

Linux on System z zFCP Performance Analysis

Stefan Raspl
IBM

2/4/2013
13112

(C) 2012, 2013 IBM Corporation



Trademarks

The following are trademarks of the International Business Machines Corporation in the United States and/or other countries.

AIX*	IBM*	PowerVM	System z10	z/OS*
BladeCenter*	IBM eServer	PR/SM	WebSphere*	zSeries*
DataPower*	IBM (logo)*	Smarter Planet	z9*	z/VM*
DB2*	InfiniBand*	System x*	z10 BC	z/VSE
FICON*	Parallel Sysplex*	System z*	z10 EC	
GDPS*	POWER*	System z9*	zEnterprise	
HiperSockets	POWER7*			

* Registered trademarks of IBM Corporation

The following are trademarks or registered trademarks of other companies.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries. Cell Broadband Engine is a trademark of Sony Computer Entertainment, Inc. in the United States, other countries, or both and is used under license there from.

Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Windows Server and the Windows logo are trademarks of the Microsoft group of countries.

InfiniBand is a trademark and service mark of the InfiniBand Trade Association.

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

ITIL is a registered trademark, and a registered community trademark of the Office of Government Commerce, and is registered in the U.S. Patent and Trademark Office.

IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency, which is now part of the Office of Government Commerce.

* All other products may be trademarks or registered trademarks of their respective companies.

Notes:

Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput improvements equivalent to the performance ratios stated here.

IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply.

All customer examples cited or described in this presentation are presented as illustrations of the manner in which some customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics will vary depending on individual customer configurations and conditions.

This publication was produced in the United States. IBM may not offer the products, services or features discussed in this document in other countries, and the information may be subject to change without notice. Consult your local IBM business contact for information on the product or services available in your area.

All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

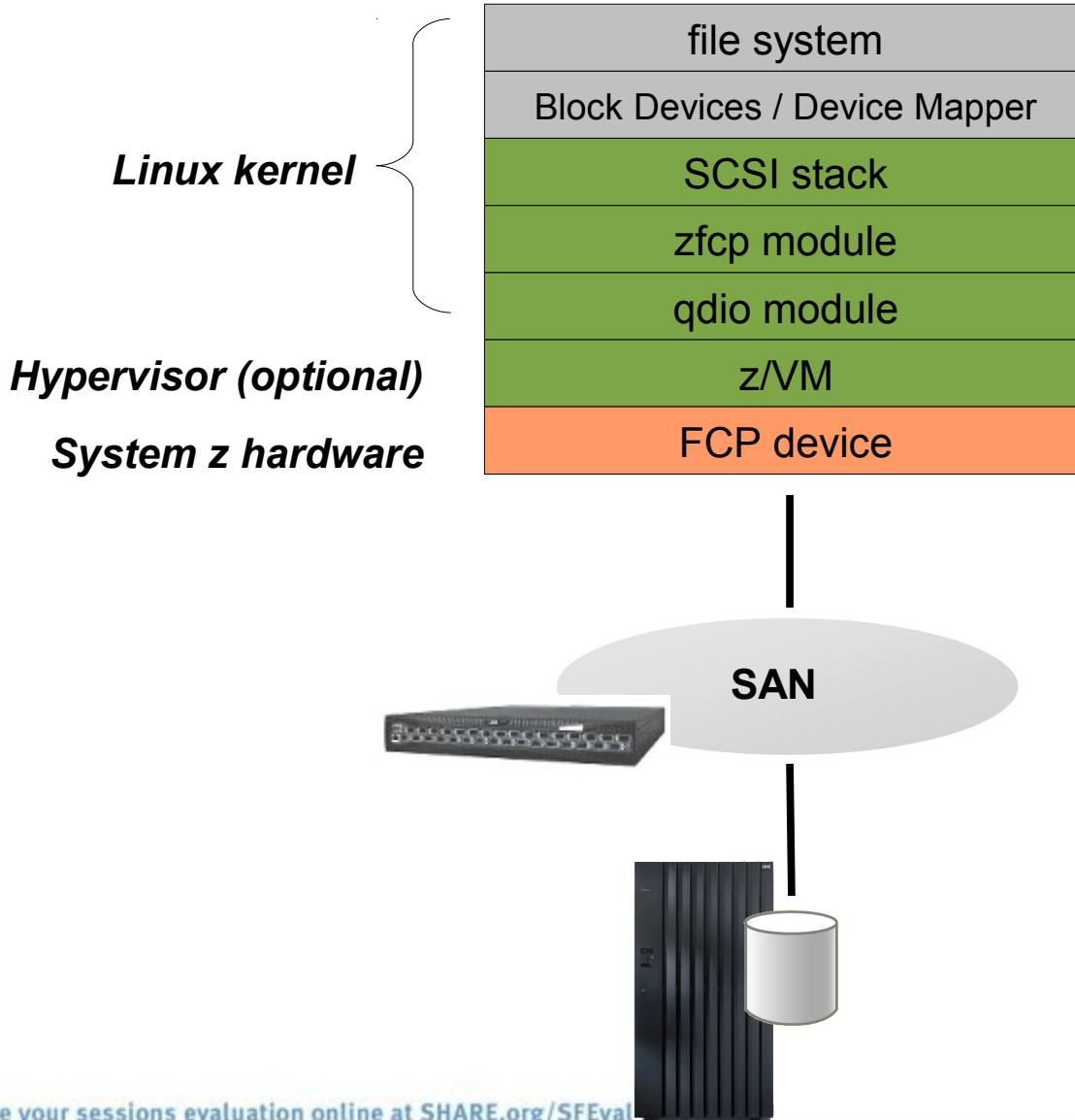
Information about non-IBM products is obtained from the manufacturers of those products or their published announcements. IBM has not tested those products and cannot confirm the performance, compatibility, or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

Prices subject to change without notice. Contact your IBM representative or Business Partner for the most current pricing in your geography.

Agenda

- Purpose
- Workflow Overview
- Collecting Data
 - Data Sources
 - Architecture
 - ziomon Usage
- Analyzing Data
 - ziorep_config
 - ziorep_utilization
 - ziorep_traffic
- Hints & Tips
- Data Analysis on Other Platforms

