Collecting CPU Measurement Facility SAMPLES on z/OS

Overview

The CPU Measurement Facility (CPU MF) allows you to collect Counter data and/or Sample data. Counter data provides counts of hardware events across the whole LPAR. This information can be used to understand how your hardware and software are interacting on the LPAR under test. CPU MF Sample data has a different purpose, allowing the user to sample instruction addresses with very low overhead, providing the information necessary to determine what executable code is consuming the most CPU resources.

Normally, CPU MF Counter data is collected over longer periods of time. Data are written to SMF at regular intervals and later post-processed. If this is the type of measurement you would like to do, collecting only CPU MF Counter information, please set this document aside and download another document, "Collecting CPU MF (COUNTERS) on z/OS" for instructions. http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/TC000041

This document describes how to collect CPU MF Sample data. In most cases, CPU MF Sample data are only collected for a few minutes (default 10 minutes) during a key time in the execution of your workload. On z/OS, CPU MF Sample data is collected using Hardware Instrumentation Services (HIS). Using HIS, you can capture sample data, plus counters and a module map using basic z/OS console commands.

This document describes how to collect CPU MF Sample data on z/OS in 4 steps:

Step 1	**Verify System and Software Requirements**
Step 2	**Configure the Server to collect CPU Measurement Facility Data**
Step 3	**Configure z/OS to collect CPU Measurement Facility Sample Data**
Step 4	**Collect CPU Measurement Facility Samples (and Counters)**

© 2010-2011 IBM Corporation

Step 1

Verify System and Software Requirements

All the following requirements must be met in order to successfully capture CPU MF Sample data.

Hardware:
The CPU Measurement Facility is only available on IBM System z10 or later machines. To use CPU MF, a System z10 machine must be at GA2 Driver 76D – Bundle #20 or higher. A System z10 BC sub-capacity machine must also be at Bundle#20 or higher**.**

Software:
The z/OS LPAR being measured must be at z/OS 1.8 or higher and PTFs for the following APARs must be applied.
- z/OS 1.8 - APARs OA25755, OA25750, OA25773, and OA27623
- z/OS 1.9 - APARs OA25755, OA25750, OA25773, OA27623, and OA32113
- z/OS 1.10 - APARs OA25755, OA25773, OA27623, OA32113, and OA34485
- z/OS 1.11 - APARs OA30429 and OA34485
- z/OS 1.12 - APAR OA34485

Step 2

Configure the Server to collect CPU MF Data

This step only needs to be done once for each LPAR of interest.

Before you collect CPU Measurement Facility data on the System z, it is necessary to enable this support for the LPARs you wish to measure. If you fail to do this, you will get a message like this.

**"HIS026I MODIFY HIS COMMAND CANNOT BE PROCESSED.
BASIC COUNTER SET IS UNAUTHORIZED".**

> **Note:** More information on the panels described in this section can be found in the System z10 PR/SM Planning Guide.
> This guide can be found at:
> http://www.ibm.com/support/docview.wss?uid=isg209611e17c3b8d419852573f700645d4d

For Activated LPARs

For most cases, the LPAR to be measured will already be activated and running. In this case a dynamic update can be used to enable CPU Measurement Facility without disrupting the LPAR to be measured.

If the LPAR to be measured is already activated, from HMC or on the SE, select the Change LPAR Security icon. Then, for the LPAR(s) of interest, check the box for each of the following settings.

Counter Facility Security Options
- Basic counter set authorization control
- Problem state counter set authorization control
- Crypto activity counter set authorization control
- Extended counter set authorization control
  (Note: the CPU MF Group Counter is not currently captured or exploited by z/OS Hardware Instrumentation Services, so this Counter option need not be selected.)

