z/OS Intelligent Resource Director (IRD)

Overview

The Intelligent Resource Director (IRD), a new feature in z/OS V1R1, extends the concept of goal-oriented resource management by allowing you to group system images that are resident on the same physical server running in LPAR mode, and in the same sysplex, into an "LPAR cluster." This gives Workload Management the ability to manage resources, both processor and DASD I/O, not just in one single image but across the entire cluster of system images.

The three functions that make up IRD are as follows:

- **LPAR CPU Management**, which lets Workload Manager distribute processor resource across an LPAR cluster by dynamically adjusting the LPAR weights in response to changesin the workload requirements. When important work is not meeting its goals, WLM will raise the weight of the partition where that work is running, thereby giving it more processing power. As part of LPAR CPU Management, WLM will also optimize the number of online logical CPUs configured online to each partition. As the LPAR weights change, the number of online logical CPUs are also changed to maintain the closest match between logical CPU speed and physical CPU speed.
  LPAR CPU Management runs on a zSeries server in z/Architecture mode, and in LPAR mode only. The participating z/OS system images must be running in goal mode. It also requires a CF level 9 coupling facility structure.
  Enabling LPAR CPU Management involves defining the coupling facility structure and then performing several operations on the hardware management console--defining logical CPs, and setting initial, minimum, and maximum processing weights for each logical partition.
- **Dynamic Channel Path Management**, which lets Workload Manager dynamically move channel paths through the ESCON Director from one I/O control unit to another, in response to changes in the workload requirements. By defining a number of channel paths as "managed," they become eligible for this dynamic assignment. By moving more bandwidth to the important work that needs it, your DASD I/O resources are used much more efficiently. This may decrease the number of channel paths you need in the first place, and could improve availability--in the event of a hardware failure, another channel could be dynamically moved over to handle the work requests.
  Dynamic Channel Path Management runs on a zSeries server in z/Architecture mode, in both basic and LPAR mode. The participating z/OS system images can be defined as XCFLOCAL, MONOPLEX, or MULTISYSTEM. If a system image running Dynamic Channel Path Management in LPAR mode is defined as being part of a multisystem sysplex, it also requires a CF level 9 coupling facility structure, even if it is the only image currently running on the CPC.
  Dynamic Channel Path Management operates in two modes: balance mode and goal mode. In balance mode, Dynamic Channel Path Management will attempt to equalize performance across all of the managed control units. In goal mode, which is available only when WLM is operating in goal mode on systems in an LPAR cluster, WLM will still attempt to equalize performance, as in balance mode. In addition, when work is failing to meet its performance goals due to I/O delays, WLM will take additional steps to manage the channel bandwidth accordingly, so that important work meets its goals.

Enabling Dynamic Channel Path Management involves defining managed channels and control units via HCD. On the hardware management console, you then need to ensure that all of the appropriate logical partitions are authorized to control the I/O configuration.

- **Channel Subsystem Priority Queuing** is an extension of I/O priority queuing, a concept that has been evolving in MVS over the past few years. If important work is missing its goals due to I/O contention on channels shared with other work, it will be given a higher channel subsystem I/O priority than the less important work. This function goes hand in hand with the Dynamic Channel Path Management described above--as additional channel paths are moved to control units to help an important workload meet goals, Channel Subsystem Priority Queuing ensures that the important workload receives the additional bandwidth before less important workloads that happen to be using the same channel.
  Channel Subsystem Priority Queuing runs on a zSeries server in z/Architecture mode, in both basic and LPAR mode. The participating z/OS system images can be defined as XCFLOCAL, MONOPLEX, or MULTISYSTEM. It is optimized when WLM is running in goal mode. It does not require a coupling facility structure.
  Enabling Channel Subsystem Priority Queuing involves defining a range of I/O priorities for each logical partition on the hardware management console, and then turning on the "Global input/output (I/O) priority queuing" switch. (You also need to specify "YES" for WLM's I/O priority management setting.)