How files end up on Storage Devices

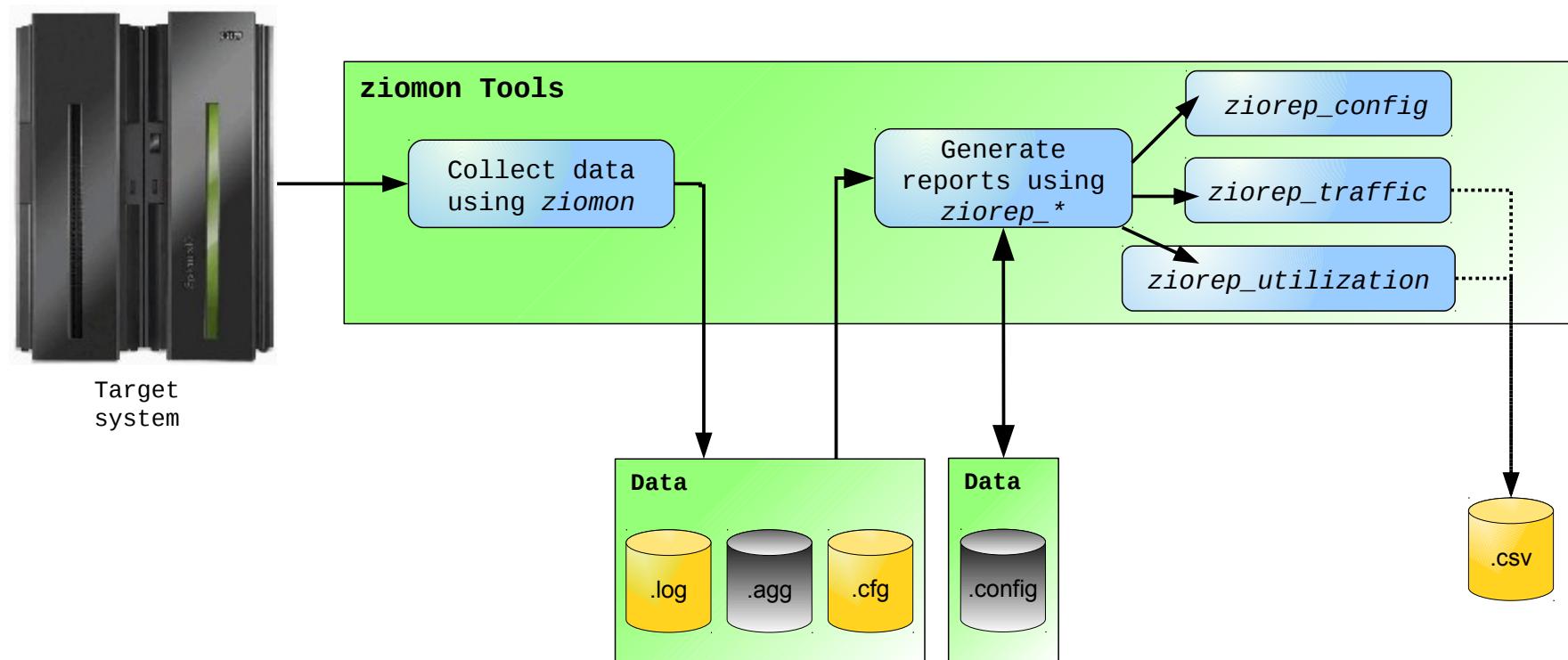


Motivation

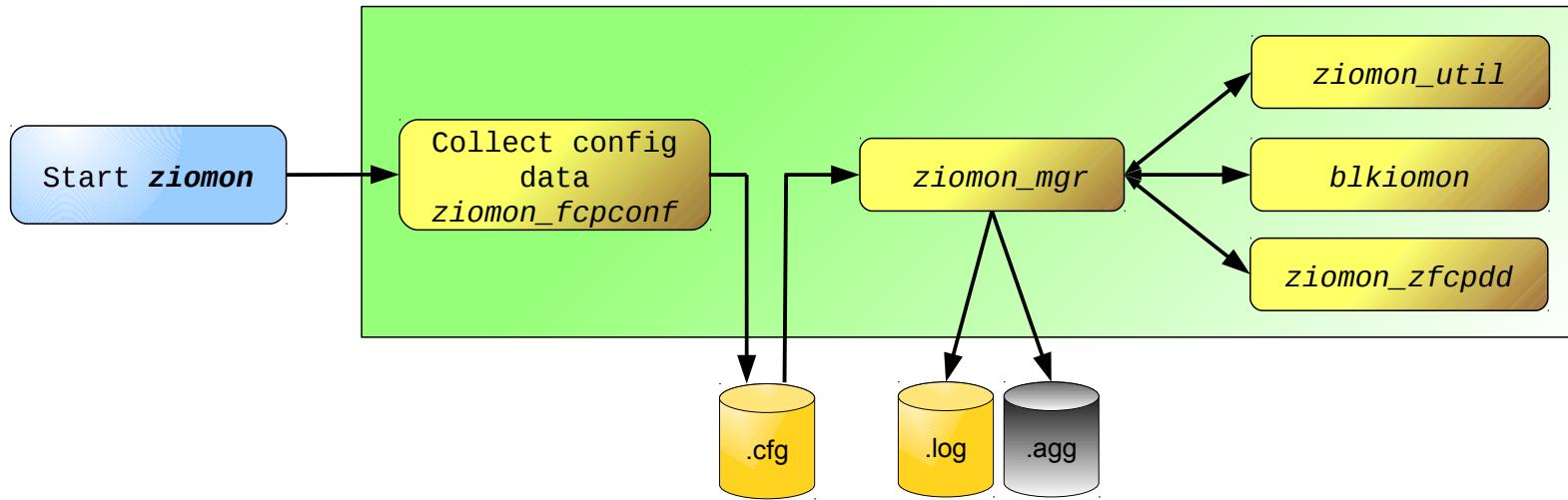
- Performance issues can potentially stem from multiple sources
- Focusing on a single component will not tell the whole story
- Holistic approach, considering all components involved (Linux, hypervisors, HBAs, SAN with storage devices) is required
- Requires advanced data analysis features, e.g. for filtering and aggregation
- *ziomon* tools aim at providing all of the above
- Availability:
 - SLES: Part of the *s390-tools* package starting with SLES10 SP3 and SLES11 SP1
 - RHEL: Part of the *s390utils* package starting with RHEL 5.4 and provided as *s390utils-ziomon* beginning with RHEL 6
- Dependencies: Requires the *blktrace* package as well as `CONFIG_BLK_DEV_IO_TRACE` enabled in the kernel config

Workflow Overview

- Collect data using *ziomon*
- Yield resulting files (.log, .agg & .cfg - all required)
- Use reporting tools to analyze data



Collecting Data: Components



- Components used by ziomon to collect data:
 - *ziomon_fcpconf* : zfcp configuration
 - *ziomon_zfcppdd*: Channel & fabric latencies, QDIO outbound queue utilization
 - *ziomon_util*: zfcp adapter utilization
 - *blkmon*: blktrace component for SCSI subsystem latencies
- Note: Data collection only starts after the configuration has been captured

Collecting Data: Invocation

- Minimum invocation:

```
user@larsson:~> ziomon -d 5 -o data /dev/sda
```

- Mandatory parameters to use:

-d : total data collection duration in minutes
-o : output file – should be on a device not sampled by *ziomon*!
<device> : SCSI device to monitor

- *ziomon* will aggregate data during collection, defaults to 60 seconds. I.e. it will not collect data on individual I/O operations.
- Use '-i' to specify an arbitrary interval length for aggregation. E.g. to invoke with highest possible resolution (2 seconds) use

```
user@larsson:~> ziomon -d 5 -i 2 -o data /dev/sda
```

- Use '-l' to specify size limit for .log file.

```
user@larsson:~> ziomon -d 5 -l 137M -o data /dev/sda
```

Will wrap around in case specified limit is too low (like in a ring buffer).
The overwritten data is aggregated into a single (bigger) interval in a .agg file.

Collecting Data: Specifying Devices

- Example:

```
user@larsson:~> multipath -ll
36005076303ffc56200000000000010cc dm-0 IBM,2107900
[size=5.0G][features=0][hwandler=0]
 \_ round-robin 0 [prio=2][active]
   \_ 1:0:0:1087127568 sdb    8:16  [active][ready]
   \_ 0:0:0:1087127568 sda    8:0   [active][ready]
```

- Multiple ways to specify devices for collecting data:

- Provide a list of single path devices:

```
user@larsson:~> zionon [...] /dev/sda /dev/sdb
```

- Specifying multipath devices will identify all underlying devices (recommended):

```
user@larsson:~> zionon [...] /dev/mapper/36005076303ff...
```

- Mixing single path and multipath devices is possible

```
user@larsson:~> zionon [...] /dev/mapper/36005076303ff... \
  /dev/sdc /dev/sdd
```

Collecting Data: Specifying Devices (continued)

- Notes:

- SCSI tape devices are supported, too
- *blktrace* (as used by *ziomon*) requires 2MB vmalloc memory per device *and* processor
- Check current configuration:

```
user@larsson:~> grep Vmalloc /proc/meminfo
```

- Use kernel parameter vmalloc to increase, e.g. vmalloc=512M

Data Analysis

- Use the *ziorep_** commands to generate reports and investigate data:
 - *ziorep_config*: Insight into the zfcp configuration
 - *ziorep_utilization*: Adapter utilization statistics
 - *ziorep_traffic*: Traffic statistics
- Note: Always pass on all files (.log and .cfg, and possibly .agg) when forwarding data to others for analysis