Sampling Facility Security Options
🔧 Basic sampling authorization control

https://9.12.16.150 - H128: Change LPAR Security - Mozilla Firefox: IBM Edition

Change Logical Partition Security - H128

Input/output configuration data set (IOCDS): a0 H128BASE

Logical Partition	Active	Performance Data Control	Input/Output Configuration Control	Cross Partition Authority	Partition Isolation	Basic Counter	Problem State Counter	Crypto Activity Counter	Extended Counter	Group Counter	Basic Sampling
H128CF1	No	☑	☑	☐	☐	☐	☐	☐	☐	☐	☐
H128CF2	No	☑	☑	☐	☐	☐	☐	☐	☐	☐	☐
H128LP01	Yes	☑	☑	☐	☐	☑	☑	☑	☑	☐	☑
H128LP02	No	☑	☑	☐	☐	☑	☑	☑	☑	☐	☑
H128LP03	No	☑	☑	☐	☐	☐	☐	☐	☐	☐	☐
H128LP04	No	☑	☑	☐	☐	☐	☐	☐	☐	☐	☐
H128LP05	No	☑	☑	☐	☐	☐	☐	☐	☐	☐	☐
H128LP06	No	☑	☑	☐	☐	☐	☐	☐	☐	☐	☐
H128USS	No	☑	☑	☐	☐	☐	☐	☐	☐	☐	☐
H128USS2	No	☑	☑	☐	☐	☐	☐	☐	☐	☐	☐
H12811	No	☑	☑	☐	☐	☐	☐	☐	☐	☐	☐
LP12	No	☑	☑	☐	☐	☐	☐	☐	☐	☐	☐
H12824	No	☑	☑	☐	☐	☐	☐	☐	☐	☐	☐
H12825	No	☑	☑	☐	☐	☐	☐	☐	☐	☐	☐
H12826	No	☑	☑	☐	☐	☐	☐	☐	☐	☐	☐
H12827	No	☑	☑	☐	☐	☐	☐	☐	☐	☐	☐
H12828	No	☑	☑	☐	☐	☐	☐	☐	☐	☐	☐
H12829	No	☑	☑	☐	☐	☐	☐	☐	☐	☐	☐
LP2A	No	☑	☑	☐	☐	☐	☐	☐	☐	☐	☐
LP21	No	☑	☑	☐	☐	☐	☐	☐	☐	☐	☐

| Save and Change | Change Running System | Save to Profiles | Reset | Cancel | Help |

Done

Click the **Save and Change** button.
The Save and Change on the lpar security panel will dyamically change the running system and also save the change to the activation profiles.

For LPARs which have not yet been Activated

For LPARs which have not yet been activated, CPU Measurement Facility can be enabled using the security tab for the activation profile.
Access the security tab and verify that ALL of those same required settings listed above are checked.

Once all the required settings are checked, proceed to activate the image.

