



European PASS Conference 2009

Microsoft SQL Server Users Conference & Expo

April 22 - 24, 2009 • Neuss, Germany • Swissôtel



Users working together

Connect. Share. Learn.



European PASS Conference 2009

Microsoft SQL Server Users Conference & Expo

April 22 – 24, 2009 • Neuss, Germany • Swissôtel

Failure is not an option

24x7 VLDB administration

=tg= Thomas H. Grohser, Senior Database Engineer,
bwin Interactive Entertainment AG



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
- First time to return more than a peta byte on query results to clients without failure SQL Server 2008
(so far all SQL Server 2008 are at 100% availability (all went online RTM + 24h)
- Hundreds of possibilities for SQL 11 ... 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 ☺)



European PASS Conference 2009

Failure is not an option

Agenda

The Mission

The Solution

- Standardizing
- Zero data loss
- High availability
- Scale up

24x7 OLTP Database
Management for VLDB

The Details

- SQL Server logins
- SQL Server jobs
- Log Shipping
- Partner databases
- Replication



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

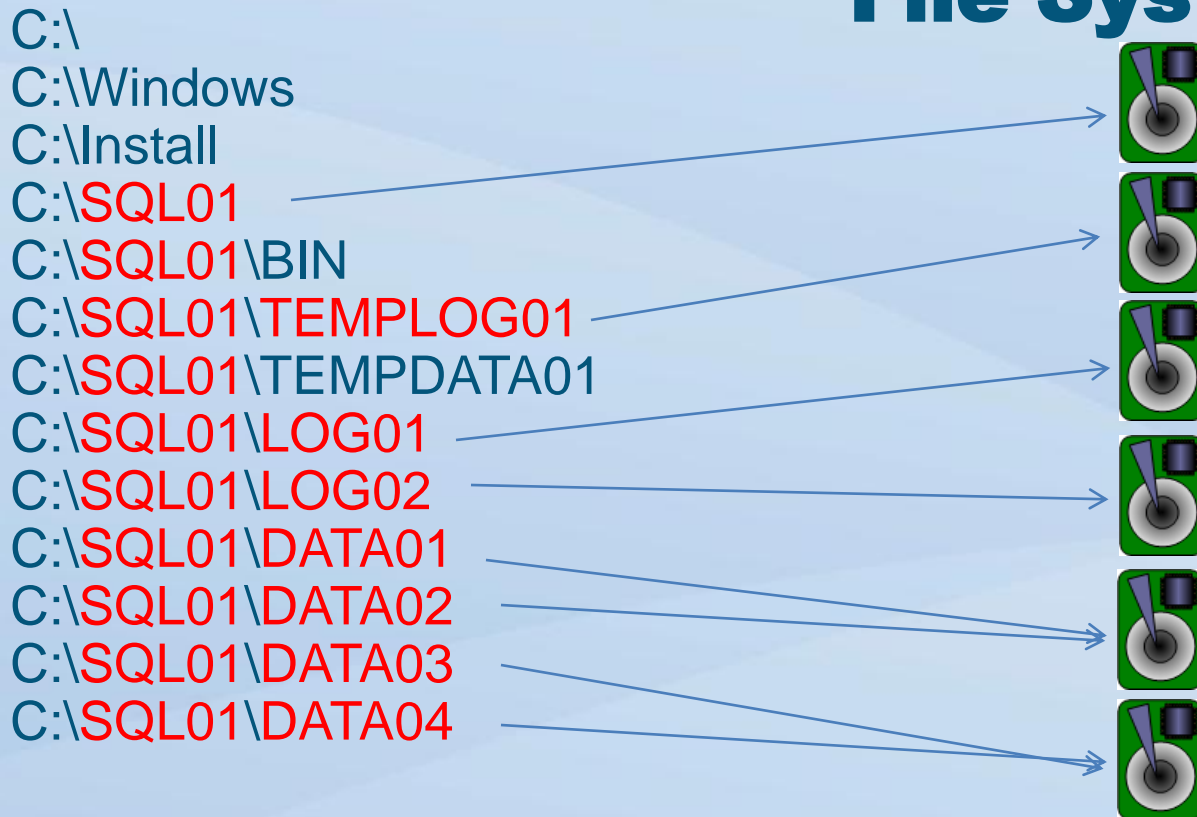
- Version, Edition, Service Pack, Patch Level