Data Analysis: Configuration Report Overview

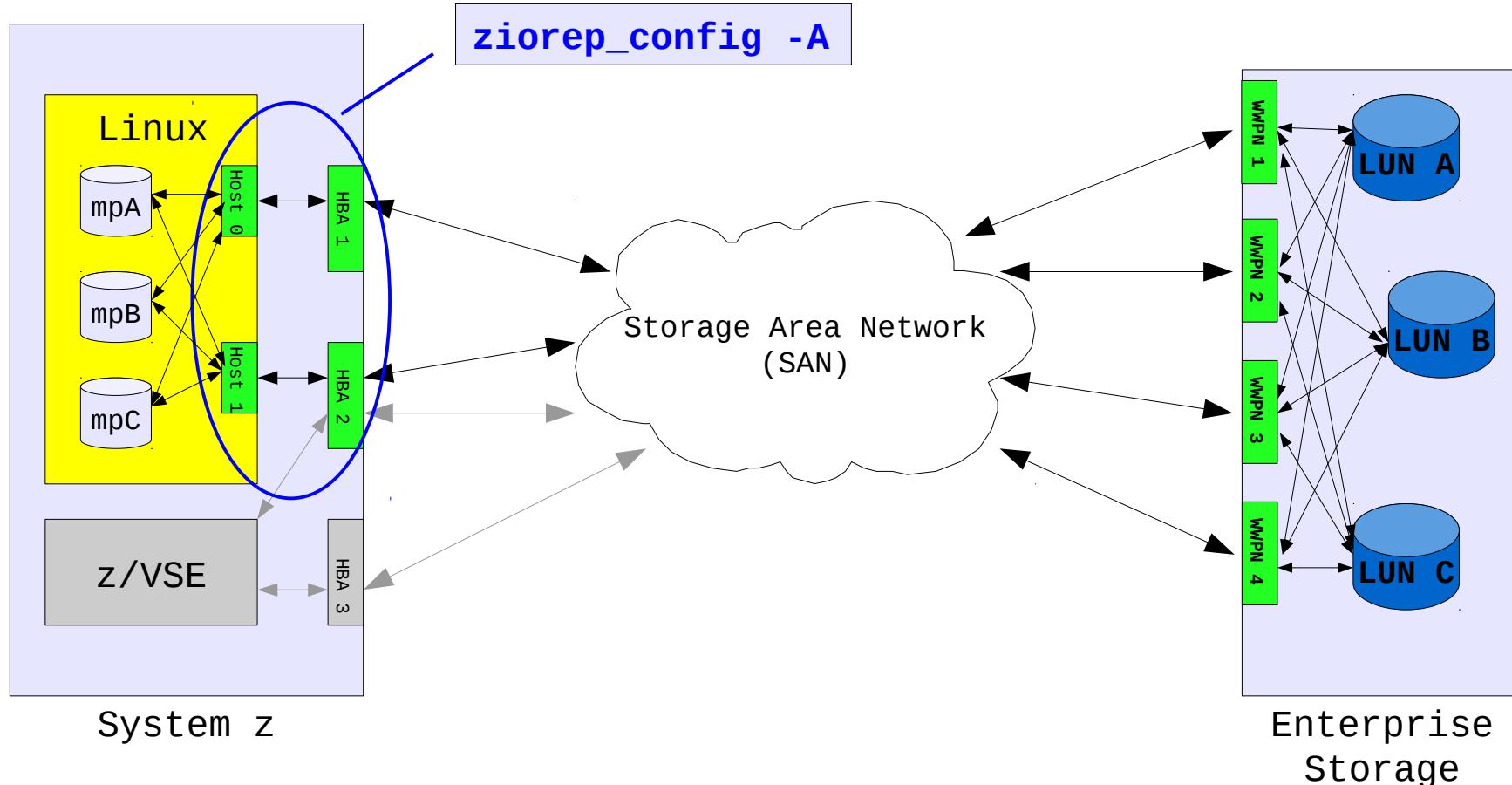
- `ziorep_config` can take a `.cfg` file as input

```
user@larsson:~> ziorep_config -i sample.cfg
```

- Displays information on zfcp-related hardware (HBAs, single and multipath devices) data was collected on.
- Skip the '`-i`' parameter for a report on the current system.
- 3 different report types available:
 - Adapter report (**default**, option '`-A`') : HBA details
 - Multipath report (option '`-M`') : Multipath/device mapper details
 - Devices report (option '`-D`') : SCSI device details
- Selection options to limit output to a subset only (can be combined):

<code>'-p <port>'</code>	limits output to specified WWPN, e.g. ' <code>-p 0x5005123456789000'</code>
<code>'-l <lun>'</code>	limits output to specified LUN, e.g. ' <code>-l 0x401040a600000000'</code>
<code>'-m <mdev>'</code>	limits output to specified multipath device, e.g. ' <code>-m 36005076303ffc56200000000000010a6'</code>
<code>'-a <adapter>'</code>	limits output to specified adapter, e.g. ' <code>-a 0.0.3c07'</code>
<code>'-s <scsi_host>'</code>	limits output to specified SCSI host, e.g. ' <code>-s host0'</code>
<code>'-d <device>'</code>	limits output to specified device, e.g. ' <code>-d sde'</code>

