



Utility Capacity on Demand: What Utility CoD Is and How to Use It

Executive overview

This IBM Redpaper document is about Utility Capacity on Demand (CoD). It provides a high-level overview to assist IT managers and business analysts who want to:

- ▶ Understand what Utility CoD is
- ▶ Understand how it works
- ▶ See how it could be valuable for their company

Utility CoD: What's in a name?

Utility CoD provides automated control over the amount of capacity used for variable business demands. It is a mechanism by which you can have extra processor capacity available within your system to meet unusually heavy workloads, but only have to activate — and pay for — the additional capacity on an as-needed basis.

As the name suggests, this behavior is similar to that of a public utility such as electricity or water. You just turn it on when you need it and turn it off when you are done, and you only pay for what you actually used.

Utility CoD: What can I do with it?

Let's say that your normal need for processing power is 100 units, and that this will nicely cover your routine daily, weekly, and monthly tasks. However, during the holiday shopping season, for instance, and at peak times such as between 11:30 AM and 1 PM, you need at least 50 units of extra processing power just to handle the online shopping workload.

If your processor maximum utilization rate is 90% during normal hours, meaning you use 90 units of processing power for your typical workload, this leaves only 10 units of available processing power, which is far short of the 50 units needed to satisfy demand during peak shopping periods.

You could build your system big enough to handle the maximum workload, but much of that processor capacity would be idle most of the time. Having all that unused capacity just sitting there would be a very expensive way to handle the situation.

You could ignore the extra demand, and suffer the consequences of poor system performance, unhappy customers, and possibly lost business, making this a potentially expensive proposition as well.

Utility CoD: A better solution

Utility CoD provides a better solution by allowing you to keep extra processing capacity in place, but only activating and paying for it on an as-needed basis. The extra capacity is in the form of *inactive processor units* located in a *Shared Processor Pool*. As workload fluctuates, these units can be put into service as required to handle the peak processing demands, then relegated to the Shared Processor Pool when the workload returns to its normal level.

Utility CoD allows you to assign inactive processors installed on the system to the Shared Processor Pool. The assigned processors will be treated as *utility processors* that are available for use by *uncapped partitions*.

Utility CoD allows use of the additional capacity on a minute by minute basis, and the service delivers a processor minute of capacity when and where you need it. And, just like electricity is accounted for as it is used, processor minutes are accounted for at the time of actual use. This similarity to other public services explains the name “Utility CoD.” In a nutshell, Utility CoD automatically delivers additional processor capacity on a temporary basis within the system’s default Shared Processor Pool. Usage of reserved processing capacity is automatically done by system, only when you need it.

Quick facts about Utility CoD

- ▶ Utility CoD is currently available for POWER6™ 570.
- ▶ Processor usage is automatically charged based on measured workload.
- ▶ Usage is measured in processor minute increments and is reported via a web interface.
- ▶ Billing is based on the reported usage.
- ▶ There are no paper contracts; registration is exclusively web-based.
- ▶ Blocks of time can be paid for post use or they can be pre-paid.
- ▶ Utility CoD optimizes processor use within the shared pool.
- ▶ This solution is especially well suited to customers with unpredictable, brief workload spikes, and who need an automated and affordable way to help assure adequate server performance is available as needed.
- ▶ Compared to an On/Off Processor day, the break even point is about 36 minutes of use in one day.
- ▶ Memory resources are not part of Utility CoD - adequate memory resources should be made available to achieve best processor utilization.

Quick comparison with other IBM CoD offerings

Utility CoD is one of the latest offerings from IBM to facilitate additional capacity on demand. Capacity on Demand (CoD) allows you to dynamically activate one or more resources on your server as your business needs dictate. You can activate inactive processors or memory units that are already installed on your server on a temporary or permanent basis.

Select IBM servers include a number of active and inactive resources. Active processors and active memory units are resources that are available for use on your server when it arrives

from the manufacturer. Inactive processors and inactive memory units are resources that are included with your server, but that are not available for use until you activate them. You activate them when you need to use them, thus increasing your system's capacity on demand.

IBM has a number of offerings to facilitate CoD; they can be categorized based on whether they are activated on a temporary or permanent basis.

