

ACHIEVING HIGH AVAILABILITY

PAULA S. MCDANIEL
SOLUTION DELIVERY MANAGER

HEWLETT-PACKARD CONSULTING

1200 EAST DIEHL ROAD

NAPERVILLE, ILLINOIS 60566

PHONE: (630) 245-3658

FAX: (630) 245-3154

paula_mcdaniel@hp.com

TABLE OF CONTENTS

Overview	1
Availability	1
Availability Continuum	1
Downtime Risk	3
Pillars of Availability	4
Technology Infrastructure	5
Hardware Reliability	5
Software Reliability	5
Monitoring and Management Tools	5
IT Processes	6
ITSM Reference Model	6
Availability Processes	8
Value of ITSM Reference Model	9
Support Partnerships	10
A.R.I.S.E.	10
Repair Contracts	11
Sustaining a Complex Environment	11
Conclusion	13

LIST OF FIGURES

<i>Figure 1: Availability Continuum</i>	2
<i>Figure 2: Relative Risk of Human Error</i>	3
<i>Figure 3: Availability Infrastructure</i>	4
<i>Figure 4: ITSM Reference Model</i>	7
<i>Figure 5: Questions Addressed by ITSM</i>	8
<i>Figure 6: Support Partnerships Provide Availability</i>	10
<i>Figure 7: HP Mission Critical Services</i>	13

Overview

Achieving high availability for your mission critical applications requires more than investments in technology. Three pillars form the foundation for an availability infrastructure that can support mission critical applications. These pillars are: technology infrastructure (the hardware, software, networks, etc.), support partnerships, and IT processes. Each of these pillars is equally important in achieving your desired availability goals. Investments must be made in each area. The amount of the investment will depend on the importance of each application to the overall business goals of the company. Many companies today make large investments in technology without regard to the other two pillars. This results in shifting the cause of downtime from hardware and software to human error. Good IT processes can go a long way to reducing the amount of human error, and strong support partnerships will help keep downtime to a minimum when a failure does occur.

Availability

Let's begin by looking at what we mean by "availability." International Data Corporation defines high availability as follows: "A computer is considered to be highly available if, when failure occurs, data is not lost, and the system can recover in a reasonable amount of time." This simplistic definition can be used as a starting point for discussion. The definition of "a reasonable amount of time" can be different for each company and for each application within the company.

From an end-user perspective, she/he needs to have reasonable, continuous access to systems that respond fast enough to accomplish the business function for which the systems were designed. Both planned and unplanned downtime detracts from the availability of the application. Most users would request continuous application availability, but availability requirements must be business decisions. The more highly available a system needs to be, the more resources it takes and the more it costs. This added cost is only justified if there is sufficient value to the business.

There is a mistaken belief in the industry that high availability is simply a hardware or software product that can be purchased, but there is no one single product that provides "high availability." Many things need to be considered when attempting to achieve availability for mission critical applications. You start by planning what levels of availability are required, and then you can design an infrastructure that can provide these levels. Once the solution is built and operational, you will need to measure it so it can be improved, and manage it to ensure that it remains viable as your business needs change.

Availability in the eyes of a user can be achieved in three simple steps: **Build it Right, Keep it Running, and Fix it Fast!** If vendors provide reliable, resilient technology and can assist in the planning and configuration of the solution, they will have **Built it Right**. Having processes in place to avoid problems, doing preventative maintenance, and monitoring for problems before they happen can **Keep it Running**. Vendors who provide a fast reactive element and good service contracts can help to **Fix it Fast**.

Availability Continuum

The process of providing availability must be a continuous cycle of improvement and change. The first step you need to take to achieve availability is determining what level you require. Availability is really a continuum, not a single state, offering different levels of availability at different costs. The highest levels of availability may be required for some of your services, but lesser levels of availability might be satisfactory for others.

