zSeries and Parallel Sysplex Performance: Sharing Resources on Uni-Processors and Migrating from ICMF

Introduction:

Any customer installing a new zSeries processor needs to include planning for the migration of their parallel sysplex environments. Two unique environments require increased planning focus.

- Configuring a Coupling Facility (CF) partition on a uni-processor which will share CP resources with z/OS or other operating system partitions. This includes environments which are running on sub-uni processors.
- Migrating from a Coupling Facility environment using ICMF (Integrated Coupling Migration Facility) to a Coupling Facility environment using internal or external links. This applies to uni-processors as well as n-way processors.

Each of these environments will be discussed in detail within this flash. The best source for additional product information on these topics will be found in:

- *The PR/SM Planning Guide*, SB10-7033
- *z/OS MVS Setting Up A Sysplex*, SA22-7625

WSC Flashes can also be read to get a broader understanding of parallel sysplex performance. For this information review WSC Flashes:

- WSC Flash W9609: LPAR Performance Considerations for Sharing Resources in a Parallel Sysplex Environment (republished in May 2002)
- WSC Flash W99037: Performance Impact of Using Shared ICFs
- WSC Flash 10159: New Heuristic Algorithm for Managing CF Request Conversion
- WSC Flash W9731A: Dynamic CF Dispatching
- WSC Flash W9828: Dynamic ICF Expansion
- WSC Flash W9846: Dynamic CF Dispatch Default set to Enabled

Configuring a Coupling Facility on a Uni-Processor

Production parallel sysplex Coupling Facility partitions should have access to sufficient CP resources. (See WSC FLASH 9609). Sharing physical processors between CF partition(s) and SCP partition(s) can be done but there will be costs in terms of responsiveness and throughput. When sharing physical CP resources it is critical the CF logical CP receive adequate physical CP resource to support the workloads. For even lightly used CF environments it is strongly recommended to ensure the physical CP resource allocated to the CF logical CP be equal to at least 50% of a physical CP.

The reason for the guideline of allocating at least 50% of a physical CP to the CF logical CP is based on the requirement for the CF to be very responsive. If a CF partition does not have sufficient access to physical CP resources requests will see increased service times, and certain

CP intensive CF operations (like recovery operations) may be impacted. So a CF partition may not need a lot of physical CP resource, but when it needs it, it needs it right away. Hence, regardless of the MIP capacity of the individual physical CP, the requirement holds true, a production CF partition should have access to at least 50% of a physical CP.

The requirements to provide sufficient resources to a production CF in a uni-processor environment obviously becomes more challenging. Customers who are running on a zSeries uni-processor are strongly encouraged to use an Internal Coupling Facility (ICF) CP to provide the needed CP capacity for the CF. ICF CPs are special purpose CPs which only run Coupling Facility Control Code and so do not contribute to the software MSU capacity metric.

If the installation only has only a uni-processor and has a requirement for a CF, but elects to not install an ICF CP to provide the physical CP resource for the CF, then the only alternative is to share the general purpose CP with the CF(s). Such a configuration carries with it significant risks of degraded performance, loss of connectivity and possible coupling link check-stops and should only be used in a production parallel sysplex with care and knowledge of the risks.

Installations which attempt to use a uni-processor and then share the physical CP between CF partition(s) and SCP partitions do so at their own risk. These risks can be mitigated somewhat by adhering to all of the following recommendations:

1. Provide the Coupling Facility partitions with an LPAR weight as close to 50% of the general purpose CP as possible. For example, if 2 z/OS LPARs and 2 CF LPARs are activated on a uni-processor, a suitable weighting would be 5 for each z/OS LPAR and 45 for each CF LPAR. This would provide each CF partition 45% of the physical CP.

Below is an example of a Partition Data Report which would provide the necessary weighting:

```
    ---- PARTITION DATA ----  ---
                              # OF
   NAME     STAT  WGT  CAP    LPs
PRODA         A     5   NO      1
PRODB         A     5   NO      1
CF1PART       A    45   NO      1
CF2PART       A    45   NO      1
```

2. Dynamic CF Dispatching must be enabled. Dynamic CF Dispatching is enabled by default for CF LPARs using shared general purpose Cps.

3. Do not cap the Coupling Facility partitions.

The adherence to these recommendations will provide the best environment for success running a parallel sysplex on a uni-processor. Dynamic CF Dispatching reduces the impact of the active wait polling loop used by the coupling facility control code (CFCC). By reducing the impact of the active wait polling loop the CF partitions will not use their entire LPAR share. The

remaining capacity will then be available to the z/OS partitions. The large relative weight given to the Coupling Facility will mean when the CF is experiencing periods of increased demand, as for example during a system reset, the CF partition will have access to sufficient physical CP resource, allowing it to handle the increased request rate. Installations need to be aware during such periods the CF will be allowed to take capacity which may have previously been used by the z/OS partitions.

