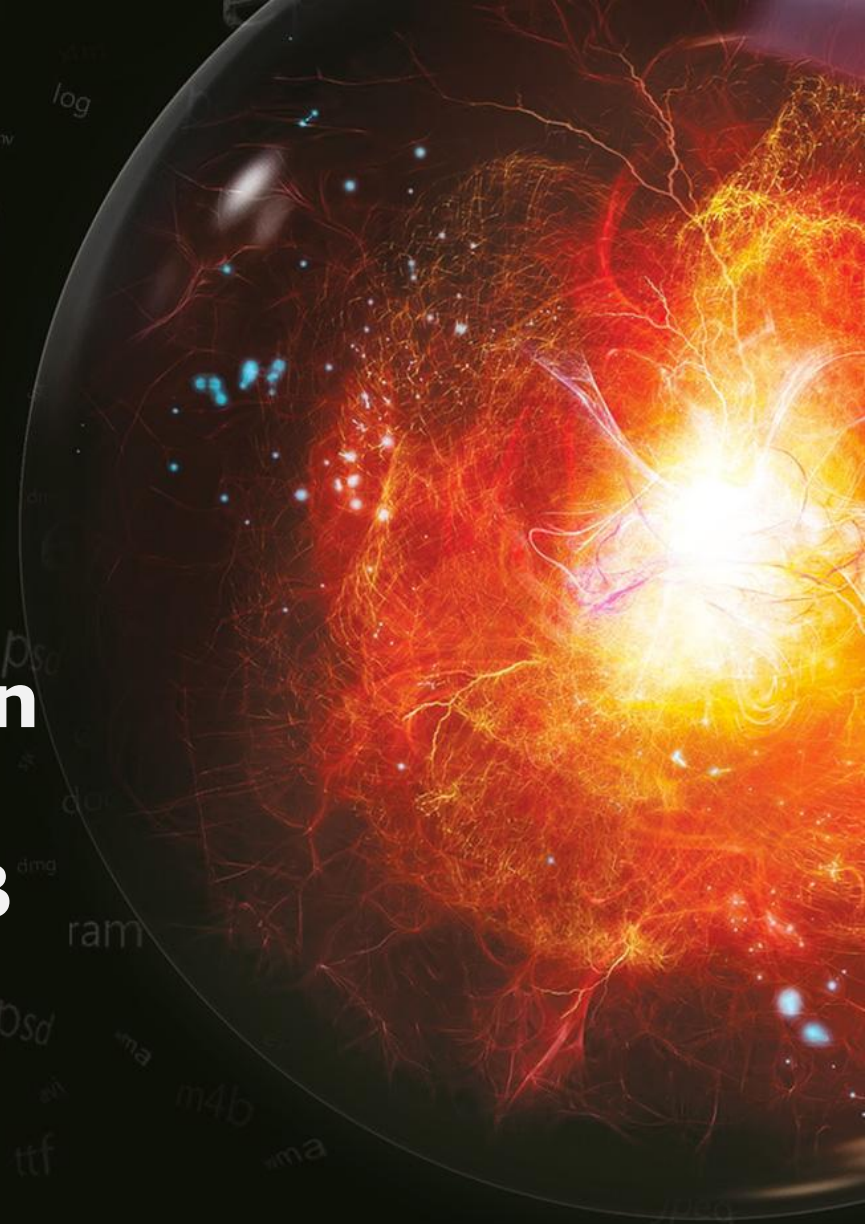


SQL Server Energy 2008

Erleben Sie eine neue Form der Energie.

Failure is not an option
24x7 OLTP database
management for VLDB

=tg= Thomas H. Grohser
Thomas.Grohser@bwin.org



Failure is not an option

but the language this presentation is given in is

```
SELECT count(*)  
      FROM Audience  
      WHERE language <> 'deutsch'
```

SELECT * FROM =tg=

- **SQL Server DBA since 1994**
 - First SQL Server Version ever used 4.21
 - First Log Shipping with failover on SQL Server 6.0
 - First SQL Server Cluster on SQL Server 6.5 (NT 4.0 + Wolfpack)
 - First time > billion rows in 1 table on SQL Server 7.0
 - First time 100% availability for more than 2 years in a row: SQL Server 2000 (936 days to be exact)
 - First time OLTP long distance database mirroring SQL Server 2005
 - Hundreds of possibilities for SQL Server 2008 ... can't wait to raise the bar again
- **Focus on SQL Server Infrastructure Architecture and Implementation**
- **Close Relationship with Microsoft**
 - SQLCAT (SQL Server Customer Advisory Team)
 - SCAN (SQL Server Customer Advisory Network)
 - TAP (Technology Adoption Program)
- Close relationship with Hardware Vendors (Focus IA64)
- **Active PASS member and PASS Summit Speaker**
- Newest project: www.sqlserver-hwguide.com
- **WARNING: I recently read a book about power point! (sorry ☺)**



Failure is not an option

Agenda

- The Mission
- The Solution
 - Standardizing
 - Zero data loss
 - High availability
 - Scale up
- The Details
 - SQL Server jobs
 - Log Shipping
 - Replication

24x7 OLTP Database
Management for VLDB

Failure is not an option

The Mission

- VLDB – A database that needs attention it's not size alone
- SLA
 - Zero data loss & 100% transactional consistency on financial transactions
 - 99.99x% availability @ 24 x 7
 - 450.000+ SQL Statements per second
 - Assumed worst case scenario: full datacenter failure with complete data loss within the datacenter
- Budget: unlimited (not kidding)

Failure is not an option

The Solution

- Standardize everything
- Work by the book
- Have some clever guys at hand
if the book runs out of pages

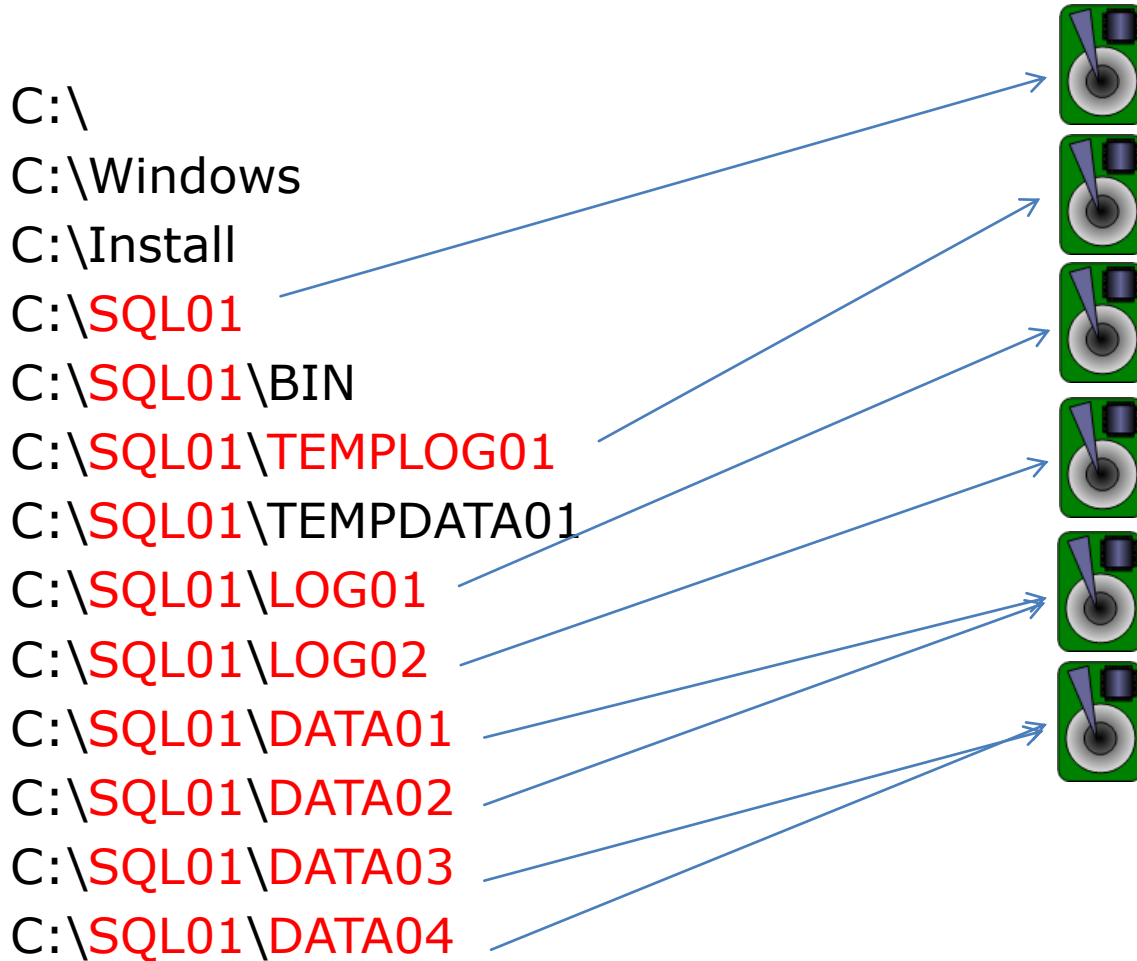
Failure is not an option

Standardizing

- Operating System
 - Version, Edition, Service Pack, Patch Level
- File System and Disks

