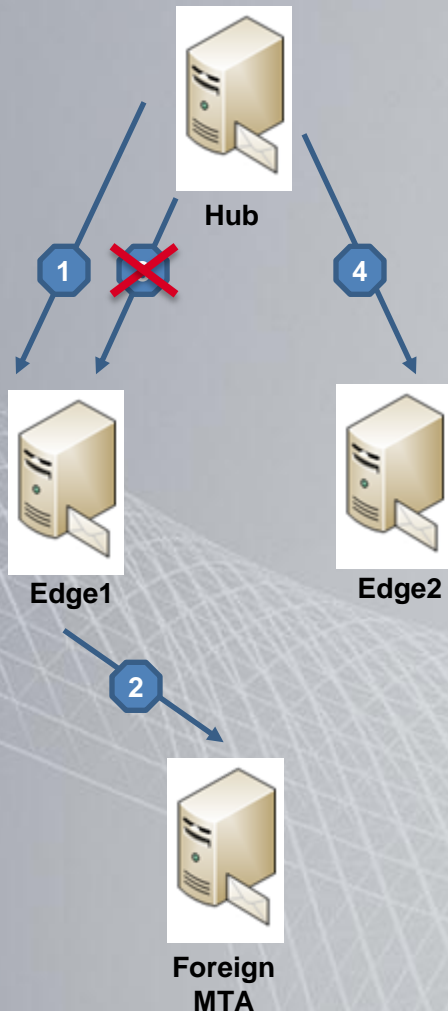   Edge1 as current, primary owner*

2. **Edge1 receives message
   (becomes "primary owner")**
   *Edge1 delivers message to next hop
   Edge1 updates discard status of the
   message indicating delivery complete
   to foreign MTA*

40

Increased Availability
How does it work?

3. Success: Hub (shadow) queries Edge1 (primary) for expiry status
Hub issues XQDISCARD command (next SMTP Session),Edge1 checks local discard status and responds with list of messages considered delivered
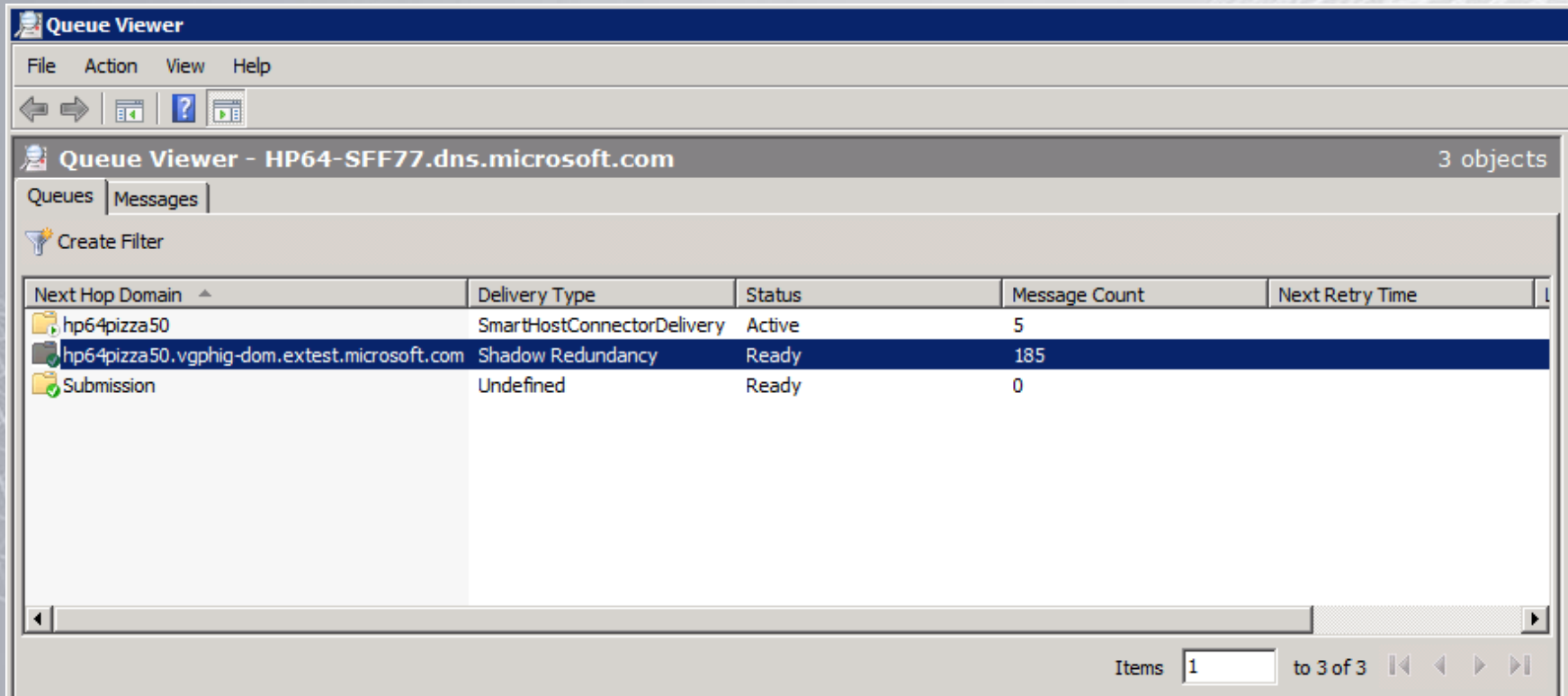*→ Hub deletes messages from its shadow queue*

