

Business Continuance Using Sun StorageTek VSM Clustering in Asynchronous and Synchronous Replication Modes

January 2008

Version 1.1

Copyright © 2007 Sun Microsystems, Inc., 4150 Network Circle, Santa Clara, California 95054, U.S.A. All rights reserved.

Sun Microsystems, Inc. has intellectual property rights relating to technology embodied in the product that is described in this document. In particular, and without limitation, these intellectual property rights may include one or more of the U.S. patents listed at <http://www.sun.com/patents> and one or more additional patents or pending patent applications in the U.S. and in other countries.

THIS PRODUCT CONTAINS CONFIDENTIAL INFORMATION AND TRADE SECRETS OF SUN MICROSYSTEMS, INC. USE, DISCLOSURE OR REPRODUCTION IS PROHIBITED WITHOUT THE PRIOR EXPRESS WRITTEN PERMISSION OF SUN MICROSYSTEMS, INC.

Use is subject to license terms. This distribution may include materials developed by third parties. This distribution may include materials developed by third parties. Parts of the product may be derived from Berkeley BSD systems, licensed from the University of California.

UNIX is a registered trademark in the U.S. and in other countries, exclusively licensed through X/Open Company, Ladson, Sun Microsystems, the Sun logo; Solaris, Sun StorageTek Crypto Key Management Station, StorageTek and StorageTek are trademarks or registered trademarks of Sun Microsystems, Inc. in the U.S. and other countries.

Products covered by and information contained in this service manual are controlled by U.S. Export Control laws and may be subject to the export or import laws in other countries. Nuclear, missile, chemical biological weapons or nuclear maritime end uses or end users, whether direct or indirect, are strictly prohibited. Export or re-export to countries subject to U.S. embargo or to entities identified on U.S. export exclusion lists, including, but not limited to, the denied persons and specially designated nationals lists is strictly prohibited. Use of any spare or replacement CPUs is limited to repair or one-for-one replacement of CPUs in products exported in compliance with U.S. export laws. Use of CPUs as product upgrades unless authorized by the U.S. Government is strictly prohibited.

DOCUMENTATION IS PROVIDED "AS IS" AND ALL EXPRESS OR IMPLIED CONDITIONS, REPRESENTATIONS AND WARRANTIES, INCLUDING ANY IMPLIED WARRANTY OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE OR NON-INFRINGEMENT, ARE DISCLAIMED, EXCEPT TO THE EXTENT THAT SUCH DISCLAIMERS ARE HELD TO BE LEGALLY INVALID.

Copyright © 2007 Sun Microsystems, Inc., 4150 Network Circle, Santa Clara, California 95054, Etats-Unis. Tous droits réservés.

Sun Microsystems, Inc. détient les droits de propriété intellectuels relatifs à la technologie incorporée dans le produit qui est décrit dans ce document. En particulier, et ce sans limitation, ces droits de propriété intellectuelle peuvent inclure un ou plus des brevets américains listés à l'adresse <http://www.sun.com/patents> et un ou les brevets supplémentaires ou les applications de brevet en attente aux Etats - Unis et dans les autres pays.

CE PRODUIT CONTIENT DES INFORMATIONS CONFIDENTIELLES ET DES SECRETS COMMERCIAUX DE SUN MICROSYSTEMS, INC. SON UTILISATION, SA DIVULGATION ET SA REPRODUCTION SONT INTERDITES SANS L' AUTORISATION EXPRESSE, ECRITE ET PREALABLE DE SUN MICROSYSTEMS, INC.

L'utilisation est soumise aux termes de la Licence. Cette distribution peut comprendre des composants développés par des tierces parties. Cette distribution peut comprendre des composants développés par des tierces parties. Des parties de ce produit pourront être dérivées des systèmes Berkeley BSD licenciés par l'Université de Californie.

UNIX est une marque déposée aux Etats-Unis et dans d'autres pays et licenciée exclusivement par X/Open Company, Ltd. Sun, Sun Microsystems, le logo Sun, Solaris, Sun StorageTek Crypto Key Management Station, StorageTek et StorageTek sont des marques de fabrique ou des marques déposées de Sun Microsystems, Inc. aux Etats-Unis et dans d'autres pays.

Ce produit est soumis à la législation américaine en matière de contrôle des exportations et peut être soumis à la réglementation en vigueur dans d'autres pays dans le domaine des exportations et importations. Les utilisations, ou utilisateurs finaux, pour des armes nucléaires, des missiles, des armes biologiques et chimiques ou du nucléaire maritime, directement ou indirectement, sont strictement interdites. Les exportations ou reexportations vers les pays sous embargo américain, ou vers des entités figurant sur les listes d'exclusion d'exportation américaines, y compris, mais de manière non exhaustive, la liste de personnes qui font objet d'un ordre de ne pas participer, d'une façon directe ou indirecte, aux exportations des produits ou des services qui sont régis par la législation américaine en matière de contrôle des exportations et la liste de ressortissants spécifiquement désignés, sont rigoureusement interdites. L'utilisation de pièces détachées ou d'unités centrales de remplacement est limitée aux réparations ou à l'échange standard d'unités centrales pour les produits exportés, conformément à la législation américaine en matière d'exportation. Sauf autorisation par les autorités des Etats-Unis, l'utilisation d'unités centrales pour procéder à des mises à jour de produits est rigoureusement interdite.

LA DOCUMENTATION EST FOURNIE "EN L'ETAT" ET TOUTES AUTRES CONDITIONS, DECLARATIONS ET GARANTIES EXPRESSES OU TACITES SONT FORMELLEMENT EXCLUES, DANS LA MESURE AUTORISEE PAR LA LOI APPLICABLE, Y COMPRIS NOTAMMENT TOUTE GARANTIE IMPLICITE RELATIVE A LA QUALITE MARCHANDE, A L'APTITUDE A UNE UTILISATION PARTICULIERE OU A L'ABSENCE DE CONTREFACON.

We welcome your feedback. Please contact STK Test Engineering at:

m.white@sun.com; kathleen.hodge@sun.com; richard.birkelo@sun.com

or

STK Test Engineering
Sun Microsystems, Inc.
One StorageTek Drive
Louisville, CO 80028-9172
USA



Sun Microsystems, Inc. 4150 Network Circle, Santa Clara, CA 95054 USA Phone 1-650-960-1300 or 1-800-555-9SUN Web sun.com

SUN™ THE NETWORK IS THE COMPUTER

©2006 Sun Microsystems, Inc. All rights reserved. Sun, Sun Microsystems, and the Sun logo; StorageTek and the StorageTek logo are trademarks or registered trademarks of Sun Microsystems, Inc. in the United States and other countries.