Permanent capacity for non-disruptive growth

These offerings let you activate inactive resources *permanently*. The result is the same as upgrading your system with larger capacity, but without any of the disruptions that can come with traditional upgrade methods. This category includes the following offering:

- ▶ **Capacity Upgrade on Demand (CUoD)**, which provides a means of planned growth for clients who know they will need increased capacity but are not sure when

Temporary capacity for fluctuating workloads

The offerings described here enable you to activate resources on a *temporary* basis. You use the resources for a set time period, paying the fee for that period, then return the resources back to inactive state. Offerings in this category include the following:

- ▶ **On/Off CoD** provides for planned and unplanned short-term growth driven by temporary processing requirements such as seasonal activity, period-end requirements, or special promotions. Activation is user managed.
- ▶ **Reserve CoD** provides automated use of processors in a shared processor pool for planned and unplanned short-term workload increases.
- ▶ **Utility CoD** is an improved version of Reserve CoD that is available on IBM POWER6 processor-based systems, including the IBM System p™ 570 and the IBM System i™ 570. It enables use of processors in a shared processor pool for planned and unplanned short-term workload increases.

Utility CoD provides the highest level of flexibility for both planned and unplanned short-term workloads. It is especially useful for customers with unpredictable, short duration workload spikes.

Utility CoD: Improved version of Reserve CoD

For those readers who are current users of Reserve CoD, or are familiar with it, this section identifies the differences between that offering and Utility CoD. Utility CoD replaces Reserve CoD on selected POWER6 processor based servers, and it offers more flexibility and granularity than its predecessor. Major points of difference are that for Utility CoD:

- ▶ Capacity is measured on a per processor minute basis, rather than on a processor day basis.
- ▶ Capacity can be paid for either before or after usage.
- ▶ Resource usage reporting is required.

Much like Reserve Capacity (which is still offered on POWER5™ servers with the HMC), Utility CoD only works with uncapped partitions, and also requires the use of an HMC for both enablement and for managing the number of Utility CoD processors that are available for use in the shared processor pool monitoring utilization.

There are no contract requirements for Utility CoD, but there is a pop-up window where the customer has to agree to the reporting requirements associated with Utility CoD. Clients must report the use of utility processor minutes and pay for that use on a timely basis or the processor utility service will be disabled.

The shared processor pool will contain both non-reserve and Utility CoD capacity. The Utility CoD capacity will only be utilized if the non-reserve capacity is fully utilized from the shared processor pool.

Utility CoD concepts

Utility CoD is for customers with unpredictable, short workload increases, and who want an automated and affordable way to help assure adequate server resource is available as needed. Once Utility CoD is enabled on your machine, you can place any number of inactive processors into the Shared Processor Pool.

The processors become available to the pool's resource manager. When the system recognizes that the combined processor utilization within the shared pool exceeds 100% of the level of base processors (purchased or active) assigned across uncapped partitions, a Utility CoD Processor Minute is charged, and this level of performance is available for the next minute of use.

If additional workload requires a higher level of performance, the system will automatically allow the additional Utility CoD processors to be used. The system automatically and continuously monitors and charges for the performance needed above the base (permanent) level.

Utility CoD enablement code

You must "enable" your system to use CoD. The Utility CoD Web site can provide a Utility CoD Enablement Code (key). To start using the inactive capacity on your system as utility capacity, you must enter the enablement code. The Utility CoD enablement code is valid for 365 server powered-on days.

During this time, your system sends reminders in the form of console messages. Console messages are sent during the last 30 days of the enablement period. This allows you time to return to the CoD Web site, accept the terms and conditions for another year, and receive a new enablement code.

After you have enabled your machine to use Utility CoD, you can move the inactive processors into the Shared Processor Pool. Then you can use them as utility processors in the pool's uncapped partitions.

Utility CoD processor minutes

You can place any number of inactive processors into the Shared Processor Pool. Once you have placed the processors in the pool, the processors become available to the pool's resource manager. When the system recognizes that all the used processors within the shared pool exceeds 100% of the level of base (purchased or active) processors assigned across uncapped partitions, a *Utility CoD Processor Minute* is charged.

This level of performance is available for the next minute of use. If additional workload requires a higher level of capacity, the system automatically allows the additional Utility CoD processors to be used. The system automatically and continuously monitors and charges for the capacity needed above the base (permanent) level.

When the system recognizes that the base processors assigned across uncapped partitions is 100% utilized, *and* at least 10% of one processor is needed, the additional processor resource is automatically applied and chargeable processor minutes start accruing. Processor minutes stop accruing when the utilization level drops and the base processors assigned can again handle the workload.

Planning for Utility CoD

Utility CoD allows you to activate processors on your server when your workload requires the additional resources. To prepare your server for Utility CoD, consider how you would like to license your software and determine when you anticipate the need to activate your resources. You also will need to plan for and set up your environment in preparation for CoD.