Failure is not an option

File System Sample

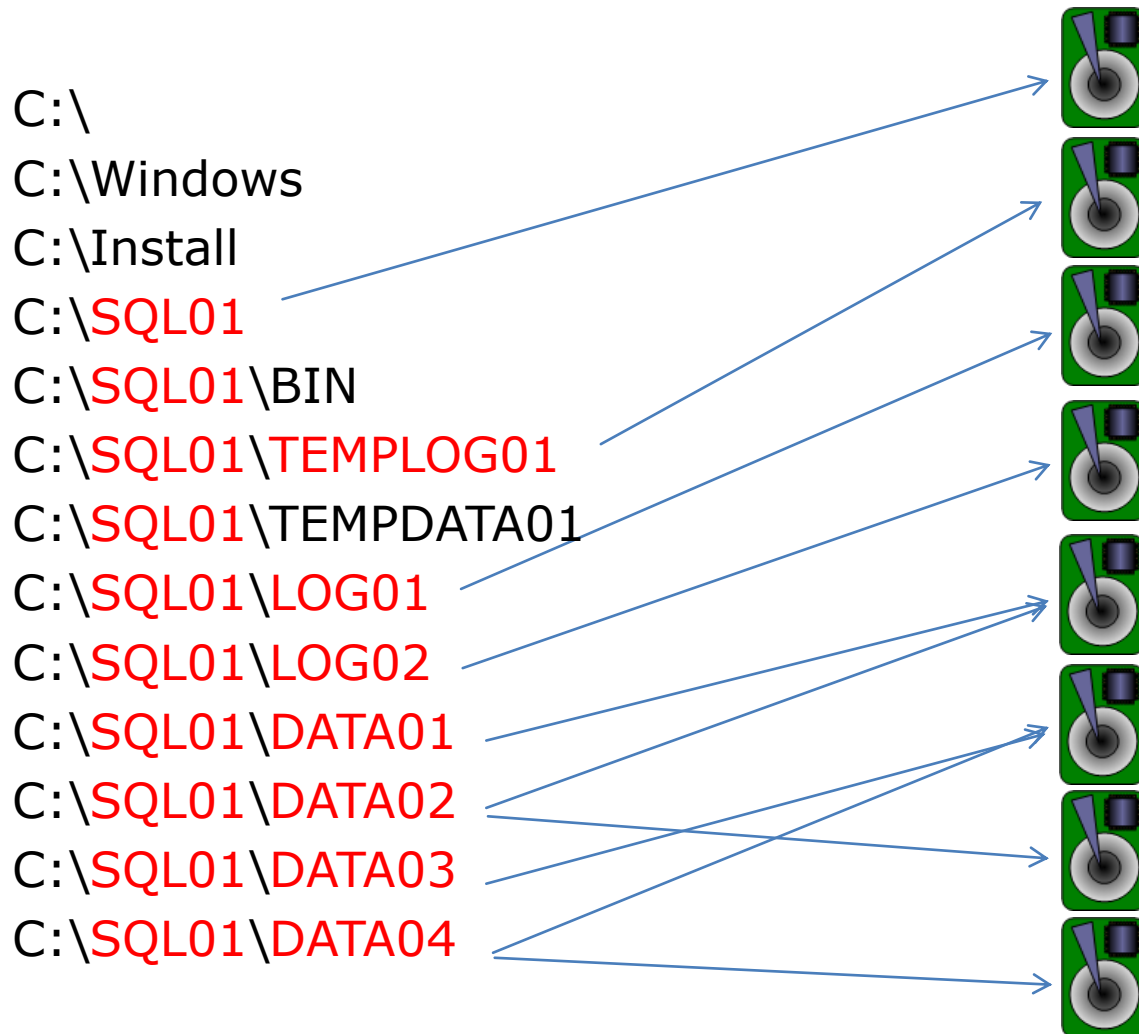


Mountpoint

Directory

Failure is not an option

File System expansion



Failure is not an option

File System settings

- Stripeseize 64/128/256 kB
depending on storage
- Partition alignment 64/128 kB
depending on storage
- NTFS Cluster size 64 kB
- 100% write cache
- 0% read cache

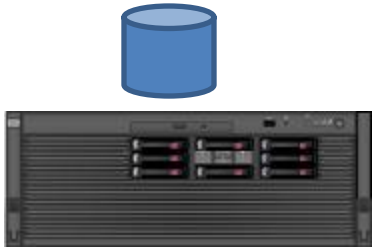
Failure is not an option

Standardizing

- Operating System
 - Version, Edition, Service Pack, Patch Level
- File System and Disks
- SQL Server
 - Version, Edition, Service Pack, Patch Level
- Network
 - Separate network for data and backup
 - IP Schema
- Documentation, Documentation,

Failure is not an option

Zero data loss



- Redundant NIC
- Redundant Power Supply
- Data files on SAN
(RAID 1/0 Multipath/2 Fabrics)
- Transaction log files on RAID 101

Failure is not an option RAID 101



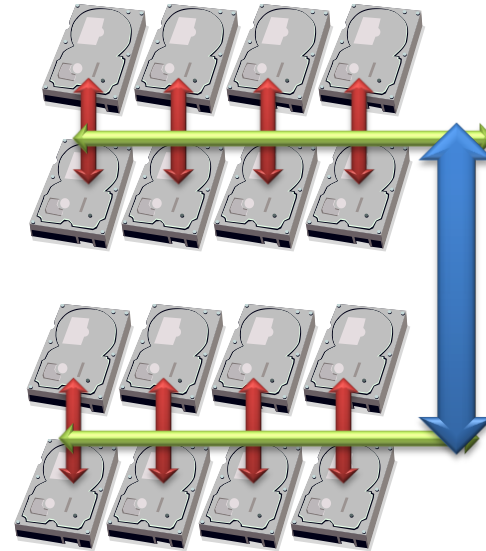
RAID Controller

RAID Controller

HW RAID 1

HW RAID 0

SW RAID 1



Failure is not an option

Zero data loss



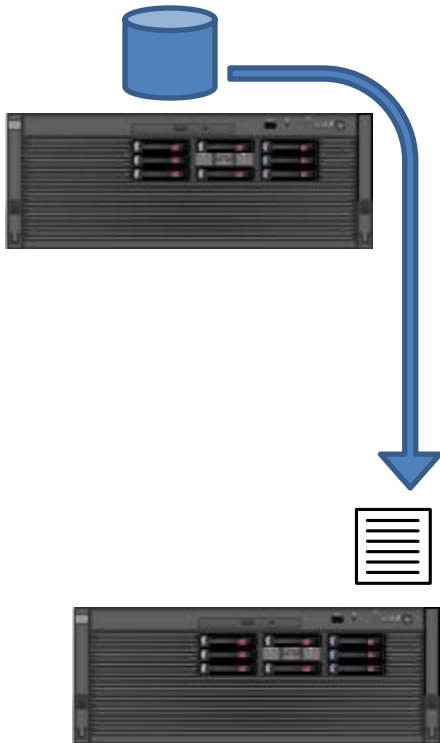
- Redundant NIC
- Redundant Power Supply
- Data files on SAN
 - (RAID 1/0 Multipath/2 Fabrics)
- Transaction log files on RAID 101

