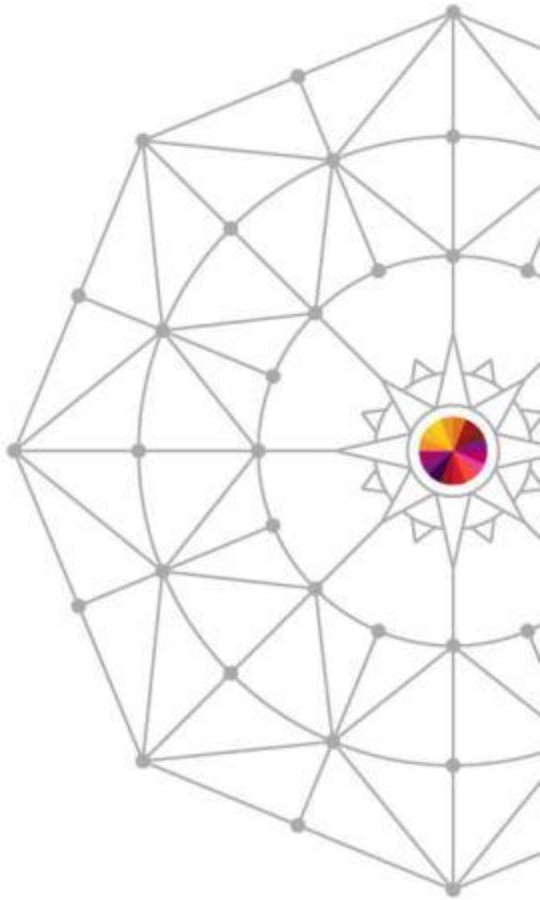
#SHAREorg

SHARE
Technology · Connections · Results

Big Data Storage in the Cloud

Russell Witt
Scott Arnett
CA Technologies

Tuesday, March 11
Session Number 15288



SHARE in Anaheim

Abstract

Need to reduce the cost of managing storage while improving storage utilization? Need a pay-as-you-go storage solution to keep up with business demands? CA Technologies product experts will discuss how you can leverage public, private or hybrid-cloud storage on-demand for your ever changing storage needs. CA Cloud Storage for System z will provide secure, affordable and scalable disaster recovery options to retrieve data remotely and without calling your disaster recovery vendor to manually deliver your tape for restore at your backup site. Join the discussion on benefits of moving to a cloud solution vs. managing and maintaining tape hardware and media on site.

Cloud Storage – Object storage vs. file storage

- File Storage – data is managed in a hierarchical format. Each file being managed has a unique name associated with it (drive:/file name or catalog data set name)
- Object Storage – data is managed as objects. It can be implemented at the device level (object storage device), system level or even the interface level. Used more often to store massive amounts of unstructured data (photos on Facebook for example).

Cloud Storage – Private / Public

- Private Cloud – A storage solution based within the corporate firewall and under control of the IT department. Usage is by departments, agencies or users within the corporation is done in a similar fashion that would be done if it was a public cloud.
  - But what about a storage that exists at a remote location with a dedicated network path versus a public network path?
  - Then, instead of a remote location, if is housed at a third-party location but still owned and controlled by the IT dept?
- Public Cloud – A storage solution outside the IT department that is contracted to supply storage on-demand to the various departments, agencies or users within the corporation.

Cloud Storage – Hybrid

- A Hybrid Cloud is when a Private Cloud is supplemented with a Public Cloud. Ideally, this implementation would be done so that it behaves as if a homogeneous solution.
- Most often this is done by using a proprietary commercial storage appliance that serves as both a storage appliance in-house (Private Cloud) as well as a gateway to a Public Cloud storage solution.
- Offers the speed of a Private Cloud implementation for in-house uses; but also the advantages of a public off-site storage repository for DR and/or long-term storage.

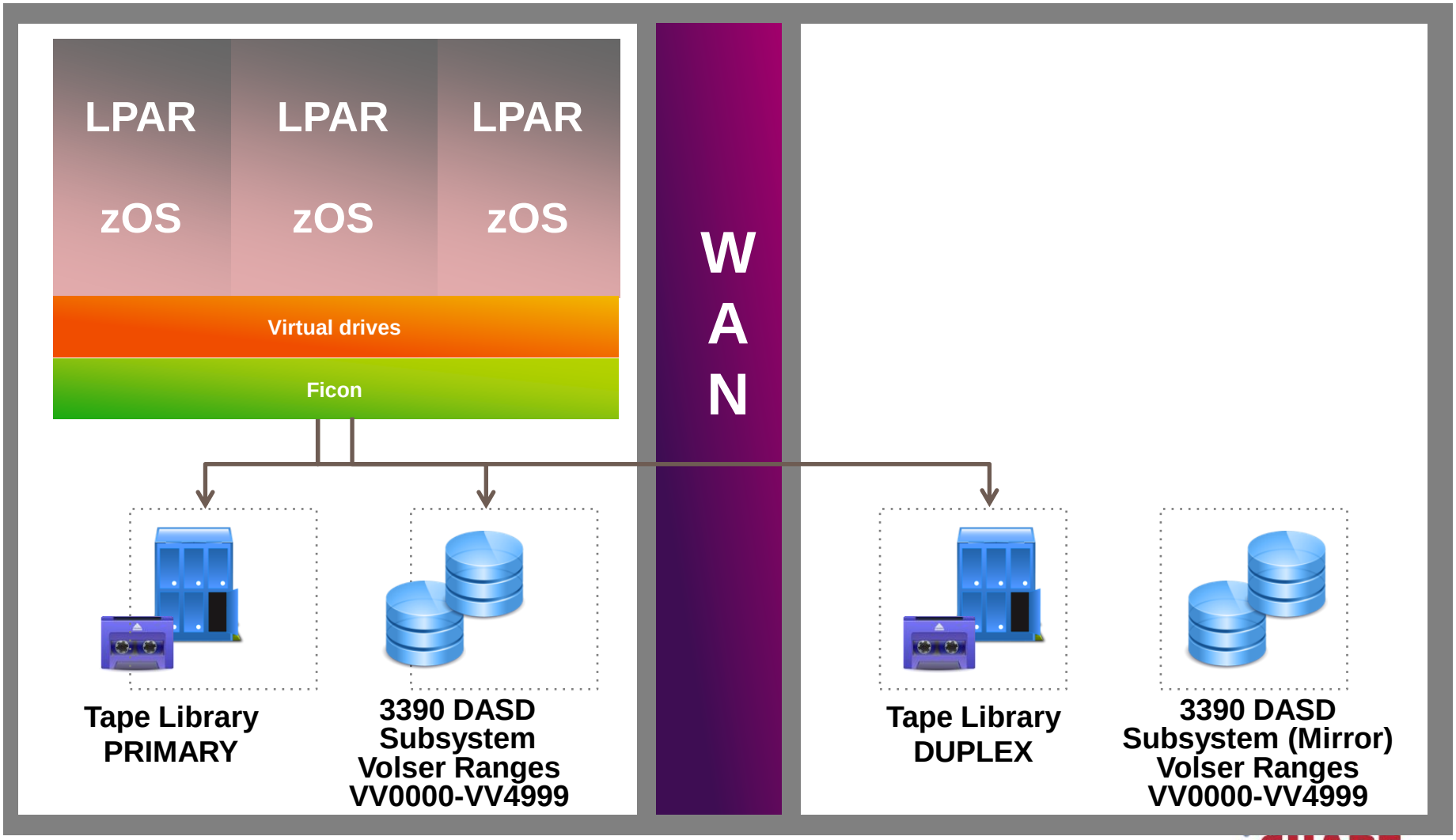
Cloud Storage – What type of z/OS Data can be sent?

