CheatSheet #62 zTidBits

z/OS' Capacity Provisioning

*z/OS MVS Capacity provisioning allows customers to set up rules defining the circumstances under which additional capacity should be provisioned in order to fulfill a specific business need. The rules are based on 3 criteria, *such as:*
a specific application, the maximum additional capacity that should be activated, time and workload conditions.
- This support provides a fast response to capacity changes and ensures sufficient processing power will be available with the least possible delay *even if* workloads fluctuate.

* z/OS Capacity Provisioning helps you manage the CP, zAAP and zIIP capacity of System z10 servers that are running one or more z/OS systems.
- Based on **On/Off CoD**, temporary capacity may be **activated** and **deactivated** with a policy you define.
- z/OS Capacity Provisioning simplifies the monitoring of critical workloads, and its automation features can help to activate additional resources much faster than operator manual operation.
NOTE: To accommodate changes of the physical capacity, the configuration of logical processors may need to be changed. You can allow Capacity Provisioning to configure logical processors online or offline whenever such resources would block a change to the hardware capacity.

* When you use Capacity Provisioning you can select different levels of automation to provide you with an appropriate level of control such as to:
- Activate and deactivate temporary capacity through operator commands
- Activate and deactivate temporary capacity based on a defined schedule, without considering workload performanc
- Have the Provisioning Manager suggest changes to the capacity of the System z10 based on the observation of workloads that you define
- Have the Provisioning Manager automatically implement changes to the capacity of the System z10 server based on the observation of workloads that you define.

* z/OS Capacity Provisioning is delivered as part of the z/OS MVS Base Control Program (BCP) component.
Capacity Provisioning includes the following:
- Capacity Provisioning Manager (Provisioning Manager) - the server program
- Capacity Provisioning Control Center (**Control Center**) - the workstation code
- Sample data sets and files: Sample jobs for setting up and customizing the Capacity Provisioning component are placed in SYS1.SAMPLIB.

* Capacity Provisioning configuration entities, such as policies and domain configurations, are defined using the Control Center on a workstation.
- Observation of workloads and the interaction with the servers defined to Capacity Provisioning is performed by the Provisioning Manager on the z/OS host.

TERMS Field of operation The scope of a z/OS Capacity Provisioning system is referred to as a Capacity Provisioning Domain, or simply a *domain*. The domain configuration describes the scope of management within a provisioning domain. The domain includes hardware and software elements. The hardware elements are one or more Central Processor Complexes (CPCs) where temporary capacity can be activated or deactivated by Capacity Provisioning. The software elements are z/OS operating systems which can run on one or more of these CPCs and which are monitored by Capacity Provisioning to determine the hardware requirements.

Rules of operation Provisioning Management is controlled by a Capacity Provisioning Policy, or simply a policy. This defines the actions to be performed on the hardware and software elements in response to the demands of the observed software elements. A policy contains rules, which define workload conditions that will trigger intervention, the resources which can be activated, and the time periods during which the rule can be applied.

NOTE: The domain is controlled by the Provisioning Manager. This runs in a z/OS system and controls the domain in real time. It observes the software elements and monitors workload demands. It can recommend hardware configuration changes to the system operator, or can be empowered to activate or deactivate hardware and software elements itself to satisfy these demands.

* The Provisioning Manager obtains z/OS Workload Manager performance metrics from z/OS' RMF.
- If the performance and capacity data indicate that a workload is *missing its goal*, and a lack of processor resources is identified as an important reason for this, the Provisioning Manager can either recommend appropriate actions, or automatically activate and deactivate additional temporary capacity if it has been suitably configured.

* Using policies for the Provisioning Manager, you can define workloads that are to be observed and also conditions that specify under what circumstances additional resources may be warranted.

* The policy describes the scope of management within a provisioning domain.

* The policy controls the provisioning of additional capacity and different policies can be created for different circumstances, but ***only one*** of these policies can be used by the Provisioning Manager at any point in time. * **The policy defines:**
- How much additional capacity may be activated
- When this additional capacity may be activated
- What triggers the activation of additional capacity

Configuring additional capacity via On/Off CoD can result in additional IBM hardware and software license charges.

* Each policy has a maximum provisioning scope which defines the total amount of resources that may be activated. **This includes:**
- Maximum amount of general purpose capacity, in MSUs
- Maximum number of Application Assist Processors (zAAPs)
- Maximum number of Integrated Information Processors (zIIPs)

Manageable objects are represented in a CIM server by instances of object oriented classes.

* General purpose capacity can be provided either by additional general purpose processors (CPs) or processor capabilities.

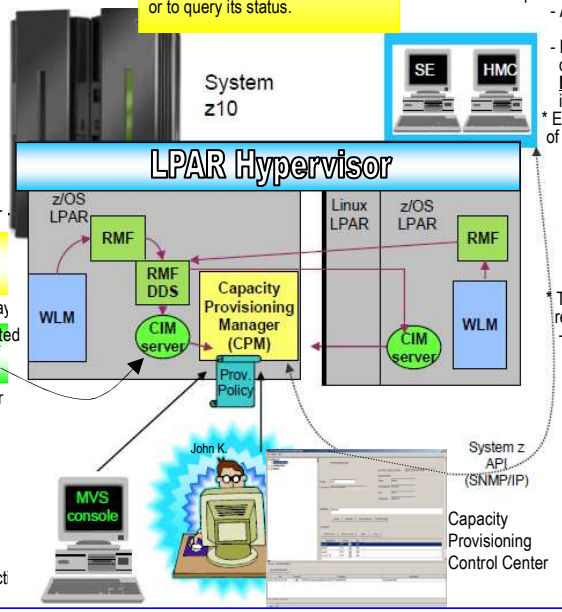
* In the policy you can optionally specify a logical processor scope, which defines the z/OS systems where the number of logical processors can be changed.
- For each system you can either specify the maximum number of processors that may be online, or specify that the limit of the LPAR definition applies.
- If you omit the logical processor definition, the number of processors for an observed system is not managed by Capacity Provisioning.