Special considerations and restrictions

General

- **Minimum Required Microcode Levels**
  To run IRD, you need the following minimum microcode levels installed:
  - 2064 zSeries Processor: Driver 36J, Bundle 9
  - 2105 Enterprise Storage Server: SC01208 (12/2000 Level)
  - ESCON Directors:
    - 9032-2: Version 4, Release 1
    - 9032-3: LIC 04.03.00
    - 9032-4: LIC 05.04.00

Dynamic Channel Path Management

- **Supported Devices**
  At the present time, the following IBM devices are supported for Dynamic Channel Path Management when connected through an ESCON or FICON Bridge channel:
  - IBM 9393 RAMAC Virtual Array
  - IBM 2105 Enterprise Storage Server

  For non-IBM devices, please contact the device manufacturer to determine if the device is supported by Dynamic Channel Path Management.
- **Unique LPAR Cluster Names**
  LPAR clusters, running on a 2064 CPC, must be uniquely named. This is the sysplex name that is associated with the LPAR cluster. Managed channels have an affinity (are owned by) a specific LPAR cluster. Non-unique naming creates problems in terms of

scope of control.

- **Disabling Dynamic Channel Path Management**
  To disable Dynamic Channel Path Management within an LPAR cluster, turning off the function by using the SETIOS DCM=OFF command is not sufficient. Although a necessary step, this does not ensure that the existing configuration is adequate to handle your workload needs, since it leaves the configuration in the state it was at the time the function was disabled. In order to ensure an adequate configuration, you must ACTIVATE to an I/O configuration that meets your workload needs across the LPAR cluster.
- **Automatic I/O Interface Reset**
  When going through all of the steps to enable Dynamic Channel Path Management, also ensure that the "Automatic input/output (I/O) interface reset" option is enabled on the hardware management console. This will allow Dynamic Channel Path Management to continue functioning in the event that one participating system image fails. This is done by enabling the option in the reset profile used to activate the CPC. Using the "Customize/Delete Activation Profiles task" available from the "Operational Customization tasks list," open the appropriate reset profile and then open the Options page to enable the option.
- **SafOS**
  When using SAfOS, care must be taken when using PROHIBIT or BLOCK on a port that is participating in Dynamic Channel Path Management.
  When blocking a managed channel port, configuring the CHPID OFFLINE is all that is required. Dynamic Channel Path Management will ensure that if the CHPID is configured to managed subsystems, then the CHPID will be deconfigured from all subsystems to which it is currently configured. When blocking a port connected to a managed subsystem, the port must first be disabled for Dynamic Channel Path Management usage. This is done using the VARY SWITCH command to take the port OFFLINE to Dynamic Channel Path Management. Disabling the port for Dynamic Channel Path Management usage will deconfigure all managed channels which are connected to the subsystem through that port. Once the port is disabled to Dynamic Channel Path Management, it can be then be blocked.
  When prohibiting a set of ports, if any of the ports are connected to managed subsystems, then the PROHIBIT operation must be preceded by the VARY SWITCH command(s) to disable the managed subsystem ports to Dynamic Channel Path Management. As in the blocking case, this will cause any managed channels currently connected to the subsystem port(s) to be deconfigured. Once the subsystem ports are disabled to Dynamic Channel Path Management, the PROHIBIT function can be invoked. This must then be followed by the VARY SWITCH command(s) to re-enable the prohibited subsystem ports to Dynamic Channel Path Management.
  When ports are unprohibited or unblocked, these operations need to be followed, as necessary, by VARY SWITCH commands to bring ports ONLINE to Dynamic Channel Path Management.
- **VARY SWITCH**
  VARY SWITCH is not supported in the COMMNDxx SYS1.PARMLIB member.
- **SWITCH Statement in CONFIGxx Member**
  Care must be taken if the SWITCH statement is included in a CONFIGxx member and a CONFIG MEMBER(xx) command is issued. Since the SWITCH statement causes a VARY SWITCH command to be issued, which is LPAR Cluster in scope, any differences in the CONFIGxx member, in terms of the SWITCH statements among the system images

within the LP Cluster, will cause the the specified switches and ports to assume the Dynamic Channel Path Management state specified within the last CONFIGxx member used. This may set/reset the Dynamic Channel Path Management state incorrectly.