- z/OS data can be broken down in many ways.
- Type of data – indexed or sequential
  - Indexed would be any type of database, VSAM files (including catalogs) even PDS or PDS/E's
  - Sequential would be any straight "flat file"
- Data Usage –
  - Is the file updated often (either DISP=MOD or individual records / blocks are updated)?
  - Is it write-once and read often
  - Is it write-once and read seldom

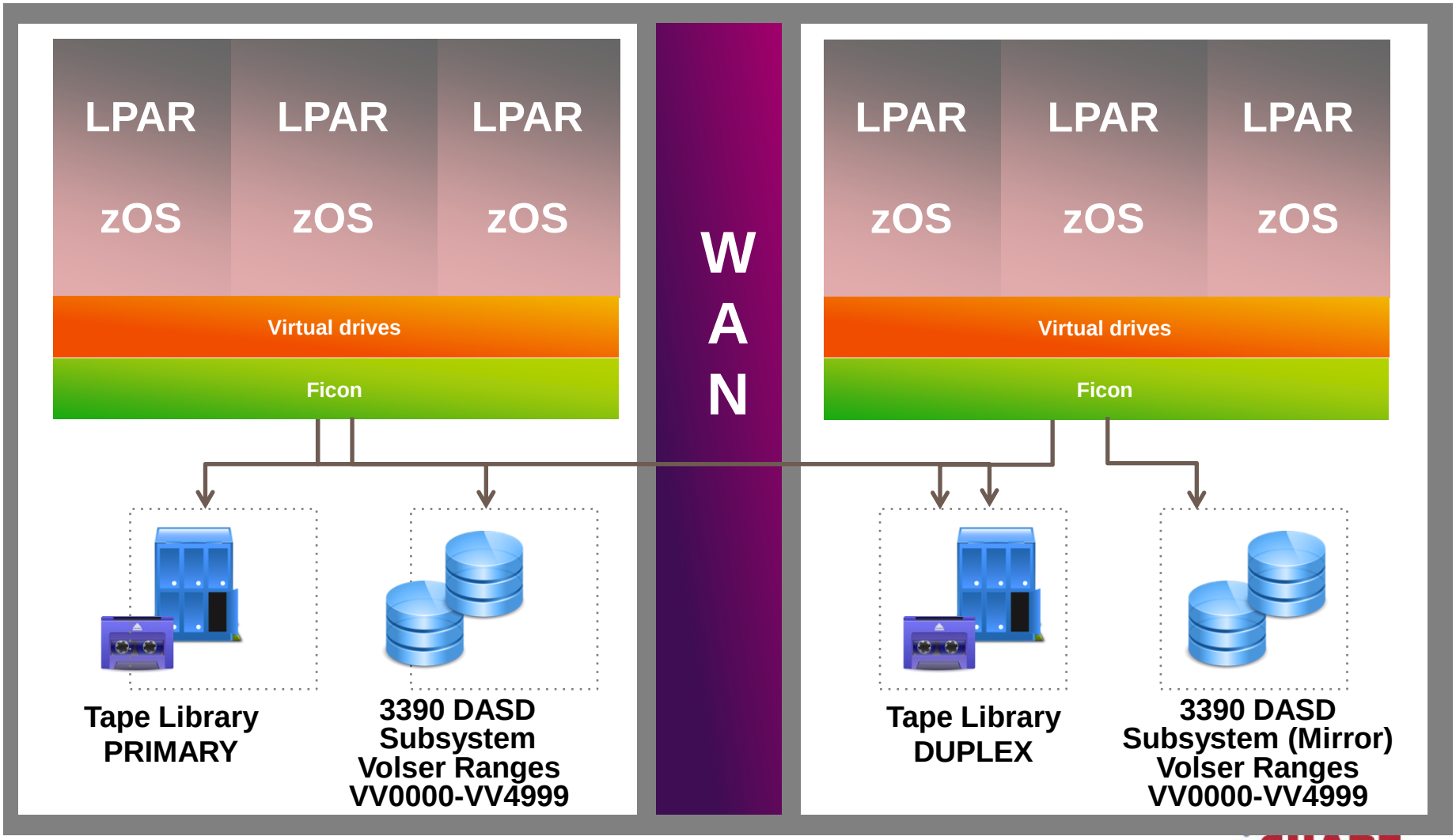
Considerations

- There are MANY considerations to take into account when looking at a Public Cloud provider. A few are;
  - Reliability – will they still be in business next year
  - Reliability – what level of RAID storage do they use
  - Encryption – how strong is the algorithm and who maintains the keys
  - Redundancy – how many copies will be maintained, and where will those copies reside (geographical dispersal may mean political dispersal)
  - Cost to store the data – how much per Gb/Tb/Pb per month
  - Cost to retrieve the data – included or extra
  - How quickly can the data be retrieved – minutes or hours