Start by taking the time to analyze the cost of downtime and the impact it has on your business. Then decide what level of availability you require. Many applications can be run on the lowest level of reliable systems. These are applications whose unavailability has little impact on the business. Other applications will require higher levels of availability. This distinction must be dictated by the needs of the business, because the higher the level of availability required, the higher the investment that is required.

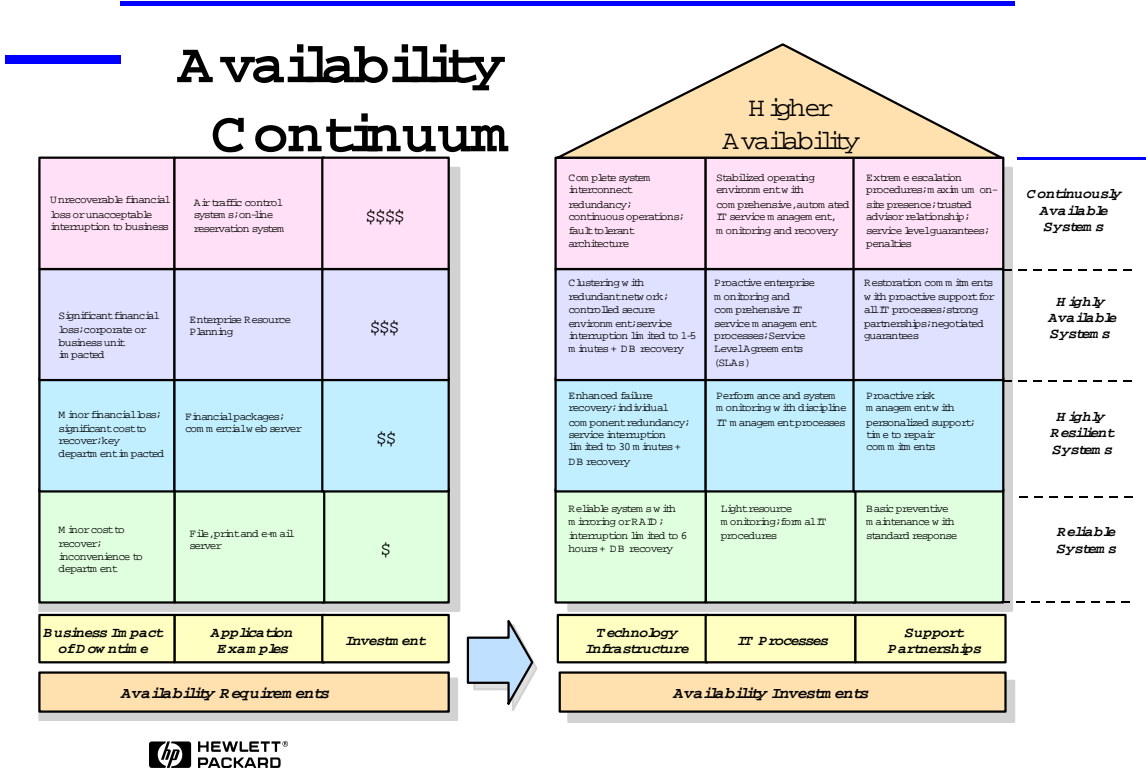


Figure 1: Availability Continuum

You can see the three pillars of availability on the right side of the availability continuum; these are all equally important. Making the highest level of investment in technology will not result in high availability if investments are not also made in IT processes and support partnerships.

The examples used in this chart are just that, examples. In some environments, the e-mail servers could be considered mission critical and require higher levels of availability. In addition, the needs on the right side may differ from customer to customer. The important thing is to understand the relationships between technology, processes, and support, and the business needs and required investments.

Downtime Risk

Let's look at how the downtime risk changes as you move up the Continuum to higher levels of availability.