Click the **Save** button.
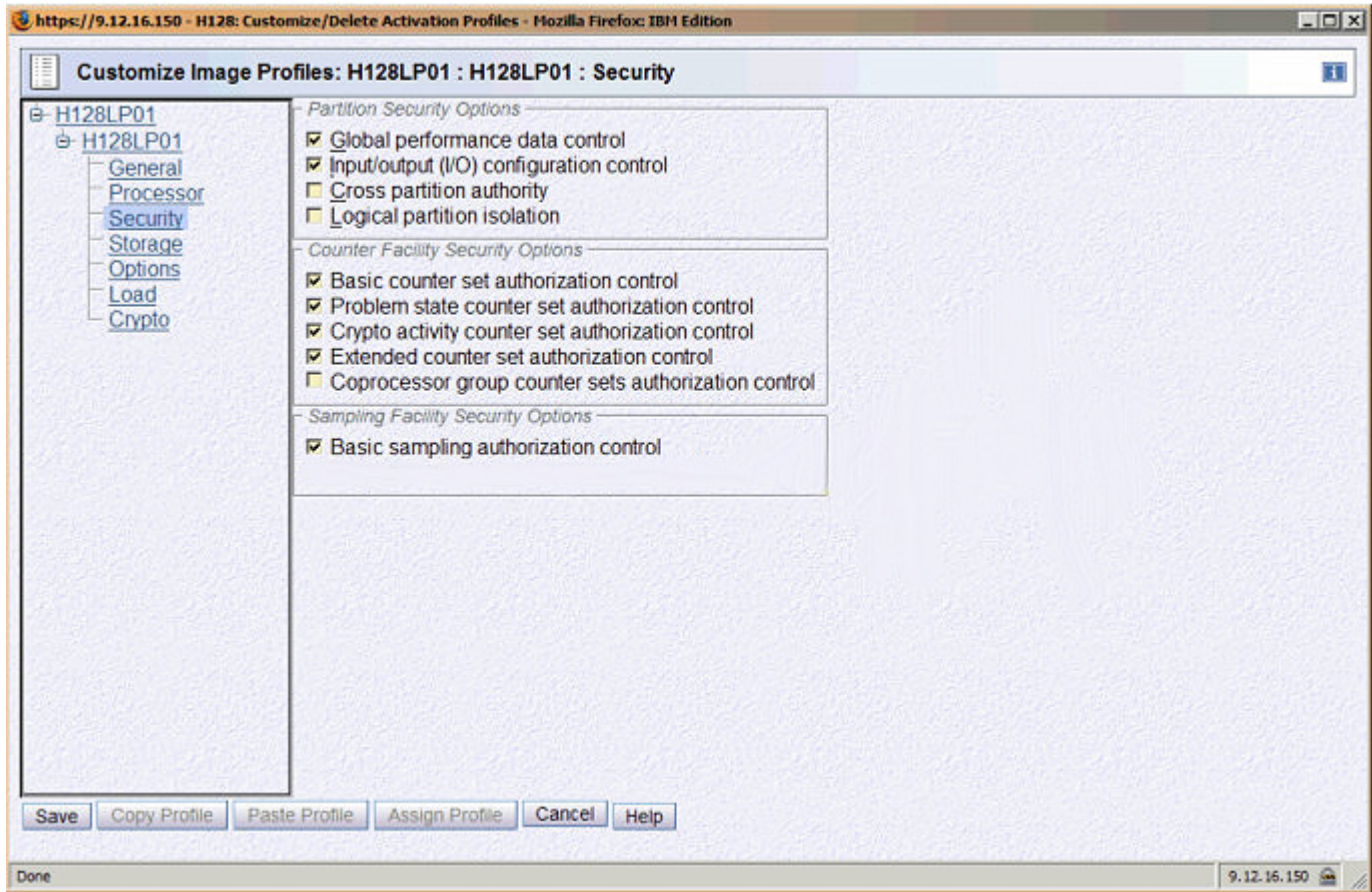
○ **Step 3** **Configure z/OS to collect CPU MF Sample Data**

This step only needs to be done once on each LPAR of interest.

1. The z/OS operating system provides access to CPU Measurement Facility data through Hardware Instrumentation Services (HIS). To use Hardware Instrumentation Services, you must be running z/OS 1.8 or higher with the PTFs for APARs OA25755, OA25750, OA25773, and OA27623 applied. In addition, for z/OS 1.9 and z/OS 1.10 systems, the PTF for APAR OA32113 should be applied and for z/OS 1.11 systems, the PTF for APAR OA30429 should be applied.
   - Note that installing these updates requires an IPL with CLPA.
   - IBM recommends that you check regularly for additional service updates for HIS.

2. If you don't have a **HIS** member in **SYS1.PROCLIB**, copy the following procedure into the member, **HIS**, in your PROCLIB. This provides a method to start Hardware Instrumentation Services (HIS).

SYS1.PROCLIB(HIS)

```
//HIS      PROC
//HIS      EXEC PGM=HISINIT,REGION=0K,TIME=NOLIMIT
//CMDFILE1 DD DUMMY
//CMDFILE2 DD DUMMY
//SYSPRINT DD SYSOUT=*
```

3. Authorize the HIS procedure to write data to the OMVS File System
    *   Add a RACF userid for the HIS procedure

    TSO Command to add a HIS userid

    adduser HIS omvs(uid(123) home('/HIS'))

    where **UID(123)** is the OMVS uid and **/HIS** is the default home directory.

    > **Note:** While OMVS access is required, there is no special authorization needed.
    > Also, any directory can be used for the HOME directory.

    Example: Define a HIS userid for the HIS-started task

    ```
    //DAEMONS  EXEC PGM=IKJEFT01
    //SYSTSPRT DD SYSOUT=*
    //SYSTSIN  DD *
     ADDUSER HIS OMVS(123) HOME('/HIS'))
    /*
    ```

4. Define the HIS-started task to RACF:
    To set up the HIS-started task to RACF, you need to define a profile for it to the RACF generic resource class called STARTED using the RDEFINE command.