Contents

Business Continuance Using Sun StorageTek VSM Clustering in Asynchronous and Synchronous Replication Modes.....	1
Contents	3
Figures	5
Change History	6
Introduction.....	7
Prerequisites.....	7
Chapter 1: Synchronous Replication Overview.....	9
Disaster Recovery Tiers:.....	9
Business Continuance Objectives:.....	10
Problem Solved with VSM Synchronous Replication:	10
Chapter 2: Operational Differences Between Asynchronous and Synchronous Replication.....	11
Asynchronous Replication.....	11
Synchronous Replication.....	11
VTSS Clustering and Immediate Migration.....	12
Mixed Asynchronous and Synchronous Modes.....	12
Synchronous Replication Messages.....	13
Chapter 3: Configurations.....	15
Workload Strategy.....	15
Distances Supported (Native and Extended).....	15
Environmental Configurations.....	15
Eight Clink Cluster Configuration.....	15
Two and Four Clink Cluster Configurations.....	16
Channel Extension Configurations.....	17
Chapter 4: Operational Considerations.....	19
Application Considerations.....	19
VTD MIH Timer Requirements.....	20
VTV Page Size.....	20
Multiple Dataset VTVs.....	20
Multiple VTCS Startup.....	20
VTD Allocation and Usage.....	21
VTSS VCF Card Configuration.....	21
Bi-directional Cluster Configuration	21
Uni-directional Cluster Configuration	22
TMS Considerations	23

[Chapter 5: Recommendations and Important Considerations.....](#) 24

[Required Steps to Enable Synchronous Replication.....](#) 24

[VTSS Microcode Levels.....](#) 24

[Serialization During Synchronous Replication.....](#) 24

[Chapter 6: Replication Job Results and Conclusions.....](#) 25

[Synchronous Replication VTV Size Recommendation.....](#) 26

[Appendix A -- References:.....](#) 27

Figures

Figure 1: Clink Replication Queue.....	13
Figure 2: 8 Clink Bi-directional Cluster.....	16
Figure 3: 2 Clink Uni-directional Cluster.....	16
Figure 4: 2 Clink Ciena Channel Extension Configuration.....	17
Figure 5: 2 Clink USD-X Channel Extension Configuration.....	17
Figure 6: 2 Clink Cisco Channel Extension Configuration.....	18
Figure 7: Local Production and Remote Backup Sites.....	19
Figure 8: FICON Bi-Directional VCF Configuration.....	22
Figure 9: FICON Uni-Directional VCF Configuration.....	23
Figure 10: Replication results by type, asynchronous and synchronous.....	25
Figure 11: Synchronous Replication time to completion percentages.....	25

Change History

Document Description	
Document owner	Kathy Hodge, Mike White, Malcolm MacAskill—Systems Integration Engineers
Organization	STK Test Engineering—Mainframe Customer Emulation & Test

Revision	Date	Description
V1.0	08/30/2007	KH/MW/MM - Initial Release
V1.1	01/29/2008	HK/MW/MM – Add Replication Test Results to Chapter 6

Introduction

This document has been compiled to provide examples and advice to companies that are interested in implementing the Sun StorageTek VSM Cluster Synchronous Replication Solution. Synchronous Replication has been added to the currently available Asynchronous Replication feature and will allow immediate copies of virtual tape volumes (VTV) to be written to a remote site before the Virtual Tape Drive (VTD) processes a rewind/unload command. To fully understand the differences between virtual tape volume asynchronous and synchronous replication and their impact on the data center both have been examined using identical job streams. Other variations examined within this paper are Clink configurations in a Bi-directional and Uni-directional environment as well as Clink, with channel extenders at various distance settings.

This whitepaper is not intended to be a step-by-step guide, but rather will serve to highlight the benefits and issues involved in a clustered Virtual Storage Manager configuration and present recommended operational techniques for using them. Use the *Sun StorageTek Virtual Tape Control System Software: Synchronous Replication Guide* for specific answers about command syntax, parmlib statements and messages.

The information presented in this document is supported by testing initiatives that were conducted in Sun’s Mainframe Customer Emulation Test (MCET) Lab. The purpose of the MCET lab is customer emulation; it is not a performance test lab and the results of this testing are variable due to common loads that exist on the shared MCET lab computer system. This initial release of the document includes VSM4 and VSM5 solutions in mixed cluster environments. Further testing is planned for a VSM4 to VSM5 cluster environment which will provide additional information for a future release of this document.

Prerequisites

The following section lists the minimum software and microcode prerequisites to enable a VSM4 and VSM5 to perform synchronous replication. You should check for later maintenance available on the Customer Resource Center.

NCS/VTCS Software –	6.1	6.2
VTCS PTF s	L1H13MI L1H13QS	L1H13QL L1H13K8 L1A0013

VTSS Microcode	VSM4	VSM5
	D02.03.xx	D02.03.xx

VTSS Hardware	VSM4	VSM5
	VCF2 Interface	VCF3 Interface

In addition to the required software and microcode upgrades, the VTSS cluster feature must also be installed on both subsystems before any cluster configuration definition can occur. VSM4 models will require FICON VCF card upgrades to the cluster configuration before synchronous replication can operate because synchronous capable Clinks must be FICON. The VTCS startup parameter for VSM Advanced Management feature is also required.

Synchronous replication is **NOT** supported on VSM3; VSM4 and beyond will support synchronous replication.

Chapter 1: Synchronous Replication Overview

The addition of the Synchronous Replication feature adds yet another important building block to use to provide robust disaster recovery solutions using the Sun StorageTek VSM4 and VSM5 products.

Disaster Recovery Tiers:

The following list of disaster recovery tiers were developed during a 1992 U.S Share conference and are listed in IBM Redbook IBM TotalStorage Solutions for Disaster Recovery, SG24-6547. The VSM Synchronous Replication Solution is designed to be used with Tier (n) solutions.

Tier 0 - No off-site data

Businesses with a Tier 0 Disaster Recovery solution have no Disaster Recovery Plan.

Tier 1 - Data backup with no Hot Site

Businesses that use Tier 1 Disaster Recovery solutions back up their data at an off-site facility.

Tier 2 - Data backup with a Hot Site

Businesses using Tier 2 Disaster Recovery solutions make regular backups on tape. This is combined with an off-site facility and infrastructure (known as a hot site) in which to restore systems from those tapes in the event of a disaster.

Tier 3 - Electronic vaulting

Tier 3 solutions utilize components of Tier 2. Additionally, some mission-critical data is electronically vaulted.

Tier 4 - Point-in-time copies

Tier 4 solutions are used by businesses that require both greater data currency and faster recovery than users of lower tiers.

Tier 4 Disaster Recovery solutions:

Batch/Online Database Shadowing and Journaling, Synchronous and Asynchronous Virtual Tape Replication, DASD Peer to Peer, SnapShot.

Tier 5 - Transaction integrity

Tier 5 solutions are used by businesses with a requirement for consistency of data between production and recovery data centers. There is little to no data loss in such solutions; however, the presence of this functionality is entirely dependent on the application in use.

Tier 5 Disaster Recovery solutions:

Software, two-phase commit

Tier 6 - Zero or little data loss

Tier 6 Disaster Recovery solutions maintain the highest levels of data currency. They are used by businesses with little or no tolerance for data loss and who need to restore data to applications rapidly. These solutions have no dependence on the applications to provide data consistency.