In a configuration where a z/OS or OS/390 partition shares the physical CP resource with the CF partition, SYNC requests are converted, unbeknown to MVS, to asynchronous requests. The requests are still reported as SYNC requests in RMF, but the SYNC requester is not accumulating task busy time. If a special purpose ICF CP is configured into this environment the SYNC requests would not be converted and the SYNC requester would incur task busy time for the duration of the SYNC request.

So the running of a parallel sysplex on a shared uni-processor without a special purpose ICF CP is a compromise on behalf of the customer. It is critical customers evaluate their capacity demands for their z/OS partitions in light of their expected use of the CF partition. **Please be advised, if enough processing power is not given to a Coupling Facility, the likelihood of exposure to loss of connectivity and/or channel check-stops is greatly magnified.**

Internal IC Links on zSeries

Internal CF links (ICs) use CP resources to perform the link operations. Though the amount of resource used to do this is very slight it is recommended for the installation to manage the number of IC links which are defined on a uni-processor. Since IC links are created using only IOCP definitions over-definition may appear to be "free". This is not the case. In a uni-processor environment where CF's and SCP partitions are sharing the CP resource it is recommended installations define no more than one IC link. The IC link capacity with this definition will be more than sufficient for most parallel sysplex configurations in a uni-processor environment.

The recommendations on limiting the number of IC links is made to conserve capacity, especially in periods of high CPU busy. Increased IC link definitions would have the potential to reduce overall SCP/CF CP capacity, and may result in channel time-outs and some performance degradation.

Migrating from ICMF Environment

The Integrated Coupling Migration Facility (ICMF) was provided as a low-cost alternative to using real coupling links for customers wanting to establish a parallel sysplex migration test facility on a single CEC. Communication between OS/390 partitions using ICMF and CF partitions using ICMF was provided by LPAR hipervisor emulation of the coupling connection. Some installations have found an ICMF configuration sufficient to meet all of their production CF requirements, not just their requirements for test. The ICMF facility is not available on zSeries processors. As installations migrate to zSeries processors additional planning will be needed to adequately replace the ICMF function.

The ICMF facility provided function which must be replaced in a migration to zSeries:
1. Allowed for the emulation of CF links.
2. Allowed the dispatching of the CF partition only when work was present, rather than using the active wait polling loop to detect when work was present. The LPAR hipervisor would monitor the emulated links and would dispatch the CF partition only when work was present.
3. SYNC requests made to the CF partition were converted to asynchronous requests. This meant senders tasks were not being charged CPU time for the operation.

The ICMF facility allowed lower cost through reduced hardware requirements (no real links, no special purpose ICF CP) but this was traded off with slower responsiveness to CF requests than was possible using dedicated CF CP resources. As installations migrate to new zSeries processors additional planning will be needed to adequately provide for migration test facilities or to replace current production CF configurations which used the ICMF facility. Both of these items will be addressed below.

CF Link Emulation

The emulation of coupling links by the ICMF Facility has been replaced by the use of Internal Coupling (IC) links. IC links are implemented using HCD. Complete information on how to configure for IC links can be found in the *PR/SM Planning Guide*, *IOCP Users Guide*, and the *HCD Users Guide*. Below is an example using one internal link (2 IC CHPIDs) for connectivity between two (2) z/OS partitions and two (2) coupling facility partitions. Internal links on the z800 and z900 processors are defined as TYPE=ICP (peer mode). A peer-mode coupling CHPID can be shared by multiple z/OS or OS/390 partitions and one (and only one) coupling facility partition. Thus each end of the link would be defined as being shared by the two z/OS partitions and one of the coupling facility partitions. See the sample IOCP statements below:

```
RESOURCE PARTITION=((CF1,3),(CF2,4),(ZOS1,1),(ZOS2,2))
CHPID PATH=(FE),SHARED,                                          *
      PARTITION=((CF1,ZOS1,ZOS2),(CF1,ZOS1,ZOS2)),CPATH=FF,      *
      TYPE=ICP
CHPID PATH=(FF),SHARED,                                          *
      PARTITION=((CF2,ZOS1,ZOS2),(CF2,ZOS1,ZOS2)),CPATH=FE,      *
      TYPE=ICP
CNTLUNIT CUNUMBR=FFFD,PATH=(FF),UNIT=CFP
CNTLUNIT CUNUMBR=FFFE,PATH=(FE),UNIT=CFP
IODEVICE ADDRESS=(FFF2,007),CUNUMBR=(FFFD),UNIT=CFP
IODEVICE ADDRESS=(FFF9,007),CUNUMBR=(FFFE),UNIT=CFP
```

Review any existing CFS, CBS, CFR, or CBR coupling channel definitions that may have been defined for those partitions (SCP and CF) that were previously using the ICMF facility. These channels, though defined, were forced offline during the activation process for partitions using the ICMF facility. With the migration to a CF not using the ICMF facility, these links may now be used for external connectivity to other CECs. If this type of external connectivity is not intended then configure the links offline if they are not needed/desired.