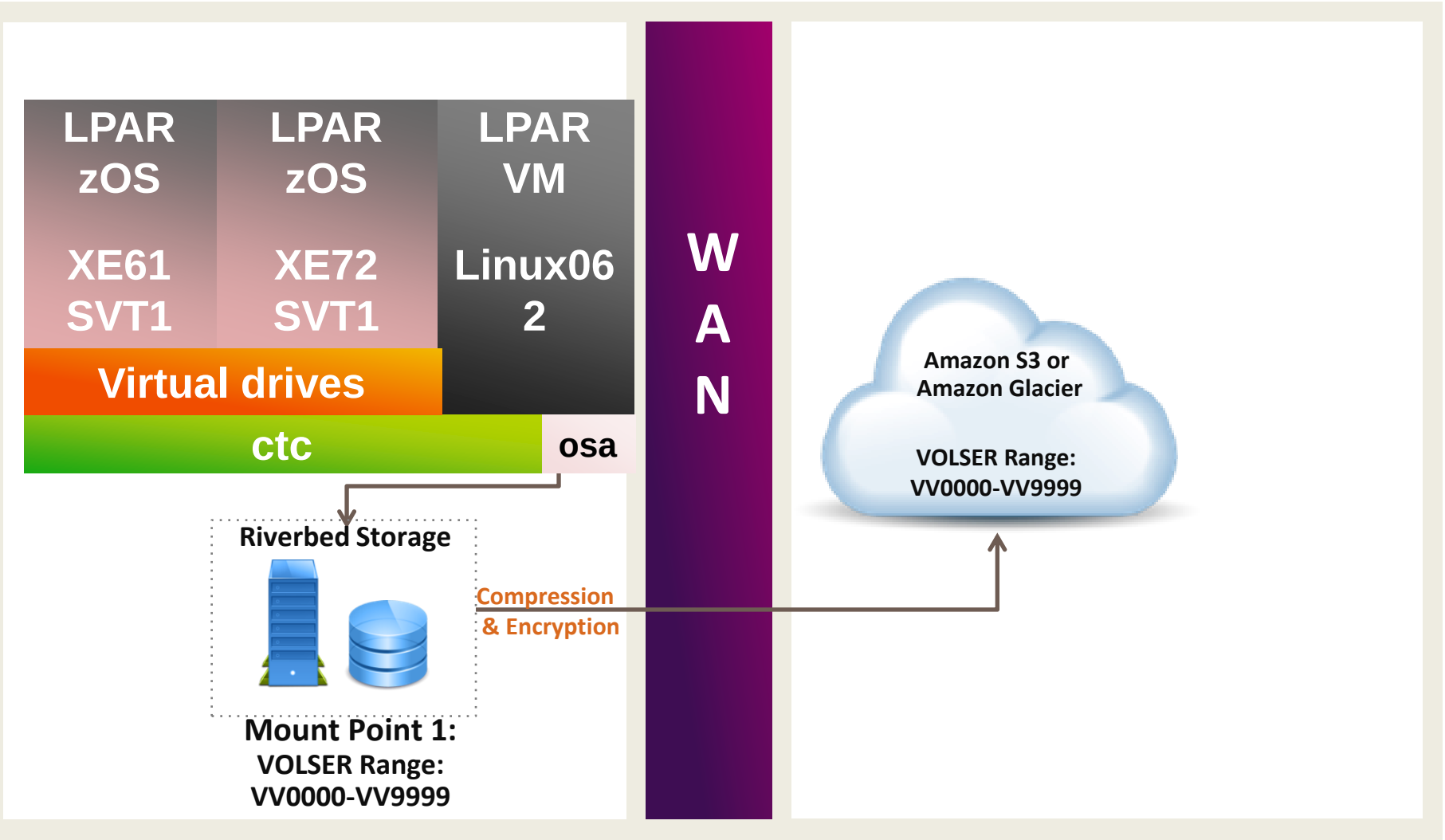
Traditional CA Vtape



Traditional Replicated CA Vtape



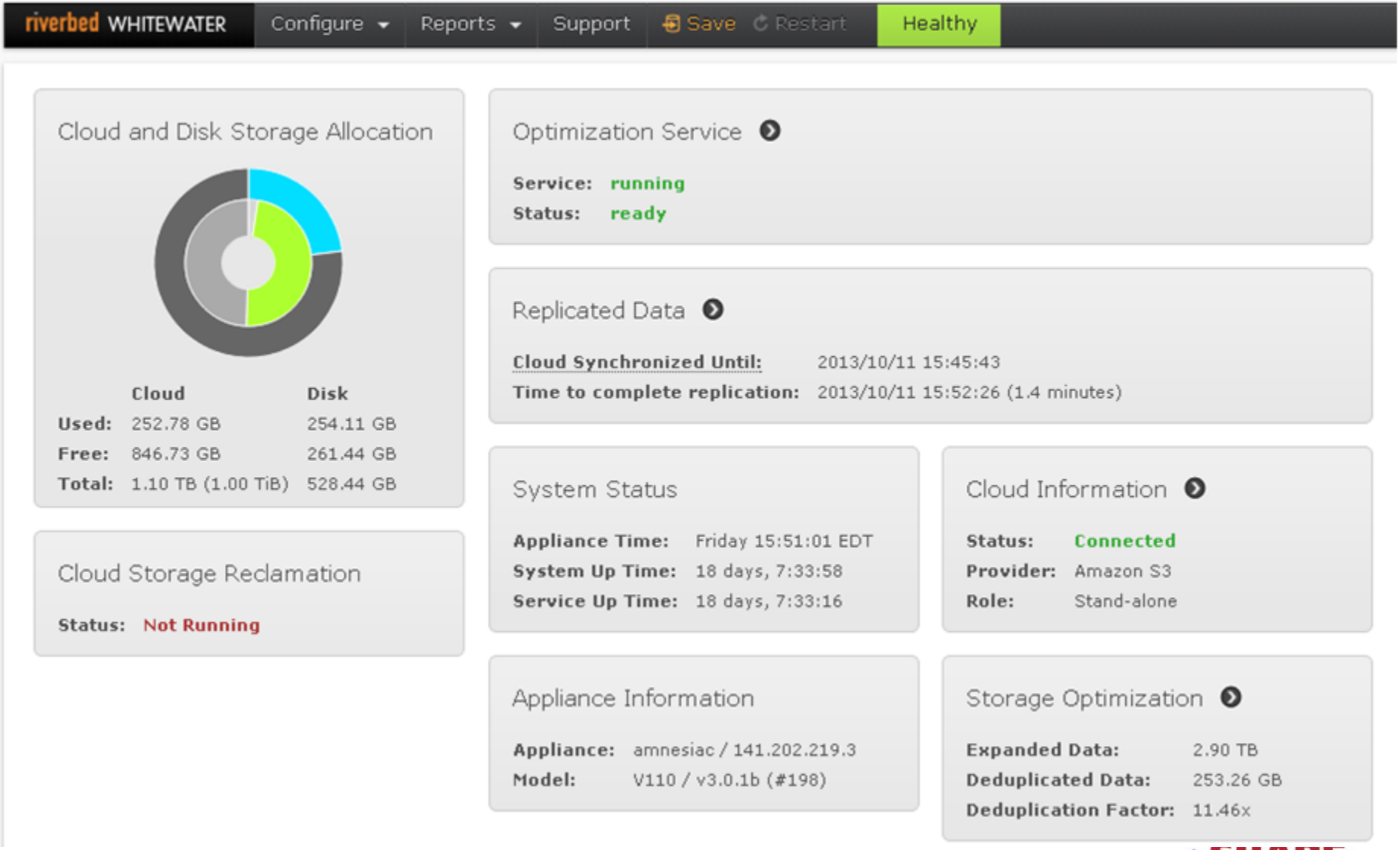
CA Cloud Storage for System z Architecture



Example Job

```
 IEF403I HELDG - STARTED - TIME=08.48.01
*IEF233A M F824,PRIVAT,SL,HELDG,STEP1,HELDA70.VTAPE4.DG1
 IEC705I TAPE ON F824,728100,SL,COMP,HELDG,STEP1,HELDA70.VTAPE4.DG1,
 MEDIA2
 IEC205I SEQOUT,HELDG,STEP1,FILESEQ=1, COMPLETE VOLUME LIST, 319
 DSN=HELDA70.VTAPE4.DG1,VOLS=728100,TOTALBLOCKS=3000
 IEF196I IEF237I 20E3 ALLOCATED TO SYS00038
 IEF196I IEF285I   SYS1.LINKLIB                                KEPT
 IEF196I IEF285I   VOL SER NOS= MVZ1CM.
 IEC205I SEQOUT,HELDG,STEP2,FILESEQ=2, COMPLETE VOLUME LIST, 323
 DSN=HELDA70.VTAPE4.DG2,VOLS=728100,TOTALBLOCKS=3000
 IEF234E K F824,728100,PVT,HELDG
 IEF404I HELDG - ENDED - TIME=08.48.25
```