Tier 6 Disaster Recovery solutions:

Batch/Online Database Shadowing and Journaling, Synchronous and Asynchronous Virtual Tape Replication, DASD Peer to Peer, SnapShot.

Tier 7 - Highly automated, business-integrated solution

Tier 7 solutions include all the major components being used for a Tier 6 solution with the additional integration of automation. This allows a Tier 7 solution to ensure consistency of data above that of which is granted by Tier 6 solutions. Additionally, recovery of the applications is automated, allowing for restoration of systems and applications much faster and more reliably than would be possible through manual Disaster Recovery procedures.

Tier 7 Disaster Recovery solutions:

Batch/Online Database Shadowing and Journaling, Synchronous and Asynchronous Virtual Tape Replication, DASD Peer to Peer, SnapShot, Disaster recovery automation software.

Source: This information on the Seven Tiers of Disaster Recovery is expanded upon in the IBM Redbook IBM TotalStorage Solutions for Disaster Recovery, SG24-6547.

Business Continuance Objectives:

In the event of a disaster, large corporations cannot afford down time and have business continuance objectives to be back online with no data lost within a minimal amount of time. Tier 7 Disaster Recovery solutions are usually implemented by these large companies to provide a high level of business continuance service. Many of the high availability and disaster recovery solutions involve replication across some distance; synchronous replication supports distance using channel extension. The decision to implement data replication in an asynchronous or synchronous mode is now a factor in business continuance capability. Sun StorageTek provides support for both synchronous and asynchronous replication of virtual tape volumes within a clustered VSM environment.

Problem Solved with VSM Synchronous Replication:

Clustered VSM support provides a means to define pairs of VTSS disk buffers that operate together providing improved virtual tape volume data recoverability. A clustered VSM allows customers to replicate virtual tape volumes from one VTSS to another. With synchronous replication, the two VTSSs operate together to provide improved data availability **without additional host or operator involvement**. VTCS, the host software component, supports asynchronous replication of a VTV within a clustered VTSS environment as well as synchronous VTV replication.

As a result of maintaining a clustered VTSS environment, failover to the receiving VTSS subsystem in the cluster will provide a backup subsystem for continued business operations. If the sending VTSS becomes unavailable it can be varied offline, which allows the workload to continue using only the receiving VTSS. Once the sending VTSS is available it can be varied back online. This will return the workload to the sending VTSS and in a bi-directional clustered environment VTCS will automatically asynchronously replicate to re-synchronize the contents of the two VTSS disk buffers.

Chapter 2: Operational Differences Between Asynchronous and Synchronous Replication

VSM Clustering is defined as two VTSS disk buffers connected together using Clinks for the purpose of creating a VTV copy or replication on both VTSSs. Clustered VSM environments support uni-directional or one way replication where the replication originates from the 'Sender' VTSS and is replicated to the 'Receiver' VTSS as well as bi-directional or two way replication where both VTSSs act as 'Sender' and 'Receiver'. Previous VSM releases supported asynchronous replication in which VTCS scheduled the VTV copy after the VTV was written and unloaded. Now, VSM supports synchronously replicating the VTV to the receiver VTSS before the VTV is unloaded and control returned to the application program.

The primary difference between synchronous and asynchronous replication is timing of the replication, synchronous is immediate and asynchronous is scheduled.

Asynchronous Replication

Asynchronous replication takes place after the VTV has been rewound and unloaded and the job step has moved on to other processing. The VTCS host software schedules the asynchronous replication immediately after the rewind/unload occurs and the replication will proceed. Asynchronous replication provides very good data consistency while allowing the greatest job step throughput by not holding up the job step while a replication completes.

Asynchronous replication, like synchronous replication, creates an exact copy of the local VTV on the receiving VTSS by transferring the compressed internal VTSS format. Unlike synchronous replication, there is no practical distance limitation because the application job is allowed to proceed to the next step while the replication is taking place. Since the application is not waiting for the remote system to respond and the application execution duration is not impacted, no matter how long the replication takes. Without a requirement for a timely response, the remote VTSS can now be placed at a greater distance away from the local VTSS.

Pros:

- ◇ No job duration impact.
- ◇ VTVs are replicated between Sender & Receiver as a scheduled event.

Cons:

- ◇ VTVs aren't immediately synchronized between the Sender and Receiver.

Synchronous Replication

Synchronous replication performs the VTV replication before the virtual tape drive completes the rewind/unload command at the sending VTSS. The replication is performed by transferring the compressed internal VTSS format directly to the receiving VTSS. This ensures that the transfer happens as quickly as possible so that the application program can resume.

A successful synchronous replication guarantees that the receiving VTSS has a copy of the VTV at the completion of the rewind/unload command, but the disadvantage is that the application must wait for the replication to complete before proceeding, potentially leading to an increased job duration time.

During synchronous replication, the sending VTSS transfers the complete VTV in compressed internal format to the receiving VTSS on a Clink for the duration of the transfer. After the VTV replication is complete, the rewind/unload is allowed to continue and control is released back to the application job step. Synchronous replication will provide the greatest data consistency in the event of a disaster because the job step is not allowed to continue processing until the VTV has been successfully written to the receiving VTSS.

The benefit of synchronous replication is realized in the case of a disaster at the primary site, the remote site can be brought online and processing can continue with a minimum of recovery processing time.

Pros:

- ◇ VTVs are replicated between Sender & Receiver as an automatic event.
- ◇ Replication completed without host intervention.

Cons:

- ◇ Application jobs run longer:
- ◇ Replication must complete before application can continue.

VTSS Clustering and Immediate Migration

Migration is the process of copying a VTV from the VTSS to the physical tape media. Migration can occur from either the sending or receiving VTSS, but is typically set up to be performed by both VTSSs. As soon as VTVs are replicated to the receiving VTSS, they are queued for immediate migration regardless of the IMMED() management class setting. In addition to enforcing an immediate migration, VSM clustering enforces mirrored Real Tape Drive (RTD) model configurations on clustered VTSSs, which ensures that if one VTSS goes offline the other can continue migrating VTVs. However, the ACS associated with the receiving VTSS is the default library for migrations resulting from replication.

Never use IMMED(YES) in a management class that specifies replication. Immediate Migrate is enforced with replication and use of the IMMED(YES) parameter causes a race condition that results in degraded migration performance.

Mixed Asynchronous and Synchronous Modes

Both Asynchronous and Synchronous replication modes can be assigned to different management classes writing to the same VTSS. By mixing replication modes you can tune your replication techniques to business continuance dataset availability requirements. The REPLICAT() parameter accepts NO for no replication, YES for asynchronous replication and YES_SYNC for synchronous replication.

The synchronous replication tasks get serviced by the Clinks using what might appear to be a bias or preferencing approach since these tasks can use any available Clink. In contrast, the asynchronous replication tasks are assigned to a specific Clink and must wait until that Clink becomes available. If a Clink becomes available and the next replication task is asynchronous, its designated Clink id is checked, if the Clink id for that task does not match, a synchronous replication will be serviced ahead of the queued asynchronous replication.