Data Analysis: Adapter Configuration Report

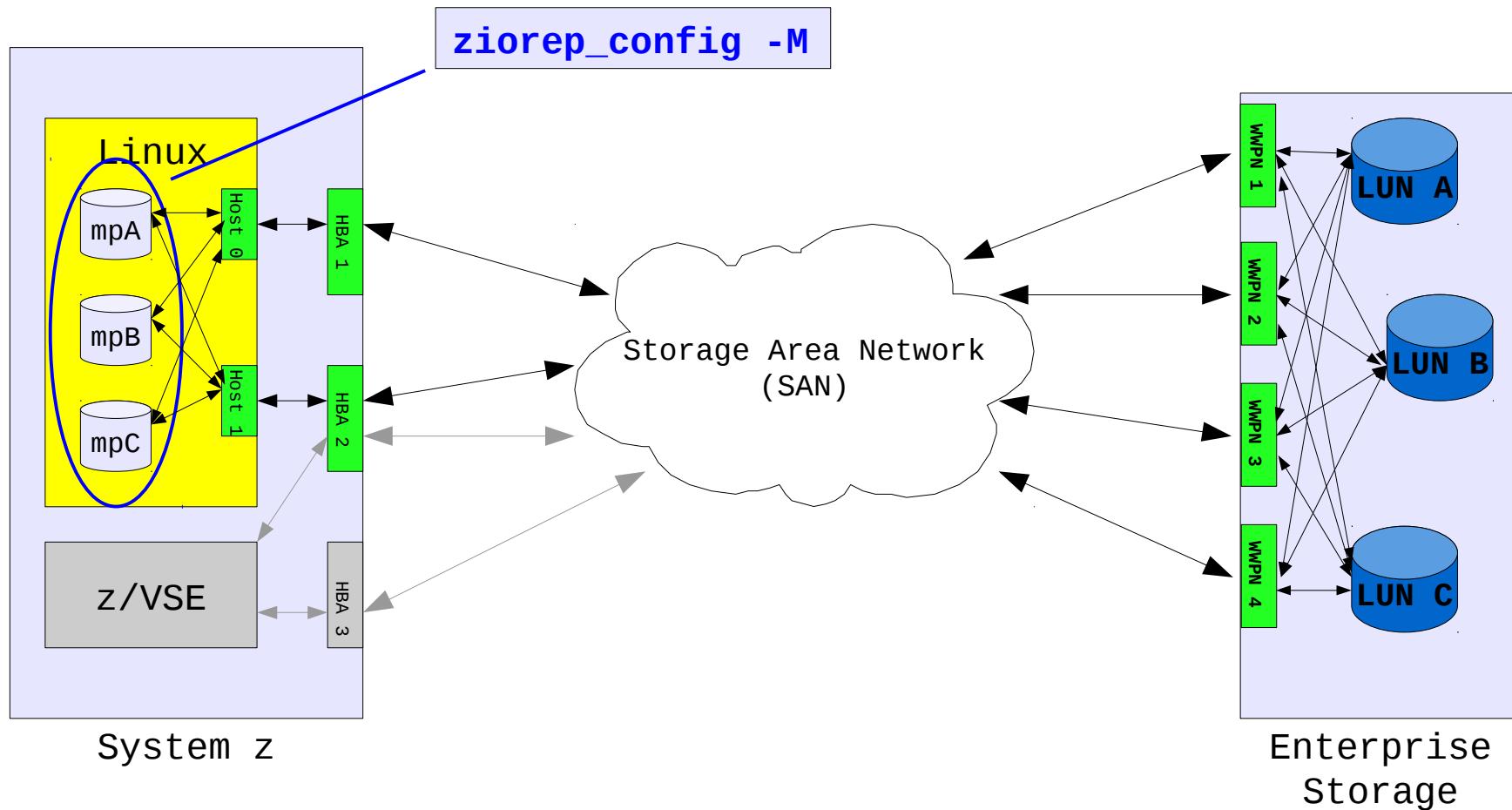


Data Analysis: Adapter Configuration Report (continued)

```
user@larsson:~> ziorep_config -A -i sample.cfg \
-a 0.0.1800 -a 0.0.1820
Host: host0
CHPID: 64
Adapter: 0.0.1800
Sub-Ch.: 0.0.0842
Name: 0xc05076ffe4805611
P-Name: 0xc05076ffe4805611
Version: 0x0005
LIC: 0x0000014c
Type: NPort (fabric via point-to-point)
Speed: 8 Gbit
State: Online

Host: host2
CHPID: 66
Adapter: 0.0.1820
Sub-Ch.: 0.0.08a2
Name: 0xc05076ffe4805e11
P-Name: 0xc05076ffe4805e11
Version: 0x0005
LIC: 0x0000014c
Type: NPort (fabric via point-to-point)
Speed: 8 Gbit
State: Online
```

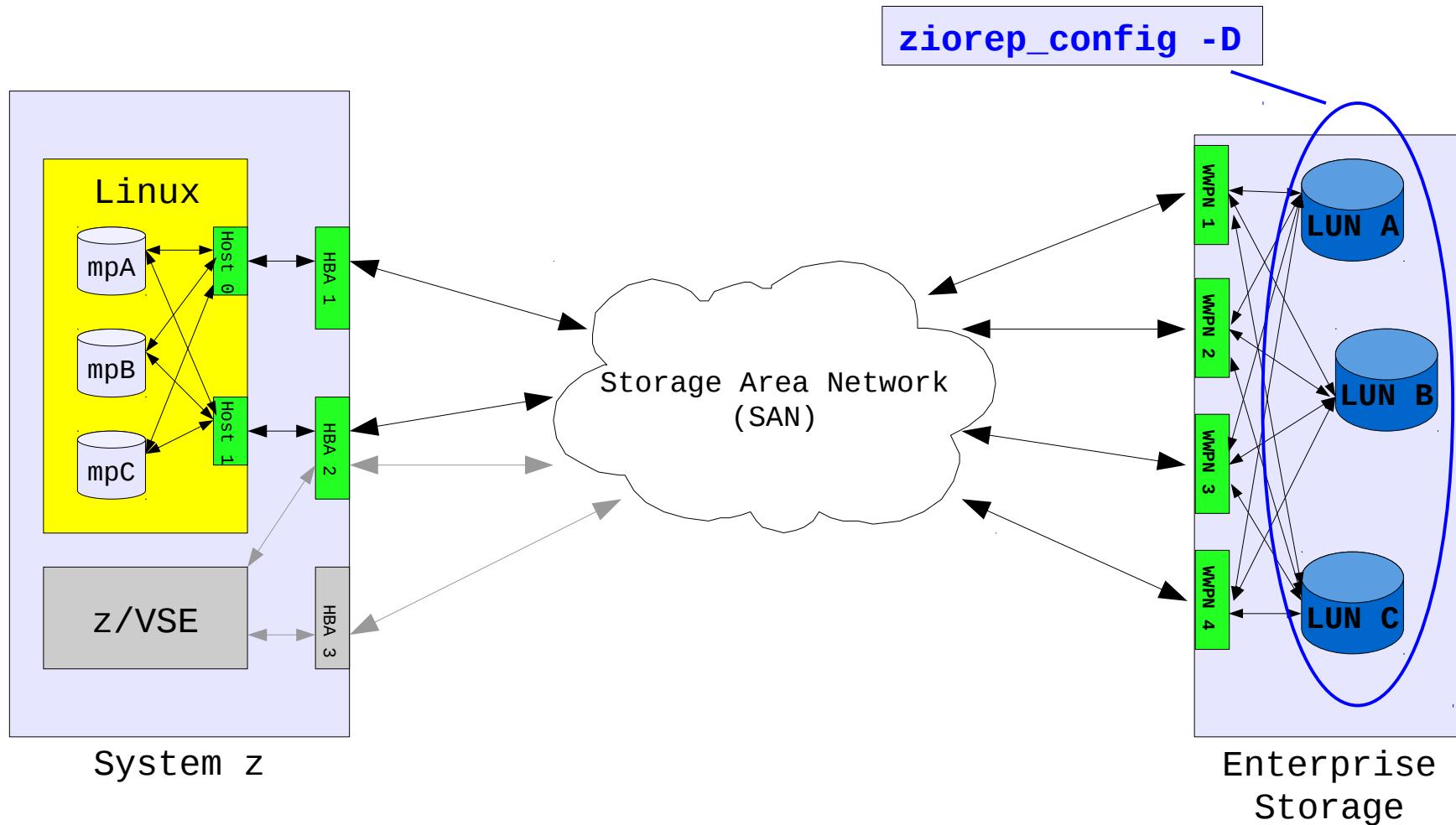
Data Analysis: Multipath Report



Data Analysis: Multipath Report (continued)

```
user@larsson:~> ziorep_config -i sample.cfg -M \
-m 36005076303ffc52a0000000000005502
0.0.1800 0x500507630303c52a /dev/sda /dev/mapper/36005076303ffc52a000000000000005502
0.0.1810 0x500507630303c52a /dev/sde /dev/mapper/36005076303ffc52a000000000000005502
0.0.1820 0x500507630303c52a /dev/sdi /dev/mapper/36005076303ffc52a000000000000005502
0.0.1830 0x500507630303c52a /dev/sdm /dev/mapper/36005076303ffc52a000000000000005502
0.0.1800 0x500507630308c52a /dev/sdd /dev/mapper/36005076303ffc52a000000000000005502
0.0.1810 0x500507630308c52a /dev/sdh /dev/mapper/36005076303ffc52a000000000000005502
0.0.1820 0x500507630308c52a /dev/sdl /dev/mapper/36005076303ffc52a000000000000005502
0.0.1830 0x500507630308c52a /dev/sdp /dev/mapper/36005076303ffc52a000000000000005502
0.0.1800 0x500507630310c52a /dev/sdc /dev/mapper/36005076303ffc52a000000000000005502
0.0.1810 0x500507630310c52a /dev/sdg /dev/mapper/36005076303ffc52a000000000000005502
0.0.1820 0x500507630310c52a /dev/sdj /dev/mapper/36005076303ffc52a000000000000005502
0.0.1830 0x500507630310c52a /dev/sdo /dev/mapper/36005076303ffc52a000000000000005502
0.0.1800 0x500507630318c52a /dev/sdb /dev/mapper/36005076303ffc52a000000000000005502
0.0.1810 0x500507630318c52a /dev/sdf /dev/mapper/36005076303ffc52a000000000000005502
0.0.1820 0x500507630318c52a /dev/sdk /dev/mapper/36005076303ffc52a000000000000005502
0.0.1830 0x500507630318c52a /dev/sdn /dev/mapper/36005076303ffc52a000000000000005502
```

