Sa Liu - Linux on IBM z Systems Service & Support <saliu@de.ibm.com>
Live Virtual Class - Jan 28, 2015

Problem Determination for Linux on IBM z Systems

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

- Introduction
- The service process
- Tools for collecting and analyzing data
- Some customer cases

Introduction

- Problem analysis may look like a straight forward process on charts, but it can take weeks to get it done.

  A problem does not necessarily show up on the place of origin

- The more information is available, the sooner the situation can be understood. Reproducing an issue and submitting data again and again usually introduces delays.

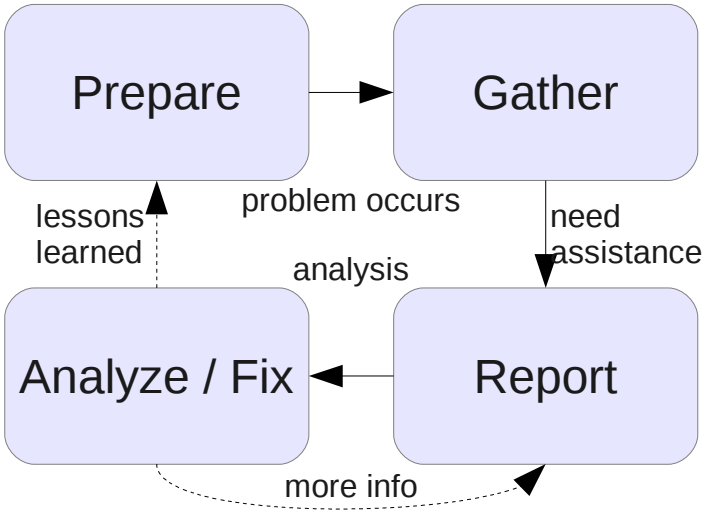
  Data of the 'healthy system' can be very helpful also

- Being prepared to collect and submit data in a problem situation will ease up the whole process.

  Data collection is required right after the occurrence of a problem

The service process

- Prepare
- Gather
- Report
- Analyze / Fix



Prepare

- Preparation for recovery
- Monitor system behavior
  - Keep healthy system data for comparison
- Preparation for gathering information:
  - Install packages for Linux tools
  - Have disks ready for system dump
  - Be aware of system settings and configuration

Gather - environment and workload

- Describe the hardware configuration
  - Machine type and resources (CPU, memory, I/O cards)
  - Storage and network attachments
- Describe the software configuration
  - Information about the used software packages
    - Operating systems
    - Middle-ware packages
  - Characteristics of workload
- Application / infrastructure setup
  - Clients
  - Network topology
  - Disk configuration

Gather - describe the problem

- What are the symptoms of the problem?
- When did it happen?
  - Date and time – as accurate as possible
  - Single occurrence or multiple times
- Any obvious reasons for the problem occurrence?
  - Is there a specific procedure to enforce the problem
  - Any recent changes to the environment
- What is impacted?
  - Single system, multiple systems, or multiple z Systems machines
  - Production, test, or development environment
- What is the expectation to have a 'healthy system'?

Gather - collect debug information

- In case of a system hang
  - Take a system dump (refer to 'Using the dump tools' for details)
  - For customers having specific kernels, please provide System.map, Kerntypes (if available) and vmlinux
  - Once the system is up again, collect also dbginfo.sh

- In case of a noticeable issue during operations
  - **Collect dbginfo.sh before doing any recovery actions!**
  - Collect the sadc/sar data
  - In case that third party programs are involved, provide the configuration, logs/traces, and – if available – performance data for the affected system(s)

In each of the two scenarios, additional data might be required for analysis such as trace data, z/VM or LPAR data, firmware data and/or data from external components such as switches, storage servers etc.