In the following figures of a hypothetical queue, the two Clinks service the queue of both synchronous and asynchronous replication tasks. On the left figure, since the first two tasks in the queue are synchronous replication tasks, the VTSS will automatically queue them for the first available Clink. In the figure on the right, the second synchronous replication task finishes first, but the next asynchronous task is assigned to Clink1. Since Clink1 is busy, Async CI1 task must wait until it becomes available; consequently the next task in the queue gets selected which in this case was synchronous. If Async CI2 had been next it would also have been selected since Clink2 is the associated Clink.

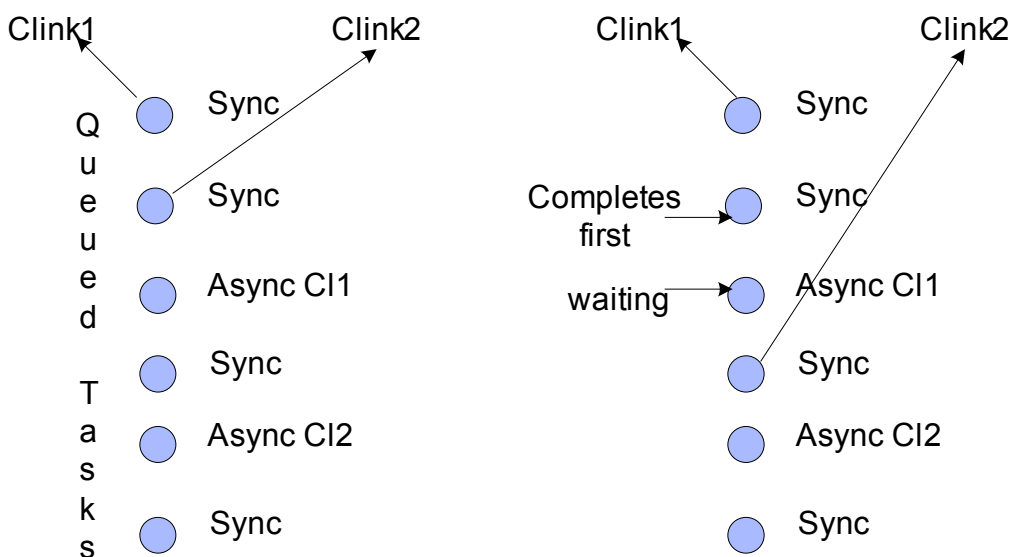


Figure 1: Clink Replication Queue

Synchronous Replication Messages

The new VTCS message SLS6900I will reflect three possible results from synchronous replication, successful, failed or dropped:

SLS6900I SYNCHRONOUS REPLICATON SUCCESSFUL FOR VTV vvvvv FROM VTSS PR1XXXXXXXXX TO VTSS SECXXXXXXXXX

SLS6900I SYNCHRONOUS REPLICATON FAILED FOR VTV vvvvv FROM VTSS PR1XXXXXXXXX TO VTSS SECXXXXXXXXX

SLS6900I SYNCHRONOUS REPLICATON DROPPED FOR VTV vvvvv FROM VTSS PR1XXXXXXXXX TO VTSS SECXXXXXXXXX

A failed synchronous replication indicates the VTSS attempted the replication but failed. A dropped synchronous replication indicates that the replicate was not attempted because

there was no Clink available or because the total elapsed time to write and replicate the VTV took greater than 40 minutes or estimated by the VTSS to take greater than 40 minutes.

When an actual failure occurs, a SLS6758I or SLS6751I message will be produced identifying the Clink that was being used and the nature of the failure. If the Clink failed then recovery is attempted and the affected Clink will be reset. Once the recovery has been completed, the Clink will be reported as online.

In the event of an unsuccessful synchronous replication attempt (failed or dropped) VTCS will drive an asynchronous replication and generate message SLS6749I accordingly. It is important to note that VTCS will not make additional replication attempts if this second, asynchronous recovery attempt fails. The following messages are generated to easily distinguish between successful synchronous and asynchronous replications in the system log message:

SLS6749I ASYNCH REPLICATION SUCCESSFUL FOR VTV *vvvvv* FROM VTSS *PRIXXXXXXXXXX* TO VTSS *SECXXXXXXXXX* ON CLINK *CLINKID*

SLS6900I SYNCHRONOUS REPLICATON SUCCESSFUL FOR VTV *vvvvv* FROM VTSS *PRIXXXXXXXXXX* TO VTSS *SECXXXXXXXXX*

The following messages are displayed in the HSC log upon Clink startup:

```
>SLS6759I SYNCHRONOUS CLINK 4 CHANIF 0M VTSS VTSS26 NOW ONLINE
>SLS6759I SYNCHRONOUS CLINK 5 CHANIF 0O VTSS VTSS26 NOW ONLINE
>SLS6759I SYNCHRONOUS CLINK 5 CHANIF 0O VTSS VTSSV NOW ONLINE
>SLS6759I SYNCHRONOUS CLINK 4 CHANIF 0M VTSS VTSSV NOW ONLINE
```

The cluster can also be displayed to verify online synchronous state:

```
NAME      VTSS  STATE DIRECTION VTSS  STATE  MODE
CLUSTER1 VTSS26 ONLINE <-----> VTSSV  ONLINE SYNC-REPLICATE
```

The Clinks can be displayed to verify online state as well:

```
VTSS  CLINK STATUS  USAGE
VTSS26  4 0M ON-SYNC  FREE
         5 0O ON-SYNC  FREE
VTSSV   4 0M ON-SYNC  FREE
         5 0O ON-SYNC  FREE
```

Chapter 3: Configurations

This chapter details synchronous clustering hardware environments, VSM Clink configurations and channel extension usage. Customer emulation workloads used during testing are also explained.

Workload Strategy

Each job suite concentrates on a single uncompressed dataset size ranging from 256KB to 1GB. Uncompressible data was used to create the workload datasets to provide a worst case scenario when evaluating replication duration times. A burst of jobs was submitted every 30 seconds with each job uniquely named to enable it to use its own initiator. Each job contains a single step which uses the IBM utility program IEBDG to generate random uncompressible records written directly to VTVs. A maximum of 64 initiators were made available to a job suite and each job suite ran sequentially. The test computer system was not dedicated and some variation in the data collected is evident but the overall results are consistent and repeatable.

Distances Supported (Native and Extended)

All distance testing described in this white paper were executed using direct connected VSM subsystems through FICON switches and an Anue distance emulator. The maximum distance tested was 3000km using a gigabit Ethernet connection.