Availability: 0,00%

Data loss: 100,00%

Failure is not an option

Zero data loss



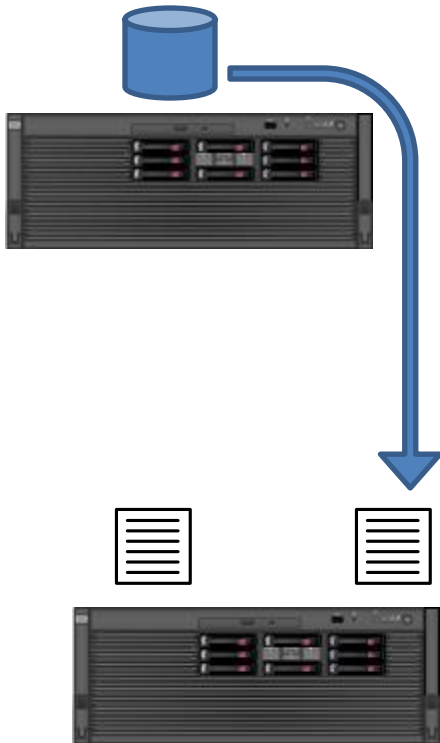
- Full backup every 24 h

Availability: 0,00%

Data loss: 100,00%

Failure is not an option

Zero data loss

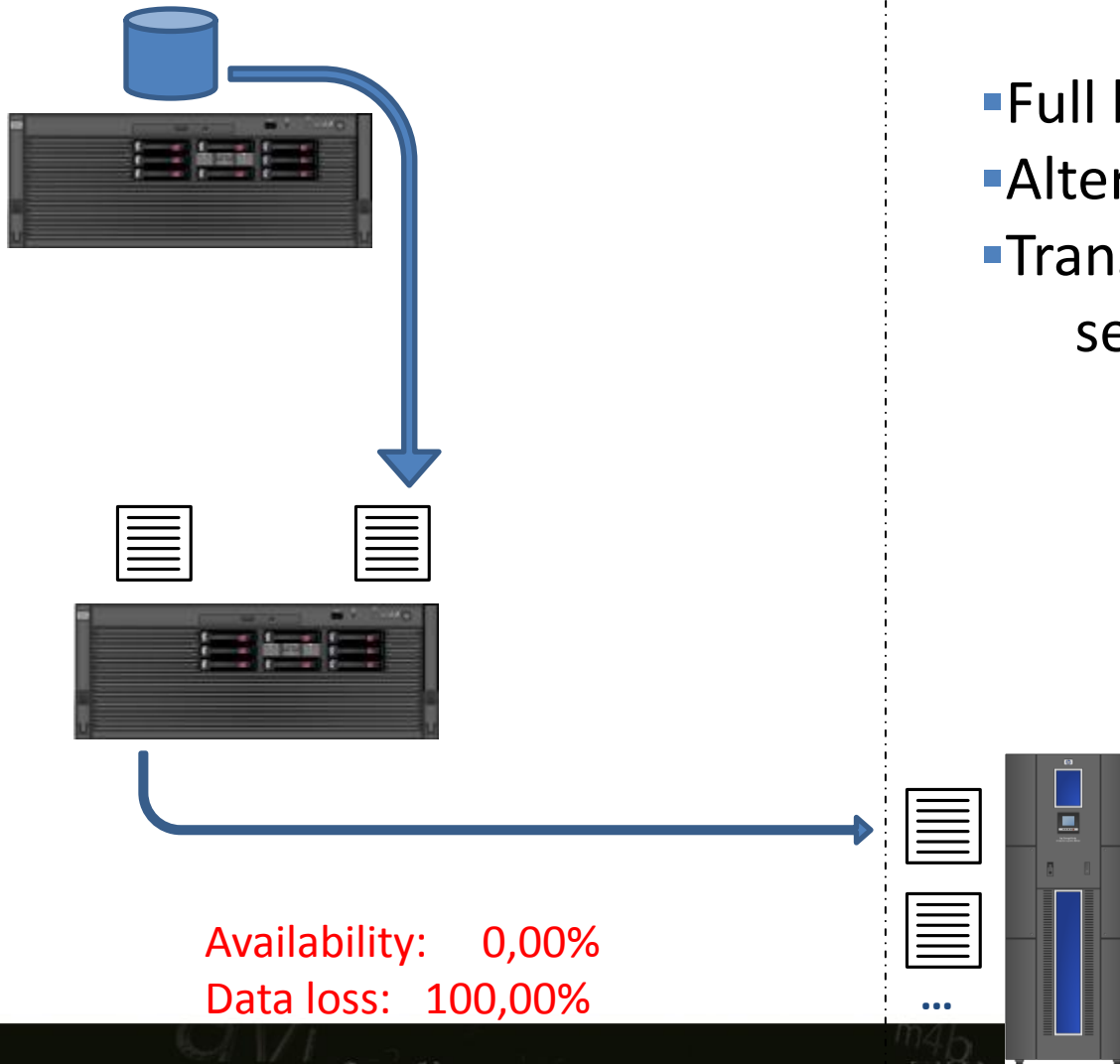


- Full backup every 24 h
- Alternating files

Availability: 0,00%
Data loss: 100,00%

Failure is not an option

Zero data loss

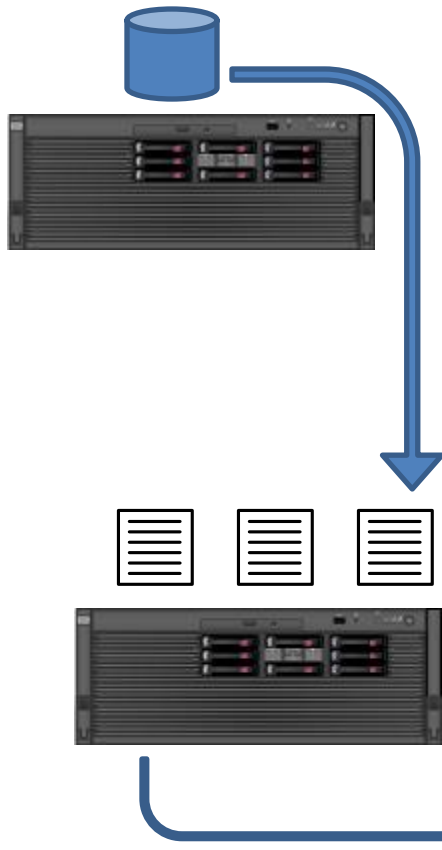


- Full backup every 24 h
- Alternating files
- Transfer of files to tape in second location

Availability: 0,00%
Data loss: 100,00%

Failure is not an option

Zero data loss

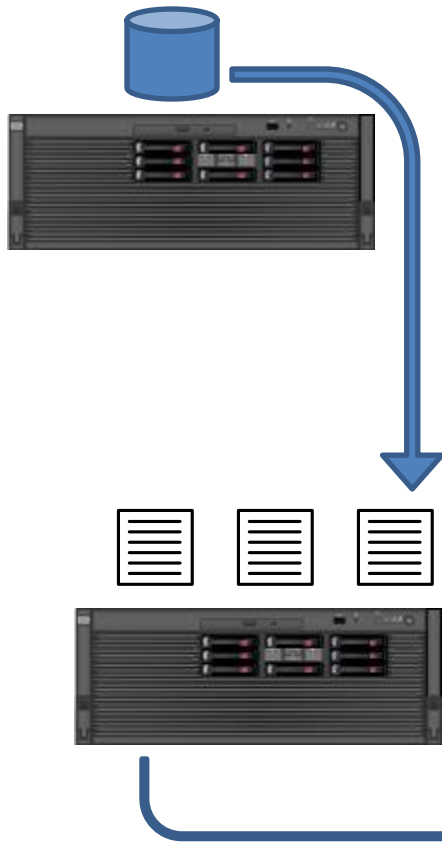


- Full backup every 24 h
- Alternating files
- Transfer of files to tape in second location
- Verify of backups (daily)

Availability: 0,00%
Data loss: 100,00%

Failure is not an option

Zero data loss

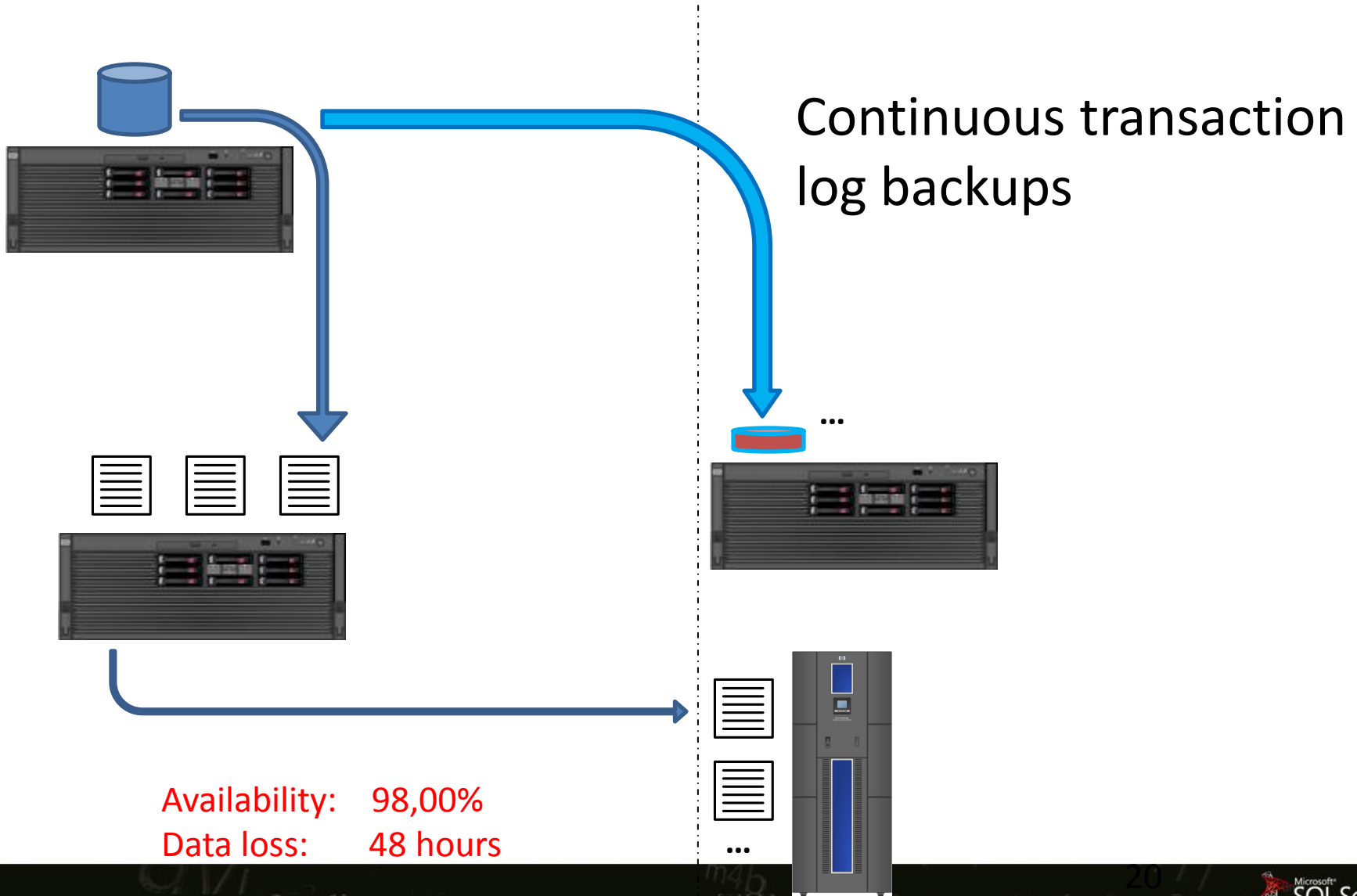


- Full backup every 24 h
- Alternating files
- Transfer of files to tape in second location
- Verify of backups (daily)
- Verify of tapes