CA Cloud Storage for System z Riverbed Dashboard



System Requirements - Software

- IBM z/OS - All generally supported releases
- CA Vtape VTS r12.6, with current maintenance
- IBM Linux on System z platform (zLinux)
  - SUSE Linux Enterprise Server 11 SP1 (s390x), kernel level 2.6.32.12-0.7
  - Running under z/VM recommended but not required

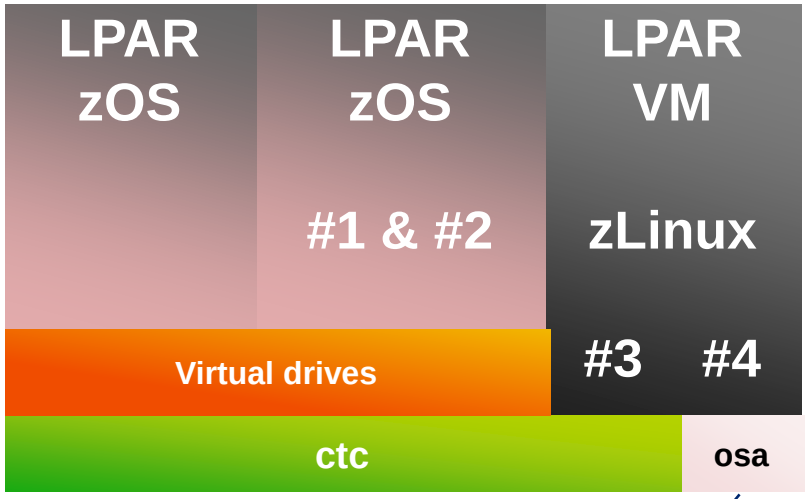
System Requirements – Hardware

- A pair of FICON CTC device addresses between the CA Vtape VTS subsystem (SVTS) running on z/OS and the zLinux server
  - Note 4 CTC addresses spread across two CHPIDs is better for throughput and availability.
- A 10-Gbps Open Systems Adapter (OSA) card connected to the Virtual Machine (VM) where the zLinux guest runs.
  - Note that a 1-Gbps OSA card is acceptable but throughput will not be as high as with the recommended 10-Gbps OSA card.
- CA Cloud Storage for System z can be configured to store data on any combination of on-premise NFS devices rather than on mainframe DASD

System Requirements – zVM & Linux on System z

- Minimum of 1 GB of free disk space (for syslog-ng, dumps, tar files, diagnostics, etc.)
- Minimum of 1 GB of memory
- Root authority to perform Network File System (NFS) mounts maintenance and installations

How it Works



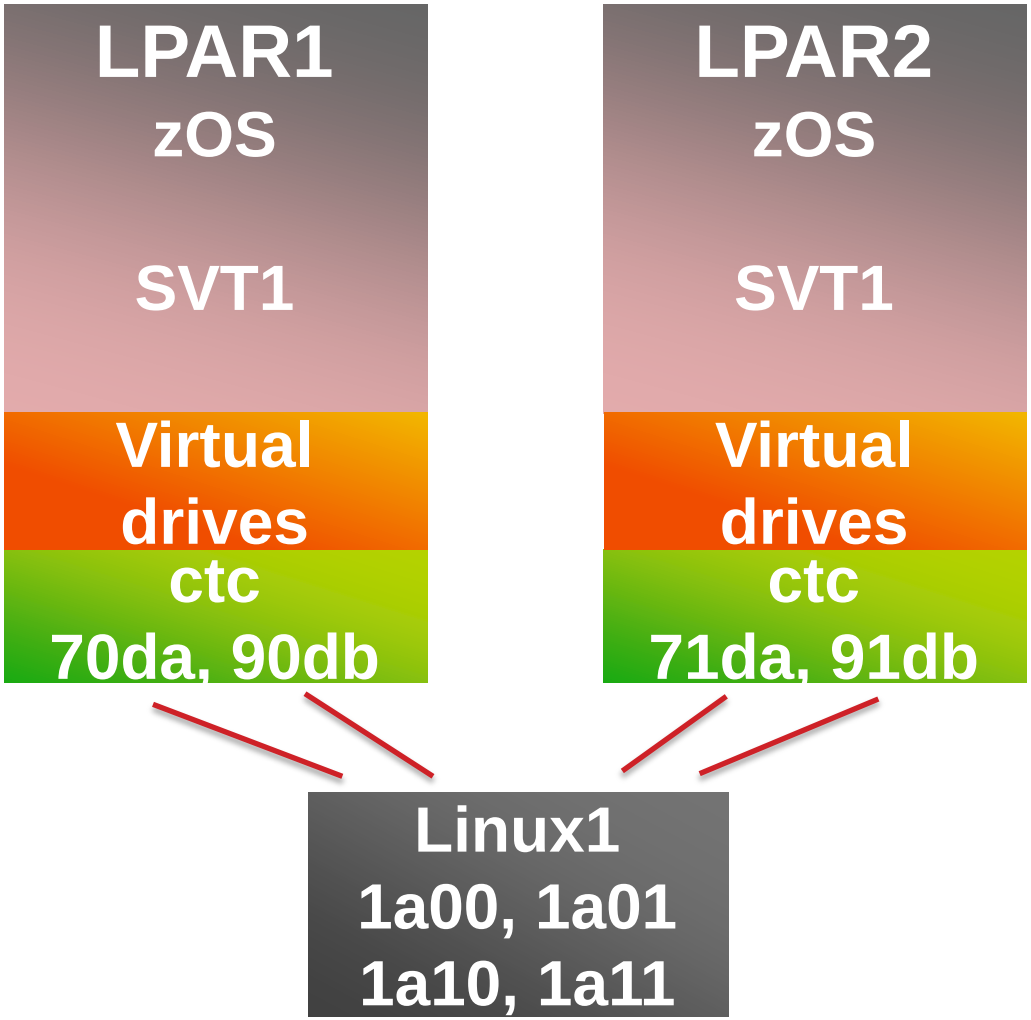
1. Mount is nearly unchanged except TCPIP is used to connect between z/OS and Linux. The Linux listener accepts the connection and forks a new process to do the work on behalf of the virtual drive.
2. The IO Engine is nearly unchanged, SSCH/IO are intercepted and the data is moved to 4MB Buffer. Only change is to use CTC pipe in place of media manager.
3. Linux receives the data over the CTC pipe.
4. Linux writes to the NFS mounted filesystem.
5. The NFS appliance is responsible for compression and encryption.

Note: This is a 100% tapeless implementation with no Backstore capability.