FICON long wavelength maximum distance (unrepeated) 10km, 20km at 1 Gb/sec or 10km at 2 Gb/sec

FICON short wavelength maximum distance (unrepeated) 500m at 1 Gb/sec or 300m at 2 Gb/sec

Environmental Configurations

The channel configurations in this section depict the test environments used to emulate customer environments by submitting workloads that write dataset sizes of 256KB, 1MB, 10MB, 100MB, and 1GB. The workloads were built using volume backups and one step jobs writing continuous data to VTVs on both VTSSs in uni-directional and bi-directional environments. Each workload was run with 400MB, 800MB, and 2GB VTV sizes as well as with large (62KB) and standard (32KB) page size. The workloads generated 1000 mounts per hour in both bi-directional and uni-directional cluster environments. Distance testing was also conducted using the following distances: zero, 100 km, 1000 km, and 3000 km.

Eight Clink Cluster Configuration

An eight Clink zero distance environment including a VSM4 to VSM4 cluster was established to allow simultaneous throughput with maximum recommended Clink configuration. Four channels in each direction are dedicated to the cluster environment leaving four interface cards for RTDs and host connections to be configured to the VTSSs. Both ESCON and

FICON host channels were used for testing in this environment. Test results did not clearly show that the tradeoff between more Clinks to fewer RTDs was an improvement.

Zero Distance

8 Channel Cluster

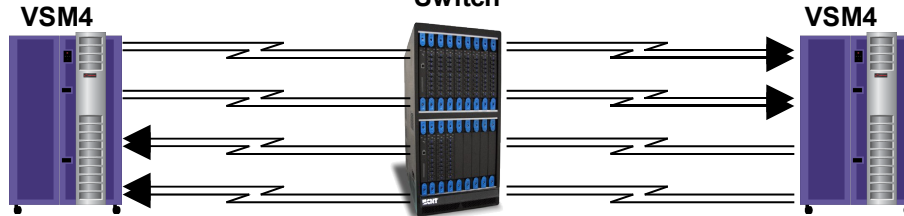


Figure 2: 8 Clink Bi-directional Cluster

Two and Four Clink Cluster Configurations

A zero distance environment was established to define a hardware configuration for a VSM5 to VSM4 cluster. Two channels in each direction are dedicated to the cluster environment leaving six interface cards for RTDs and host connections to be configured to the VTSSs. This same environment was used for uni-directional and bi-directional testing, but reduced to two Clinks for the uni-directional tests. The uni-directional tests always established the VSM5 as the sending VTSS.

Zero Distance

4 Channel Cluster

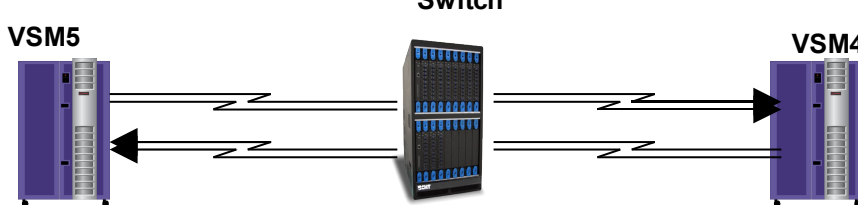


Figure 3: 4 Clink Bi-directional Cluster

Zero Distance

2 Channel Cluster

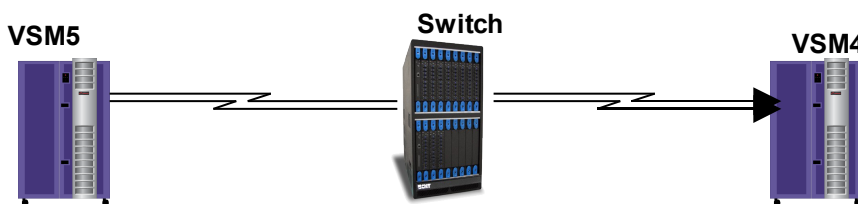


Figure 3: 2 Clink Uni-directional Cluster

Channel Extension Configurations

The following distance emulation environments were tested across the VTSS clusters including channel extension hardware from three vendors. Two channels are dedicated to the cluster environment leaving six interface cards for RTDs and host connections to be configured to the VTSSs.

Channel extension hardware vendor interoperability was validated with Brocade USD-X, Ciena 2000, and Cisco 9513 Channel Extension. Distance was emulated using an Anue distance emulator for all of the following configurations. The distance emulation tests revealed that A Gigabit Ethernet network infrastructure is best for use with synchronous replication.

Ciena Channel Extension

The maximum distance tested was 1000km using an OC-3 (155.52 Mbps) connection between two Ciena 2000 channel extenders.

100 km / 1000 km / 3000 km Distances

2 Channel Cluster



Figure 4: 2 Clink Ciena Channel Extension Configuration

Brocade/McData Channel Extension

The maximum distance tested was 3000km using a Gigabit Ethernet connection between two USD-X (9112) channel extenders.

100 km / 1000 km / 3000 km Distances

2 Channel Cluster



Figure 5: 2 Clink USD-X Channel Extension Configuration

Cisco Channel Extension

The maximum distance tested was 3000km using a Gigabit Ethernet connection between two Cisco 9513 switches.

100 km / 1000 km / 3000 km Distances

2 Channel Cluster



Figure 6: 2 Clink Cisco Channel Extension Configuration

Chapter 4: Operational Considerations

This chapter focuses on the operational aspects of application types, job throughput, VTCS startup, VTD usage, and VCF card configuration.

Application Considerations

The best tape applications for synchronous replication are Disaster Recovery and Records Retention oriented. Several examples of these applications include: Database Backup/Restore; Infrastructure Backup/Restore; Data Archival; and Data Vaulting. The backup datasets can be replicated to the second VTSS located at a remote site immediately using synchronous replication. The backup datasets are then copied to physical tape for tertiary storage, and then finally archived in the remote ACS Vault. Storage Management applications like HSM are good candidates for virtual tape, which provide faster access to the backed up data for recovery purposes. Any tape application that is taking advantage of tape virtualization features can also benefit from the synchronous replication capability.

The figures below depict a local production environment in Denver and a remote data archival and vaulting site in Reno. The remote site can also be used to store backups for disaster recovery purposes.