    > **Note:** If the STARTED class is not active, RACLISTed, and GENERIC profile checking is
    > not activated, use the RACF SETROPTS command to activate (CLASSACT), RACLIST,
    > and GENERIC the STARTED class first before you can define a new profile to it. In most
    > environments, this might already have been done. Therefore, you might not have to include
    > the SETROPTS command to CLASSACT, RACLIST, and GENERIC the STARTED class.
    > In this example, we include these SETROPTS commands for completeness. Running these
    > commands when the STARTED class is already activated has no effect.

    Example: RACF commands to define HIS-started task to RACF

    ```
    //DAEMONS  EXEC PGM=IKJEFT01
    //SYSTSPRT DD SYSOUT=*
    //SYSTSIN  DD *
     SETROPTS CLASSACT(STARTED)
     SETROPTS RACLIST(STARTED)
     SETROPTS GENERIC(STARTED)
     RDEFINE  STARTED  HIS.* STDATA(USER(HIS) TRUSTED(YES))
     SETROPTS RACLIST(STARTED) REFRESH
     SETROPTS GENERIC(STARTED) REFRESH
    /*
    ```

5. Specify where to store the HIS output file:
- Create the HOME directory in a local USS file system by issuing the "**mkdir**" command under USS. Also assign read/write/exec authority.

```
tso omvs

mkdir /HIS
```

In this example, /**HIS** will be the default directory where the HIS output file will be stored.

> **Note:** If you plan to capture lots of sample data, this output directory for HIS data needs to be large enough.
> For sample data, a directory with 1 GB available is recommended

6. Optional: Configure for collection of SMF 113 subtype 2 records.
- Counter data can be captured in SMF records as well as in the *.CNT file.
  If SMF data is desired (available for Counter data only), do the following.
  - Start collecting SMF Type 113 records by adding 113 to the SYS statement and if necessary to the SUBSYS(STC(TYPE statements in the **SMFPRMxx** member(s) for each LPAR for which data is being collected.
    SMF type 113, subtype 2 records are provided by the Hardware Instrumentation Services (HIS) support.
  - Dynamically pick up the changes to the **SMFPRMxx** member(s) by issuing **SET SMF=xx** command on each LPAR of interest.
  - Verify that the SMF type 113 records are being collected by issuing **D SMF,O** command on each LPAR of interest. Look for output similar to this:

    SYS(TYPE(0,2,3,6:10,14,15,22:24,26,30,32,33,41,42,47:48,59,60:69,70:79,80:83,85,
    88,89,90:91,94,98,100:103,108,110,113,115:117,130,134,148:151,161,200,
    244,245)) – PARMLIB
    SUBSYS(STC,TYPE(23,42,70:79,99:103,110,113)) -- PARMLIB

Step 4

Collect CPU Measurement Facility Samples (and Counters)

> **Note:** The z/OS LPAR being measured must be at z/OS 1.8 or higher with the PTFs for APARs OA25755, OA25750, OA25773, and OA27623 applied.
> In addition, for z/OS 1.9 and z/OS 1.10 systems, the PTF for APAR OA32113 should be applied and for z/OS 1.11 systems, the PTF for APAR OA30429 should be applied.

Perform this step each time you want to collect HIS sample data.

1. First, for each LPAR where you will be collecting HIS sample data, determine the maximum sampling frequency to ensure that you don't introduce significant overhead on your system. Please see Appendix A in this document for instructions on how to determine the maximum recommended sampling frequency for your LPAR, assuming that

the processor will be active for the duration of the measurement.

2. Issue the following z/OS Command to start the Hardware Instrumentation Services (HIS) Address Space :

System Console
```
s his
```

The following message indicates a successful start:
HIS002I HIS INITIALIZATION COMPLETE

> **Note:** If the correct Driver is not installed on the system, error message
> "**HIS031I REQUESTED FUNCTION IS UNSUPPORTED BY HARDWARE**"
> will be reported upon unsuccessful completion of the z/OS start HIS Address Space Command.