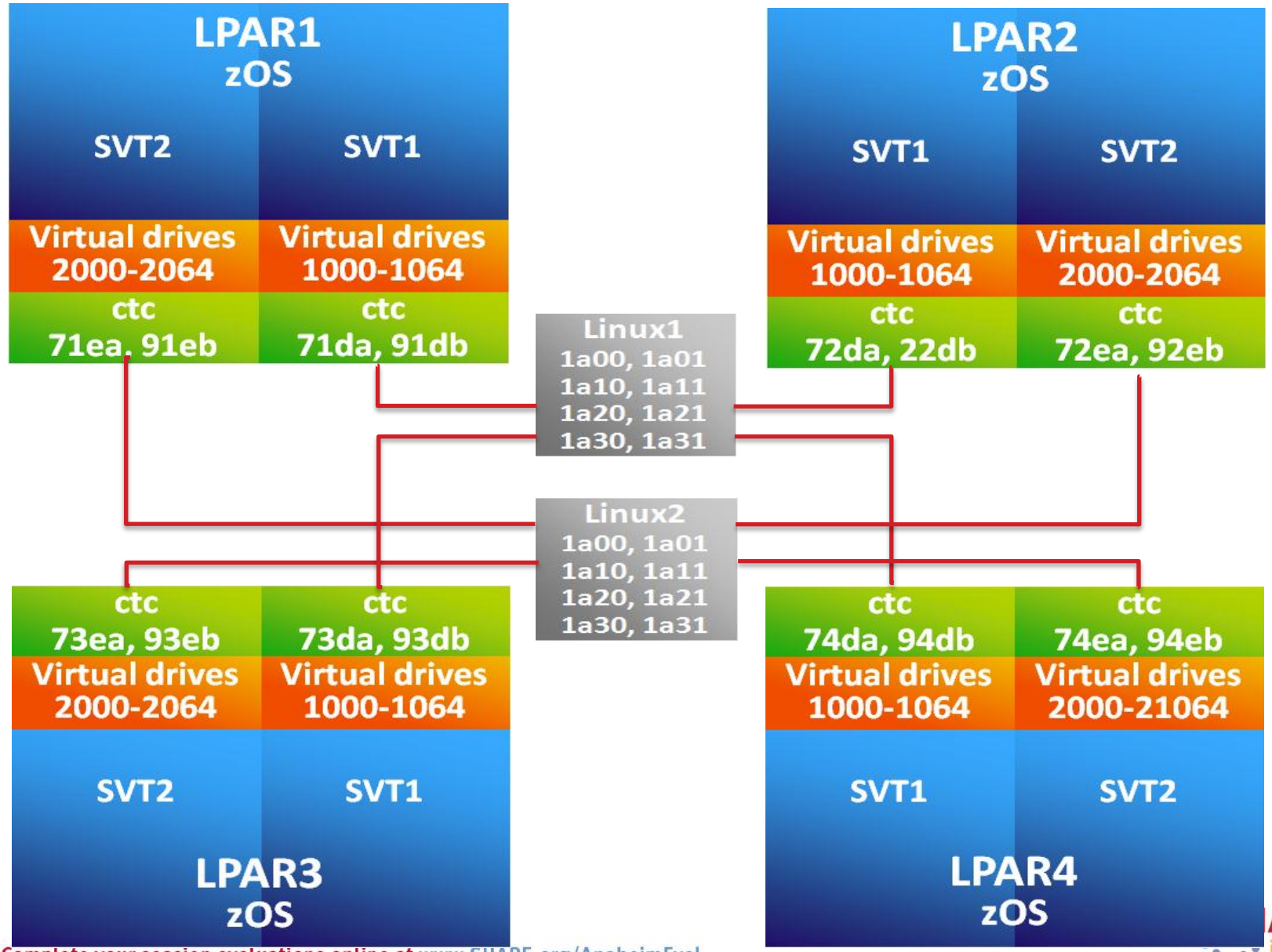
Why do I Need Linux?

- Objective is for the solution to use as little general processor as possible
  - Utilize zIIP when available
  - Utilize IFL when available
  - If specialty engines are not available then general processors are used
- Why not go directly from z/OS to NFS?
  - The z/OS NFS Client adds 20x CPU which is not zIIP eligible
- Ok, then why not use x86 Linux instead of Linux on z?
  - This is under consideration. We know there will be a general CPU increase but it should not be as costly as with z/OS NFS Client