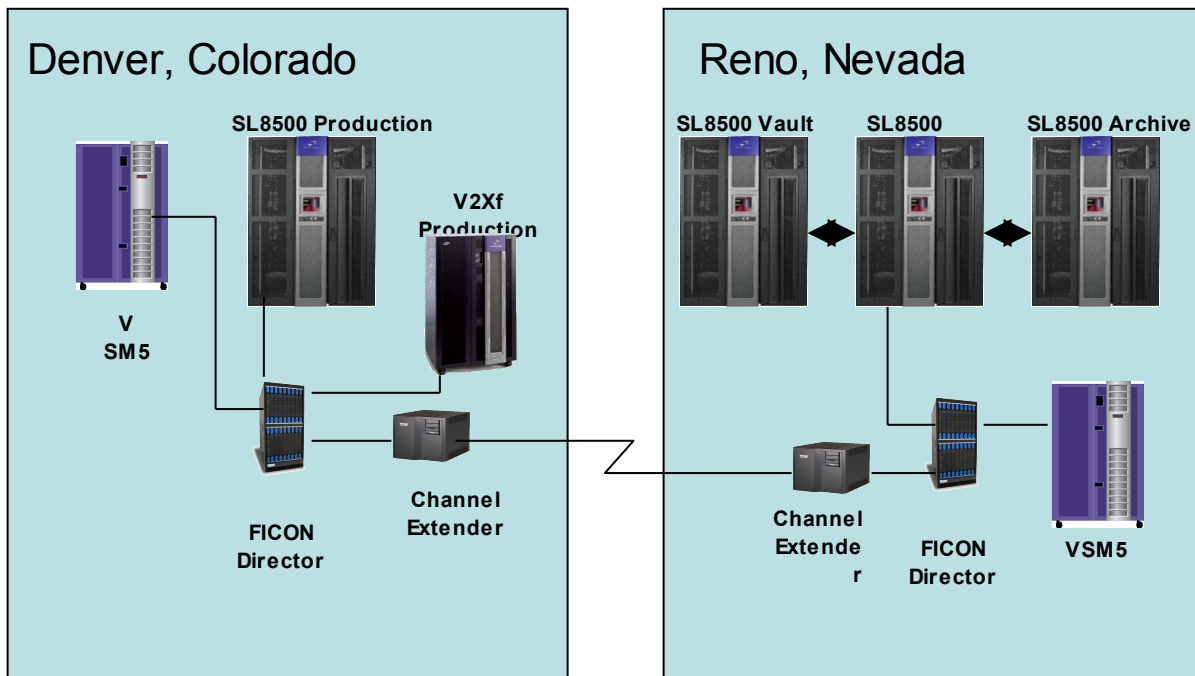


Figure 7: Local Production and Remote Backup Sites

Depending on your applications and throughput requirements there may be a practical distance limit for using synchronous replication since the wait time may affect the application. This practical limit is evaluated on a case by case basis.

VTD MIH Timer Requirements

Synchronous replication transfers are not considered complete until the VTV has been replicated and rewind/unload is complete. To ensure that the Clink is not tied up indefinitely a 40 minute timer is set inside the sending VTSS which is applied to the synchronous replication task. In the event that the replication can't complete within the time limit, the VTV is unloaded and deleted from the receiver VTSS and then rescheduled by VTCS as an asynchronous replication.

The VTD device MIH time out must be set to 45:00 minutes on each host before implementing Synchronous replication to ensure an MIH time out is not encountered during synchronous replication. If an MIH time out occurs, the results are unpredictable and can include job failure and replication failure, even though the replication may appear to have occurred. The reason the replication may appear to have occurred is because the VTSS is unaware of the host MIH condition.

In order to verify that MIH is set properly, issue the following command:

- D IOS,MIH,DEV=(xxxx-xxxx)

Use the following command to set the MIH timeout value:

- SETIOS MIH,DEV=(xxxx-xxxx),TIME=45:00

VTV Page Size

The VTSS microcode level that supports synchronous replication also supports large page size; previous bench marks show an increased performance of 20% can be achieved with sustained writes when large page size is used. See *VSM Large Tape-Page Feature* white paper for details of this benchmark. In order to take advantage of this feature from a host use NCS/VTCS 6.2 and a G level CDS.

NCS/VTCS 6.1 does not support large page size but does support standard page size with synchronous replication and is compatible with 6.2 and a shared F level CDS on separate LPARs.

Multiple Dataset VTVs

Synchronous replication supports multiple dataset VTVs; however, the VTV replication does not occur until the VTV is unloaded. The benefit to this approach is that a single job may be established to associate a group of datasets on a VTV which are only replicated at the end of the job, therefore, the replication delay occurs. This approach may not always be a desirable condition; in which case don't stack multiple datasets on a single VTV that will be synchronously replicated.

Multiple VTCS Startup

Data centers using multiple HSC/VTCS images sharing a single CDS must quiesce any active Clinks to offline before starting or restarting a HSC/VTCS on another image to prevent synchronous replications from flushing and being converted to asynchronous replications.

Synchronous replication queued transfers may be flushed if a second VTCS subsystem is started on another LPAR because VTCS will flush all current synchronous replications so that initialization can complete in a timely sequence. All synchronous replications that were flushed are re-driven as asynchronous replications.

When VTCS starts it will issue an I/O to every VTD device that is defined. If the MVS UCB for a VTD is marked as busy - MVS will not forward the I/O to the VTSS. Normally the delay would not be noticed during startup but if the device is busy performing an extended rewind/unload for synchronous replication the delay is potentially 40 minutes.

VTD Allocation and Usage

Enabling the cluster configuration requires the first 16 VTDs offline, these VTD addresses are reserved for use by the Clinks and cannot be used for mount requests. If these VTDs are varied online to MVS, ASE (assigned elsewhere) and ECAM errors will occur for this VTD range. Consequently, unless these VTDs are offline, the Clinks won't vary online.

All hosts running VTCS must have access to at least one of the first 16 VTDs, failure to do so will disable synchronous replication. MVS Guest systems require these 16 VTDs to be attached to the MVS Guest, but in the offline state.

VTSS VCF Card Configuration

VCF2 and above interface cards support FICON channel and Clink connectivity, VSM4 will support mixed mode, FICON and ESCON host channels and RTDs, VSM5 only supports FICON interface cards.

Bi-directional Cluster Configuration

The bi-directional cluster replication feature became available with NCS/VTCS 6.1 product release. In a bi-directional cluster configuration, the storage administrator connects two VTSSs together using FICON Clinks in order to perform replication in both directions. Both VTSSs in the cluster perform replications so VTV replication flow will go in both directions simultaneously. Each VTSS in the cluster can provide VTV migration support for the other while servicing a production workload for local applications.

Configure Clink ports on VCF cards for Clinks or Host connectivity, it is not recommended to combine Clinks and RTDs on the same VCF port since the activity on the card is serialized and the RTD traffic would interfere with the Clink data transfer activity.