Data Analysis: Device Configuration Report



Data Analysis: Device Configuration Report (continued)

```
user@larsson:~> ziorep_config -i sample.cfg -D \
```

```
-1 0x4055400500000000 -1 0x4055400000000000
```

0.0.1800	0x500507630303c52a	0x4055400000000000	host0	/dev/sg49	/dev/sdax	67:16	Disk	2107900	IBM	0:0:0:1073758293
0.0.1800	0x500507630303c52a	0x4055400500000000	host0	/dev/sg108	/dev/sdde	70:192	Disk	2107900	IBM	0:0:0:1074085973
0.0.1800	0x500507630318c52a	0x4055400000000000	host0	/dev/sg74	/dev/sdbw	68:160	Disk	2107900	IBM	0:0:1:1073758293
0.0.1800	0x500507630318c52a	0x4055400500000000	host0	/dev/sg118	/dev/sddo	71:96	Disk	2107900	IBM	0:0:1:1074085973
0.0.1800	0x500507630310c52a	0x4055400000000000	host0	/dev/sg56	/dev/sdbe	67:128	Disk	2107900	IBM	0:0:2:1073758293
0.0.1800	0x500507630310c52a	0x4055400500000000	host0	/dev/sg122	/dev/sdds	71:160	Disk	2107900	IBM	0:0:2:1074085973
0.0.1800	0x500507630308c52a	0x4055400000000000	host0	/dev/sg87	/dev/sdcj	69:112	Disk	2107900	IBM	0:0:3:1073758293
0.0.1800	0x500507630308c52a	0x4055400500000000	host0	/dev/sg89	/dev/sdcl	69:144	Disk	2107900	IBM	0:0:3:1074085973
0.0.1810	0x500507630303c52a	0x4055400000000000	host1	/dev/sg80	/dev/sdcc	69:0	Disk	2107900	IBM	1:0:0:1073758293
0.0.1810	0x500507630303c52a	0x4055400500000000	host1	/dev/sg113	/dev/sddj	71:16	Disk	2107900	IBM	1:0:0:1074085973
0.0.1810	0x500507630318c52a	0x4055400000000000	host1	/dev/sg103	/dev/sdcz	70:112	Disk	2107900	IBM	1:0:1:1073758293
0.0.1810	0x500507630318c52a	0x4055400500000000	host1	/dev/sg124	/dev/sddu	71:192	Disk	2107900	IBM	1:0:1:1074085973
0.0.1810	0x500507630310c52a	0x4055400000000000	host1	/dev/sg46	/dev/sdau	66:224	Disk	2107900	IBM	1:0:2:1073758293
0.0.1810	0x500507630310c52a	0x4055400500000000	host1	/dev/sg105	/dev/sddb	70:144	Disk	2107900	IBM	1:0:2:1074085973
0.0.1810	0x500507630308c52a	0x4055400000000000	host1	/dev/sg86	/dev/sdci	69:96	Disk	2107900	IBM	1:0:3:1073758293
0.0.1810	0x500507630308c52a	0x4055400500000000	host1	/dev/sg107	/dev/sddd	70:176	Disk	2107900	IBM	1:0:3:1074085973
0.0.1820	0x500507630303c52a	0x4055400000000000	host2	/dev/sg110	/dev/sddg	70:224	Disk	2107900	IBM	2:0:0:1073758293
0.0.1820	0x500507630303c52a	0x4055400500000000	host2	/dev/sg120	/dev/sddq	71:128	Disk	2107900	IBM	2:0:0:1074085973
0.0.1820	0x500507630310c52a	0x4055400000000000	host2	/dev/sg69	/dev/sdbr	68:80	Disk	2107900	IBM	2:0:1:1073758293
0.0.1820	0x500507630310c52a	0x4055400500000000	host2	/dev/sg71	/dev/sdbt	68:112	Disk	2107900	IBM	2:0:1:1074085973
0.0.1820	0x500507630318c52a	0x4055400000000000	host2	/dev/sg109	/dev/sddf	70:208	Disk	2107900	IBM	2:0:2:1073758293
0.0.1820	0x500507630318c52a	0x4055400500000000	host2	/dev/sg116	/dev/sddm	71:64	Disk	2107900	IBM	2:0:2:1074085973
0.0.1820	0x500507630308c52a	0x4055400000000000	host2	/dev/sg92	/dev/sdco	69:192	Disk	2107900	IBM	2:0:3:1073758293
0.0.1820	0x500507630308c52a	0x4055400500000000	host2	/dev/sg117	/dev/sddn	71:80	Disk	2107900	IBM	2:0:3:1074085973
0.0.1830	0x500507630303c52a	0x4055400000000000	host3	/dev/sg83	/dev/sdcf	69:48	Disk	2107900	IBM	3:0:0:1073758293
0.0.1830	0x500507630303c52a	0x4055400500000000	host3	/dev/sg125	/dev/sddv	71:208	Disk	2107900	IBM	3:0:0:1074085973
0.0.1830	0x500507630318c52a	0x4055400000000000	host3	/dev/sg97	/dev/sdct	70:16	Disk	2107900	IBM	3:0:1:1073758293
0.0.1830	0x500507630318c52a	0x4055400500000000	host3	/dev/sg99	/dev/sdcv	70:48	Disk	2107900	IBM	3:0:1:1074085973
0.0.1830	0x500507630310c52a	0x4055400000000000	host3	/dev/sg114	/dev/sddk	71:32	Disk	2107900	IBM	3:0:2:1073758293
0.0.1830	0x500507630310c52a	0x4055400500000000	host3	/dev/sg128	/dev/sddy	128:0	Disk	2107900	IBM	3:0:2:1074085973
0.0.1830	0x500507630308c52a	0x4055400000000000	host3	/dev/sg112	/dev/sddi	71:0	Disk	2107900	IBM	3:0:3:1073758293
0.0.1830	0x500507630308c52a	0x4055400500000000	host3	/dev/sg127	/dev/sddx	71:240	Disk	2107900	IBM	3:0:3:1074085973

Data Analysis: Adapter Utilization Report

- Ficon Express card with FCP CHPIDs has a CPU, bus and adapter chip
- Utilization report provides two sections:
 - **FCP channel** report provides details on
 - CPU utilization
 - Bus utilization
 - Adapter utilization
 - **Virtual HBA** report provides details on
 - QDIO utilization, queue full and failure conditions
 - Throughput
 - I/O requests numbers
- Purpose: Identify potential bottlenecks within the adapter
- Note: Percentages reported are for entire FCP channel – no fraction according to utilization given when FCP channel is shared by multiple OS images!
I.e. high utilization during phases of inactivity point to activity on other OS images sharing the same channel.