The relative risk of human error

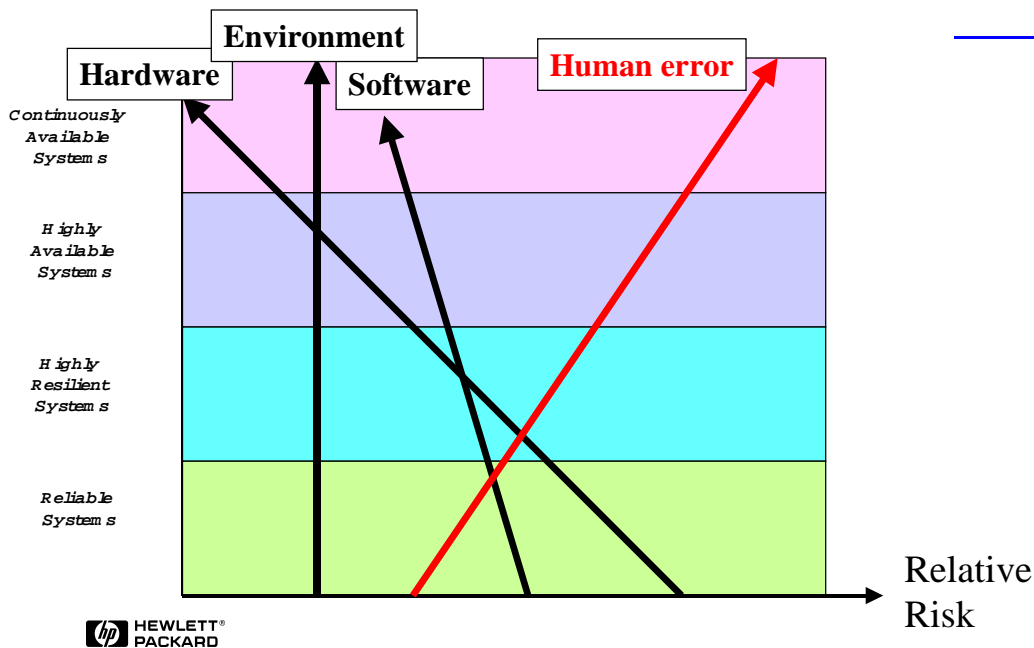


Figure 2: Relative Risk of Human Error

In a reliable systems environment the level of hardware redundancy is minimal so most failures will cause an outage. As you travel up the Continuum the greater levels of redundancy (e.g. mirrored disks, redundant systems, etc.) reduce the likelihood of a hardware failure causing an outage. In a Continuously Available Systems environment, complete system interconnect redundancy reduces the risk of an outage due to hardware to virtually nil.

In the area of software risk, at the reliable systems environment the level of software resiliency is relatively minimal so most failures will cause an outage. As you travel up the Continuum the greater levels of software resiliency (e.g. journaled file systems, database recoverability, etc.) reduce the likelihood of a software failure causing an outage.

Again, as you travel up the continuum, improvements in security and environmental controls reduce the risk of an outage due to the environment. However, when contrasted with the software and hardware risk reductions, the relative risk is fairly constant. Once you have implemented computer room security, redundant power supplies, etc., there is little more you can do to greatly reduce the risk.

The real difference in risk comes in the area of relative risk due to human error. Creating the redundancy and resiliency required for highly available systems means an increase in the complexity of the configurations. With greater complexity, the relative risk of failures due to human error increases significantly. Good management of change, formal IT processes, and proactive services are the keys to success in this area.

Pillars of Availability

Building an effective, highly available infrastructure requires a strong foundation. We look at the three components of availability as pillars, providing the foundation. It is through a continuous process of investment and quality improvement in the areas of people, process, and technology, that you can achieve the availability infrastructure necessary to meet your essential business goals. It is important to realize that every pillar is necessary for a complete mission-critical solution. HP has set an industry standard by helping you understand each pillar's importance.