File System and Disks



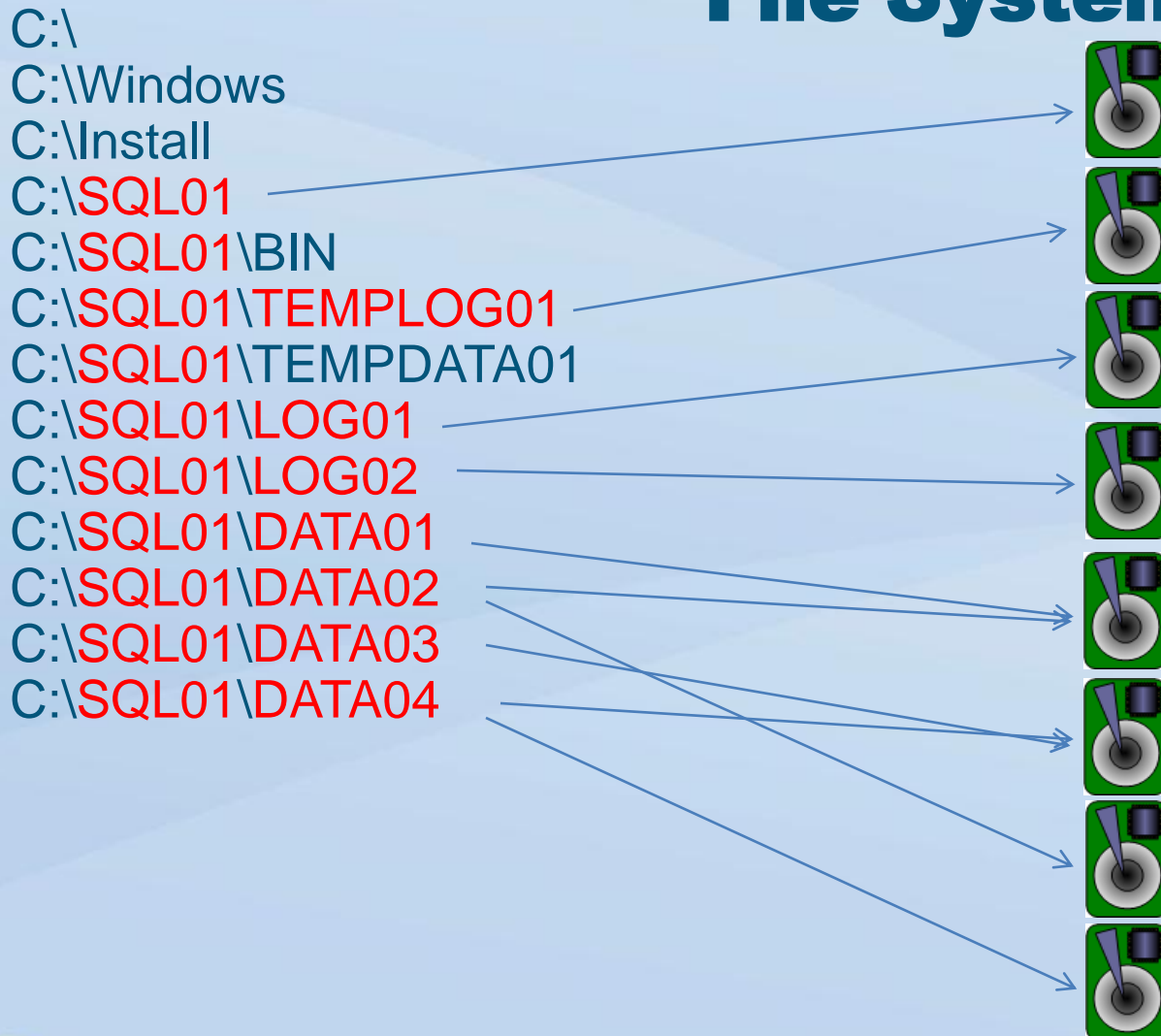
Failure is not an option

File System Sample



Failure is not an option

File System expansion



Failure is not an option

File System settings

Stripesize 64/128/256 kB

depending on storage

Partition alignment 64/128 kB

depending on storage

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

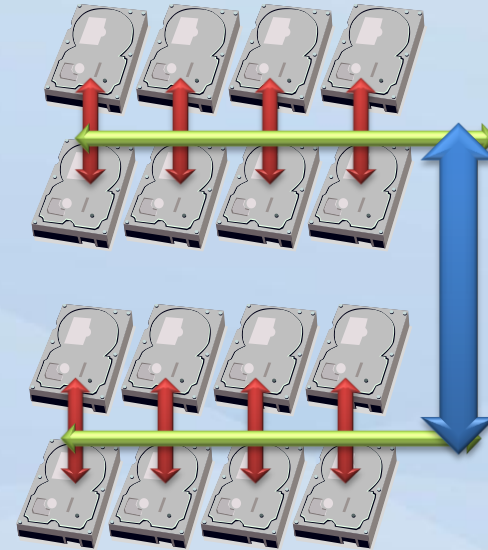


Failure is not an option RAID 101



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%

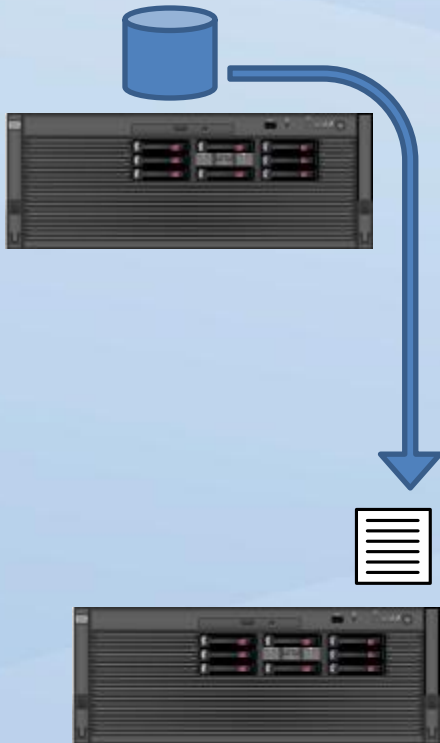


European PASS Conference 2009

Session Code – Failure is not an Option- 24x7 VLDB Administration

Failure is not an option

Zero data loss



- Full backup every 24 h

Availability: 0,00%

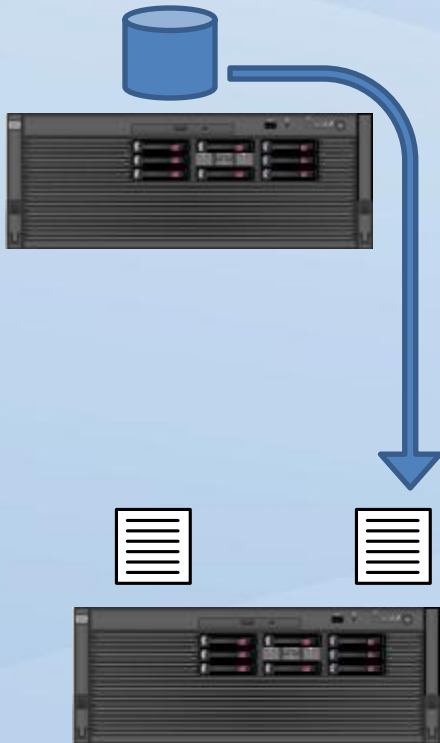
Data loss: 100,00%



European PASS Conference 2009

Failure is not an option

Zero data loss



- Full backup every 24 h
- Alternating files

Availability: 0,00%

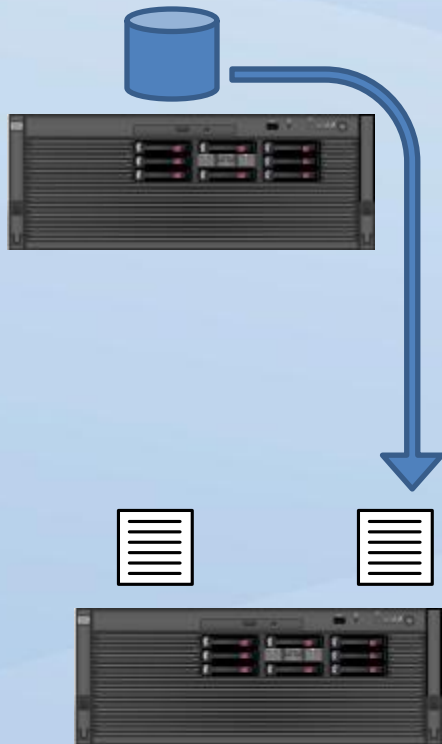
Data loss: 100,00%



European PASS Conference 2009

Failure is not an option

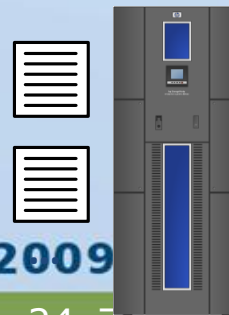
Zero data loss



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