3. At the appropriate time during the measurement **(which should take place during specific time(s) such as Peak Hours)**, issue the following z/OS Command to start CPU Measurement Facility Data Collection:

System Console
```
f his,b,tt='runid',path='/HIS',ctr=(b,p,c,e),mas=all,buf=1024,st=b,sf=800000,dur=10
```

Where:

- **B** tells HIS to Begin collecting the MF Data described hereafter.
- **TT** allows you to specify a unique label or **'runid'** to identify this measurement.
- **PATH** identifies where the HIS output files will get written. If the user does not specify the **PATH** option, files will be written to the HIS userid's HOME directory specified in the OMVS RACF segment.
- **CTR** indicates to HIS that Counters are to be collected. If the user specifies "**ALL**", all available counter sets will be captured. Optionally, the user can select the counter sets of interest : "**B**" for the **B**asic counter set; "**P**" for the **P**roblem counter set; "**C**" for the **C**ryptographic counter set; and/or "**E**" for the **E**xtended counter set.
- **MAS or (MAPASID)** indicates to HIS what address spaces should be included in the map. The "**ALL**" option (recommended) requests a map of all swapped in address spaces, Optionally, the MAPASID or MAPJOB option can be used to capture a map of only the address spaces of interest. If you specify MAPASID (MAS) or MAPJOB (MJOB) you will always map the Nucleus and LPA areas as well.
- **BUF** identifies the number of 4K buffers for each logical processor where the Samples will be written out to during the measurement. The recommended value is **1024** for minimal overhead at the expense of more storage (1024 4K blocks per processor).
- **ST** identifies the sample type. The possible values are **B**asic OR **D**iagnostic. Unless instructed by IBM, use ST=Basic.
- **SF** is your sampling frequency. **This value indicates the number of samples per minute** that will be collected during the measurement. The permitted frequency values are 1 to 2,400,000 (Samples/Minute). The default sampling rate is 800,000 samplers per minute for 10 minutes, providing 8 million samples. On smaller machines, however, this default sampling rate could cause excessive overhead.   Please see appendix A in this document to determine the maximum recommended sampling rate for your machine.
- **DUR** is an optional parameter for the duration of the measurement in minutes. The permitted range is 1 to 1,440 minutes. If the HIS is instructed to collect Samples (BASIC OR DIAGNOSTIC) and the measurement is expected to last longer than 10 minutes, the user will need to specify this parameter, since the default duration for a measurement with samples is 10 minutes. The **DUR** value can be set to the exact duration of the measurement or set to a time longer than

the measurement interval for those cases where the user would prefer to control the ending of the data collection with the **"f his,e"** command.

Successful start of CPU Measurement Facility data collection is indicated by the following message:
HIS011I HIS DATA COLLECTION STARTED

> **Note:** If any set up errors are encountered such as an option error, then, data collection can be stopped by issuing an "**f his,e**" command and restarting data collection with the correct options. To recover in some cases, it may be necessary to stop HIS ("**stop his**") and restart HIS ("**s his**").

> **Note:** For more information about collecting HIS data, please see the manual, z/OS V1R10.0 MVS System Commands, which explains the details about using HIS commands.
> **http://publibz.boulder.ibm.com/cgi-bin/bookmgr_OS390/BOOKS/ IEA2G191/4.21.17?SHELF=&DT=20090114020956**
> or
> **https://submit.boulder.ibm.com/zos390/zosr10/ga/iea2g192.pdf**

4. If the measurement needs to be ended before the 10 minutes or the specified DUR= value, issue the following z/OS command to stop CPU Measurement Facility data collection:

System Console
> f his,e

Successful stop is indicated by a message similar to the following:

HIS022I HIS DATA COLLECTION IS ENDING. 464
OUTPUT FILE PREFIX: SYSHIS20080413.223225

The file which contains the Counters will have a name whose format is similar to this :
SYSHIS20100301.143520.000.CNT
The file(s) which contain the Samples will have name(s) whose format is similar to this :
SYSHIS20100301.143520.000.SMP.01
There will be ONE of these files PER Logical PU for the LPAR.
The file which contains the module map will have a name whose format is similar to this:
SYSHIS20100301.143520.000.MAP

Note the output file names. These Unix System Services files can be found in the directory specified using the PATH option, or in the HOME directory associated with the HIS proc (the /**HIS** directory in our example).