Availability: 98,00%
Data loss: 48 hours

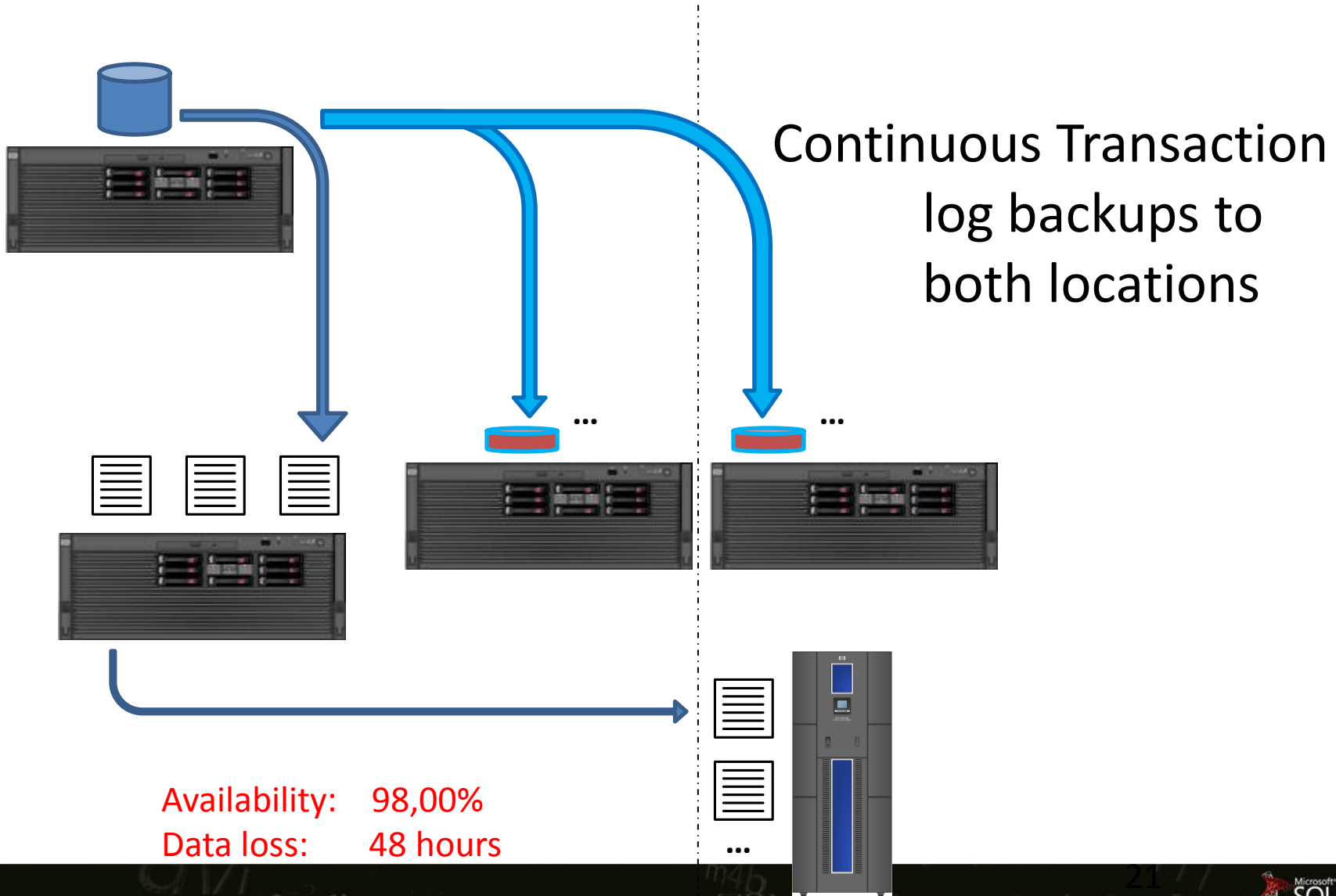
Failure is not an option

Zero data loss



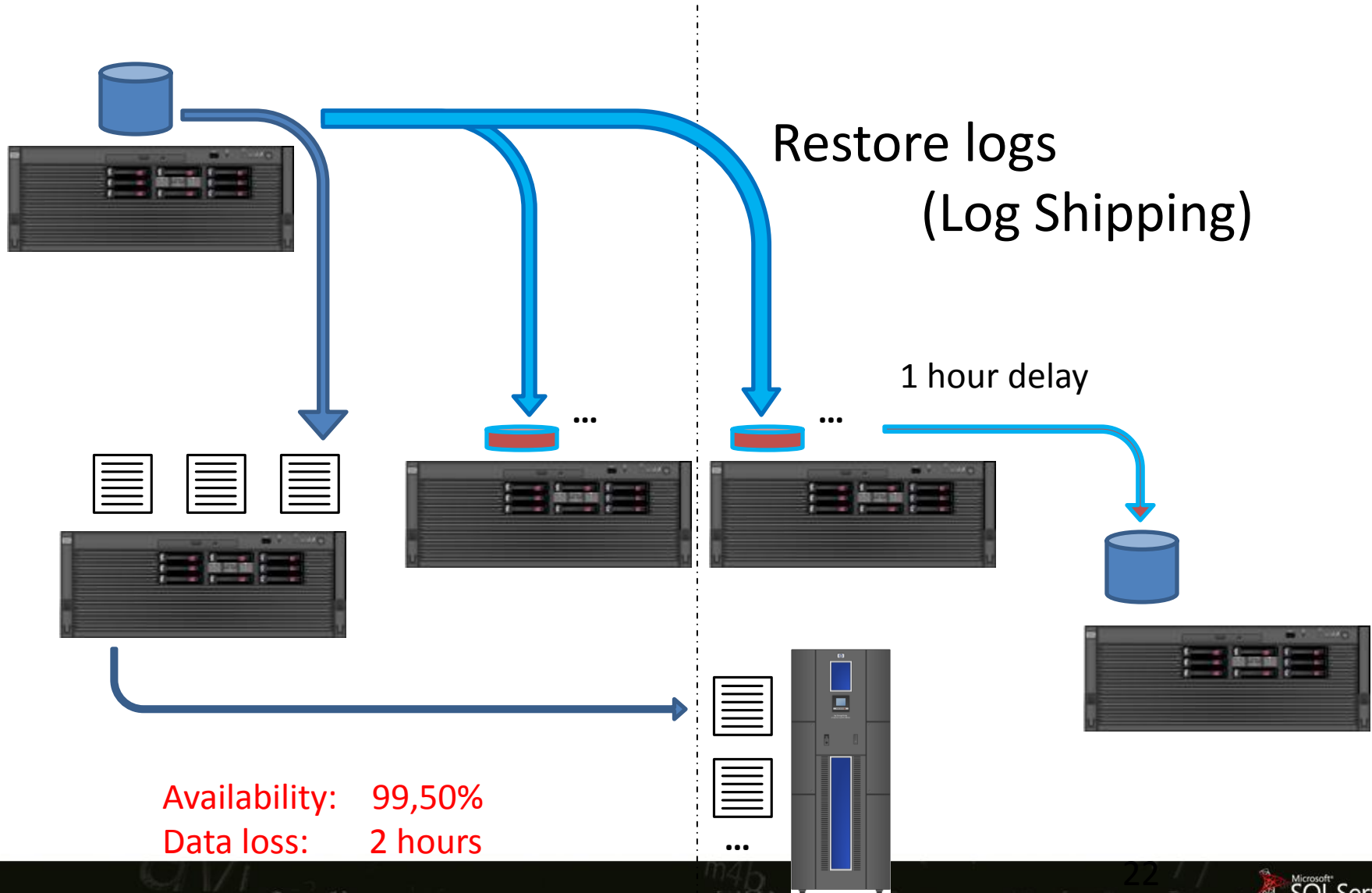
Failure is not an option

Zero data loss



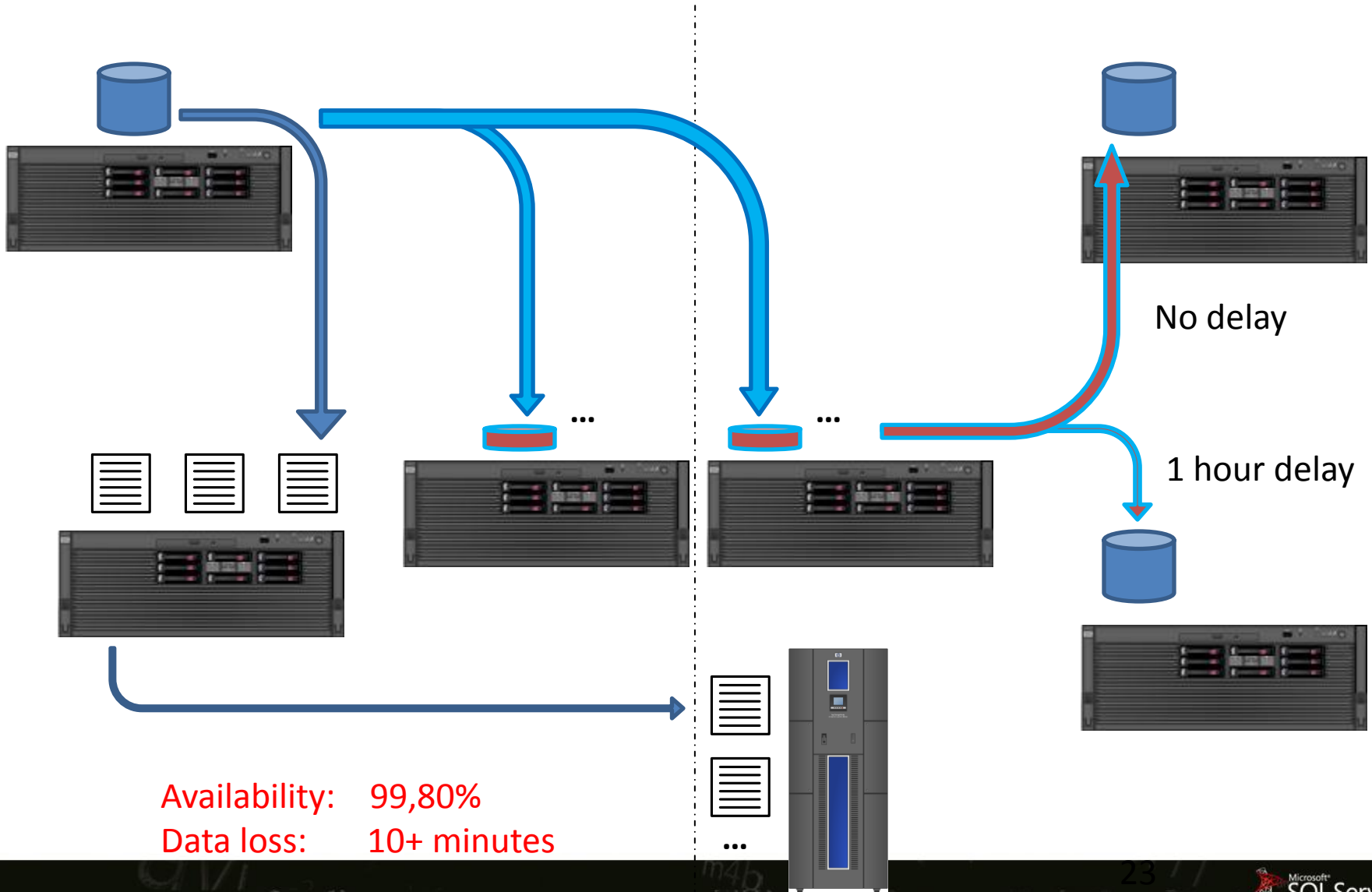
Failure is not an option

Zero data loss



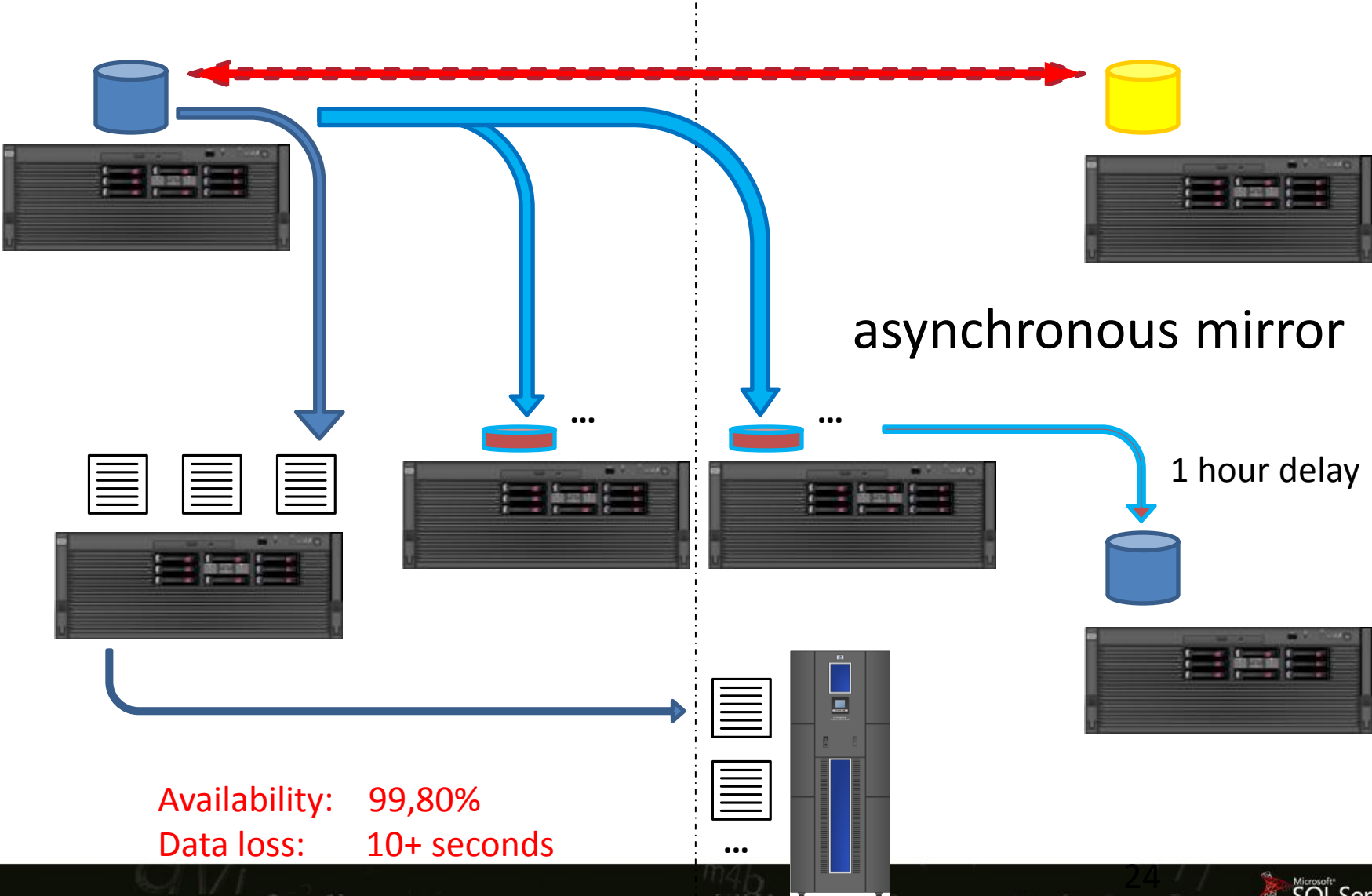
Failure is not an option