Software licensing considerations

As you select software for activating Utility CoD resources, you must consider how you would like to license your software. A number of methods are used to license software, such as per user, by software tier, or per processor. Typically a tool, such as a license manager, is used to manage the licenses. A license manager detects use of the software, compares it to the entitlement, and then takes action based on the results. A license manager can be provided by IBM (for example, Software License Manager) or one can be made available by the software provider.

Per user licensing

This applies to IBM and non-IBM middleware, and independent software vendor (ISV) offerings.

For Utility CoD, there is No Charge for this category. User entitlement does not change when inactive processors are temporarily activated.

Software tier licensing

This applies to IBM and non-IBM middleware, and independent software vendor (ISV) offerings.

For Utility CoD, there is No Charge for this category. Tier entitlement does not change when inactive processors are temporarily activated.

Per processor licensing for i5/OS, AIX, and Linux

For Utility CoD, there is No Charge. Processor entitlement does not change when inactive processors are temporarily activated.

Note: This rule might not apply to Linux® distributions. See your Linux distributor for details.

Per processor licensing for IBM middleware

For Utility CoD, the Daily User Charge applies. One processor day of entitlement must be purchased each time that any number of inactive processors are temporarily activated.

Determining when to activate resources

To determine when to activate additional processors and how many new resources you require, monitor your trends in CPU utilization by using a performance tool. Several performance tools are available to report CPU utilization information. Performance tools such as System i Navigator provide you with detailed information about how resources are being used, and alert you when utilization reaches a predefined critical level.

When computing average usage of all available processors, the system functions that report CPU utilization do not include the inactive processors in the total amount of CPU capacity. The inactive processors are not considered “available” within the various system functions that report CPU utilization percentages. The percentage of used CPU capacity is a calculated metric based on the amount of time that the processor was active within an elapsed time.

This is typically reported as a percentage, where 100% indicates that the processor was busy for the entire elapsed time. When multiple processors are present, CPU time must be adjusted to represent the average usage of all processors, so utilization is always reported as the percentage of total available capacity.

Setting up your environment for Capacity on Demand

Before you order any activation features, prepare your environment for integrating the additional capacity to ensure that your server is able to fully utilize the activated processors. To set up your environment for Capacity on Demand (CoD), you should:

1. Prepare any logical partitions (LPAR).
2. Perform any I/O conditioning.
3. Perform any disk upgrades.

Newly activated processors are immediately available for use by uncapped logical partitions. You can choose to assign these processors to one or more logical partitions. If there are no uncapped partitions, then Utility CoD processors cannot be used.

Using Utility CoD

Utility CoD provides the following services:

- ▶ Enable Utility CoD
- ▶ Disable Utility CoD
- ▶ Report use of Utility CoD processor minutes
- ▶ Review Utility CoD history
- ▶ Lookup Utility CoD keys

Figure 1 illustrates the process flow of Utility CoD. As you can see, every step is Web-based, there are no paper contracts, and no manual intervention is required.