Data Analysis: Adapter Utilization Report

- Use '-s' to get a quick overview of the available data

```
user@larsson:~> ziorep_utilization -s sample.log
Data Summary
-----
Aggregated range: none
Detailed range: 2012-12-11 10:51:55 to 2012-12-11 11:51:55
Interval length: 2 seconds
HBA/CHPID: 0.0.591d/42
              0.0.5b1d/43
WWPN/LUN (dev): 0x5005076303000104/0x4021402100000000 (/dev/sdi)
                  0x5005076303000104/0x4021402100000000 (/dev/sdj)
                  0x50050763030b0104/0x4021402100000000 (/dev/sdk)
                  0x50050763030b0104/0x4021402100000000 (/dev/sd1)
                  0x5005076303100104/0x4021402100000000 (/dev/sdm)
                  0x5005076303100104/0x4021402100000000 (/dev/sdn)
[...]
```

- Use '-b <begin>' and '-e <end>' to limit output to specified timeframe
- Use '-i <length>' to aggregate data to bigger intervals
- Use '-c <chpid>' to limit output to specified adapter
- Notes:
 - Check the interval length to figure out available aggregation values
 - Aggregated range only listed in case of wrap-around in .log file (.agg exists)

Data Analysis: FCP Channel Utilization Report

```
user@larsson:~> ziorep_utilization sample.log -b "2012-12-11
10:58:55"
```

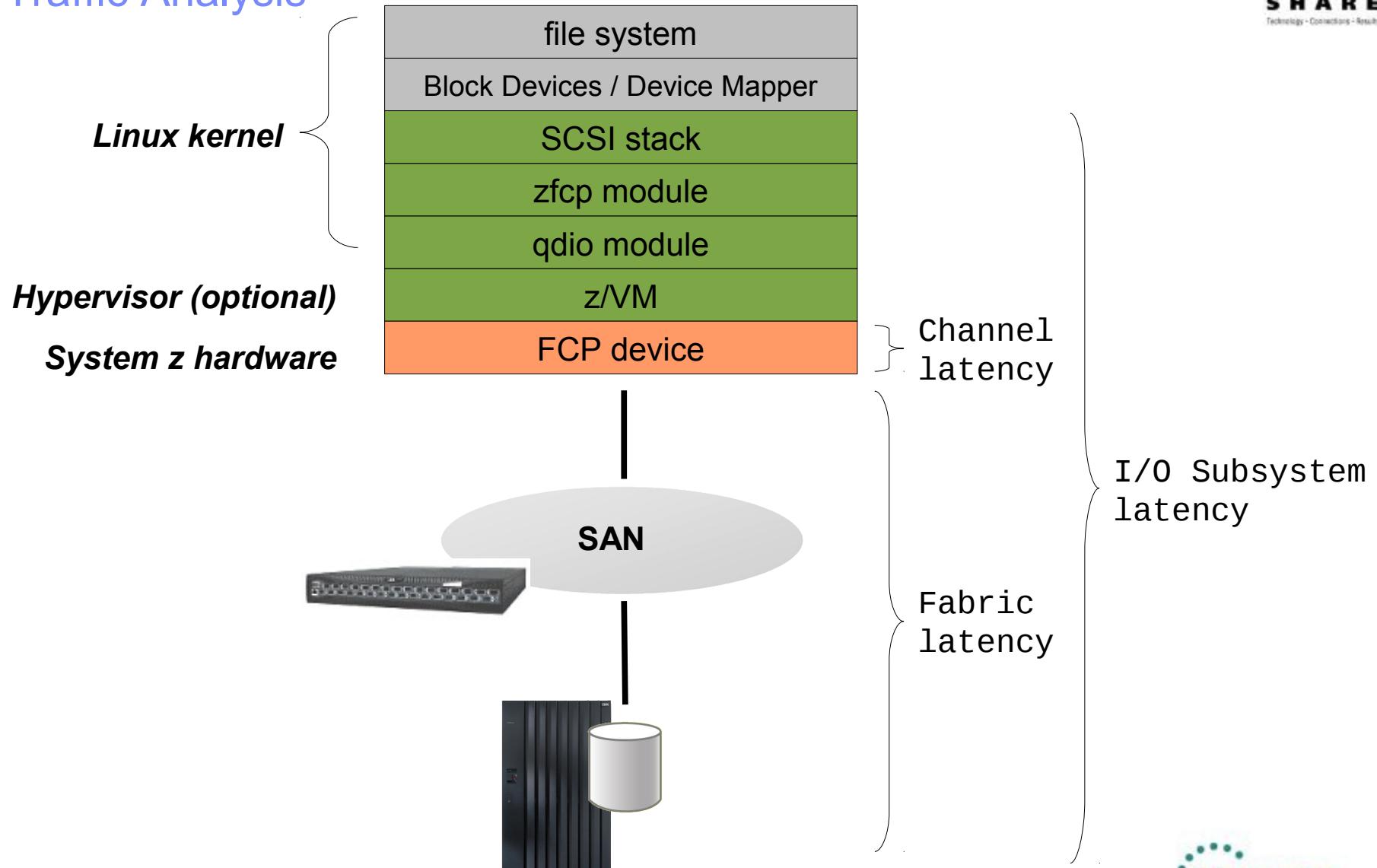
CHP adapter	in %	--bus in %	--cpu in %						
ID	min	max	avg	min	max	avg	min	max	avg
2012-12-11 10:58:55									
42	3	3	3.0	20	20	20.0	2	2	2.0
43	3	3	3.0	12	12	12.0	1	1	1.0
10:58:57									
42	3	3	3.0	18	18	18.0	2	2	2.0
43	4	4	4.0	15	15	15.0	1	1	1.0
10:58:59									
42	2	2	2.0	16	16	16.0	2	2	2.0
43	3	3	3.0	17	17	17.0	1	1	1.0
10:59:01									
42	3	3	3.0	16	16	16.0	2	2	2.0
43	5	5	5.0	17	17	17.0	1	1	1.0
10:59:03									
42	1	1	1.0	16	16	16.0	2	2	2.0
43	4	4	4.0	15	15	15.0	1	1	1.0
[...]									
CHP Bus-ID	qdio util.% queu fail -thp	in MB/s-	I/O reqs-						
ID	max	avg	full	erc	rd	wrt	rd	wrt	
2008-12-11 10:58:55									
42/0.0.591d	99.2	14.0	12	0	1.0	2.7	5	692	
43/0.0.5b1d	91.4	9.3	0	0	0.5	3.3	4	517	
[...]									

Data Analysis: Virtual HBA Utilization Report

```
user@larsson:~> ziorep_utilization sample.log -i 60 -c 43
[...]
```

CHP Bus-ID	qdio util.%	queu	fail	-thp in MB/s-	I/O reqs-	rd wrt	rd wrt	
ID	max	avg	full	erc	rd	wrt	rd wrt	
2012-12-11 10:52:55								
10:53:55	43/0.0.5b1d	99.2	13.4	302	0	0.2	2.6	880 15K
10:54:55	43/0.0.5b1d	98.4	10.0	203	0	0.3	3.1	163 1.3K
10:55:55	43/0.0.5b1d	99.2	12.6	178	0	0.2	3.1	1.3K 10K
10:56:55	43/0.0.5b1d	99.2	4.2	79	0	0.3	3.0	729 4.6K
10:57:55	43/0.0.5b1d	0.0	0.0	0	0	0.0	0.0	168 0
10:58:55	43/0.0.5b1d	99.2	8.9	136	0	3.7	1.9	2.9K 11K
10:59:55	43/0.0.5b1d	100.0	12.6	177	0	0.6	2.8	335 15K
11:00:55	43/0.0.5b1d	99.2	14.0	177	0	0.3	3.1	367 19K
11:01:55	43/0.0.5b1d	99.2	13.9	188	0	0.3	3.8	1.5K 14K
11:02:55	43/0.0.5b1d	99.2	14.1	162	0	0.4	3.4	1.8K 14K
[...]	43/0.0.5b1d	99.2	8.5	150	0	0.3	3.1	1.1K 8.2K