Zero data loss



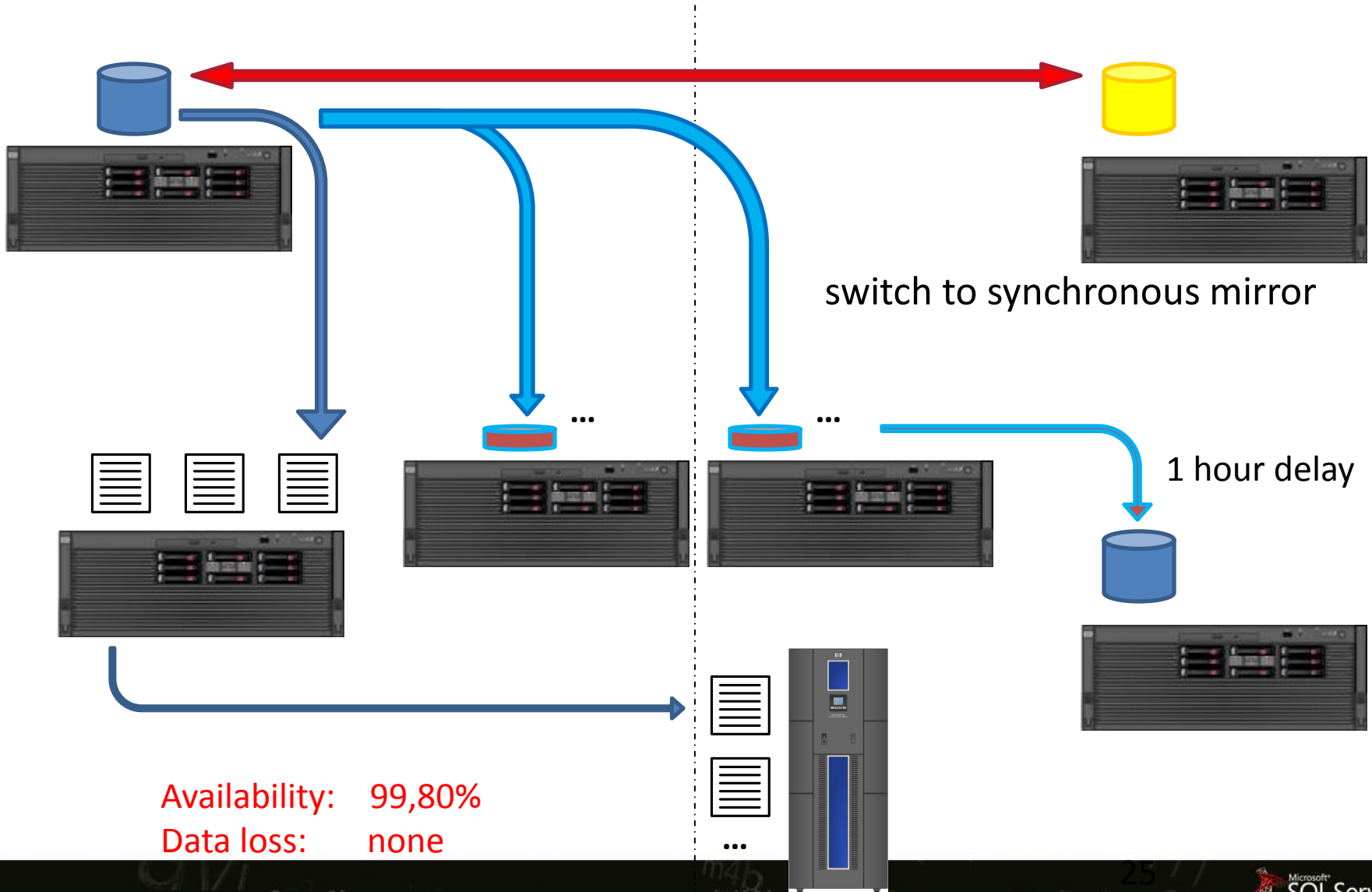
Failure is not an option

Zero data loss



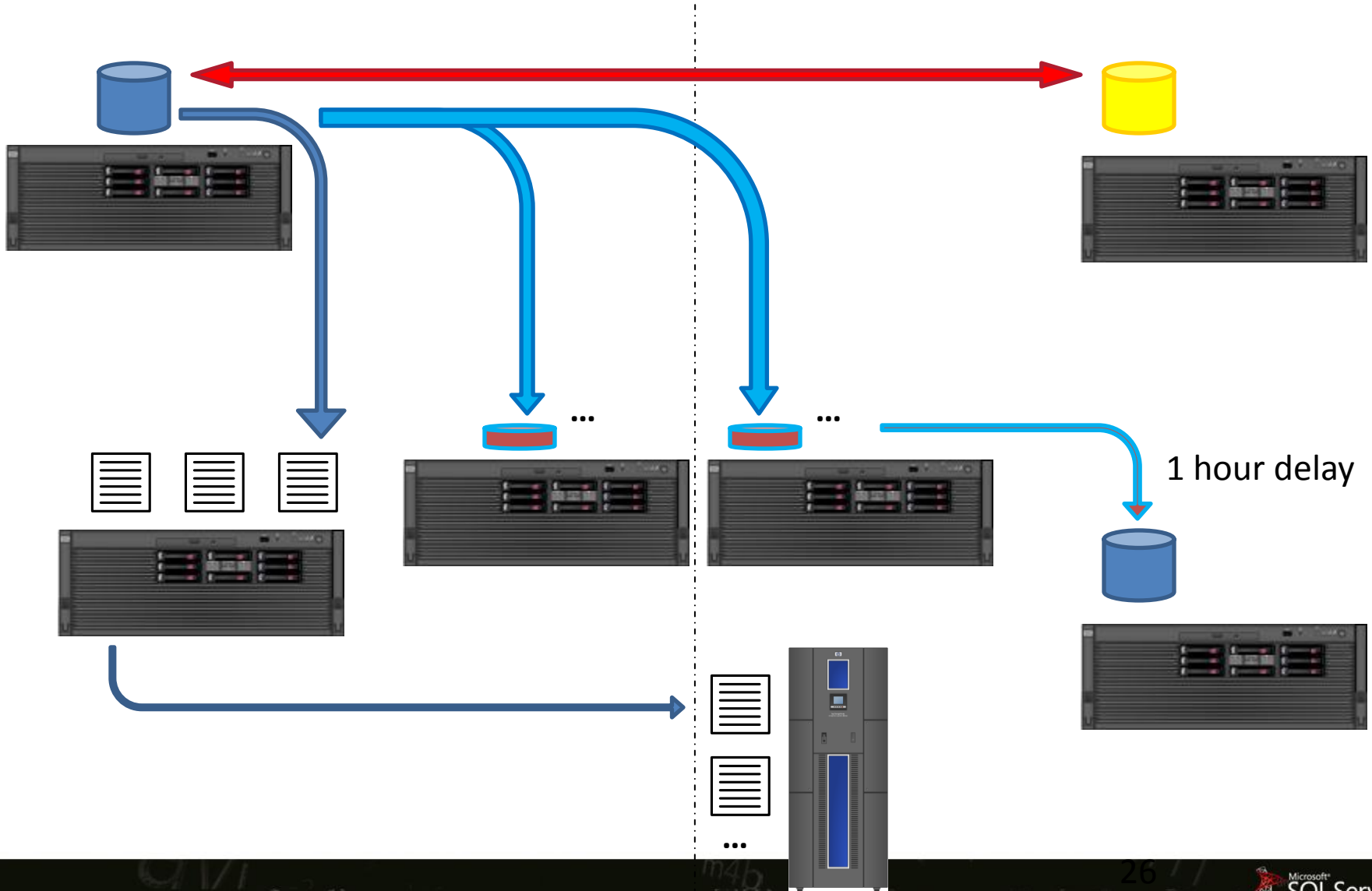
Failure is not an option

Zero data loss

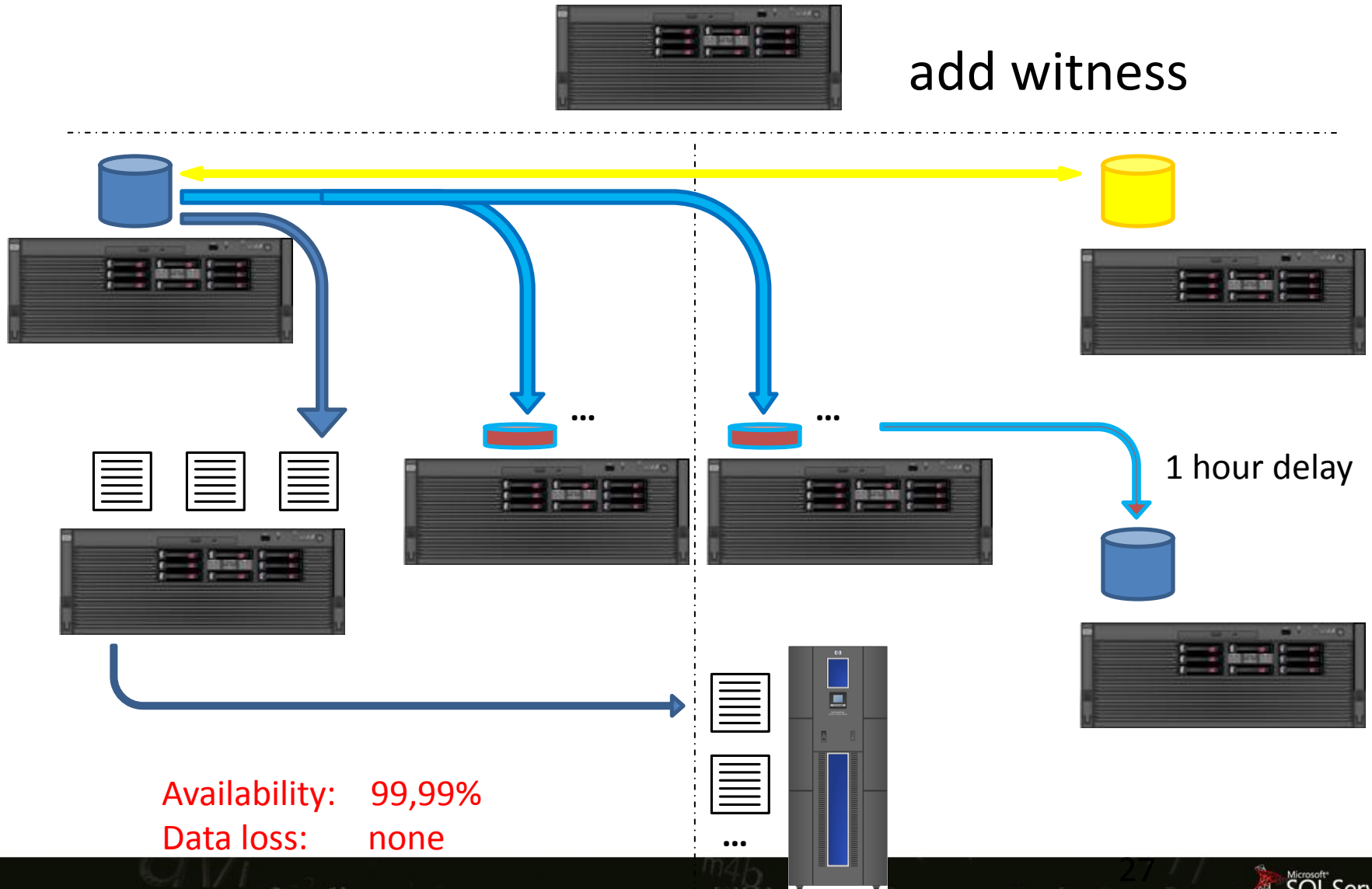


Failure is not an option

Availability



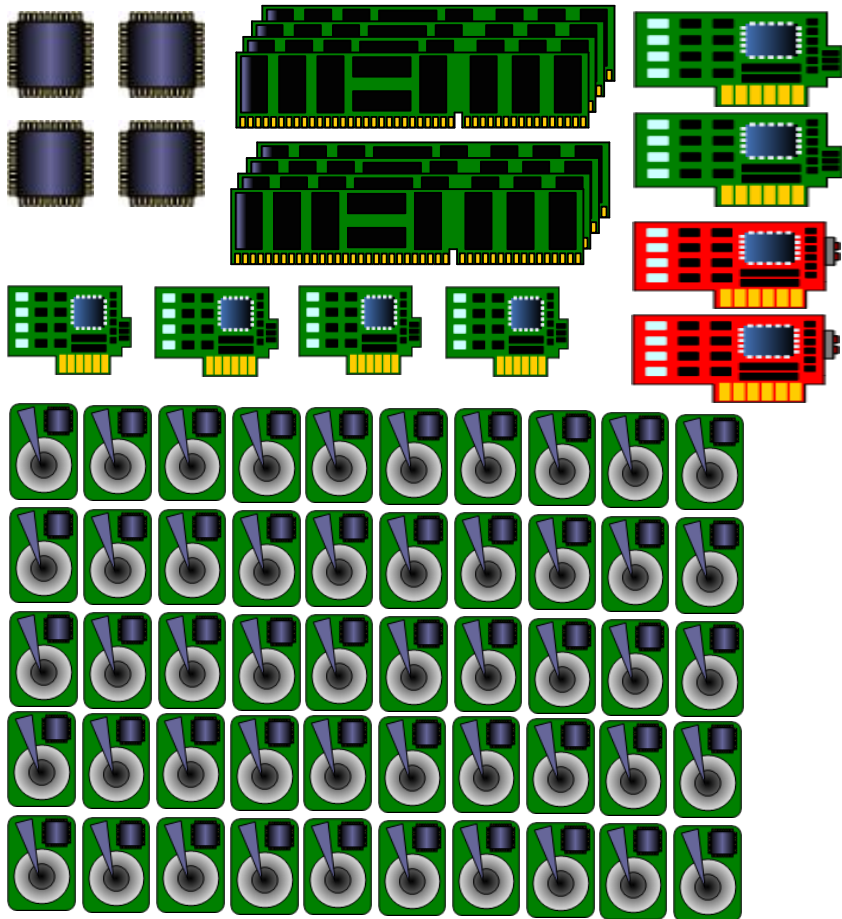
Failure is not an option Availability



Availability: 99,99%