CF Dispatching

The capability to reduce the impact of the CFCC active wait polling loop may be provided through the use of Dynamic CF Dispatching. DyCP resource consumption by a CF using the ICMF facility and a CF using Dynamic CF Dispatching are not equivalent. The major difference being with the ICMF facility, the LPAR hipervisor was aware of requests being sent over the emulated links, and as a result would dispatch the CF partition. This method of dispatching gave fairly good service, (i.e. responsiveness) to the CF requests with minimal CP resource expended.

With dynamic CF dispatching when the CF is not used or lightly used the coupling facility control code will suspend activity for short periods, allowing the other partitions to use the CP resource. As the traffic increases, the CFCC code is run more often and more of the CP resource will be consumed by the CF partition. This allows the installation to define a CF partition which consumes minimal CP resource until the time when it is needed as was true with a CF partition using the ICMF facility. But with dynamic CF dispatching there is no notification when a CF request is placed on the link and so a request will wait until the CF partition is dispatched. When activity rates are low the CF is suspended more often to save CP resource and so CF requests will experience longer service time as they wait for the suspended CF to come active to handle the request.

The more often LPAR dispatches the CF and the CF has work present (load) the CFCC will dynamically increase the CF activity interval. This means the CF partition will be dispatched more often, and correspondingly the amount of time it takes for the CF partition to recognize it has an outstanding request will go down. As the CF becomes more active it will use more CP capacity. So a request rate to a CF which is sporadic in nature will tend to see CF request service times which are much larger and have a much higher standard deviation than the same request rate to a CF partition using the ICMF facility. As activity grows the service times will improve but the CPU capacity required will also increase. This ability to suspend activity to save resource is why dynamic CF dispatching was designed to support test and hot standby configurations and not production environments.

So dynamic CF dispatching may provide the same test capability as was true with ICMF facility. But for installations which used the ICMF facility to provide a low cost production CF, it is important to note it is **not** recommended to use dynamic CF dispatching in a production CF.

In order for a CF to provide reasonable request service times the CF partition must have access to at least 50% of a physical CP. With access to 50% of a physical CP it is expected the CF request service times may begin to approach the service times of a CF which was using the ICMF facility. In order to get to the previous service times of a CF using the ICMF facility, dynamic CF dispatching cannot be used. The best method of providing this CF capacity is through the use of a special purpose ICF CP.

SYNC Requests Converted to ASYNC

In a CF partition using the ICMF facility, or a CF partition sharing general purpose CPs, the impact on SYNC service times should be understood. In order to avoid deadlock situations if

either of the above two configurations are present the LPAR hipervisor, unknown to MVS, will change the form of a SYNC request to an ASYNC request. Requests will be reported in RMF as SYNC requests but they are actually issued asynchronously. This means SYNC requests in either of these two configuration will appear to have higher service times and larger standard deviations than SYNC requests which are made to CF partitions which have dedicated resources (ICF special purpose CPs or dedicated general purpose CPs). For more information on this please see WSC FLASH W99037 and W9731A.

As customers migrate to the new environment from their previous environment using the ICMF facility the choice of the new platform will become important. If the customer chooses to provide a CF with special purpose ICF CPs or with dedicated general purpose CPs, they need to be aware the SYNC requests issued in such a configuration will be true SYNC requests and sender CP time will be charged to the task issuing the CF SYNC request.

Summary

Installations migrating from an environment using the ICMF facility will need do increased planning to best configure a CF environment which meets the capacity and performance requirements of their parallel sysplex environment. It is important to plan the capacity requirements for the CF environment for both normal operations and for recovery or restart periods when CF request rates may be much higher. It is especially important installations understand how configuration choices which result in the sharing of CF CP resources can impact performance, throughput and availability. Installations which need to provide a test CF environment may wish to explore the use of dynamic CF dispatching, using either special purpose ICF CPs or general purpose CPs. Installations which need to provide a production CF environment are strongly urged to use special purpose ICF CPs.

Special Notices