Availability: 0,00%

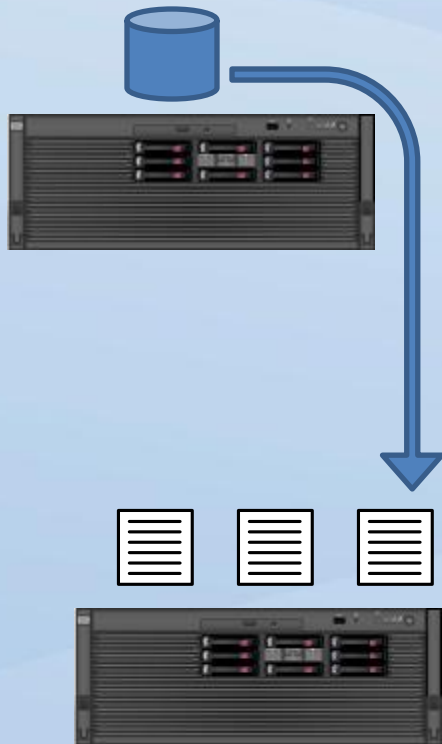
Data loss: 100,00%



European PASS Conference 2009

Failure is not an option

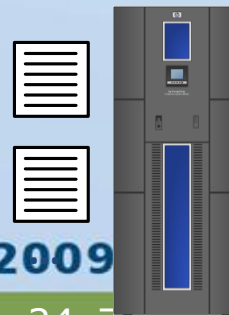
Zero data loss



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

Availability: 98,00%

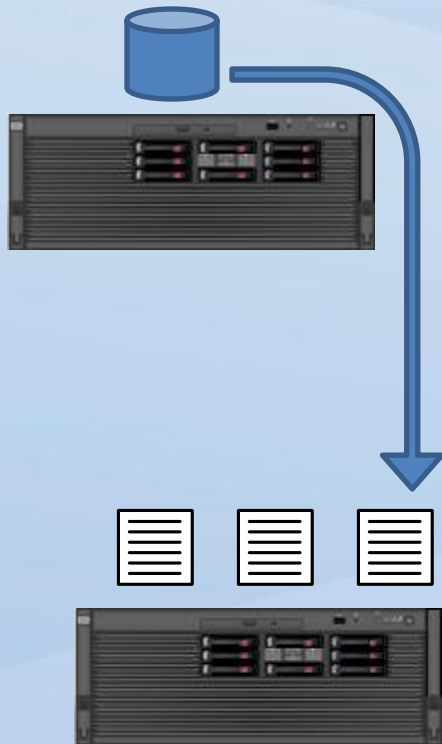
Data loss: 100,00%



European PASS Conference 2009

Failure is not an option

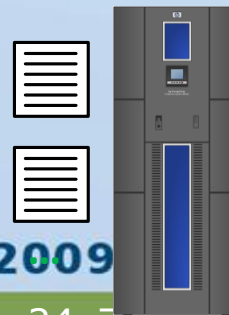
Zero data loss



- 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%

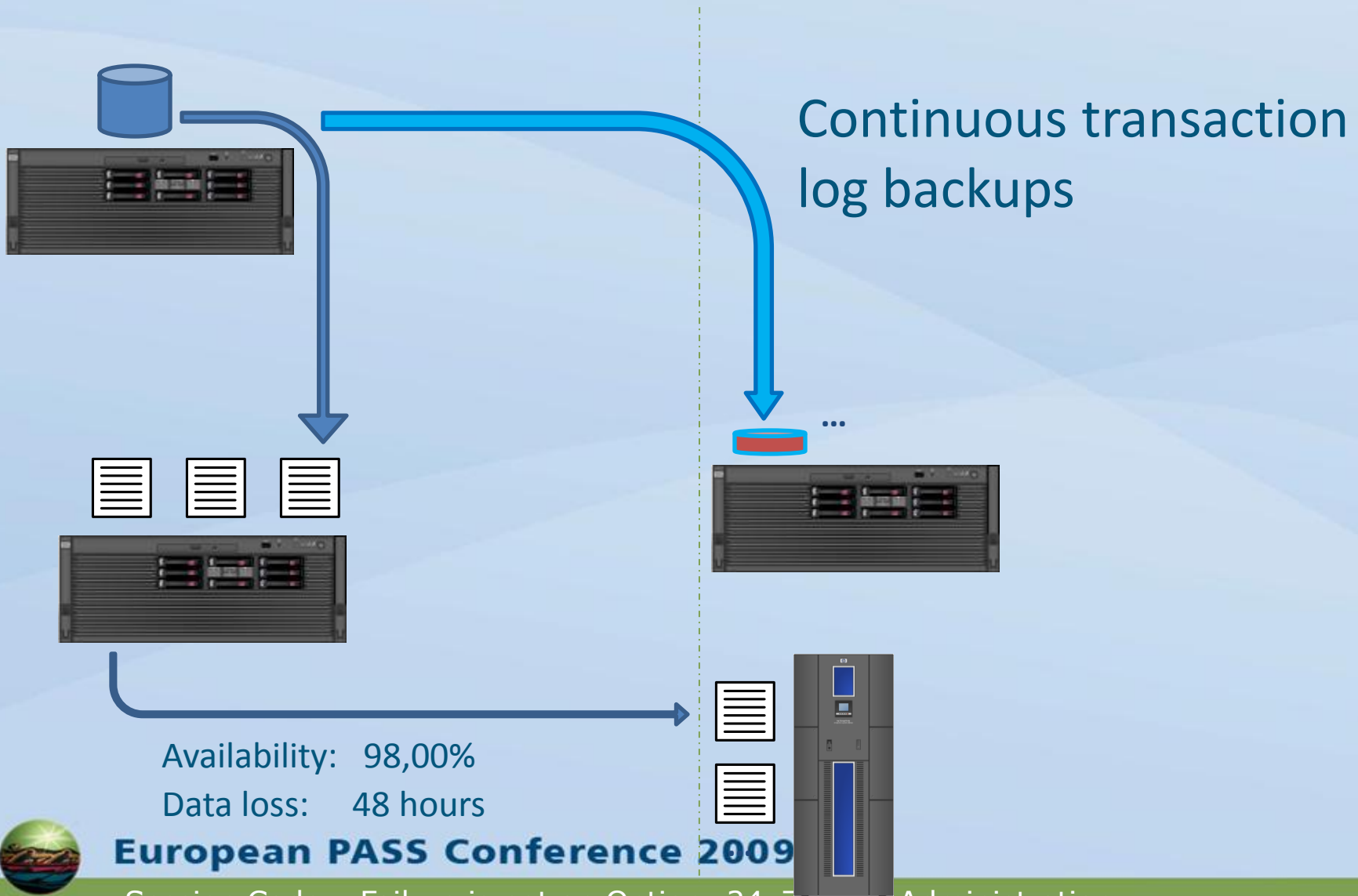
Data loss: 48 hours



European PASS Conference 2009

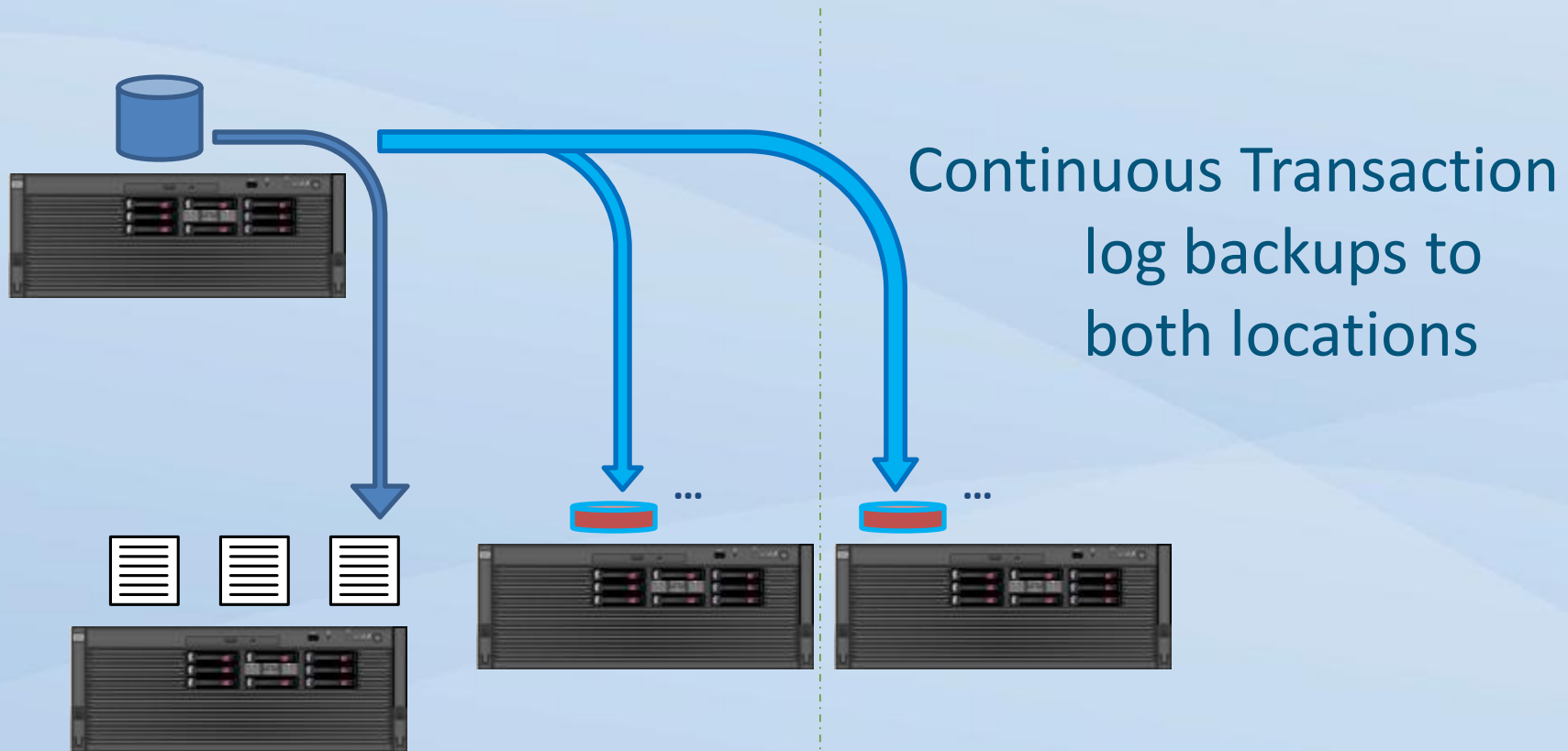
Failure is not an option

Zero data loss



Failure is not an option

Zero data loss



Continuous Transaction
log backups to
both locations

Availability: 98,00%

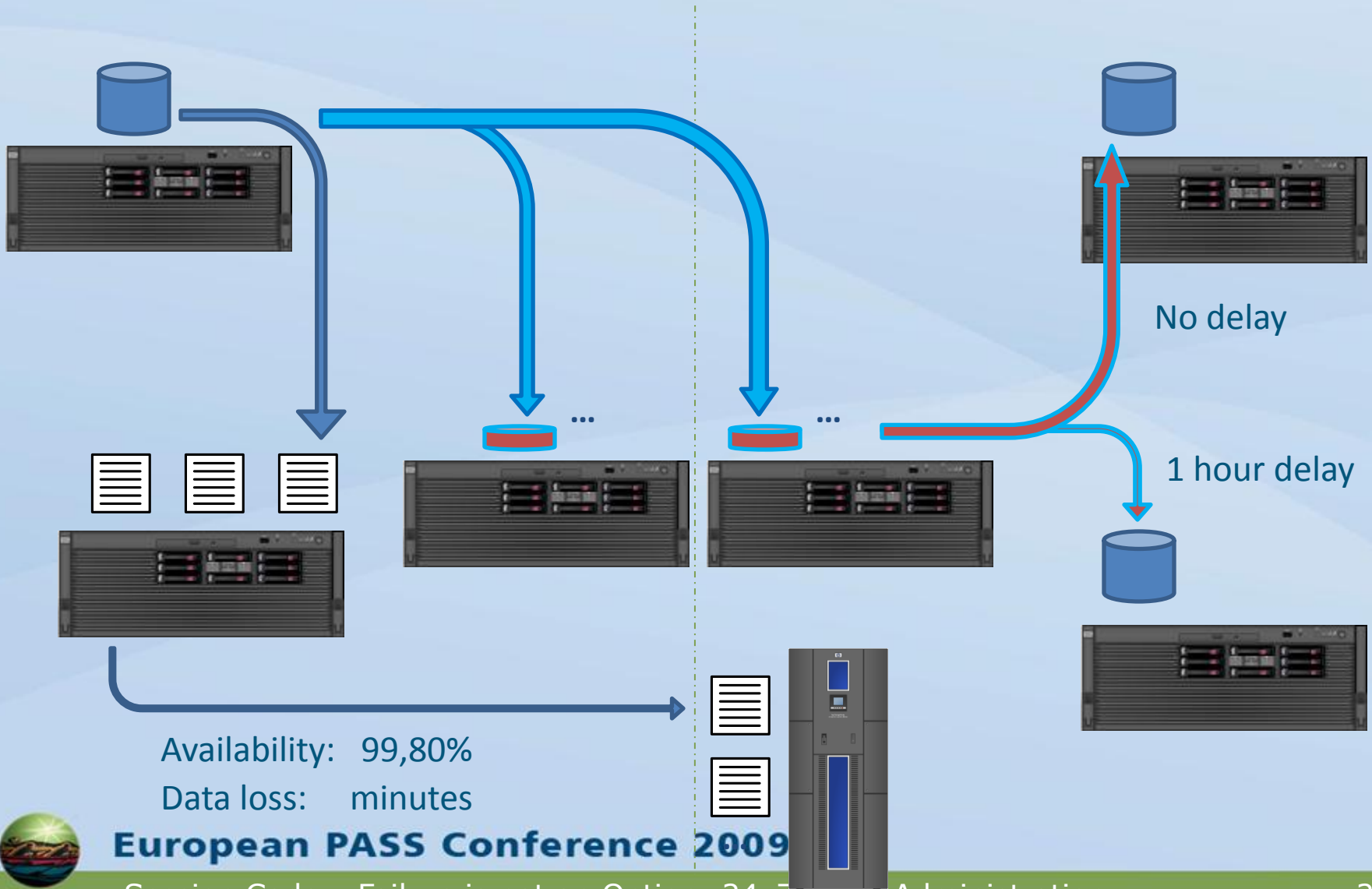
Data loss: 48 hours



European PASS Conference 2009

Failure is not an option

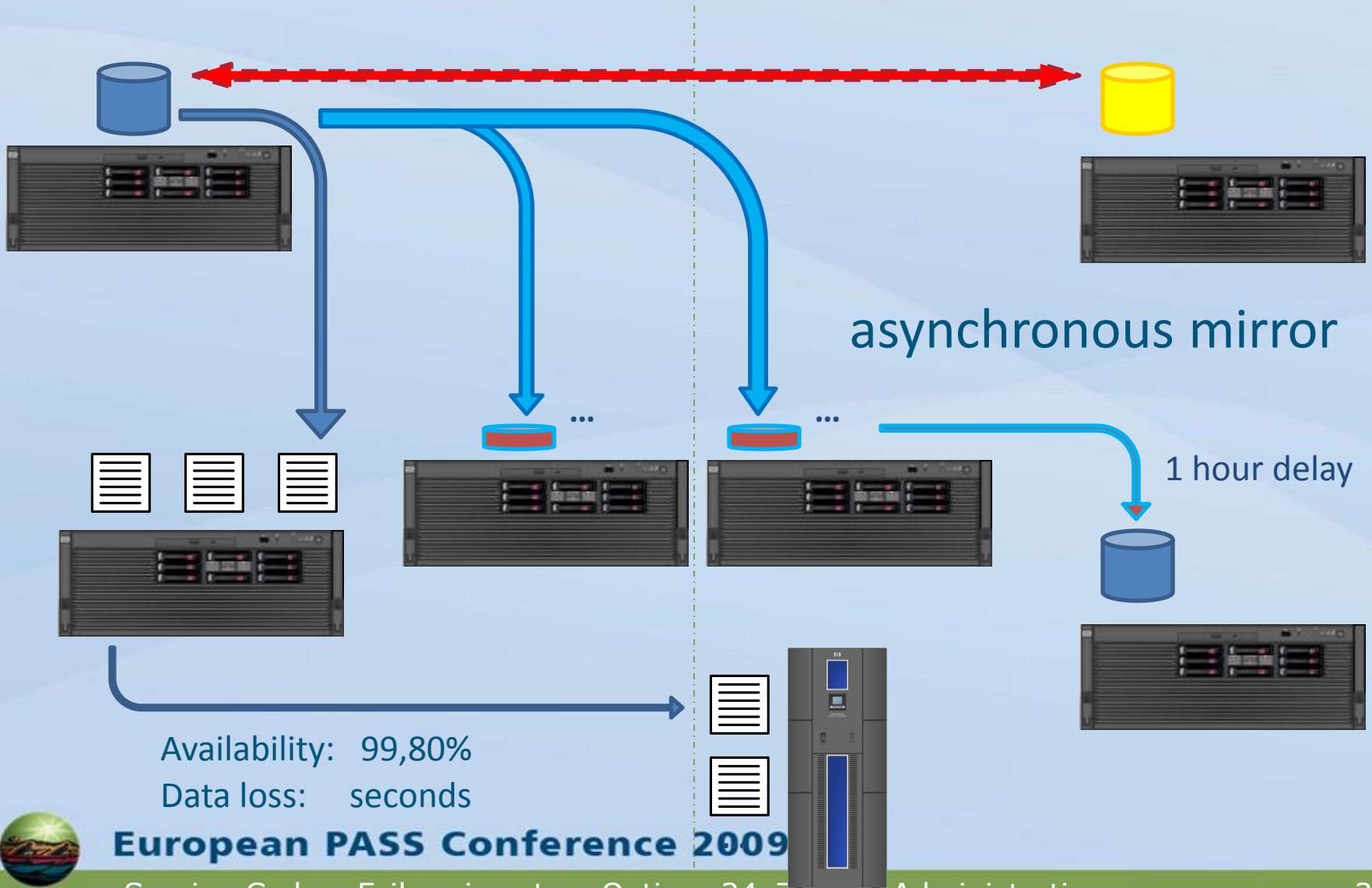
Zero data loss



European PASS Conference 2009

Failure is not an option

Zero data loss



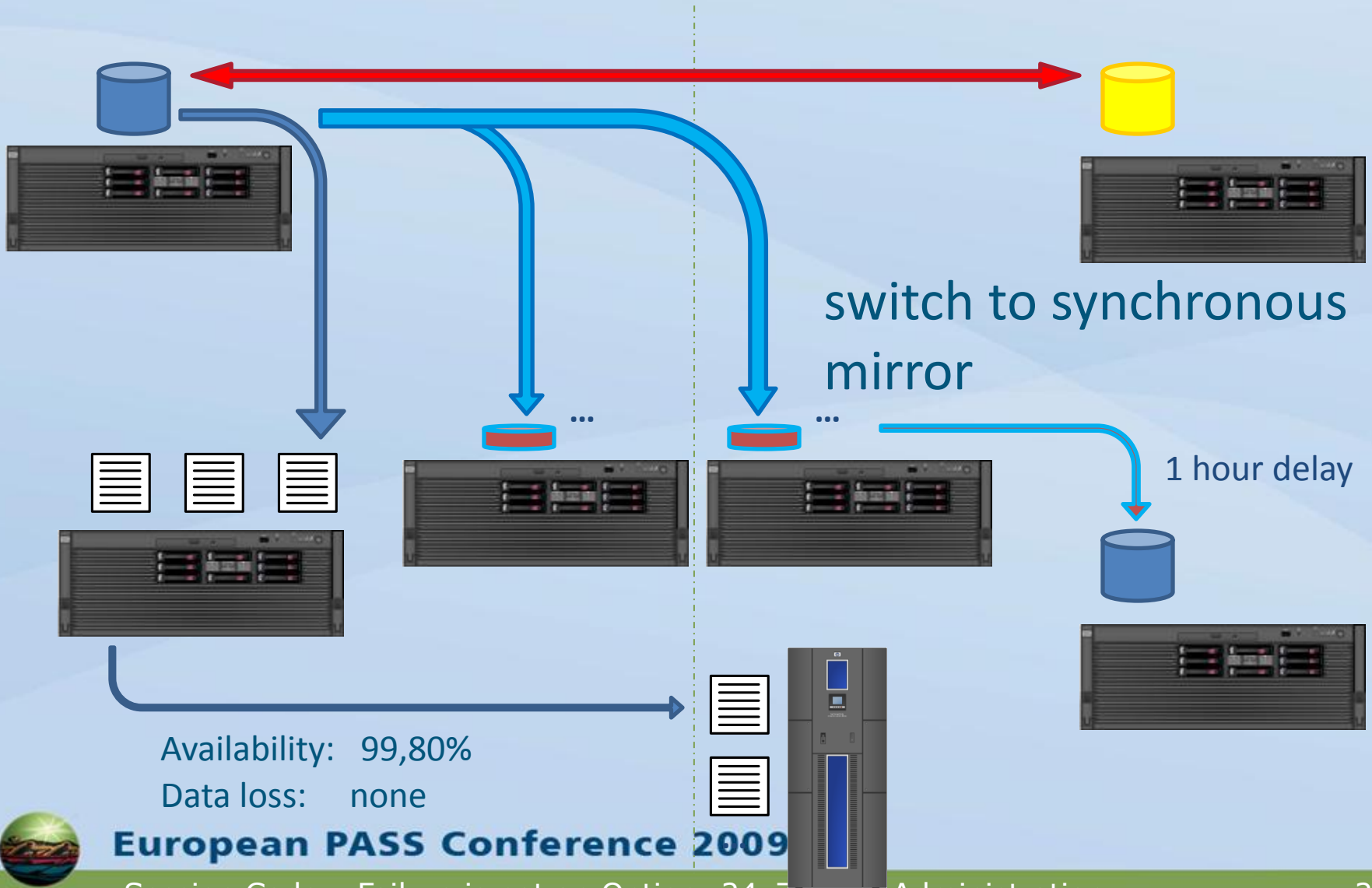
Availability: 99,80%