Data loss: none

Failure is not an option

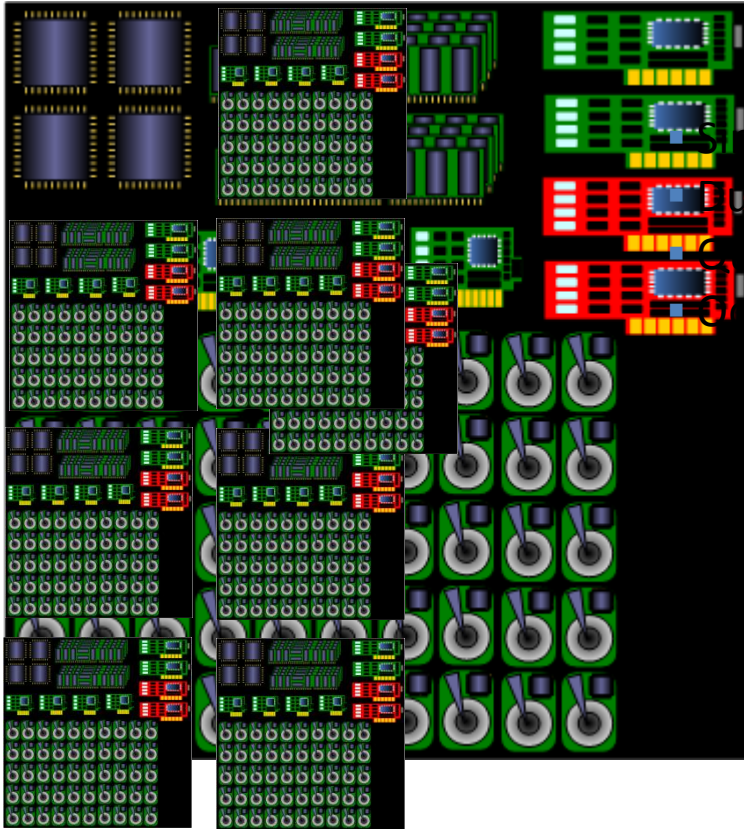
Scale Up – Single NUMA node



- 4 x Dual Core ITANIUM 2 CPUs
24 MB cache each
- 64 GB memory
- 4 x dual port 1 Gb/s network card
- 2 x dual port HBA (4Gb/s)
- 2 x P800 RAID controller
- 50 x 72 GB 15kRPM SAS disks
- SAN storage as needed
n x 512GB (on 64 spindles each)