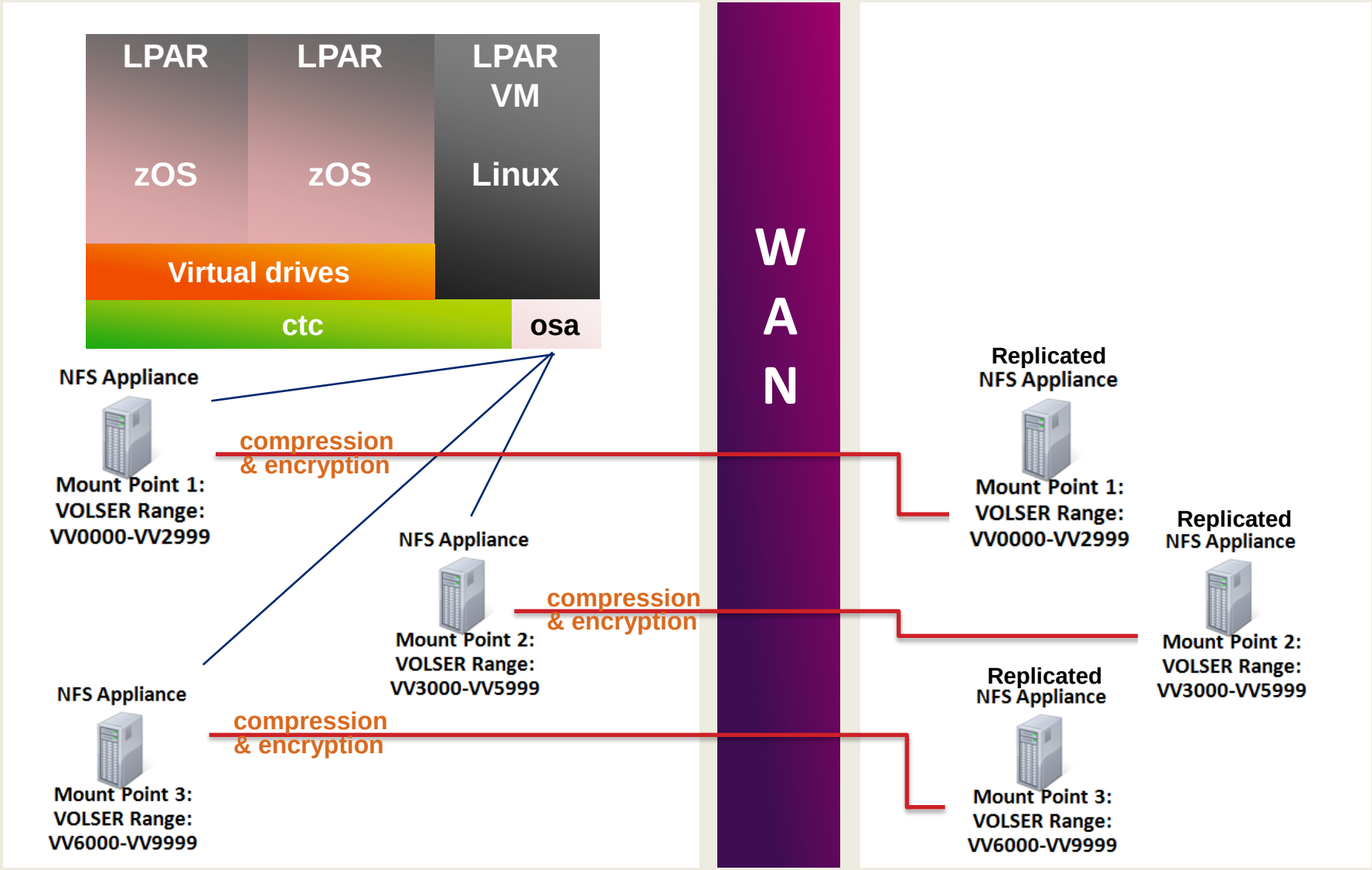
Most Common Configuration



Highly Redundant Environment



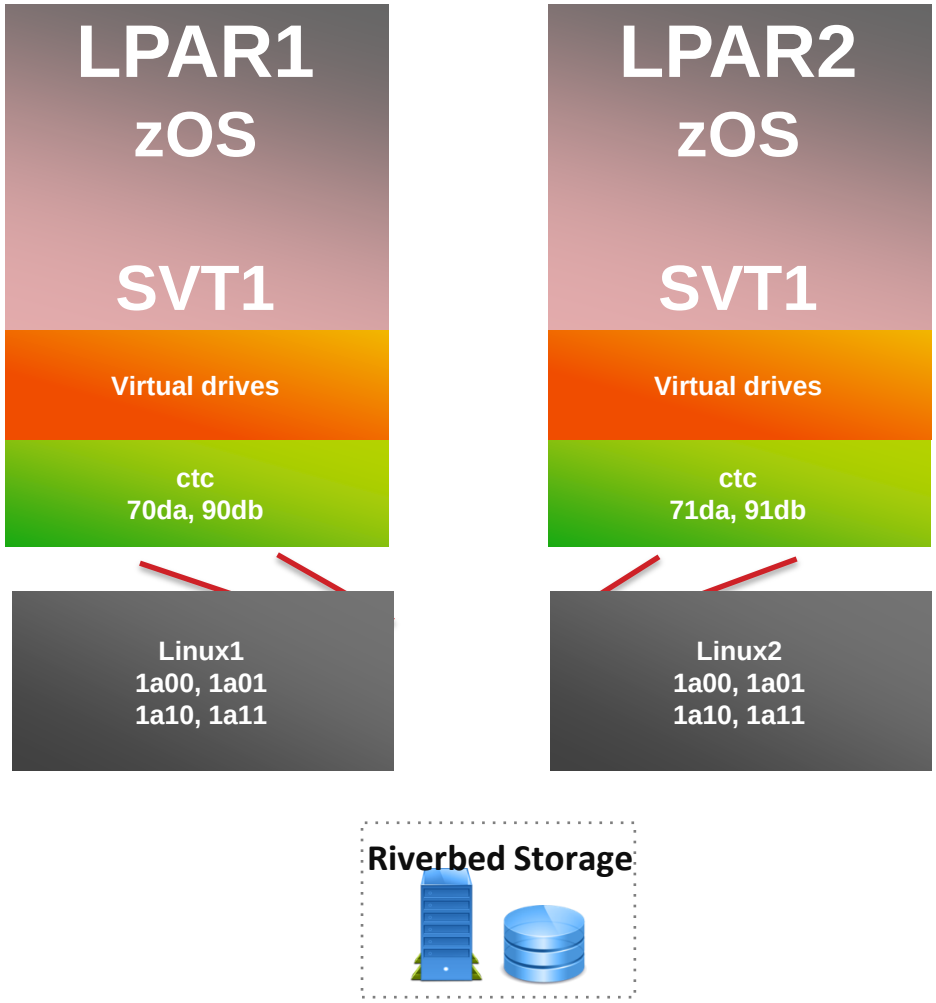
Multiple Mount Point Configuration



Benchmark Environment

- zEC12 2817-615 utilizing 2 z/OS LPARs and 1 z/VM with 2 Linux guests
- Each z/OS LPAR used 4 CTC addresses on 2 shared CHPIDs
- Both Linux guests shared a 10 Gbe OS4 Express4s configured to use Jumbo frames on a VSWITCH LAN
  - *TCPIP Segmentation Offload was not configured because the OSA must be configured dedicated*
- Both Linux guests wrote to the same 2030 Riverbed Whitewater

Benchmark Configuration



Benchmark Test

- Each z/OS LPAR ran 5 concurrent jobs running IEBDG to write 8GB virtual volumes
  - The 10 jobs ran continuously by resubmitting themselves over a 15+ minute timeframe to span a SMF interval
- CS4z z/OS started task information was collected in a WLM reporting class
- Linux on z IFL resources measured at the z/VM LPAR level
- Throughput reported by the OSA CHPID report
- Collected SMF 7n records and ran RMF reports
  - SYSRPTS(WLMGL(RCLASS(RVTSTC)))
  - REPORTS(CHAN)
  - REPORTS(CPU)

General Processor and zIIP Usage

- LPAR 1

```
REPORT BY: POLICY=PLEXT1                                REPORT CLASS=RVTSTC
                                                        DESCRIPTION =VTAPE Started Tasks

-TRANSACTIONS-  TRANS-TIME HHH.MM.SS.TTT  --DASD I/O--  ---SERVICE---   SERVICE TIME  ---APPL %---
AVG        3.00 ACTUAL                 0  SSCHRT   0.1  IOC    135688   CPU  101.387  CP       2.61
MPL        3.00 EXECUTION              0  RESP     0.7  CPU      4349K  SRB    6.925  AAPCP    0.00
ENDED         0 QUEUED                 0  CONN     0.5  MSO    772874   RCT    0.000  IIPCP    0.15
END/S      0.00 R/S AFFIN              0  DISC     0.2  SRB    297038   IIT   14.279
#SWAPS        0 INELIGIBLE             0  Q+PEND   0.1  TOT      5555K  HST    0.000  AAP       N/A
EXCTD         0 CONVERSION             0  IOSQ     0.0  /SEC     6172   AAP      N/A  IIP      6.98
AVG ENC    0.00 STD DEV                0                                IIP   62.813
REM ENC    0.00                                         ABSRPTN  2057
MS ENC     0.00                                         TRX SERV 2057
```

```
REPORT BY: POLICY=PLEXT1                                REPORT CLASS=RVTSTC
                                                        DESCRIPTION =VTAPE Started Tasks

-TRANSACTIONS-  TRANS-TIME HHH.MM.SS.TTT  --DASD I/O--  ---SERVICE---   SERVICE TIME  ---APPL %---
AVG        3.00 ACTUAL                 0  SSCHRT   0.2  IOC    148214   CPU  103.320  CP       2.57
MPL        3.00 EXECUTION              0  RESP     1.0  CPU      4834K  SRB    7.529  AAPCP    0.00
ENDED         0 QUEUED                 0  CONN     0.5  MSO    638444   RCT    0.000  IIPCP    0.03
END/S      0.00 R/S AFFIN              0  DISC     0.5  SRB    352243   IIT   14.416
#SWAPS        0 INELIGIBLE             0  Q+PEND   0.1  TOT      5973K  HST    0.000  AAP       N/A
EXCTD         0 CONVERSION             0  IOSQ     0.0  /SEC     6636   AAP      N/A  IIP      7.19
AVG ENC    0.00 STD DEV                0                                IIP   64.695
REM ENC    0.00                                         ABSRPTN  2212
MS ENC     0.00                                         TRX SERV 2212
```