Report – report the problem to service team

- Problem Report
  - Include the environment and workload description
  - Include the description of the problem situation
  - Upload the collected data (such as system dump, dbginfo.sh, sadc/sar data)
- Customers with an IBM service contract can open a Problem Management Record – PMR
  - Provide all the details that you have gathered
  - Upload the gathered data to Enhanced Customer Data Repository (ECuREP)
    - See the EcuREP web pages for further details
- In case multiple tickets are opened e.g. at IBM and the Linux Distribution Partners, make them aware of each other

Analyze / Fix – providing a fix for an issue

- The following is a usual process to deliver a fix for an issue:
  - IBM receives a problem report
    (from customers or Linux Distribution Partners - LDPs)
  - Analyze the data and identify the root cause
  - Prepare a patch (includes usually verification tests)
  - Deliver the patch to the LDPs
  - LDPs pick up the patch and deliver the updated packages to customers

Tools for problem determination

- General tools
- Performance tools
- Dump tools
- Special features and tools

General tools

- dbginfo.sh – collects debugging information and system configuration
- supportconfig – SLES
- sosreport - RHEL

General tools – dbginfo.sh (cont'd)

Sample output of the script dbginfo.sh

```
[root@system]# dbginfo.sh
dbginfo.sh: Debug information script version 1.15.0-0.136.3
Copyright IBM Corp. 2002, 2013

Hardware platform     = s390x
Kernel version        = 3.0.76 (3.0.76-0.7-default)
Runtime environment   = z/VM

1 of 7: Collecting command output

2 of 7: Collecting z/VM command output

3 of 7: Collecting procfs

4 of 7: Collecting sysfs

5 of 7: Collecting log files

6 of 7: Collecting config files

7 of 7: Collecting osa oat output skipped - not available

Finalizing: Creating archive with collected data

Collected data was saved to:
>>  /tmp/DBGINFO-2014-06-20-10-42-42-system-123456.tgz  <<
```

General tools – dbginfo.sh (cont'd)

- dbginfo.sh script is required to run before rebooting the system
- dbginfo.sh script continues to run even on issues during data collection
- dbginfo.sh script mounts the debugfs/s390dbf automatically to collect the s390dbf traces
- Collecting the sysfs can take some time dependent on the number of devices being attached
- Running dbginfo.sh script requires 'enough' disk space under /tmp
- The latest version (V1.26.0) allows to specify a directory for data collection

Check out:
http://www.ibm.com/developerworks/linux/linux390/s390-tools.html

Performance tools

- sadc – system activity data collector
- iostat – monitors I/O device load and the CPU utilization
- z/VM MONWRITE – collects CP *MONITOR data
- dasdstat – displays DASD performance data
- ziomon/ ziorep – collects FCP performance data and generate reports

Performance tools – sadc / sar

- sadc – System Activity Data Collector (data gatherer)
  - Syntax of script execution
    `/usr/lib64/sa/sadc [options] [interval [count]] [binary_outfile]`
- Think about the right interval
  - Too small - too much data & overhead, can mask the issue
  - Too large - values are too “averaged”, peaks no longer visible
- Option '-d' enables collection of disk statistics
  - See man page for all options
- Data is stored in binary format and requires translation through 'sar'
  - Translation requires same environment as data collection

```
[root@system]# /usr/lib64/sa/sadc -d 10 30 sadc_output
```

Performance tools – sadc / sar (cont'd)

- sar – System Activity Report (reporting tool)
  - Syntax of script execution
    `sar [options] sadc_outfile [> sar_outfile]`
- Option '-A' generates all available reports
  - Please refer to the man page for all options
- Allows to generate specific reports only
  - – e.g. for CPU, memory, and disk I/O
- Start and end times can be used to generate a problem specific report

```
[root@system]# sar -A -f sadc_output > sar_output
```

Dump tools

When the system hangs, create a kernel dump

- DASD dump tool
  - writes the dump directly to a DASD partition.
- Tape dump tool
  - writes the dump directly to a tape device.
- SCSI dump tool
  - writes the dump to a file system
  - or to a SCSI partition (kernel 3.12)
- kdump
  - preloaded kdump kernel and initrd in memory
  - writes the dump to storage or transfer it over the network
- VMDUMP (for z/VM guest operating systems)
  - writes the dump to z/VM spool space (z/VM reader).