Failure is not an option

Scale Up



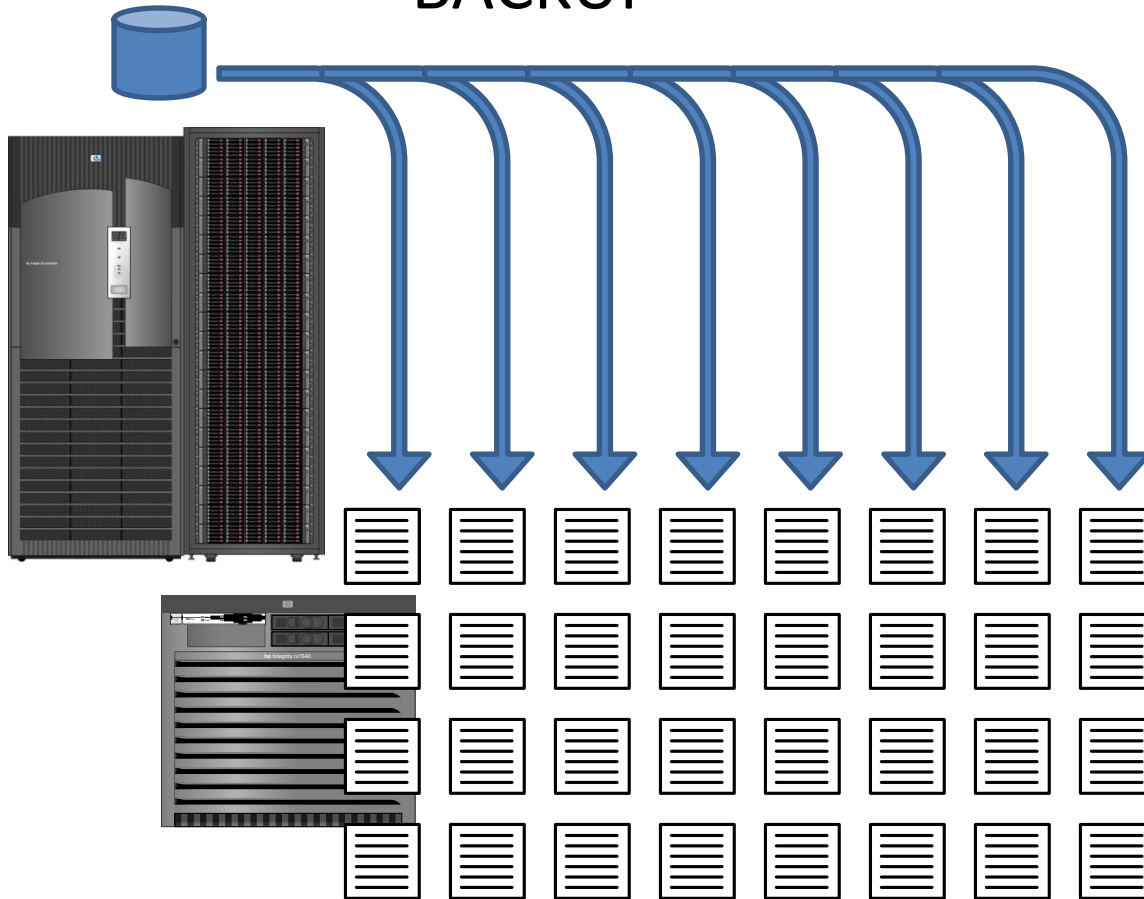
	cores	GB	disks	NIC	HBA
Single	8	64	50	8	4
Dual	16	128	100	16	8
Quad	32	256	200	32	16
Octal	64	512	400	64	32

Almost linear scaling

Failure is not an option

Scale Up

BACKUP



- Use eight parallel one GB/s sec network interface cards (one physical network, eight subnets)


- Use 32 parallel backup files each on a separate set of spindles with aligned partitions

- Transfer four files per network interface card

Failure is not an option

Scale Up

SQL Server	IP Address	Network Mask
Network Card 1	192.168.1.2	255.255.255.0
Network Card 2	192.168.2.2	255.255.255.0
File Server	IP Address	Network Mask
Network Card 1	192.168.1.1	255.255.255.0
Network Card 2	192.168.2.1	255.255.255.0



Failure is not an option

Scale Up

```
BACKUP DATABASE MyVLDB
TO
    DISK='\\192.168.1.1\backup\MyVLDB_1.bak',
    DISK='\\192.168.2.1\backup\MyVLDB_2.bak'
WITH
    BLOCKSIZE = 8192
```

- Use Jumbo Frames if you can (+100%) with about 9016 bytes frame size
- With 4088 byte frames `BLOCKSIZE = 65536` gives better results

Failure is not an option

The Details

- Mirroring (not yet) as simple to manage as clustering
- SQL Server jobs
- Log Shipping
- Replication

Failure is not an option Magic Job



Jobs, Logins, ...

SQL
Agent
Magic job

A purple starburst icon with a white outline.

Jobs, Logins, ...

SQL
Agent
Magic job

A purple starburst icon with a white outline.

Failure is not an option Magic Job



Jobs, Logins, ...

SQL
Agent
Magic job

A purple starburst icon with a white outline.

Jobs, Logins, ...

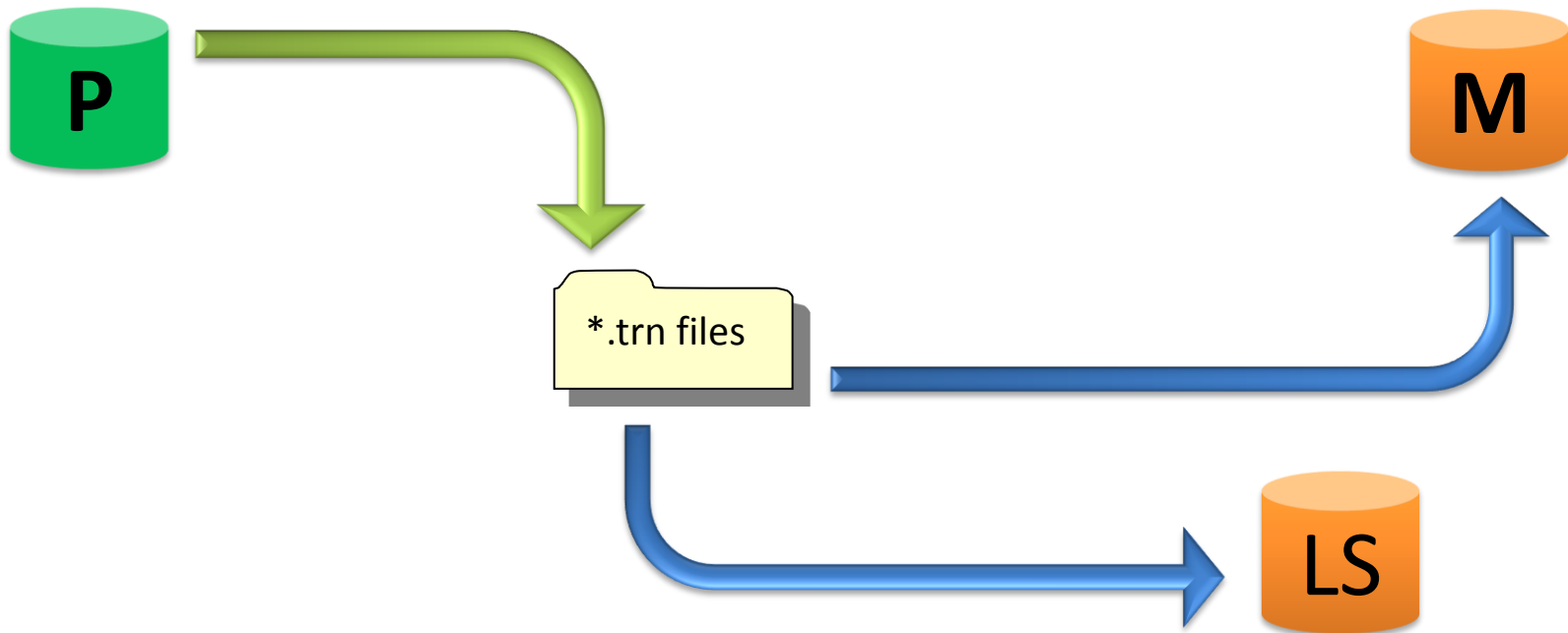
SQL
Agent
Magic job

A purple starburst icon with a white outline.

Failure is not an option

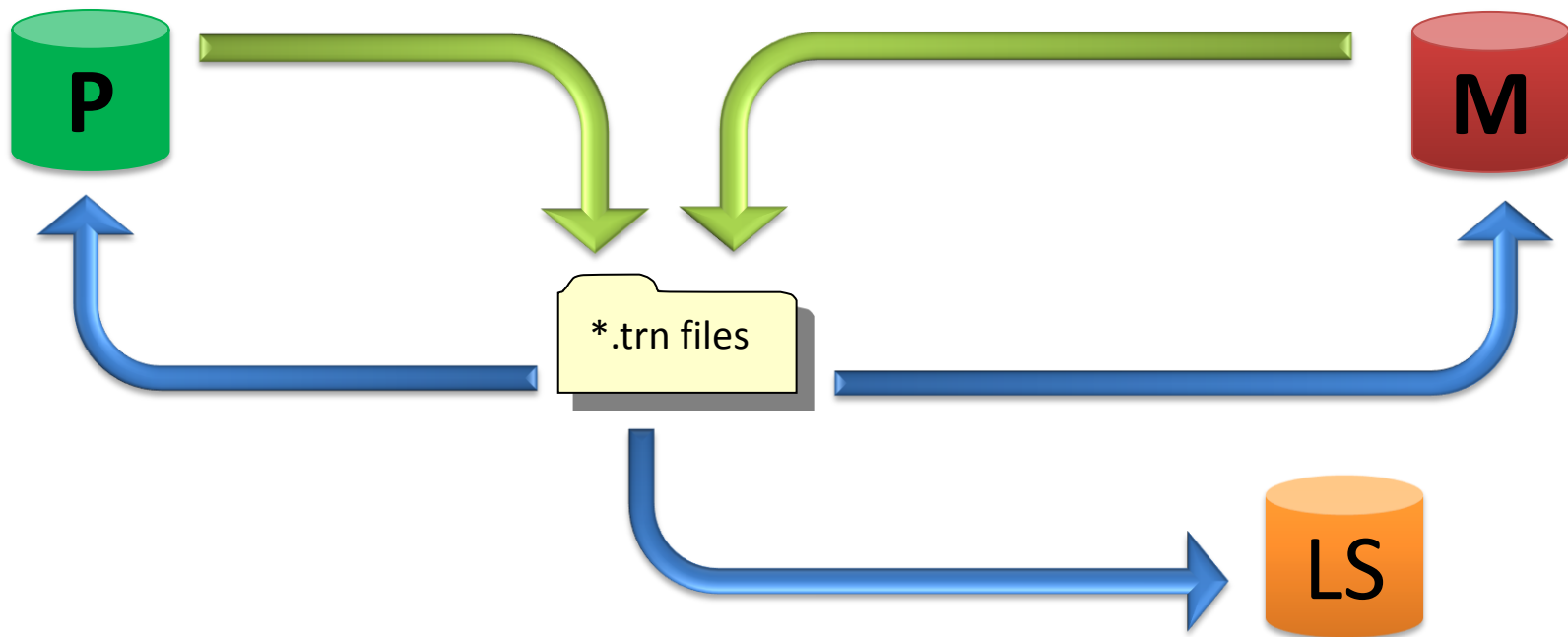
Log Shipping

- Create the backup jobs on the principal server
- Create the restore jobs on the log shipping server
- Create the restore jobs on the mirror server