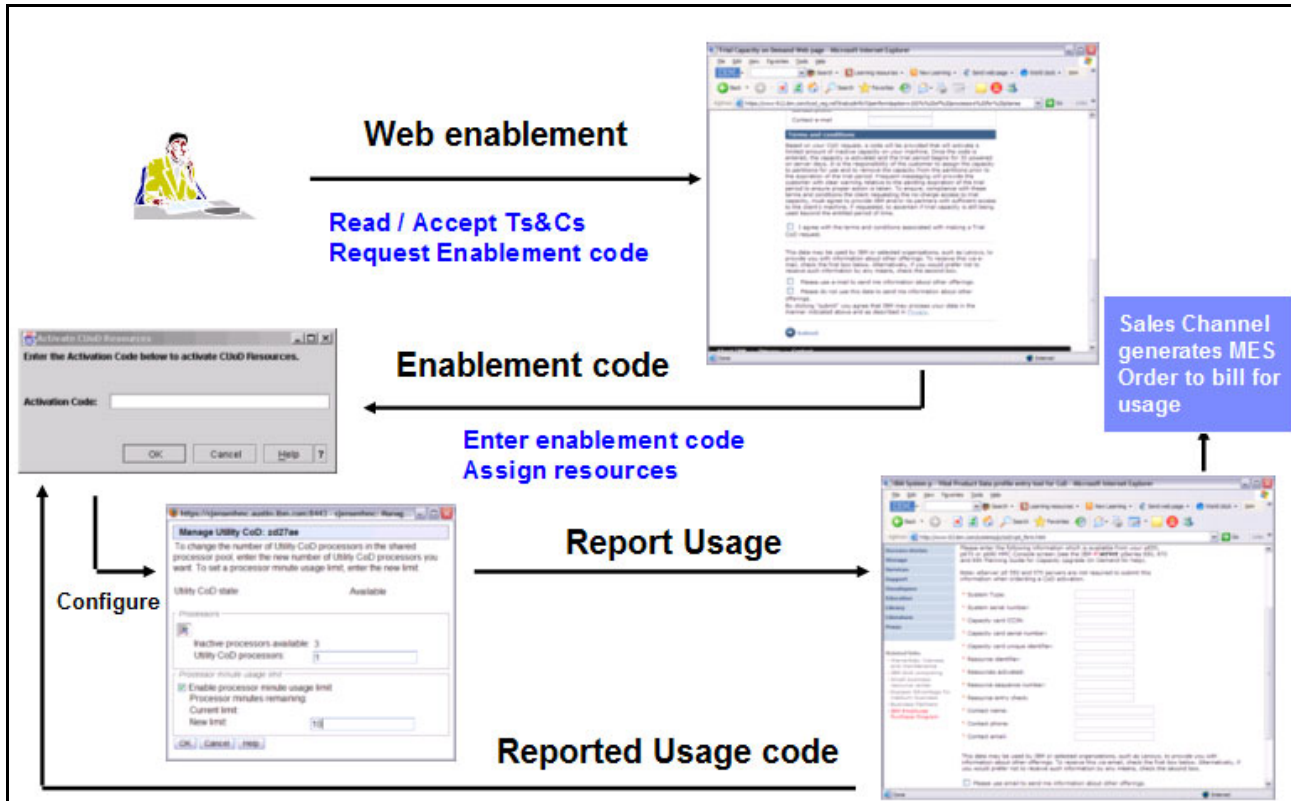


Figure 1 Utility CoD flow

This section provides an overview of the three main steps for using Utility CoD, specifically, how to:

- ▶ Set it up
- ▶ Use it
- ▶ Pay for it

Setting up Utility CoD

1. First, you need to enable Utility CoD. To enable it, visit the following link:
<http://www-304.ibm.com/jct01004c/systems/support/planning/capacity/index.html>
 Read the terms and conditions on the CoD Web site, then select the **Click to Accept** button.
2. A Utility CoD enablement code is provided by the Web site.
3. Enter the Utility CoD enablement code into the system.
4. Assign a quantity of inactive processors to the shared processor pool using HMC.
5. Set condition/state parameters for automated activation using HMC.

Note: The enablement code must be renewed every year via the CoD Web site.

Using Utility CoD

1. When the system recognizes that the base processors assigned across uncapped partitions are 100% utilized *and* at least 10% of one processor is needed, additional

processor resource is automatically applied and chargeable processor minutes start accruing.

2. Minutes stop accruing when the utilization level drops and the base processors assigned can handle the workload.
3. As needed, report usage to the CoD Web site, receive reporting authorization codes, and enter reporting codes into the system.

Note: Only processor activations are automated, not memory activations.

Paying for Utility CoD

1. You, as the user of the Utility CoD feature, do the following:

- Access the CoD Web site.
- Report minutes usage on the CoD Web site. Report in blocks of 100 minutes.

2. The CoD Web site provides a reported usage code. Enter this usage code into the system.

Reporting insight: There is a limited quantity of processor minutes enabled on the system. If the customer forgets to enter their usage code, the system will stop providing utility CoD function when that limit is reached. Once entered, the reported usage code extends the number of processor minutes available.

3. IBM® sends a statement of reported usage to the sales organization associated with you as a customer.
4. Upon receipt of the usage information, the sales organization executes the appropriate order as follows:
 - System i: Order one #5404 feature for each 100 processor minutes.
 - System p: Order one #5640, or #5641, or #5404 feature, depending on the processor feature, for each 100 processor minutes.

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.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs.

This document REDP-4416-00 was created or updated on May 28, 2009.

Send us your comments in one of the following ways:


- ▶ Use the online **Contact us** review Redbooks form found at:
ibm.com/redbooks
- ▶ Send your comments in an email to:
redbooks@us.ibm.com
- ▶ Mail your comments to:
IBM Corporation, International Technical Support Organization
Dept. HYTD Mail Station P099
2455 South Road
Poughkeepsie, NY 12601-5400 U.S.A.



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:

Redbooks (logo) ®
AIX®
i5/OS®

IBM®
POWER5™
POWER6™

System i™
System p™

The following terms are trademarks of other companies:

Linux is a trademark of Linus Torvalds in the United States, other countries, or both.

Other company, product, or service names may be trademarks or service marks of others.