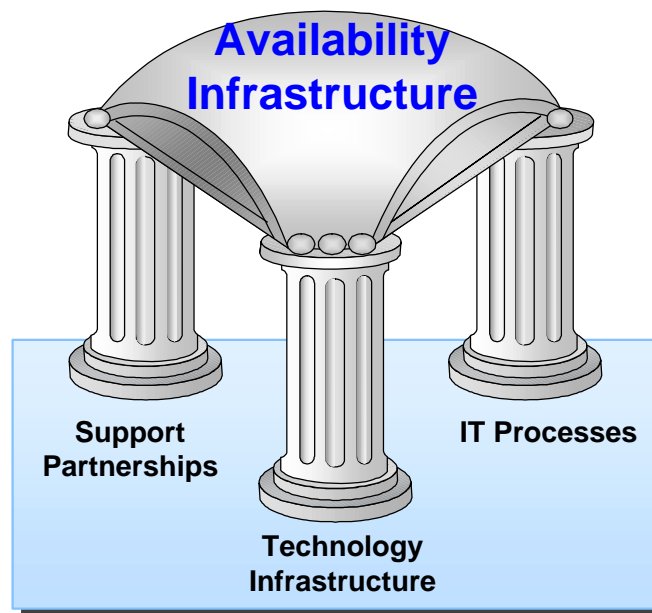


Figure 3: Availability Infrastructure

First, the technology must be reliable and resilient enough to provide maximum performance while still being flexible enough to respond to changing business needs.

Second, the supportability must be designed into the technology. Supportability encompasses reactive services, predictive capabilities, remote diagnostic tools, training, and education. Having the highest level of reactive service, built-in supportability, and a well-educated staff, will ensure that you can diagnose and resolve problems quickly.

Third, the processes that control the environment must be highly flexible and integrated, implement robust process measurements, define needed linkages, and provide required reporting mechanisms. Having easy, automated processes for management will minimize human error, which, as we have seen, causes increasing amounts of downtime in complex environments.

Now, let's look at the foundation of availability one pillar at a time.

Technology Infrastructure

Technology is usually the first place people invest money when trying to achieve high availability. There are many products on the market that will help you reach your availability goals. The area of technology is the area where you need vendors to **Build it Right**. We'll look at the broad categories of hardware, software, and monitoring tools.

Hardware Reliability

Systems featuring high availability are created with a solid IT architecture, and are designed and built around high-quality, durable hardware. Diagnostic tools that minimize time to diagnose and repair problems are essential. On-line fault detection should also be available on the hardware you choose.

The goal of the architecture is to keep the system running. This requires an effort to find and remove, or minimize, all single points of failure. Increasingly common ways to do this are with hot swappable disk drives; disk mirroring; shared disks and memory; redundant power supplies; redundant networks; server clusters; and extensive diagnostic technology.

Software Reliability

Software reliability depends first on a robust operating system capable of handling errors. You also need reliable applications that have features such as restartability and recoverability built in. The architecture should also incorporate failover capability, such as HP's MC/ServiceGuard and Microsoft's Cluster Server. These software products allow an application to be started on an alternate server in the case of a failure. They also provide the additional feature of rolling software upgrades, which allow you to keep running an application on one system while another system in the cluster is being upgraded or patched.

Things such as journaled file systems, application restartability and database recoverability minimize the amount of time it takes to resume operations following a failure event. Load balancing tools can also be utilized to increase availability by making better use of underutilized systems and by ensuring that system overload does not render an application unusable.

Online database backup tools provide a way to minimize planned downtime by permitting an application to continue running while an off-line copy of the database is being backed up.

Monitoring and Management Tools

An often-overlooked part of the technology infrastructure is the monitoring and management tools. To prevent unscheduled downtime, it is important to constantly monitor the environment and take action when needed. You need automated tools that assist you in managing your environment and in detecting potential problems before a failure occurs. These tools should allow you to monitor, administer, and maintain your systems from one central management station.

Tools are also required to measure the performance levels of your systems. Remember, if you don't measure your current performance levels you cannot improve them, and providing availability is a continuous cycle of improvement.