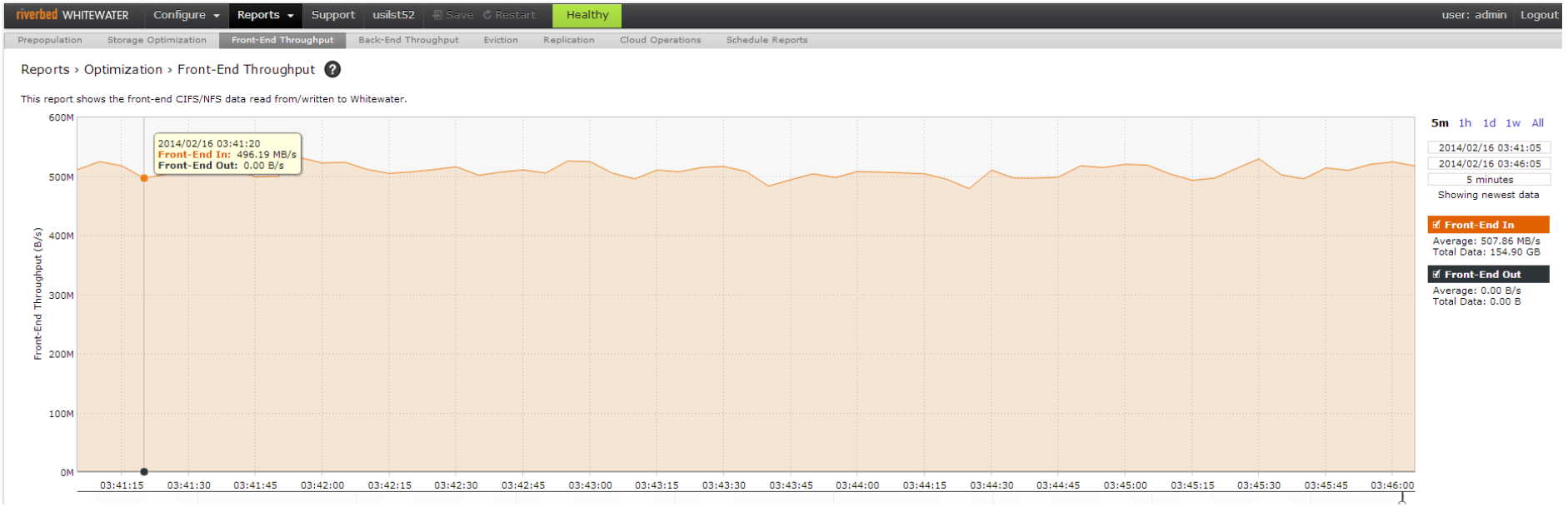
IFL Usage

--------- PARTITION DATA ----------								-- LOGICAL PARTITION PROCESSOR DATA --			-- AVERAGE PROCESSOR
			----MSU----		-CAPPING--		PROCESSOR-	----DISPATCH TIME DATA----		LOGICAL PROCESSORS	
NAME	S	WGT	DEF	ACT	DEF	WLM%	NUM TYPE	EFFECTIVE	TOTAL	EFFECTIVE	TOTAL
SYSB	A	100			NO		2 IFL	00.21.52.517	00.21.53.482	72.92	72.97
SYSC	A	65			NO		2 IFL	00.00.05.008	00.00.06.239	0.28	0.35
VMZ	A	100			NO		3 IFL	00.20.40.129	00.22.16.530	45.93	49.50
ZAWARE	A	50			NO		2 IFL	00.00.01.721	00.00.01.791	0.10	0.10
PHYSICAL									00.00.21.759		
								------------	------------		
TOTAL								00.42.39.377	00.44.39.803		

- Measured the entire z/VM LPAR
- The report tells you how busy each IFL was on average
- With TCP Segmentation Offload enabled, IFL utilization is expected to drop by up to 30%

Throughput Achieved 500+ MB/sec

CHANNEL PATH ID	TYPE	G	SPEED	SHR	UTILIZATION(%) PART	TOTAL	BUS	READ(MB/SEC) PART	TOTAL	WRITE(MB/SEC) PART	TOTAL	FICON OPERATIONS RATE	ACTIVE	DEFER
17	OSD			Y	0.00	29.60	25.19	0.00	1.92	0.00	509.84			
41	FC_S	12		Y	30.44	64.83	15.79	0.00	0.63	127.01	251.97	1910.4	4.1	0.0
48	FC_S	12		Y	30.79	61.59	16.22	0.65	2.11	139.52	257.40	2262.2	3.0	0.0



Results

System	CPU Model	General CPU Engine Speed	Specialty Engine Speed	Peak Hourly Data Transfer (GB)	Projected %zIIP processor @ Peak	Projected %GP processor @ Peak	Projected %IFL processor @ Peak (Jumbo=Y, TSO=N)
Baseline	2827-615	947	1514	1830	14.17	5.18	148.5
zEC12	2827-7xx	1514	1514	1830	14.2	3.2	148.5
z196	2817-6xx	768	1202	1830	17.9	6.3	187
z10	2097-6xx	616	889	1830	24.2	7.9	252.8
z9	2094-6xx	468	581	1830	37	10.4	386.8

- The solution is capable of writing 500+ MB/sec which is 1.8 TB/hr
  - This should scale by replicating resources
- Minimizes general processor resource consumption by offloading to specialty engines
- IFL is expected to come down when exploiting TCPIP Segmentation Offload