Data loss: seconds

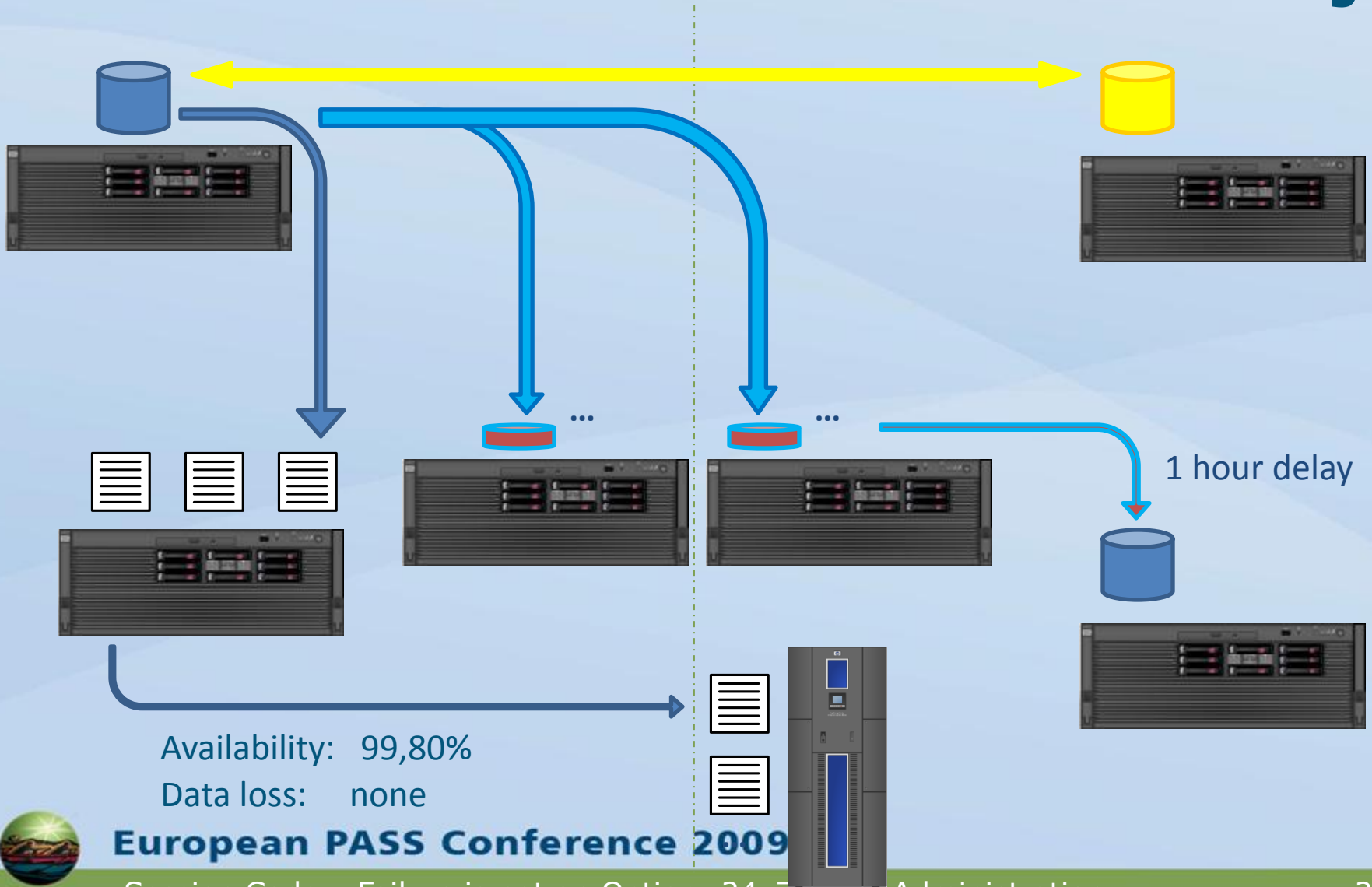
European PASS Conference 2009

Failure is not an option

Zero data loss

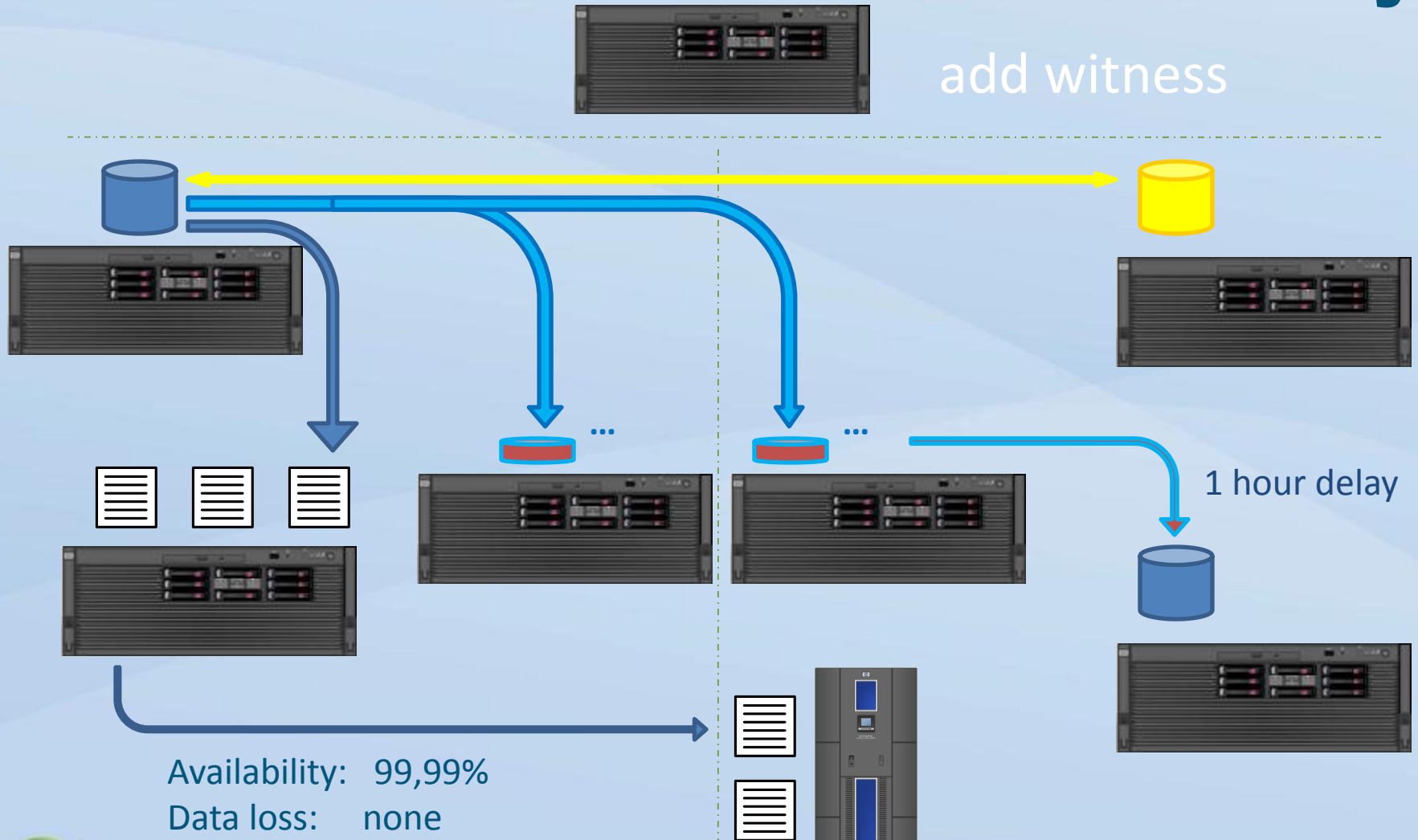


Failure is not an option Availability



European PASS Conference 2009

Failure is not an option Availability



European PASS Conference 2009

Failure is not an option

Scale Up

Selected CPU

IA64 / Itanium 2

Selected server/memory architecture

SMP / NUMA

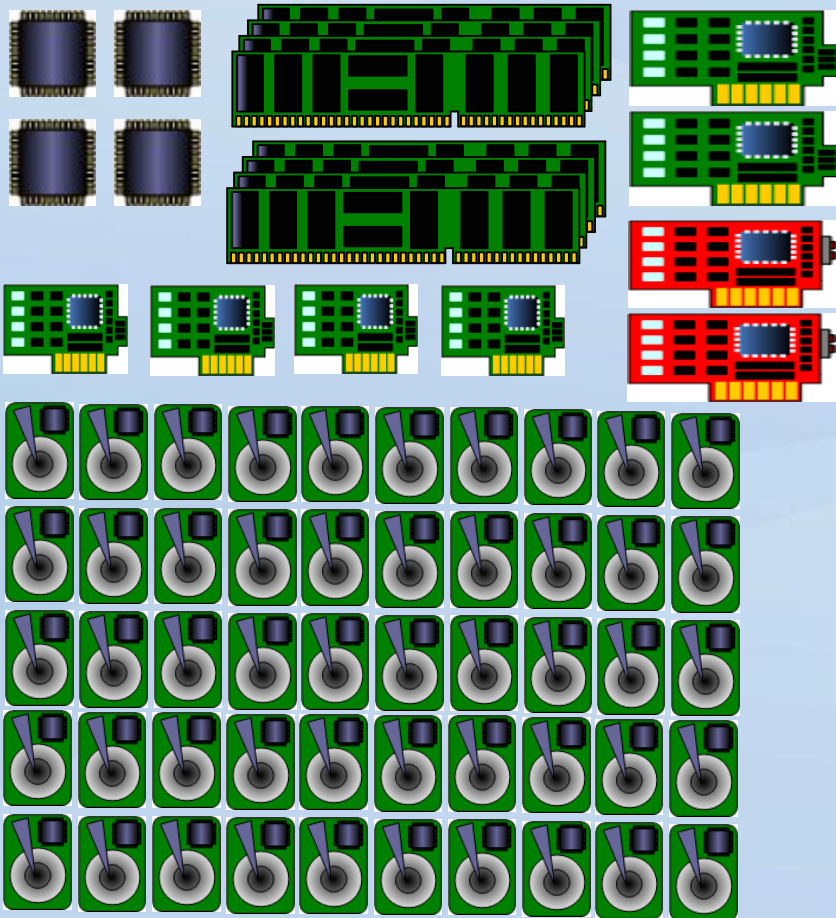
SQL Server 2008 Enterprise Edition

Windows Server 2008 for Itanium-Based Systems



Failure is not an option

Scale Up – Single NUMA node

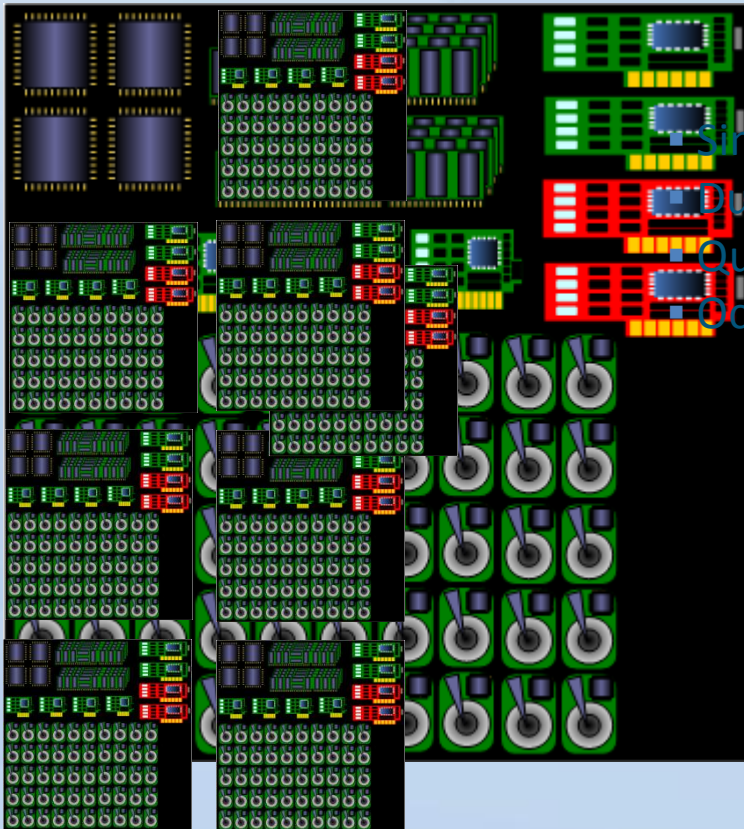


- 4 x Dual Core ITANIUM 2 CPUs
24 MB cache each
- 64 GB memory
- 4 x dual port 1 Gb/s network card
- 2 x dual port HBA (4Gb/s)
- 2 x P800 RAID controller
- 50 x 72 GB 15kRPM SAS disks
- SAN storage as needed
n x 512GB (on 64 spindles each)



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

1 NUMA Node Server

- 2 x NUMA node basic configuration

plus

- 2 x P600 (512MB cache)
- 16 x 72 GB 15kRPM SAS disks

2 NUMA Node Server

- 2 x NUMA node basic configuration