Traffic Analysis



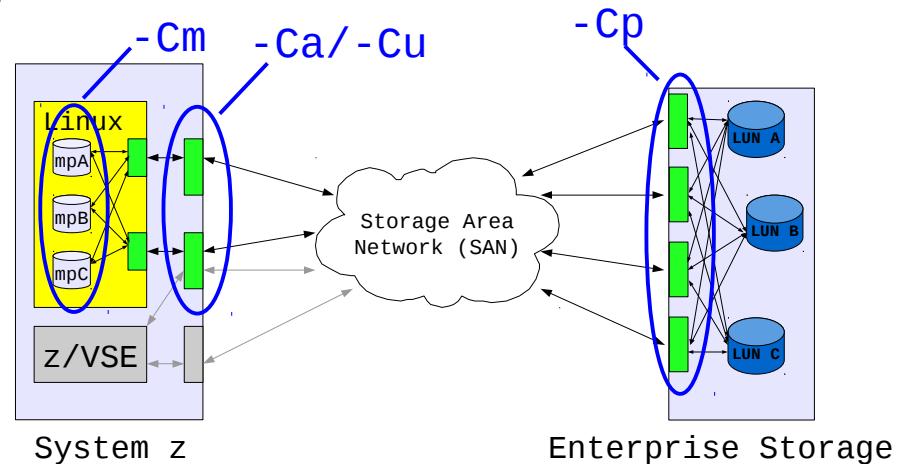
Data Analysis: Traffic Report

- *ziorep_traffic* aims at identifying latencies of all components involved
- Provides two views on data
 - Regular view listing min/max/avg/std (default)
 - Detailed view providing histograms
- Supports option '-s' for a data summary (just like *ziorep_utilization*)
- Supports device selection options '-c/-u/-p/-l/-d/-m'
(similar to *ziorep_config*)
- Provides data aggregation capabilities
- **Notes:**
 - Numbers reported consider *serialized* traffic! I.e. parallelism of requests is not accounted for. Hence throughputs are likely to divert from what other tools will report!
 - *ziomon* has a performance impact of approx. 5 percent on CPU utilization

Data Analysis: Traffic Aggregation

- Use option '-C' to specify an aggregation level:

- 'a'/'u' aggregates by CHPID / bus-ID
- 'p' aggregates by WWPN
- 'm' aggregates by multipath device
- 'A' aggregates **all** data



- Can be combined with device selection options '`-c/-u/-p/-l/-d/-m`'
- Note: Don't confuse *device selection* with *aggregation* options:
 - **Device Selection**: restricts use of data records according to the scope specified (e.g. data records for specific LUNs only)
 - **Aggregation**: generates combined results for given level of detail (e.g. combined result for all I/O handled by specified adapter)

→ Filtering applies to *input* data, aggregation shapes the resulting *output*

Data Analysis: Traffic Aggregation Examples

- Aggregate by bus-ID

```
user@larsson:~> ziorep_traffic sample.log -i 60 -Cu
Bus-ID |I/O rt MB/s|thrp in MB/s|---I/O requests---|I/O subs. lat. in us---|channel lat. in us---|fabric lat. in us---
min max avg stdev #reqs rd wrt bidi min max avg stdev min max avg stdev min max avg stdev
2012-12-11 10:52:55
0.0.591d 0.0 115.5 2.8 1.154K 12.8K 338 12K 0 169 6.7M 112.0K 126.9K 20 23K 3.877K 3.513K 72 969K 100.8K 74.16K
0.0.5b1d 0.0 76.1 2.6 1.112K 15.8K 880 15K 0 166 8.8M 127.1K 198.4K 20 23K 4.262K 3.668K 72 584K 110.7K 77.29K
10:53:55
0.0.591d 0.0 104.1 9.2 2.113K 183 20 163 0 272 191K 39.98K 51.86K 20 23K 4.460K 4.129K 73 387K 110.1K 67.69K
0.0.5b1d 0.0 12.4 3.1 1.224K 1496 163 1.3K 0 329 827K 115.6K 62.18K 20 16K 4.179K 3.406K 73 262K 108.4K 61.32K
10:54:55
0.0.591d 0.0 115.9 2.9 1.193K 14.1K 1.6K 12K 0 166 574K 120.9K 68.88K 20 24K 4.378K 3.793K 73 725K 119.8K 69.71K
0.0.5b1d 0.0 91.8 3.1 1.215K 11.7K 1.3K 10K 0 168 702K 116.6K 67.06K 20 24K 4.448K 3.916K 71 696K 116.4K 68.05K
[...]
```

- Aggregate all available devices in detailed traffic analysis mode

```
user@larsson:~> ziorep_traffic sample.log -i 60 -CA -D
|---I/O request sizes in KBytes---|
| 0   1   2   4   8   16  32  64  128 256 512 1K  2K  4K  8K >8K |
|---I/O subsystem latency in us---|
| 0   8   16  32  64 128 256 512 1K  2K  4K  8K 16K 32K 64K 128K 256K 512K 1M  2M  4M  8M 16M 32M >32M |
|---channel latency in us---|
| 0   1   2   4   8   16  32  64  128 256 512 1K  2K  4K  8K 16K 32K 64K 128K>128K |
*|---fabric latency in us---|
| 0   8   16  32  64 128 256 512 1K  2K  4K  8K 16K 32K 64K 128K 256K 512K 1M  2M  4M  8M 16M 32M >32M |
2012-12-11 10:52:55
* 293 0 0 7.0K 154 138 260 575 983 1.9K 17K 0 0 0 0 0 0
0 0 0 0 0 0 208 84 111 130 116 245 1.1K 3.4K 4.3K 5.9K 12K 747 59 40 12 6 2 0 0
0 0 0 0 0 0 1.0K 2.2K 266 224 404 2.4K 4.3K 6.9K 7.8K 3.9K 191 0 0 0 0
0 0 0 0 0 0 214 77 60 253 124 126 359 1.6K 4.2K 4.2K 6.1K 11K 777 4 0 0 0 0 0 0
10:53:55
* 0 0 0 261 6 10 24 40 66 128 1.1K 0 0 0 0 0 0
0 0 0 0 0 0 20 33 48 27 32 69 115 163 317 852 2 1 0 0 0 0 0
0 0 0 0 0 0 335 108 14 24 63 376 658 960 1.1K 664 53 0 0 0
0 0 0 0 0 0 32 149 38 147 53 44 68 111 258 381 921 2.2K 2 0 0 0 0 0 0
[...]
```

Data Analysis: Traffic Aggregation Examples (continued)

- Aggregate a selection of devices into a single interval

```
user@larsson:~> ziorep_traffic sample.log -i 0 -c A -d sdp \
-p 0x50050763030b0104 -c 42
* |I/O rt MB/s|thrp in MB/s|---I/O requests---|-I/O subs. lat. in us--|--channel lat. in us---|---fabric lat. in us---|
  min   max   avg   stdev #reqs    rd   wrt   bidi   min   max   avg   stdev   min   max   avg   stdev   min   max   avg   stdev
2012-12-11 11:51:55
*   0.0 115.9   2.9  1.181K  192K   42K 150K     0   154 315M 95.63K 1.048M    19   25K 3.725K 4.056K    68 969K 85.31K 109.5K
[...]
```

- Same selection, but detailed view

```
user@larsson:~> ziorep_traffic sample.log -i 0 -c A -d sdp \
-p 0x50050763030b0104 -c 42 -D
-----I/O request sizes in KBytes-----
|-----I/O subsystem latency in us-----
|-----channel latency in us-----
* -----fabric latency in us-----
2012-12-11 11:51:55
*   13K    0    0  55K   614   663 1.6K 3.2K 5.8K   12K 100K    0    0    0    0
  0    0    0    0    0   12K 8.0K 2.1K 3.3K 6.1K 7.6K 5.5K   18K 23K 36K 70K 826 157 21    7    4    1    0    3
  0    0    0    0    0   38K 11K 1.0K 963 2.4K 13K 23K 36K 41K 25K 2.4K    0    0    0
  0    0    0    0    0   11K 8.4K 2.6K 2.2K 3.6K 6.3K 7.7K 7.2K 21K 20K 39K 64K 693 126    0    0    0    0    0
[...]
```