* A policy contains one or more provisioning rules.
- These rules define limits (*provisioning scope*) to the capacity that can be activated, time periods (time conditions) when activation is possible, and triggers (workload conditions) that can cause act
See bottom right side of this slide.

Contained in z/OS base component free of charge based on z10 On/Off CoD feature running z/OS R9 & above

The **Provisioning Manager** monitors the workload on a set of z/OS systems and organizes the allocation of additional capacity to these Systems when required. The systems to be observed are defined in a domain configuration file. Details of additional capacity and the rules for its allocation are held in a policy file. These two files are created and maintained using the Control Center.

The **Control Center**, installed on a workstation, is the graphical user interface to Capacity Provisioning . Through this interface administrators work with provisioning policies and domain configurations, and can transfer these to the Provisioning Manager. You can set up a direct connection from the Control Center to the Provisioning Manager, and use this to transfer provisioning policies and domain configurations files to the Provisioning Manager, or to query its status.

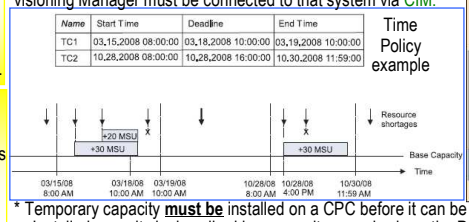


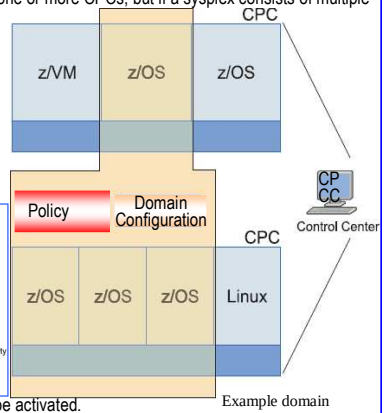
* Rules contain provisioning conditions which describe the situations in which the Provisioning Manager can activate temporary resources under the rule.
-These situations can include time conditions indicating periods in which provisioning is allowed, and workload conditions indicating demand that can trigger activation.
* Workload conditions are expressed in terms of the z/OS WLM service class model.
* Additional capacity may only be activated by z/OS Capacity Provisioning when business critical work is suffering.
- This work should be identified at the planning stage and must be specified in the workload conditions of a policy.
* The Provisioning Manager controls a domain monitoring the observed systems, and can activate or propose manual activation of temporary capacity, based on the settings in your active domain configuration and policy.
* The specifications of the CPCs to be managed and the systems to be observed are held in a domain configuration.
* The Provisioning Manager must be able to access these CPCs from every host system it can run on. Information from available CPCs is obtained through a connection to the hardware console.
- This console can be a service element (SE) or a hardware management console (HMC).

NOTE: HMC can be replaced throughout by SE if your system uses this.

* This illustration shows an example domain configuration with two CPCs to be managed and four systems to be observe
- This displays that Capacity Provisioning can be performed for one or more CPCs, but if a sysplex consists of multiple systems then all observed systems of this sysplex must belong to the same Capacity Provisioning domain.
- It is **not** supported that one CPC belongs to multiple domains.
- Each system in the domain configuration is identified by its z/OS system name and by the name of the sysplex it belongs to.
- Systems keep the same name even if they move across LPARs and CPCs.

* To get information about the workload running on a system or to manage the number of processors of that system the Provisioning Manager must be connected to that system via CIM.





Example domain

* Temporary capacity **must be** installed on a CPC before it can be activated.
- Installed capacity is described in a capacity record where the Provisioning Manager can only activate the residual capacity in this record within limits which were defined during the order process of the record.
- It is possible that some CPCs in a domain may not have temporary capacity.
NOTE: In this case the Provisioning Manager can still report resource shortages on these CPCs.

* The provisioning domain contains a set of logical partitions (LPARs).
- These can be parts of a stand-alone system (a monoplex) or can be parts of a sysplex.
- A z/OS system runs within each LPAR so when you define the domain configuration you specify the set of z/OS systems to be observed.
- Each z/OS system is identified by name, and if it is running in a sysplex this name is further qualified by the name of the sysplex.
NOTE: A system can only be observed and considered for capacity changes if it runs on a CPC in the provisioning domain.

* Each CPC in the domain configuration is identified by its logical name which defines it on the SE of that processor complex (see Time Policy example above).
- For each CPC there is an enabled attribute which specifies whether the CPC is to be considered for temporary capacity changes.
- If it is ***enabled*** the Provisioning Manager is allowed to perform changes to the temporary capacity of that CPC.
- If it is ***disabled*** only manual capacity changes using Provisioning Manager commands are allow
NOTE: You can switch the enabled attribute on or off at runtime, in the same way as for systems by using the Provisioning Manager commands ENABLE CONFIGURATION and DISABLE CONFIGURATION

* The temporary capacity that can be activated on a CPC can be described by multiple capacity records, identified by unique record IDs.
- Only ***one*** of these records can be in use by the Provisioning Manager ***at any one time***.

Defines the systems where Capacity Provisioning should Manage the number of logical processors and which processor limits apply for those systems.

Limits the amount of additional resources that may be activated on behalf of the contained rules.

Two types of conditions are supported (RULES):
1. *Time condition* specifies time periods during which additional capacity can be activated.
2. *Workload condition* identifies work which is eligible to cause an activation, and the conditions under which activation can be triggered. Eligible work is specified according to the workload model of the z/OS Workload Manager (WLM).