Check the documentation 'Using the dump tools'

Special features and tools

- s390dbf traces – uses the kernel debug feature
- top – shows resource usage
- ps – reports a snapshot of the current processes
- netstat – shows information about the Linux networking subsystem
- tcpdump – collects traffic information for a network interface
- oprofile – profiling of all running code on Linux systems

Special features – s390dbf traces

- z Systems specific driver tracing environment
  - Uses ring buffers
  - Available in live system and in system dumps
- Must be mounted for live view:
  - 'mount -t debugfs /sys/debug /sys/kernel/debug'
- Each component has these control interfaces
  - level – controlling the trace detail between 0 <--> 6 (default: 2)
    - 'echo 6 > level'
  - pages – shows and defines the pre-allocated space:
    - 'echo 20 > pages'
  - flush – cleans the ring buffer:
    - 'echo 1 > flush'
- And one of these output files
  - hex_ascii – output is not that human readable, but very useful for debugging
  - sprintf – human readable output, usually an event log

Customer cases

- Case 1: SCLP console
- Case 2: RHEL7 installation fails on zEnterprise
- Case 3: DASD partition recognition

Customer cases - #1 SCLP console

- Problem description
  - Fail-over of Linux machine with no obvious reason
- Environment
  - SLES11 SP2, RHEL6 U4 running on LPAR
- Tools used for problem determination
  - dbginfo.sh
  - Hardware traces of z Systems machine
- Result of analysis
  - Fail-over triggered due to Linux machine being unresponsive
  - Pressure on SCLP console identified, which blocks Linux instance
- Solution
  - Feature has been implemented for the SCLP console device driver that allows to drop messages in case the Linux instance issues more messages than the SCLP interface of the z Systems machine can accept

Customer cases - #2 RHEL7 installation fails on zEnterprise

- Problem description
  - On a z196, a customer tried to install RHEL7 in a z/VM guest, which is part of an SSI cluster.
  - The RHEL7 installation failed with the following message:

```
The Linux kernel requires more recent processor hardware
```

- Environment
  - RHEL7, but also applies to SLES12
  - z/VM 6.2

- Tools used for problem determination
  - z/VM commands

Customer cases - #2 RHEL7 installation fails on zEnterprise (cont'd)

- Result of analysis
  - The error message indicated that the installation was not performed on a z196 or later system
  - Requested the output of the z/VM command:

```
#CP q cpuid
CPUID = FF1F0B8220978000
```

  - The command returned the machine type of a System z10 -> 2097
  - The z/VM LPAR in question was part of an SSI cluster with a default domain, which included members on a System z10. This downgraded the SSI cluster architecture level to a System z10.

- Solution
  - Disable relocation of the z/VM guest, to allow using the native member architecture
  - To run a RHEL7 or SLES12 system as z/VM guest in an SSI cluster, the architecture level must be z196 or later

Customer cases - #3 DASD partition recognition

- Problem description
  - After reboot one of the DASD devices did not show up correctly
- Environment
  - SLES11 SP2 running on LPAR
- Tools used for problem determination
  - dbginfo.sh
  - supportconfig
- Result of analysis
  - From a code review, a raise condition has been recognized in the DASD device driver. The ioctl for partition detection can fail, in case another process tries to open the device in parallel
- Solution
  - The DASD device driver has been improved for recovery actions in case the partition detection fails for specific reasons.

Summary

- There is no reason to worry about Linux on IBM z Systems
- Only a small number of issues result in a code fix
- Be prepared for the worse case to speed up analysis and
  … make yourself familiar with the tools
- The delivered information and data are the key for success
  - System and environment information
  - Workload characteristics
  - Clear problem description
- Immediate data collection is important
  - dbginfo.sh to collect the system configuration, logs, traces and runtime information
  - sadc/sar to collect the 'initial set' of performance data
  - System dump in case the system hangs

References

- IBM Linux on IBM z Systems
  www.ibm.com/systems/z/os/linux/
- IBM developerWorks – Linux on IBM z Systems
  ('*Device Drivers, Features and Commands*', '*Using Dump Tools*', '*Troubleshooting Guide*')
  www.ibm.com/developerworks/linux/linux390/
- IBM KnowledgeCenter – Linux on System z
  www.ibm.com/support/knowledgecenter/linuxonibm/liaaf/lnz_r_main.html
- IBM Linux on System z – Live Virtual Classes
  www.vm.ibm.com/education/lvc/zlinlvc.html
- IBM Redbook: *Problem Determination for Linux on System z*
  http://www.redbooks.ibm.com/abstracts/sg247599.html
- IBM Techdocs
  www.ibm.com/support/techdocs/atsmastr.nsf/Web/Techdocs

Questions?

Sa Liu

Linux on IBM z Systems
Service & Support

IBM Systems

Schoenaicher Strasse 220
D-71032 Boeblingen
Mail: Postfach 1380
D-71003 Boeblingen

Phone (+49)-7031-16-3104
saliu@de.ibm.com