Data Analysis: Hints & Tips

- Use '-i 0' to get a good starting point to identify anomalies quickly

```
user@larsson:~> ziorep_utilization sample.log -i 0
```

```
CHP|adapter in %---bus in %---cpu in %---|
  ID min max avg min max avg
2012-12-11 11:51:55
  42 0 7 0.3 0 29 2.1 1 2 1.1
  43 0 10 0.4 0 32 2.3 0 2 0.1

CHP Bus-ID |qdio util.%|queu|fail|-thp in MB/s-|I/O reqs-|
  ID max avg full erc rd wrt rd wrt
2008-12-11 11:51:55
  42/0.0.591d 99.2 1.8 1.6K 0 0.5 2.9 33K 124K
  43/0.0.5b1d 100.0 1.9 1.8K 0 0.5 2.9 31K 126K
```

```
user@larsson:~> ziorep_traffic sample.log -i 0
```

```
WWPN          LUN          |I/O rt MB/s|thrp in MB/s-|----I/O requests---|I/O subs. lat. in us---|channel lat. in us---|fabric lat. in us---|
               min      max     avg   stdev #reqs    rd   wrt   bidi   min      max     avg   stdev   min      max     avg   stdev
2012-12-11 11:51:55
0x5005076303000104:0x4021402100000000 0.0 54.5 2.9 1.195K 3094 421 2.7K 0 165 731K 115.9K 73.45K 21 23K 4.934K 4.689K 72 725K 109.3K 71.47K
0x5005076303000104:0x4021402100000000 0.0 46.4 2.9 1.192K 3106 419 2.7K 0 170 558K 117.3K 73.41K 20 24K 5.140K 4.621K 72 513K 110.5K 71.49K
0x50050763030b0104:0x4021402100000000 0.0 17.1 2.8 1.182K 2218 336 1.9K 0 171 544K 112.7K 70.57K 20 22K 4.563K 4.364K 72 543K 106.2K 68.57K
0x50050763030b0104:0x4021402100000000 0.0 47.8 3.3 1.269K 1514 228 1.3K 0 172 333K 95.84K 65.99K 22 20K 4.190K 3.936K 74 323K 89.22K 63.65K
0x5005076303100104:0x4021402100000000 0.0 48.3 1.2 750.4 1784 797 987 0 163 178M 189.0K 4.210M 19 21K 3.192K 4.309K 72 717K 78.39K 98.46K
[...]
```

```
user@larsson:~> ziorep_traffic sample.log -i 0 -D
```

		I/O request sizes in KBytes																									
		0	1	2	4	8	16	32	64	128	256	512	1K	2K	4K	8K	>8K										
WWPN	LUN	I/O subsystem latency in us																									
		0	8	16	32	64	128	256	512	1K	2K	4K	8K	16K	32K	64K	128K	256K	512K	1M	2M	4M	8M	16M	32M	>32M	
		channel latency in us																									
		0	1	2	4	8	16	32	64	128	256	512	1K	2K	4K	8K	16K	32K	64K	128K	256K	512K	1M	2M	4M	8M	16M
		fabric latency in us																									
		0	8	16	32	64	128	256	512	1K	2K	4K	8K	16K	32K	64K	128K	256K	512K	1M	2M	4M	8M	16M	32M	>32M	

```
2012-12-11 11:51:55
0x5005076303000104:0x4021402100000000 179 0 0 591 14 7 21 37 103 168 2.0K 0 0 0 0 0 0 0 0 41 65 214 308 596 1.5K 40 1 0 0 0 0 0 0
0 0 0 0 0 0 160 23 46 35 36 41 65 214 308 596 1.5K 40 1 0 0 0 0 0 0 0
0 0 0 0 0 0 314 110 7 9 36 251 361 604 763 549 92 0 0 0 0 0 0 0 0
0 0 0 0 0 0 160 38 14 75 38 32 37 61 240 282 671 1.4K 33 1 0 0 0 0 0 0 0
0x5005076303100104:0x4021402100000000 179 0 0 725 3 3 13 28 27 76 730 0 0 0 0 0 0 0 0 48 56 97 125 176 578 9 29 0 3 0 0 0 0 0
0 0 0 0 0 0 178 431 24 11 18 48 56 97 125 176 578 9 29 0 3 0 0 0 0 0
0 0 0 0 0 0 658 139 20 11 12 75 137 205 254 261 22 0 0 0 0 0 0 0 0
0 0 0 0 0 0 163 451 15 18 10 21 55 56 123 118 185 545 8 26 0 0 0 0 0 0 0
[...]
```

Data Analysis on Other Platforms

- Reporting tools do not depend on s390 architecture
- To compile on other platforms:
 - Extract source of *s390-tools* package, either from your distribution's source rpm, or download from
<http://www.ibm.com/developerworks/linux/linux390/s390-tools.html>
and extract via

```
user@larsson:~> tar xvfj s390-tools-1.17.0.tar.bz2
```
 - Change to *zimon* subdirectory, build and install

```
user@larsson:~> cd s390-tools-1.17.0
user@larsson:~> make
user@larsson:~> sudo make install
```
- Run reporting tools on any given data

Data Analysis on Other Platforms (continued)

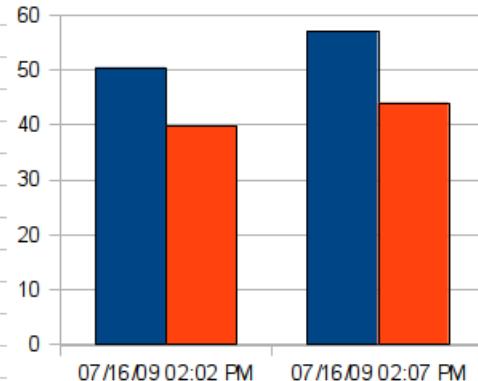
- Use option '-x' in `ziorep_utilization` and `ziorep_traffic` to export data to .csv format, e.g.

```
user@larsson:~> ziorep_traffic data.log -x -i 300 -b "2012-11-27
08:45:21" -c a
```

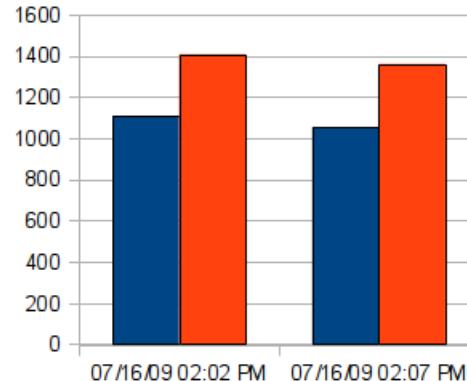
- All data filtering options specified will apply to exported data
- Import in spreadsheet and perform further processing

	A	B	C	D	E	F
1	timestamp	aggregated	CHPID	I/O rate in MB/s min	I/O rate in MB/s max	throughput in M
2	07/16/09 02:02 PM	0	3a	1.49	88.2	
3	07/16/09 02:02 PM	0	47	1.44	67.25	
4	07/16/09 02:07 PM	0	3a	0.53	103.14	
5	07/16/09 02:07 PM	0	47	0.27	68.86	
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						
21						
22						

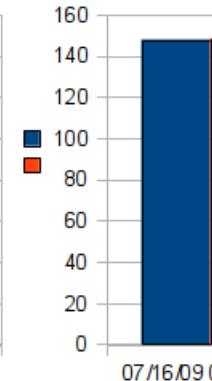
Throughput Avg
(higher is better)



I/O Latency Avg
(lower is better)



Ch:



References

- Linux on System z on DeveloperWorks

<http://www.ibm.com/developerworks/linux/linux390>

- Linux on System z Documentation

http://www.ibm.com/developerworks/linux/linux390/distribution_hints.html

Specifically:

- Device Drivers, Features, and Commands
- How to use FC-attached SCSI devices with Linux on System z (chapters 11&12)

- Linux on System z – Downloads

http://www.ibm.com/developerworks/linux/linux390/development_recommended.html

Questions?



Stefan Raspl

*Linux on System z
Development*

Schönaicher Strasse 220
71032 Böblingen, Germany

Phone +49 (0)7031-16-2177
stefan.raspl@de.ibm.com