Defining Availability
Jaqui Lynch

When discussing availability, most of us think of issues that relate to disaster recovery and failover. However, the general meaning of "availability" encompasses far more than this. According to dictionary.com, the definition of availability is:

1. Present and ready for use; at hand; accessible
2. Capable of being gotten
3. Qualified and willing to be of service or assistance

This definition fits well with the IBM On Demand Business initiative and, specifically, the Capacity on Demand (CoD) and Capacity BackUp offerings for IBM* eServer pSeries* and eServer p5 servers. In this article, I'll explain these features and ways to apply them.

Planning for Growth
CoD allows you to have additional processors and memory configured in the system but only pay for the extra resources when you start using them. There are several flavors of this capacity, and they each work a little differently, as I'll explain later. CoD's purpose is to help ensure the resources are there when they're needed as the workload grows. Two of the key advantages to implementing CoD are:

1. The resources are already there, so no disruptive installations are needed to activate them.
2. Spare processors in the system are used for dynamic processor sparing. This means that if a processor starts to fail, it's deactivated and one of the CoD processors is activated in its place.

So, not only do you have the capability to grow into the capacity, but it's also provided in a seamless, non-disruptive manner and can be used to provide failover for processors.

CoD comes in four flavors—Capacity Upgrade on Demand (CUoD), On/Off CoD, Reserve CoD and Trial CoD. Capacity Backup for Disaster Recovery is an additional offering.

CUoD—This involves installing inactive processors and memory in the server when it's purchased, and then activating these resources as the workload grows. In the eServer p5 server world, the activation increments are one processor or 1 GB of memory. This provides a very granular level of control for workload growth. Once processors or memory are activated in this manner, they're permanently activated. Activation is done when the extra capacity is needed by ordering and paying for an activation code, which is used dynamically to activate the processors and/or memory. Thus, CUoD's primary purpose is to allow for predictable, permanent growth in memory and processors.

On/Off CoD—This can be viewed as a rental agreement paid in retrospect. Effectively you have the capability to turn processors and memory on and off as needed. Activation has a minimum usage of one day, even if you only use five minutes. When On/Off CoD is activated, you set up the system to monitor the amount and length of the activations, and then send a monthly usage report to IBM. You're billed quarterly for the activations used based on those reports. When signing up for this offering, ensure that you understand what happens if the report isn't sent in. Additionally, once On/Off CoD is activated, special steps must be followed before it can be removed from the system, thus negating the need for those monthly reports. On/Off CoD is a good option for those systems where you have clear, predictable peak processing needs (e.g., the day after Thanksgiving in the retail industry).

Reserve CoD—This allows the user to pre-purchase a 30-day block of processor-usage time. This only applies to inactive processors in the shared processing pool. Management and usage of Reserve CoD processors is handled automatically by the system. When the workload needs to exceed the total of permanently active processors, the system starts to activate the reserve processors.

Trial CoD—Every server that has CUoD capabilities provides the option to take advantage of Trial CoD, which is the ability to activate all processors and memory at one time for up to 30 continuous days. This first activation is for all CUoD processors and memory. Once you purchase permanent processor activation, you have the option of another 30-day trial for up to two processors and 4 GB of memory. Trial CoD is requested by registering at the pSeries CoD Web site.

As explained, there are four ways customers can plan for growth, whether it's predictable or not. Additionally, by having inactive processors and memory in the system, you get the benefit of having redundancy if a processor starts to fail. Rather than taking a hard outage, the system can deactivate the failing processor and instantly activate a replacement.

Capacity BackUp—An additional offering that offers great value in the availability space is Capacity BackUp. When planning for disasters, it normally becomes necessary to duplicate everything in the main datacenter at the disaster-recovery site. This can become an expensive option and is often cost-prohibitive. IBM's solution to this challenge is the Capacity BackUp offering. In a nutshell, you purchase your regular production server (IBM eServer p5 570, eServer p5 590 or eServer p5 595), and then you purchase a Capacity BackUp version for your disaster-recovery site. I'll use the eServer p5 570 as an example: Your production server might be a full 16-way, but your Capacity BackUp server at the disaster-recovery site would have two active processors that you pay for and 14 inactive ones. It comes with 42 no-charge processor days that can be used in a disaster—additional processor days can be activated using On/Off CoD. So now you have a server fully provisioned at your disaster-recovery site, and you can use two of those processors for real workloads during normal times. Then in the event of a disaster, you can activate the whole complex and run your production while recovering your main datacenter. In comparison, the eServer p5 590 comes with four active processors and 28 inactive processors, and the eServer p5 595 comes with four active processors and either 28 or 60 inactive processors.

Capacity BackUp changes the landscape when planning for disaster-recovery sites. The ability to bring in a fully provisioned environment at a much lower cost now exists. However, this requires some planning. The inactive processors in a Capacity BackUp solution can't be permanently activated, and additional processor-day usage over the base allocation incurs charges. If not used as designed, it's possible to incur serious costs. Capacity BackUp isn't designed for use in a High Availability Cluster Multi-processing (HACMP*) or similar environment. However, it can be an attractive solution for implementing warm-site disaster recovery.

One to Grow On
Take advantage of the capacity offerings IBM has to facilitate business continuity. Availability can grow as your business grows. The livelihood of your business depends on it.

Jaqui Lynch, *eServer Magazine, IBM edition for UNIX* technical editor senior systems engineer focusing on pSeries and Linux* at Mainline Information Systems, has worked in the IS industry for more than 26 years. She's been responsible for a wide variety of projects and OSs across multiple vendor platforms, including mainframes, UNIX systems, midrange systems and personal workstations. Jaqui can be reached at* jaqui.lynch@mainline.com*.*