Problem Reporting

If you have reason to believe that IRD is not functioning properly, IBM will need CTRACEs, SMF 99 and 70-79 records, and an SVC dump for diagnosis. The following are IBM's recommendations for collecting this data:

- **SMF Records**
  Ensure that SMF record 99 (all subtypes) and SMF records 70-79 (all subtypes) are enabled within SMFPRMxx. This information is used for Intelligent Resource Director analysis and contains the required WLM and RMF information. This information should be collected by all systems within the LPAR cluster as well as for all systems not part of the LPAR cluster but within the same sysplex as the images within the LPAR cluster.
  When reporting to IBM, please provide a list of system names in the cluster that produced the data plus a short description of which major workloads were running at the time the data was collected.

- **CTRACE**
  Ensure that Component Trace (CTRACE) is enabled for IOS with the DCM option specified. This traces information beyond the default and is specific to Dynamic Channel Path Management. The information is used in conjunction with the SMF records listed above. CTRACE with the DCM option should be run in conjunction with an external writer. Use of the external writer ensures that important entries are not lost in the event the internal CTRACE buffers wrap.
  The following is a simple example of how to set up external CTRACE writers using preallocated datasets.

  ```
  Dataset: SYS1.PROCLIB(OPSWRTR)
   //OPSWRTR  PROC
   //IEFPROC   EXEC PGM=ITTTRCWR,REGION=256K,TIME=1440
   //TRCOUT01  DD DSNAME=TEST1.OPSWRTR0,DISP=SHR

   //TRCOUT02  DD DSNAME=TEST1.OPSWRTR1,DISP=SHR
   //TRCOUT03  DD DSNAME=TEST1.OPSWRTR2,DISP=SHR
  ```

  Notes:
  - ○ If you use this sample code, tailor it to your system. We recommend that you have at least three datasets. To add more, just add more TRCOUTnn entries.
  - ○ Different systems in a sysplex cannot write to the same dataset.

  To start the external writer for IOS, issue:

  ```
       trace ct,wtrstart=opswrtr
  ```

  To start CTRACE, issue:

  ```
       trace ct,on,comp=sysios
       r nn,wtr=opswrtr,asid=(nn),jobname=(iosas),options=(DCM),end
  ```

  Alternatively, you can set up a SYS1.PARMLIB member with the trace options.
  To stop the trace and get the remaining in storage trace records to the trace datasets,

issue:

```
 trace ct,off,comp=sysios
```

To stop the writer, issue:

```
     trace ct,wtrstop=opswrtr
```

In order to prevent paging auxiliary storage shortage, ensure that sufficient auxiliary storage has been allocated. The additional CTRACE entries generated by IOS for Dynamic Channel Path Management may require up to 500 megabytes of auxiliary storage.

Please refer to z/OS V1R1.0 MVS Diagnosis: Tools and Service Aids for details on how CTRACE options can be specified to IOS, and how an external writer for CTRACE can be established.

- **SVC Dumps**
  Please also collect and archive SVC dumps on all members of the cluster at the end of a test. Use the following parameters when generating an SVC dump for IRD:

  ```
  Dataset: SYS1.PARMLIB(IEADMCxx)
   SDATA=(PSA,NUC,SQA,LSQA,RGN,LPA,TRT,CSA,SWA,SUM,
   GRSQ,COUPLE,XESDATA,WLM,SERVERS),
   JOBNAME=(IOSAS,*MASTER*,WLM),
   DSPNAME=('IOS*'.*,'*MASTER*'.*),
   END
  ```

  When you take the dump, issue the dump command with the PARMLIB parameter:

  ```
  DUMP COMM=(title),PARMLIB=xx
  ```

  Do not send these dumps to IBM unless requested to do so.

Post-GA Support Program

In order to assure customer satisfaction with IRD, IBM has created a Post-GA Support Program. The purpose of this program is to identify customers that are considering exploiting IRD functions and proactively monitor their progress.

To learn more about the program, please send a note to IRD@us.ibm.com or call (845) 435-8844 in the US. We will contact you to help review your configuration plans.

More information

For more detailed information on the Intelligent Resource Director, see z/OS V1R11 MVS Planning: Workload Management, and the IBM Redbook z/OS Intelligent Resource Director