Failure is not an option Log Shipping

- Establish mirror and failover the database
- Create the backup job on the mirror server
- Create the restore job on the principal server



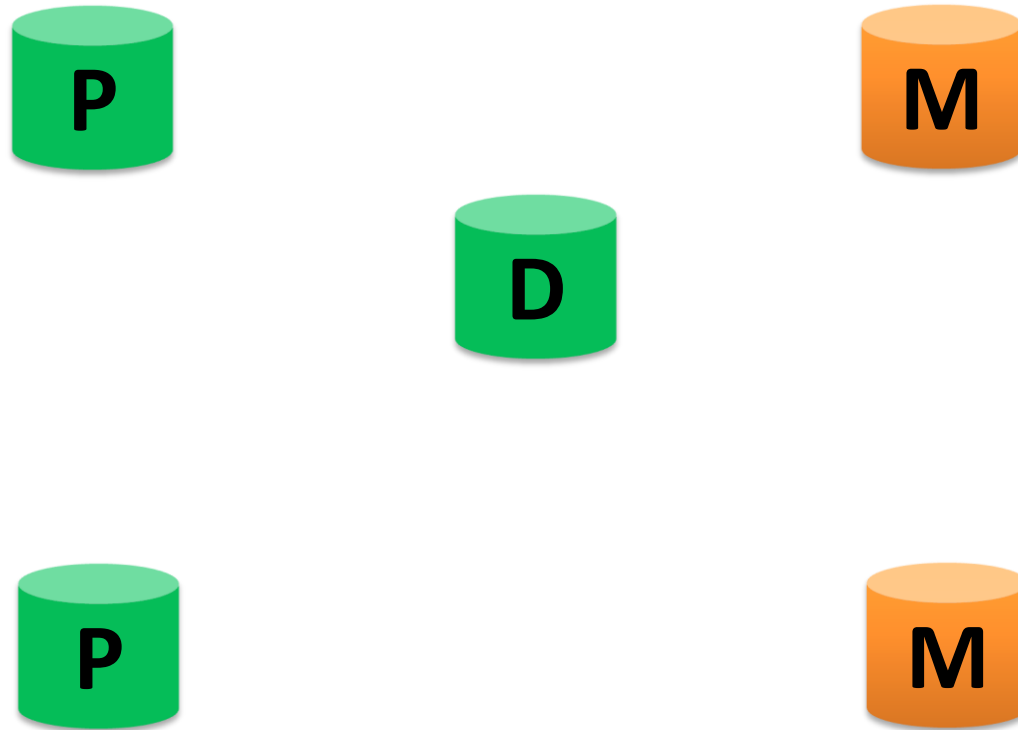
Failure is not an option

Replication

- Transactional Replication
 - from a mirrored publisher database
 - (2005 and 2008 supported)
 - (distributor must be 2008)
 - into a mirrored subscriber database
 - (2005 and 2008 possible)
 - (distributor can be 2005 or 2008)

Failure is not an option

Replication



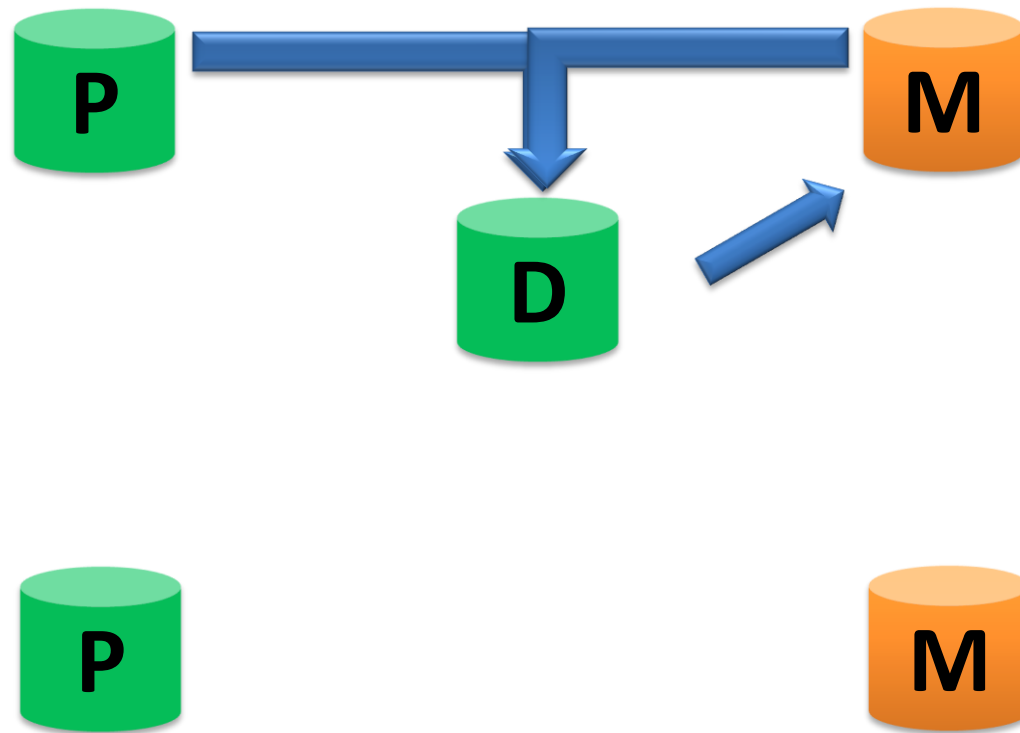
Failure is not an option

Replication

- Publishing from a mirrored database
(SQL Server 2005 or 2008)
 - Publisher : Create the publication as always
 - Distributor (SQL Server 2008):
In the Agent Profile you must add a
`-PublisherFailoverPartner`
entry

Failure is not an option

Replication

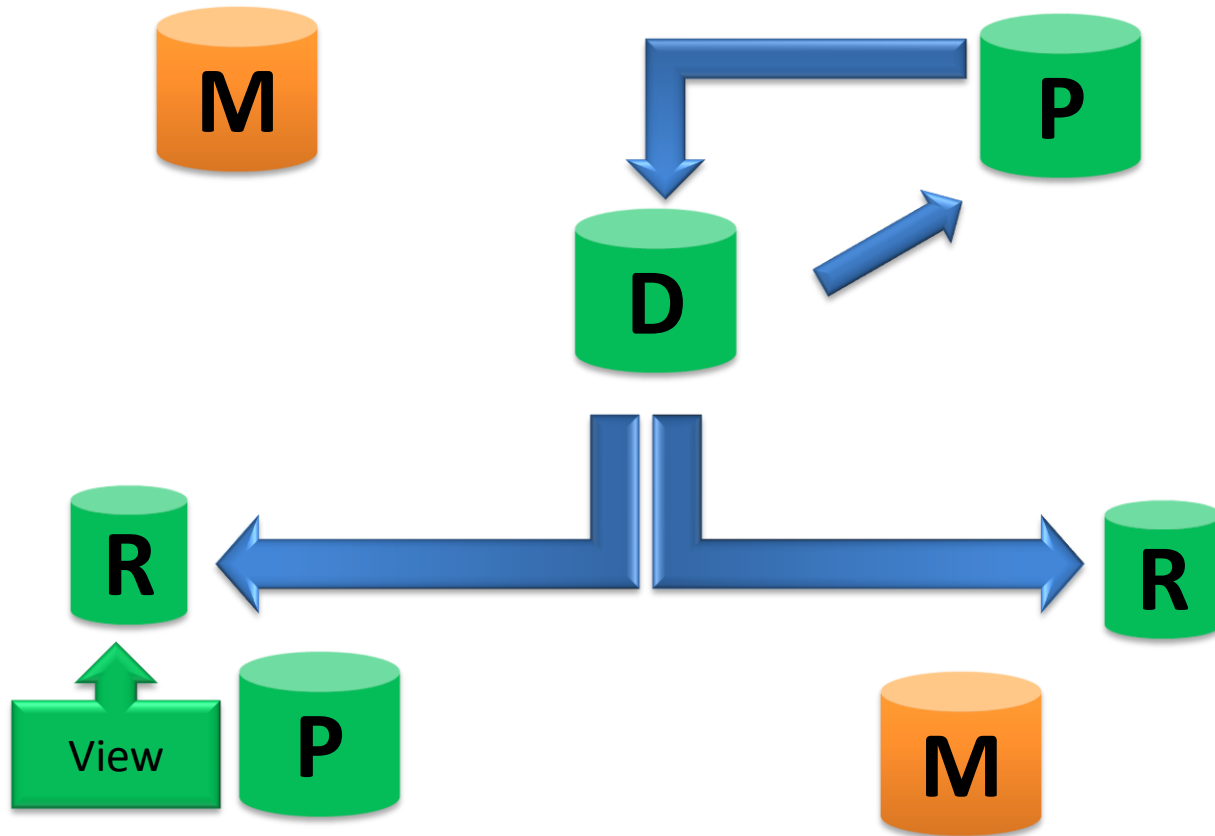


Failure is not an option

Replication

- Subscribing into a mirrored database (2005 or 2008)
- Publisher /Distributor: Create the publication as always
- Subscriber:
 - Create a helper database on each server
(principal and mirror)
(they must have the same name !!!)
 - Create two subscription one into each of the two helper databases
 - On principal database create a view to the replicated data

Failure is not an option Replication



Use **P**;
`CREATE VIEW dbo.MyTable as`
`SELECT * FROM R.dbo.MyTable;`