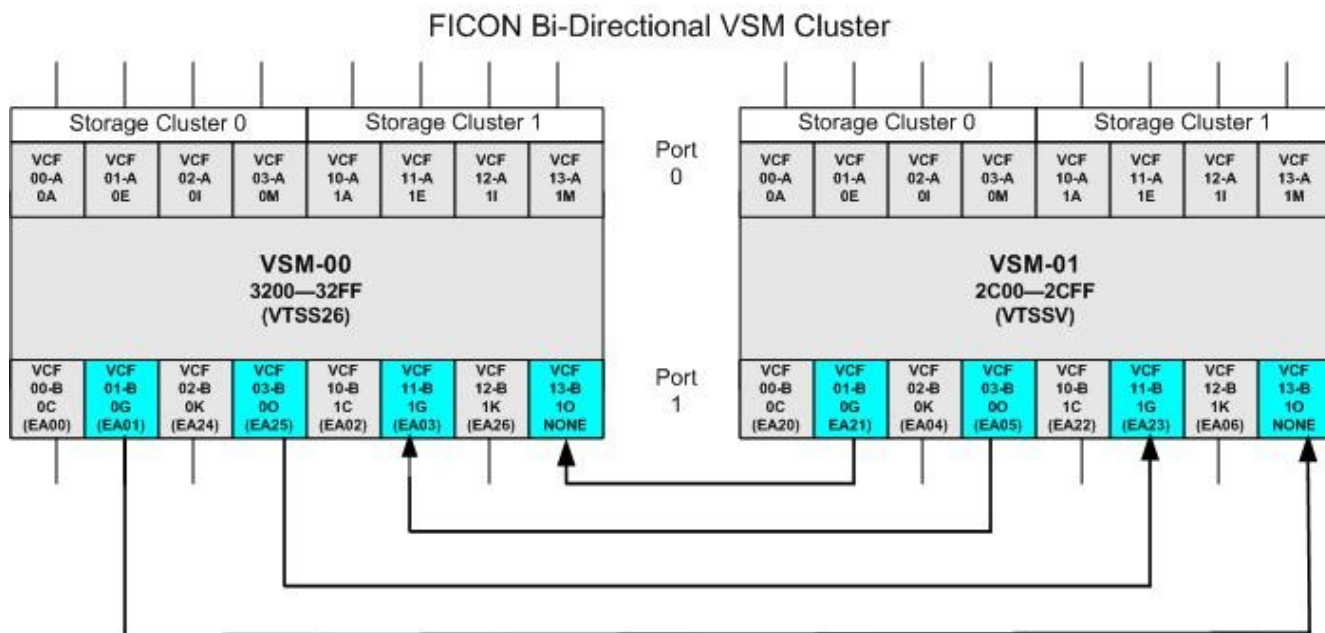


Figure 8: FICON Bi-Directional VCF Configuration

The following VTCS config parmlib cluster statements are required to define the above bi-directional cluster. Both VTSSs send and receive data across the designated Clinks. The Clinks are configured to opposite storage clusters on the VTSSs. Failure to setup the Clinks with the opposite clusters will result in timeouts and errors during replication.

```

CLUSTER NAME=CLUSTER1 VTSS (VTSS26,VTSSV)
  CLINK VTSS=VTSS26 CHANIF=1G
  CLINK VTSS=VTSS26 CHANIF=1O
  CLINK VTSS=VTSSV CHANIF=1G
  CLINK VTSS=VTSSV CHANIF=1O
    
```

Uni-directional Cluster Configuration

A uni-directional cluster defines one way communication between two VTSSs, the storage administrator can connect VTSSs for replication in one direction. The sending VTSS replicates to the receiver VTSS so the VTVs flow in only one direction.

In this configuration, each VTSS services a subset of the production workload and functions as a “warm standby” for the other.

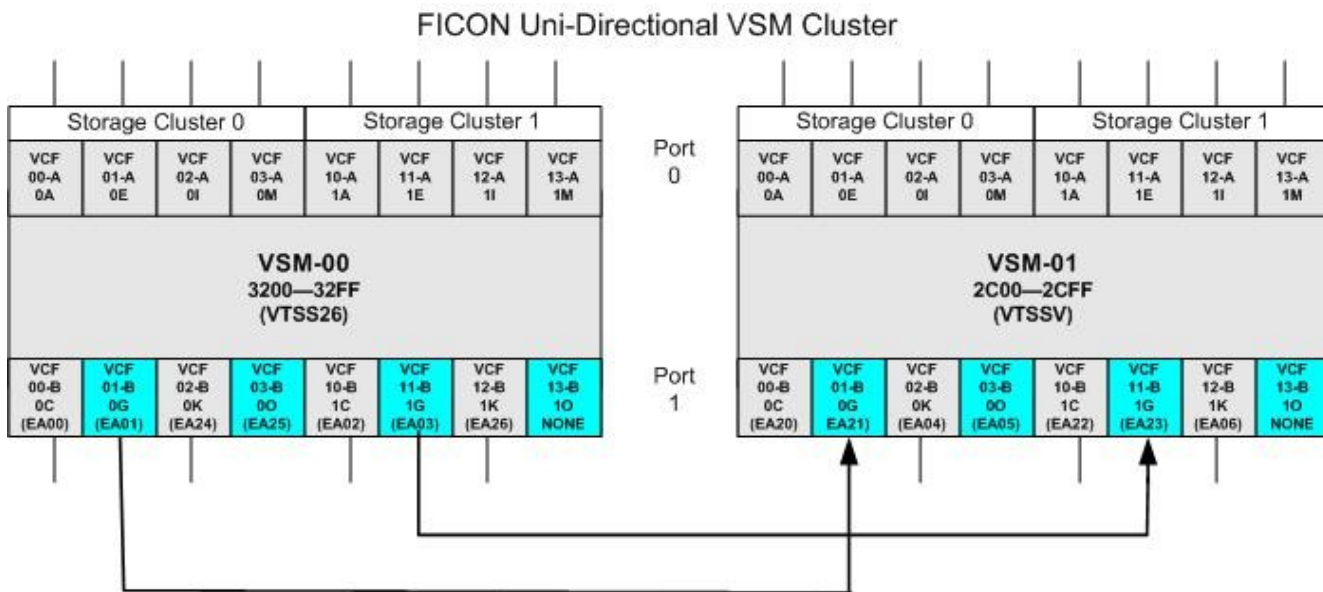


Figure 9: FICON Uni-Directional VCF Configuration

The following VTCS config parmlib cluster statements are required to define the above uni-directional cluster. VTSS26 will send and VTSSV will receive replicated VTVs across the designated Clinks. The Clinks are configured to opposite storage clusters on the VTSSs for maximum redundancy.

```
CLUSTER NAME=CLUSTER1 VTSS (VTSS26,VTSSV)
    CLINK VTSS=VTSS26 CHANIF=0G
    CLINK VTSS=VTSS26 CHANIF=1G
```

TMS Considerations

No change is required to TMS when using synchronous replication. The TMS catalog is updated with the change in scratch status and dataset name assignment immediately after the VTV mount whether the dataset is to be cataloged or not. If more than one VTV is required to accommodate the dataset size, each of the required TMS volumes are allocated as requested and the TMS catalog is updated when each VTV mount request is satisfied. If the tape dataset was cataloged, the system catalog is updated during dataset close.

If the job fails during a VTV write, the TMS and the system catalogs are updated. If multiple VTVs are required to accommodate the dataset and then the write fails; the VTVs are allocated up to the point of failure. Finally, the VTV is replicated even though it may be incomplete.

Synchronous replication doesn't update the TMS and system catalogs on the remote site, consequently, additional procedures such as PPRC or backups are needed to ensure that replicated VTVs, TMS and system catalogs are synchronized.

Chapter 5: Recommendations and Important Considerations

This section presents some recommendations and considerations when enabling synchronous replication as well as some potential pitfalls to avoid.

Required Steps to Enable Synchronous Replication

Review the following steps when establishing a synchronous replication environment and consult the *Sun StorageTek Virtual Tape Control System Software: Synchronous Replication Guide* for details.