These files are required to reduce and analyze the data from the measurement. It is also worth noting that there is no need to pre-allocate these files. z/OS will create the files as needed. However, make sure there is enough space in the HFS directory for the .SMP files which are created when collecting is started and are continually written to during the duration of the measurement. Also, the MAP file, written when data collection is stopped, can be large if a map of all address spaces is requested (as recommended). An HFS directory with at least 1GB of space should suffice for most sample measurements.

Determining your maximum recommended sampling rate

Hardware Instrumentation Services provides a default sampling rate of 800,000 samples per minute. In 10 minutes, this allows you to capture up to 8 million samples, plenty of samples for very accurate estimates of where CPU time is being consumed. However, on smaller sub-capacity machines, this high sampling rate could severely degrade system performance. The approach below to determine your maximum recommended sampling rate is intended to allow you to adjust the sampling frequency and duration of the measurement as needed to keep the system overhead low, while collecting enough samples to get accurate estimates of what code is consuming the most CPU time.

1. First, determine the number of logical processors, including zAAPs and zIIPs on your LPAR. This number is important because Hardware Instrumentation Services tries to reach your requested sampling frequency across all the processors. Therefore, the sampling frequency on any individual processor should be about Requested_Sampling_Frequency / Number_of_Logical_Processors.

2. Note that the weight of your LPAR may affect the total samples collected, but this weight is not a factor in selecting a maximum sampling frequency. The sampling frequency set for each CPU is calculated assuming that the processor will be active for the duration of the measurement. Samples are only captured when the LPAR under test is running on the CPU.

3. Use the table below to locate the machine model on which your LPAR is running. Once your machine model, not the maximum recommended samples per minute per logical processor in column 3. Let's call this value MaxRecommendedSamples.

4. To keep the overhead low, the Sampling Frequency (SF) value used to collect HIS sample data should not be more than MaxRecommendedSamples * Number_of_Logical_Processors.

Example:
- LPAR runs on 2098-N05 machine
- LPAR has 4 general CPUs and no zAAPs or zIIPs
- Number of CPUs is 4
- From the table below, the MaxRecommendedSamples value for a 2098-N01 is 149,002.
- Therefore, the SF value used for your LPAR should not exceed 149,002 * 4 = 596,008

Machine	PCI	Maximum recommended samples per minute per logical processor
2817-701	1202	1,066,075
2817-601	768	681,153
2817-501	588	521,508
2817-401	240	212,860
2097-701	902	800,000
2097-601	626	555,211
2097-501	462	409,756
2097-401	214	189,800
2098-Z01	661	586,253
2098-Y01	604	535,698
2098-X01	528	468,293
2098-W01	471	417,738
2098-V01	421	373,392
2098-U01	373	330,820
2098-T01	333	295,344
2098-S01	297	263,415
2098-R01	264	234,146
2098-Q01	237	210,200
2098-P01	211	187,140
2098-O01	189	167,627
2098-N01	168	149,002
2098-M01	147	130,377
2098-L01	127	112,639
2098-K01	108	95,787
2098-J01	96	85,144
2098-I01	86	76,275
2098-H01	76	67,406
2098-G01	69	61,197
2098-F01	58	51,441
2098-E01	52	46,120
2098-D01	45	39,911
2098-C01	37	32,816
2098-B01	29	25,721
2098-A01	25	22,173

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785 U.S.A.

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of 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.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. These and other IBM trademarked terms are marked on their first occurrence in this information with the appropriate symbol (® or ™), indicating US registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at http://www.ibm.com/legal/copytrade.shtml

The following terms are trademarks of the International Business Machines Corporation in the United States, other countries, or both:
CICS®
DB2®
IBM®
IMS™
MVS™
Parallel Sysplex®
RACF®
RMF™
S/390®
System z10®
System z®
WebSphere®
z/OS®
z/VM®
z10™

The following terms are trademarks of other companies:
Java is a trademark or registered trademark of Oracle and/or its affiliates in the US and other countries.
UNIX is a registered trademark of The Open Group in the United States and other countries.
Other company, product, or service names may be trademarks or service marks of others.