4. Failure: Hub (shadow) queries Edge1 (primary) discard status and resubmits
Hub opens SMTP session, issued XQDISCARD command (heartbeat)—if Hub can't contact Edge1 within timeout, resubmits messages in shadow queue—resubmitted messages are delivered to Edge2 (go to #1)

Hub

1

4

Edge1

Edge2

2

Foreign MTA

41

Shadow Redundancy
Queue Viewer

Increased Availability
Automated Service Recovery

- Exchange Server 2007 memory resource pressure results in decreased service availability
  - Exchange Server 2010 implements signal to generate Dr. Watson report (determine cause of failure) and restarts
  - Exchange Server 2010 Alert can send to System Center to further analyze resource pressure
- Exchange Server 2007 queue database corruption results in downtime until administrator can perform manual recovery
  - Exchange Server 2010, transport will detect queue database corruption, move/delete DB, and continue operation
    - Shadow redundancy provides data resiliency

Shadow Redundancy & Transport ASR

demo

Online Move Mailbox

Limit User Disruption During Mailbox Moves And Maintenance

E-Mail Client

Client Access Server

Mailbox Server 1
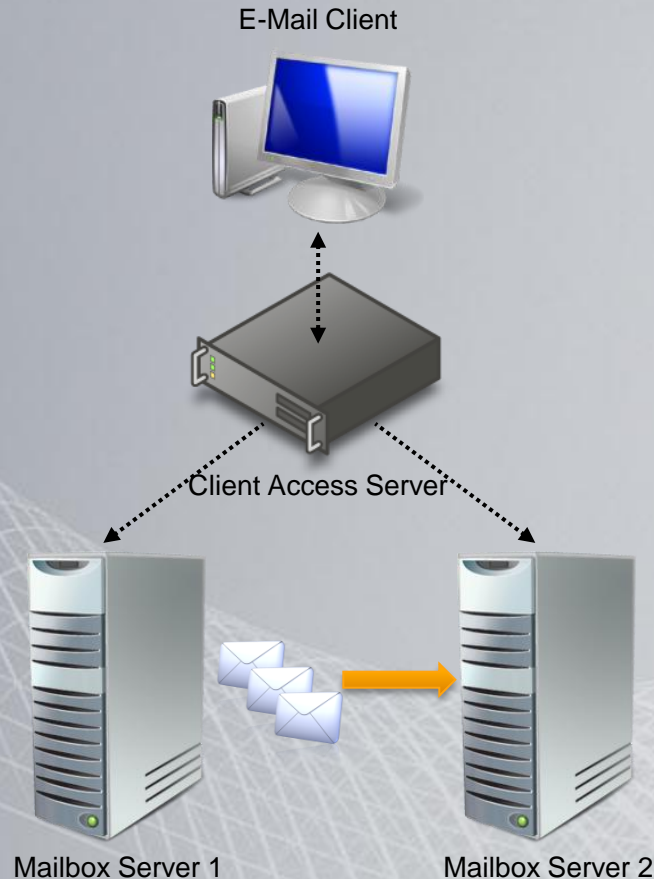
Mailbox Server 2

- Users remain online while their mailboxes are moved between servers
  - ✓ Sending messages
  - ✓ Receiving messages
  - ✓ Accessing entire mailbox
- Administrators can perform migration and maintenance during regular hours
- Also can be used to migrate users from on-premise server to Exchange Online