4/8/16 NUMA Node Server

- 4/8/16 x NUMA node basic configuration

plus

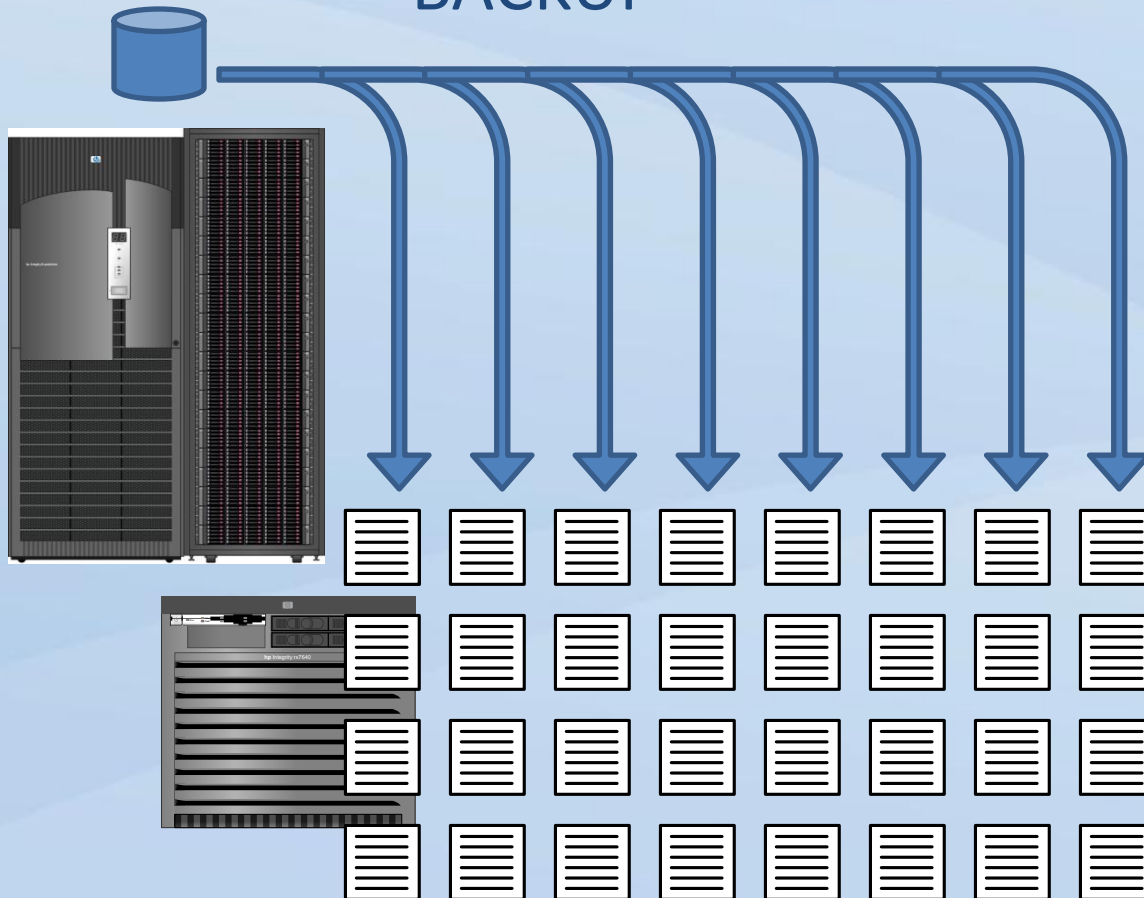
- 2 x single port 10 GE network card



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



Failure is not an option

The Details

Mirroring (not yet) as simple to manage as clustering

SQL Server logins

SQL Server jobs

Log Shipping

Replication

Partner databases



Failure is not an option

Magic Job

On **both** the principal and the mirror server:

Create a helper job (e.g.: Manage Mirrors)

with a schedule to run it once every minute

Create a helper database (e.g.: admin)

to store info like the last state of a database



Failure is not an option Magic Job



Jobs, Logins, ...

SQL
Agent
Magic job



Jobs, Logins, ...

SQL
Agent
Magic job



Failure is not an option Magic Job



Jobs, Logins, ...



Jobs, Logins, ...



Failure is not an option

Magic Job

```
USE Admin;  
GO  
CREATE SCHEMA FailoverHandler;  
GO  
CREATE TABLE FailoverHandler.DBStatus  
(  
    database_id                int,  
    lastStatus                  varchar(16),  
    lastStatusUpdateUTC        datetime  
);
```



Failure is not an option