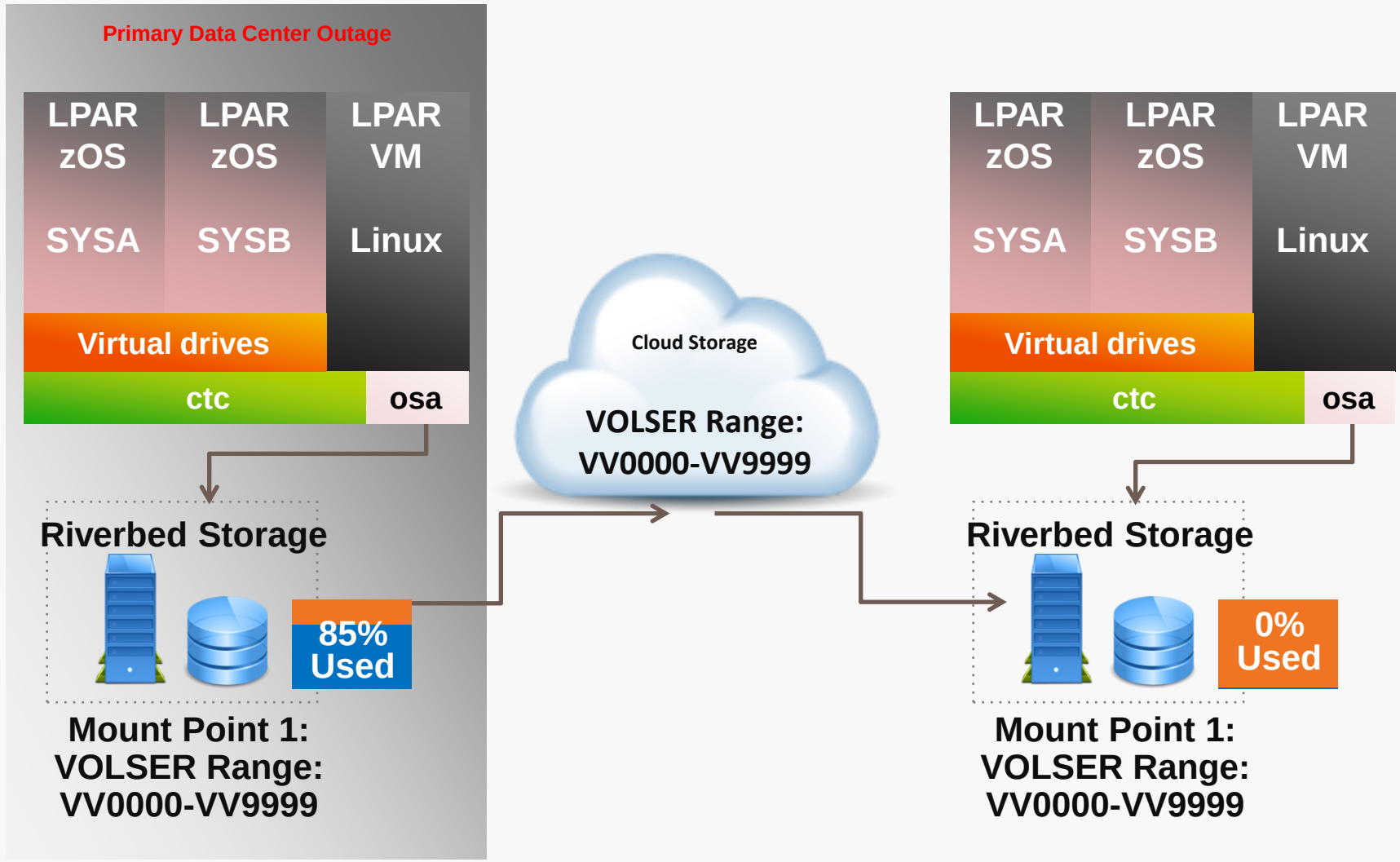
Recovery Time Objective (RTO)

- When your RTO is measured in hours then you can decrease cost and conserve network bandwidth by replicating to/recovering from the cloud
  - At your DR site you import your cloud settings from your cloud provider into a new Riverbed appliance and prioritize the volumes you recover
  - This is where information in the form of a report available at the DR location from the Tape Management System is critical. To know WHICH volumes contain which files is critical to the DR plan. Just like it is now with “pickup truck vaulting”.

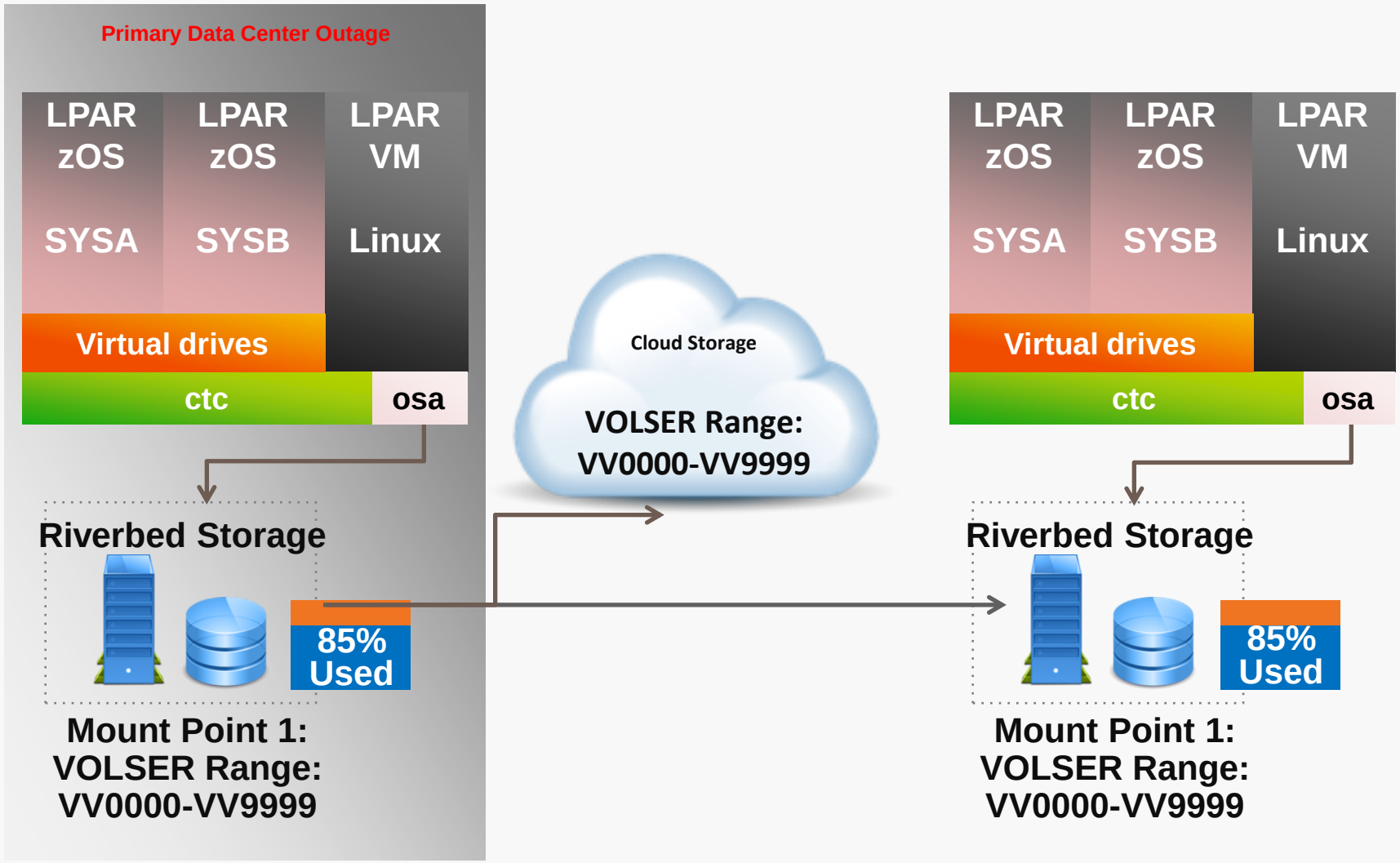
Recovery Time Objective (RTO)

- When your RTO is measured in minutes then you can have a 2nd Riverbed appliance in “warm” mode receiving replication updates the same as when the cloud receives them
  - At your DR site your Riverbed appliance already holds a mirror of your primary appliance and recovery can begin immediately
  - You can “pin” critical data in the appliance so that it is available at the recovery site

Cold DR Mode with RTO Measured in Hours



Warm DR Mode with Shortest RTO



Summary

Develop a Tapeless Cloud enabled Enterprise Storage Virtual Tape Solution that reduces cost and provides a flexible disaster backup and recovery solution

QR Code