- Exchange 2010 & Exchange 2007 SP2 Online
- Exchange 2003 Offline

High Availability Management

Online Mailbox Move

demo

Agenda

- Exchange 2010 High Availability Fundamentals

- High Availability Management

- Storage Improvements

- End-to-End Availability Improvements

- High Availability Design Examples

High Availability Design Example
Branch Office or Smaller Deployment

Hardware Load Balancer

CAS/HUB/ MAILBOX 1

CAS/HUB/ MAILBOX 2

DB1 → DB1

DB2 ← DB2

DB3 → DB3

Mailbox servers in a DAG can host other Exchange server roles

2 server configurations, should always use RAID
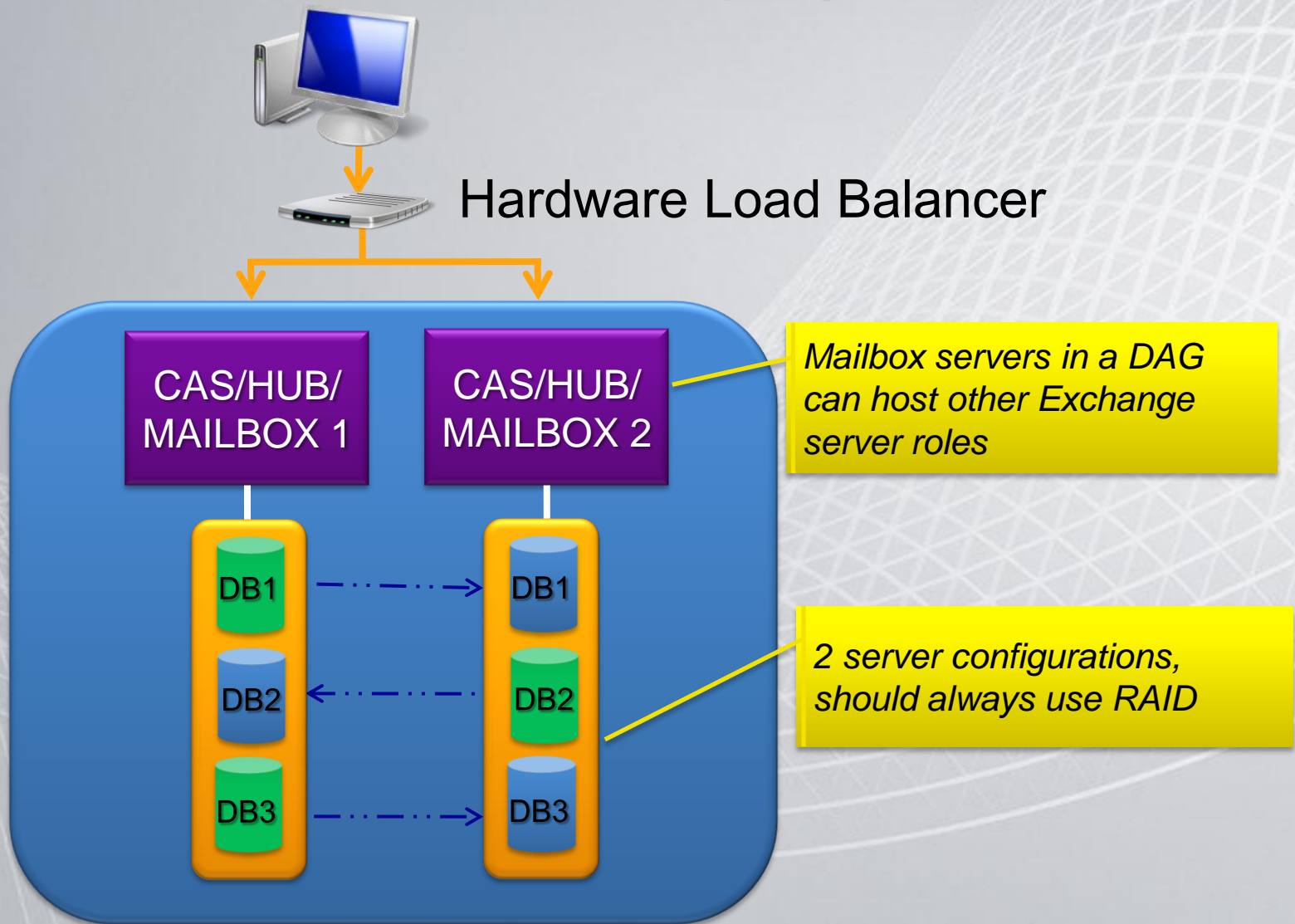
High Availability Design Example
Double Resiliency

AD: Copenhagen

CAS NLB Farm

- ~~Upgrade Site~~ Single Server 1
- ~~Server 2~~ 4-Node fails
- ~~SCHA Copies~~ ServerA upgrade is done
- ~~2BOTv copies idle~~ Active > physical Copies

Mailbox Server 1

Mailbox Server 2

Mailbox Server 3

Mailbox Server 4

DB1 DB2 DB3

DB4 DB5 DB6

DB7 DB8 DB1

DB2 DB3 DB4

DB5 DB6 DB7

DB8 DB1 DB2

DB3 DB4 DB5

DB6 DB7 DB8

Database Availability Group (DAG)

High Availability for Other Server Roles

	Client Access	• Hardware load balancer (recommended) or Windows Network Load Balancing (NLB)
	Hub Transport	• No special configuration required (load balancing and failover is automatic)
	Edge Transport	• Use DNS round robin, Multiple MX records
	Unified Messaging	• Configure IP gateway to point to more than one UM server

Site Resilience
Datacenter Failover: Basics

- Customers can <u>evolve</u> to site resilience
- Standalone → Local Redundancy → Site Resilience
  - Keep extending the DAG
- No single subnet requirements
- Normal administration remains unchanged
- Disaster recovery usually requires manual intervention
- Standby datacenter is "always live"

Takeaways

- With each release, our goals are to make highly available Exchange
  - Easier and cheaper to deploy
  - Easier and cheaper to manage
  - Support better SLAs with faster and more granular recoveries
  - Improve site resiliency support

Thankyou

John Ferris

john.ferris@microsoft.com

http://blogs.technet.com/ferris/

Microsoft *TechNet*