Magic Job

```
CREATE PROC AutoFailoverHandler.CleanupDB @dbID as int
AS BEGIN
    DECLARE @currentStatus AS varchar(16) =
        (SELECT state_desc FROM sys.databases WHERE database_id = @dbID);
    DECLARE @lastStatus AS varchar(16) =
        (SELECT isnull((SELECT lastStatus FROM Admin.FailoverHandler.DBStatus
            WHERE database_id = @dbID
            AND lastStatusUpdateUTC>dateadd(minute,-5,GetUTCdate())),'N/A'));
    IF (@lastStatus <> @currentStatus)
    BEGIN
        -- Here we place the stuff to update
    END
    UPDATE FailoverHandler.DBStatus
        SET lastStatus = @currentStatus,lastStatusUpdateUTC=GetUTCDate()
        WHERE database_id = @dbID;
END
```



Failure is not an option

Magic Job

Execute the stored procedure for each mirrored database in a job step in our helper job on each server

```
EXEC AutoFailoverHandler.CleanupDB
    @dbID = db_id('myVLDB')
EXEC AutoFailoverHandler.CleanupDB
    @dbID = db_id('myOtherVLDB')
```



Failure is not an option SQL Server logins

Windows integrated logins must just be created on the mirror server, they use the Windows SID to map to the Database User.