- VSM4 or VSM5 hardware with FICON Interfaces.
- Install required microcode level on both VTSSs (see prerequisites).
- Ensure VTD MIH is set to 45:00 on all hosts.
- Enabled cluster feature on both VTSSs.
- Ensure PCAP on receiving VTSS can accommodate replication capacity.
- Define FICON Clinks on the VTSSs.
- Install Synchronous Replication VTCS PTFs on all hosts (see prerequisites).
- Configure RTDs to both VTSSs.
- Configure VTCS Global statement SYNCHREP(YES).
- Create Management Class with REPLICAT(YES_SYNC).
- Vary the first 16 VTDs offline on both VTSSs.
- Start host software subsystem and observe Clink status.

VTSS Microcode Levels

Both the sending and receiving VTSS must have synchronous replication capability otherwise an error condition exists and VTCS will not perform any VTV replication, see prerequisites in this white paper. It should be noted that in the case where only one VTSS supports synchronous replication the cluster is still capable of asynchronous replication and the configuration must be set up for asynchronous clustering to avoid configuration errors.

Serialization During Synchronous Replication

When using Synchronous Replication ensure your applications are not holding datasets or experiencing database enqueues and locks across tape dataset writes. If the tape dataset is multi-volume and goes to a new volume, that new volume will have to wait on the synchronous replication to complete before continuing regardless of any 'UNIT=' DD statement parameters.

Attempts to expedite processing with more than one VTD, by use of the UNIT=TAPE,2 parameter on the JCL data definition, will not cause the next volume request to use another VTD while the previous VTD is replicating.

Chapter 6: Replication Job Results and Conclusions

During the process of understanding how to efficiently use synchronous replication many jobs were executed and that data was summarized in the following chart and table.

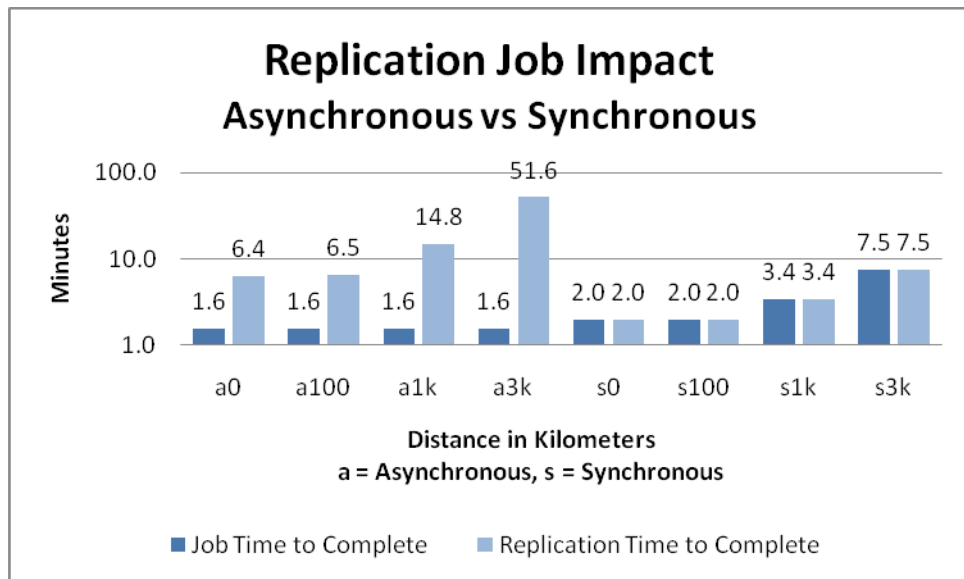


Figure 10: Replication results by type, asynchronous and synchronous

Figure 10 graphs the replication’s impact on job duration. During synchronous replication there is a direct correlation between job duration and replication duration since the job has to wait for the replication to finish before the job can end. During asynchronous replication, the replication is an independent operation, therefore it has no impact on the job’s duration.

Distance in Kilometers	Synchronous Replication Job Duration Increase
Direct Connect	25.74%
100 Kilometers	26.69%
1k Kilometers	118.19%
3k Kilometers	380.82%

Figure 11: Synchronous Replication time to completion percentages

Figure 11 provides in percentages, the change in job duration between a job that was run using asynchronous replication verses a job that was run using synchronous replication.

The data for figures 10 and 11 was collected in the Sun Mainframe Performance lab using controlled conditions on a dedicated z/OS processor. Ten job streams of twenty jobs running serially were submitted. Each bar in figure 10 represents a complete job within the stream set’s average execution time. Each job wrote 1 full 800MB VTV compressed at a 4:1 rato.

The zero distance runs were direct Clink connections and the distance runs used two FICON channel extenders with a distance emulator between them.

Synchronous Replication is indicated in the following situations:

- When the replication copy must be on the receiver VTSS before the application continues. If the tape dataset is 1GB or larger use smaller VTVs to ensure the synchronous replication does not convert to asynchronous.
- When synchronous replication is desired for increased business continuance and the data center can afford extended job durations. If the jobs are spread out by time then the VTSS does a very good job of scheduling and performing the replication quickly.
- In many cases and for smaller tape dataset sizes the difference between using synchronous replication and asynchronous replication is not noticeable. In the case of large tape datasets 1GB or greater we recommend running a test job synchronously before permanently committing to synchronous replication.

Asynchronous replication is indicated in the following situations:

- The job must complete in a timely manner due to batch window restrictions.
- An immediate replication copy is not required, VTCS will usually have the replication scheduled and completed soon after the job completes.
- Synchronous replication distance becomes prohibitive, 100 to 1000km distance is not prohibitive but 3000km starts to really extend the job duration.

The distance emulation tests revealed that A Gigabit Ethernet network infrastructure is best for use with synchronous replication.

Synchronous Replication VTV Size Recommendation

Workloads with varied sizes of VTVs revealed that in general, by using smaller VTV sizes with large tape datasets the chance of a dropped synchronous replication can be reduced or eliminated. The duration of the replication is dependent on the size of the VTV. Instead of transferring large VTVs which take longer, creating several smaller VTV transfers will complete more quickly resulting in more wait periods of a shorter duration allowing the data to reach the receiving VTSS sooner. *Note: this technique may elongate the job step.*

Appendix A -- References:

- **Sun StorageTek Virtual Tape Control System Software: Synchronous Replication Guide**
- **VSM Large Tape-Page Feature White Paper, May 17, 2007**
Version 1.0 Bill Gray John Warner Kevin Zwack
- **IBM Redbook IBM TotalStorage Solutions for Disaster Recovery, SG24-6547.**
- **Risky Thinking – On Risk Management, Disaster Recovery, and Business Continuity** <http://www.riskythinking.com/>
- **Computer Technology Review -- Data Replication** by Jim McKinstry
http://www.wmpi.com/index.php?option=com_content&task=view&id=446&Itemid=44
- **Recovery Specialties Storage and Business Continuity consulting for z/Series Environments** <http://recoveryspecialties.com/tape05.html>
- **Brocade**
<http://www.brocade.com/index.jsp>
- **Cisco**
<http://www.cisco.com>
- **Ciena**
<http://www.ciena.com>