For each user / login pair where the login is a SQL Server login map the user with the login using

```
exec sp_change_users_login
```



Failure is not an option

SQL Server logins - Code

```
DECLARE @user AS TABLE (username sysname);
DECLARE @username as sysname;
INSERT INTO @user
    SELECT u.name as username
        FROM sys.sysusers u
            left outer join sys.syslogins l ON (u.sid = l.sid)
        WHERE u.islogin = 1 AND u.isntname <> 1 and u.isntgroup <> 1
            and u.hasdbaccess = 1 AND l.sid is null;
WHILE ((SELECT COUNT(*) FROM @USER) > 0)
BEGIN
    SET @username = (SELECT TOP(1) username from @user);
    EXEC sp_change_users_login
        @Action = 'Auto_Fix', @UserNamePattern = @username;
    DELETE FROM @user where username = @username
END
```



Failure is not an option

SQL Server jobs

Have a first job step that checks if the database is online

Check in every step

Enable / Disable with

```
EXECUTE msdb.dbo.sp_update_job  
    @job_id = @jobID,  
    @enabled = 1;
```

Special care must be taken for jobs that job starts when the server starts and therefore must start with database (used for forever running jobs)



Failure is not an option

SQL Server jobs

```
DECLARE @jobID as uniqueidentifier
WHILE 1=1
    BEGIN
        SET @jobID = (SELECT TOP(1) job_id from msdb.dbo.sysjobs
            WHERE name like '%#MyDBName' and [enabled] = 0);
        IF (@jobID IS NULL) BREAK;
        EXECUTE msdb.dbo.sp_update_job @job_id=@jobID, @enabled=1;

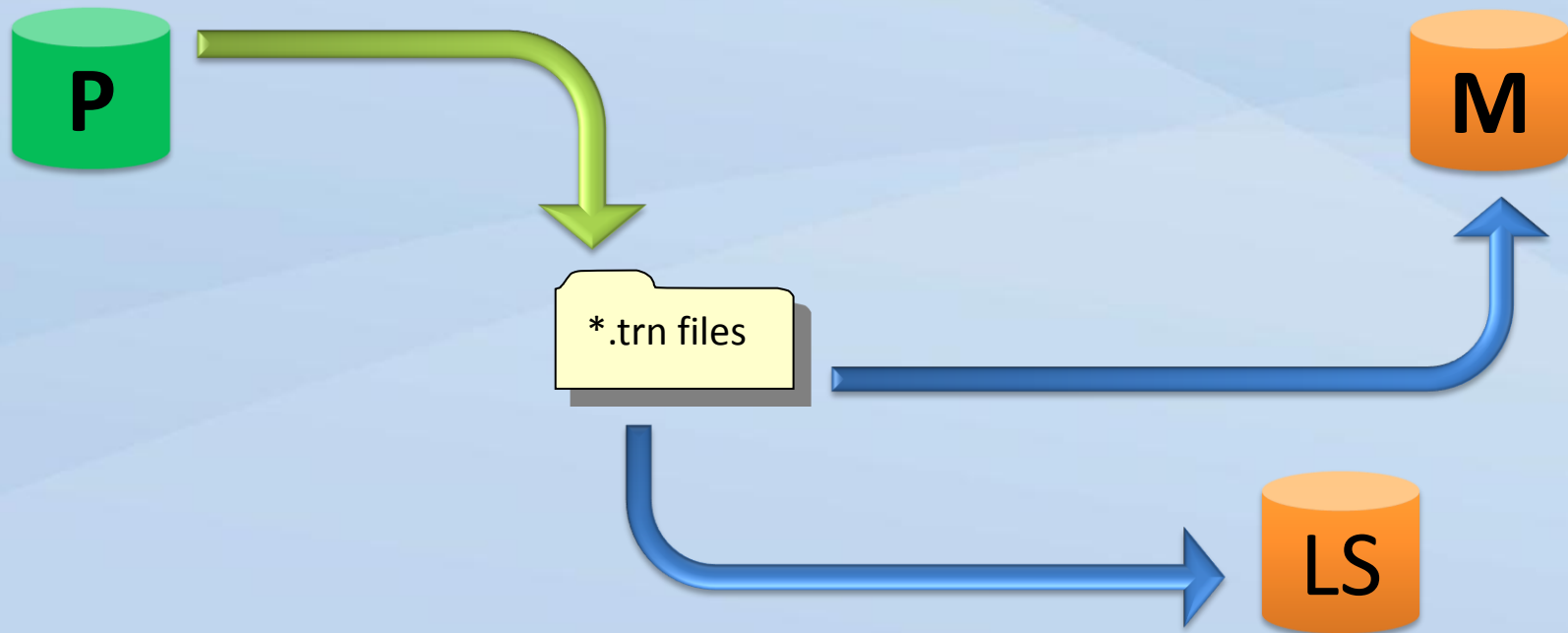
        IF ((SELECT freq_type FROM msdb.dbo.sysjobs j
            inner join msdb.dbo.sysjobschedules js ON (j.job_id=js.Job_id)
            inner join msdb.dbo.sysschedules s ON (js.schedule_id =
                s.schedule_id)
            WHERE j.job_id = @jobID) = 64)
            EXECUTE msdb.dbo.sp_start_job @job_id = @jobID;
    END
```



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



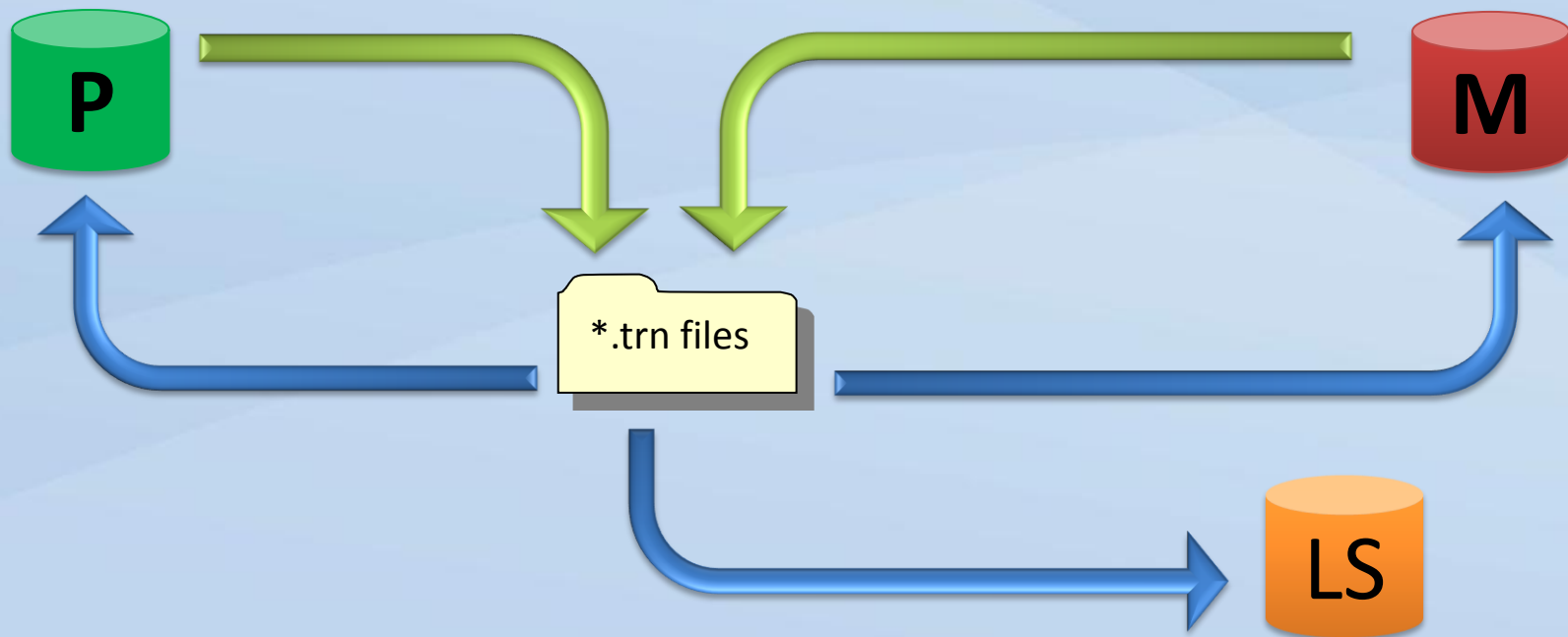
Failure is not an option

Log Shipping

Failover the database

Create the backup job on the mirror server

Create the restore job on the principal server



Failure is not an option

Partner databases

Partner databases are a databases that must be online on the same server

Therefore if one database fails over to the mirror all others must failover too

No problem if the server fails, because all db's will failover

Otherwise we must help a little with

```
ALTER DATABASE myOtherDB SET PARTNER FAILOVER
```

The code can be found in the demo scripts



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



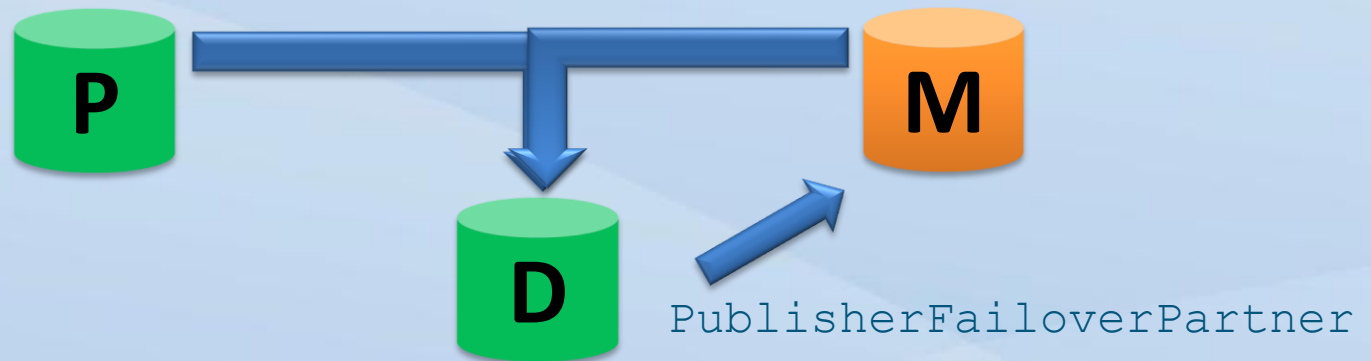
Failure is not an option Replication

Publishing from a mirrored database (2005 or 2008)

- Publisher : Create the publication as always
- Distributor (2008): In the Agent Profile you must add a
-PublisherFailoverPartner



Failure is not an option Replication



Failure is not an option

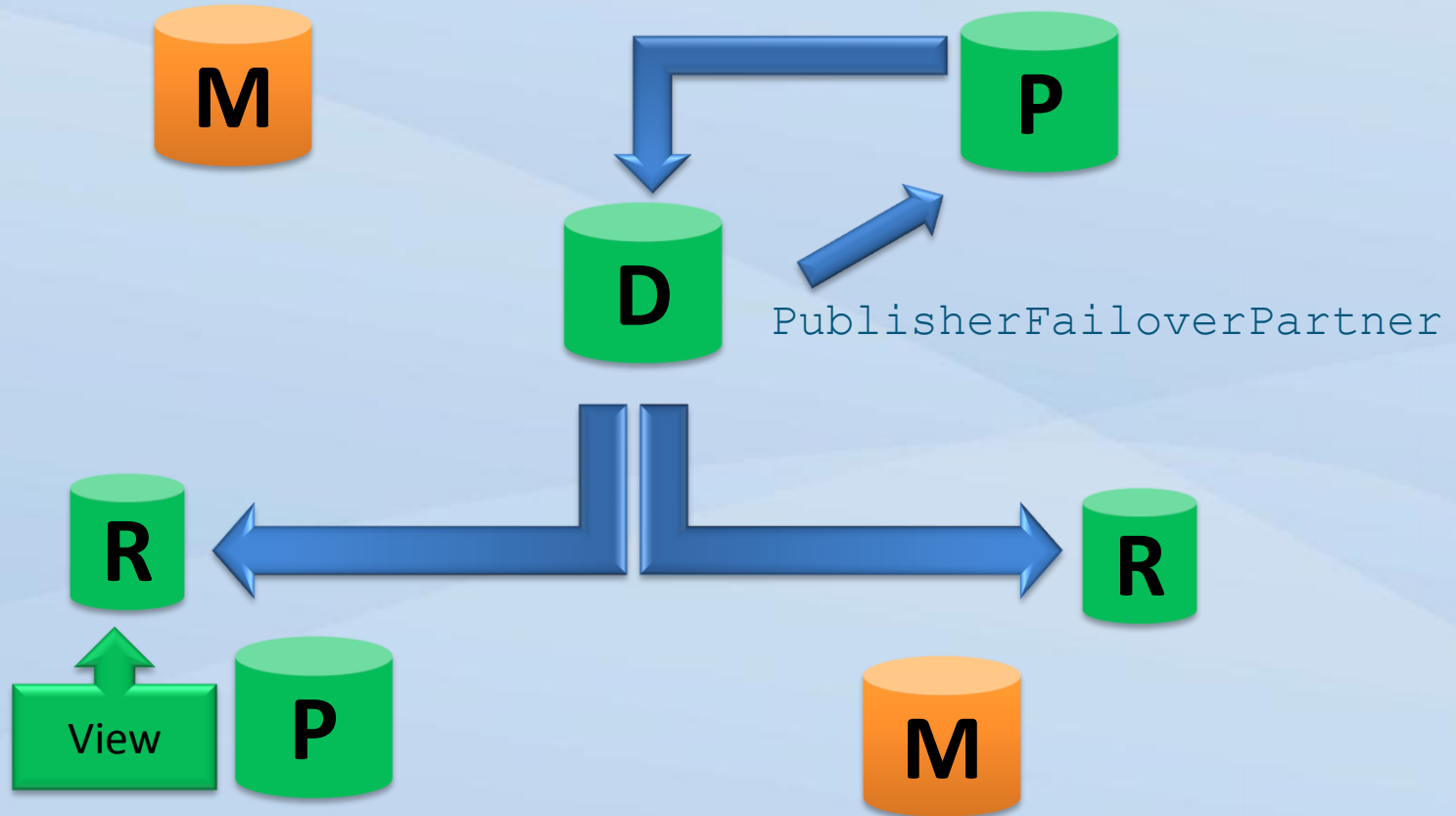
Replication

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) (same name)
 - Create two subscription one into each of the 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;
```



Failure is not an option Replication

Subscriber: Alternative method:

- Implement reinitialize form LSN

- White paper from Gopal Ashok (Microsoft Corporation) and Paul S. Randal (SQLskills.com)

<http://download.microsoft.com/download/d/9/4/d948f981-926e-40fa-a026-5bfcf076d9b9/ReplicationAndDBM.docx>



Demo

Putting it all together

The Demo

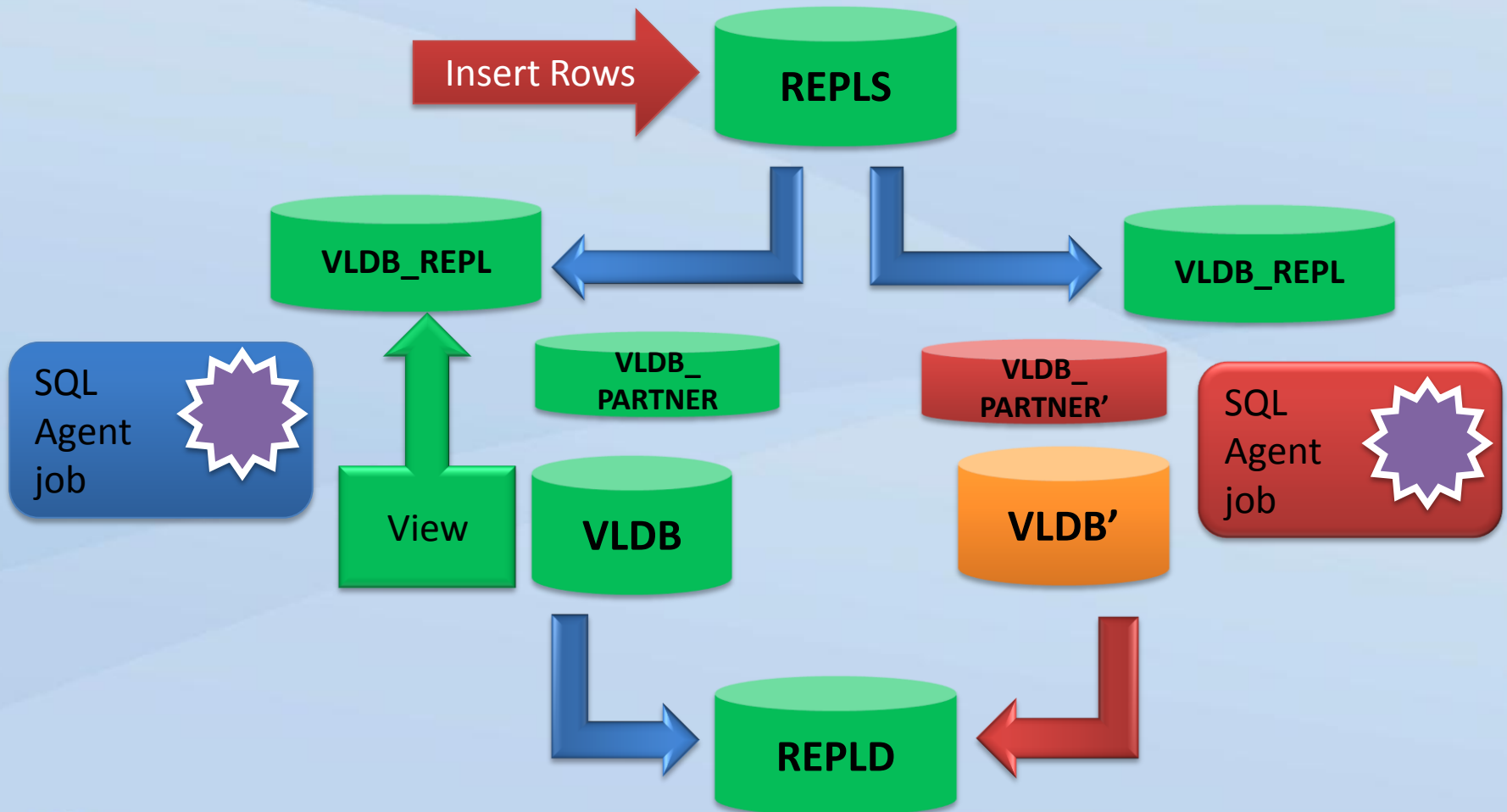


European PASS Conference 2009

Session Code - Session Title

Failure is not an option

The Demo



Failure is not an option

The Demo

You can download the demo scripts from

www.sqlserver-hwguide.com



Failure is not an option Call to action

Establish a SLA

Standardize your environment

Use your knowledge to build

- Reliable
- Highly Available
- Extreme performing

Solutions

SQL Server

fulfilling the SLA



Don't Forget to Fill Out Your Evaluations





Questions?



tg@grohser.com

Next sessions:

14:30



How to Optimize tempdb Performance

Brad McGehee (Red Gate Software)

Track 1

Standardizing and Centralizing Report Design

Andreas Wolter

Track 2

Making CLR Fly in SQL Server

Simon Sabin

Track 3

Analysis Services Performance Tuning

Thomas Kejser (SQLCAT)



Thank you!

tg@grohser.com