IT Processes

Let's look at the second pillar now, IT processes. In many organizations, the current situation is that processes are undefined or vague, there are poor linkages between processes, and there are unclear roles and responsibilities. This results in inconsistent service delivery, poor customer satisfaction, multiple reworks, periodic work stoppages, communications breakdowns, and duplicate work efforts. The IT processes are not helping to **Keep it Running**.

That's why HP has developed the HP IT Service Management (ITSM) Reference Model. The reference model, which uses a series of high-level process relationship diagrams, allows HP to help you assess your existing IT environments and to design, implement, and manage new environments with fully integrated processes and technologies. It also helps determine which processes and technologies should be managed internally and which should be outsourced.

ITSM Reference Model

The HP ITSM Reference Model is based on a set of industry standard concepts defined in the Information Technology Infrastructure Library (ITIL). ITIL was originally created by the United Kingdom government to better understand the management of services delivered to an end-user community by IT professionals. Hewlett-Packard leveraged its own wealth of IT experience by involving Hewlett-Packard's internal IT organization as an active participant in the model's development and verification.

HP ITSM Reference Model

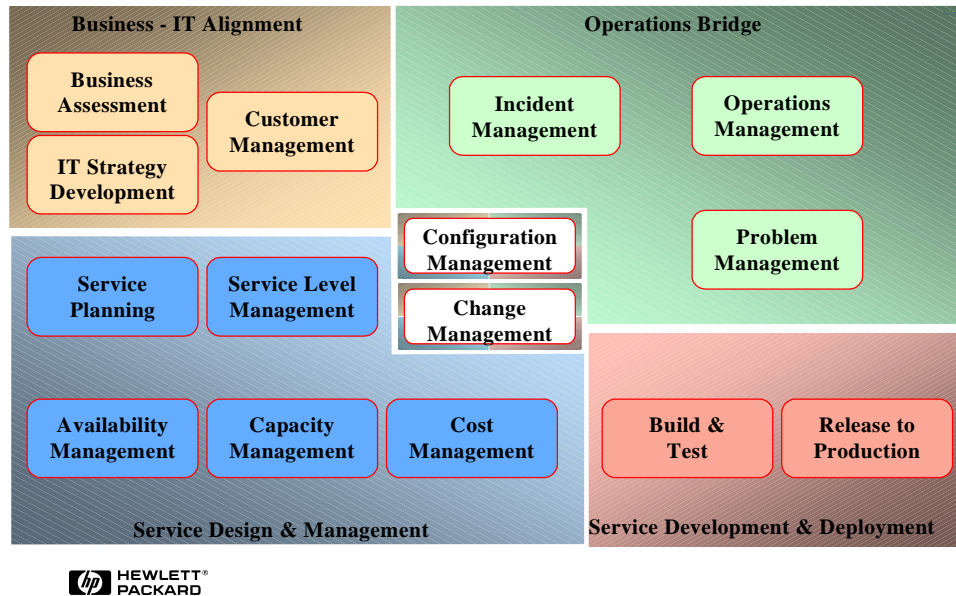


Figure 4: ITSM Reference Model

HP's IT Service Management Reference Model divides the required processes up into 4 major process groups, each with a specific purpose. Business-IT Alignment understands the business and customer requirements to formulate an IT strategy that optimizes IT's added value. Service Design and Management translates the IT strategy into planned IT services, designs and re-designs these services on an on-going basis to promote continuous improvement, and defines and manages the service levels within cost constraints or service budget. Service Development and Deployment develops, tests, and deploys the services according to the service design. The Operations Bridge interfaces with end users, runs the services, monitors and maintains the service infrastructure, and resolves and prevents incidents.

There are two central processes, Change Management and Configuration Management, that are the heart of the model. Change Management coordinates and controls production changes. Configuration Management controls assets and documents the asset relationships.

Each process group within the Reference Model addresses a specific high level question. The answers to these questions define how an IT organization provides value to the business and computing services to its end users.

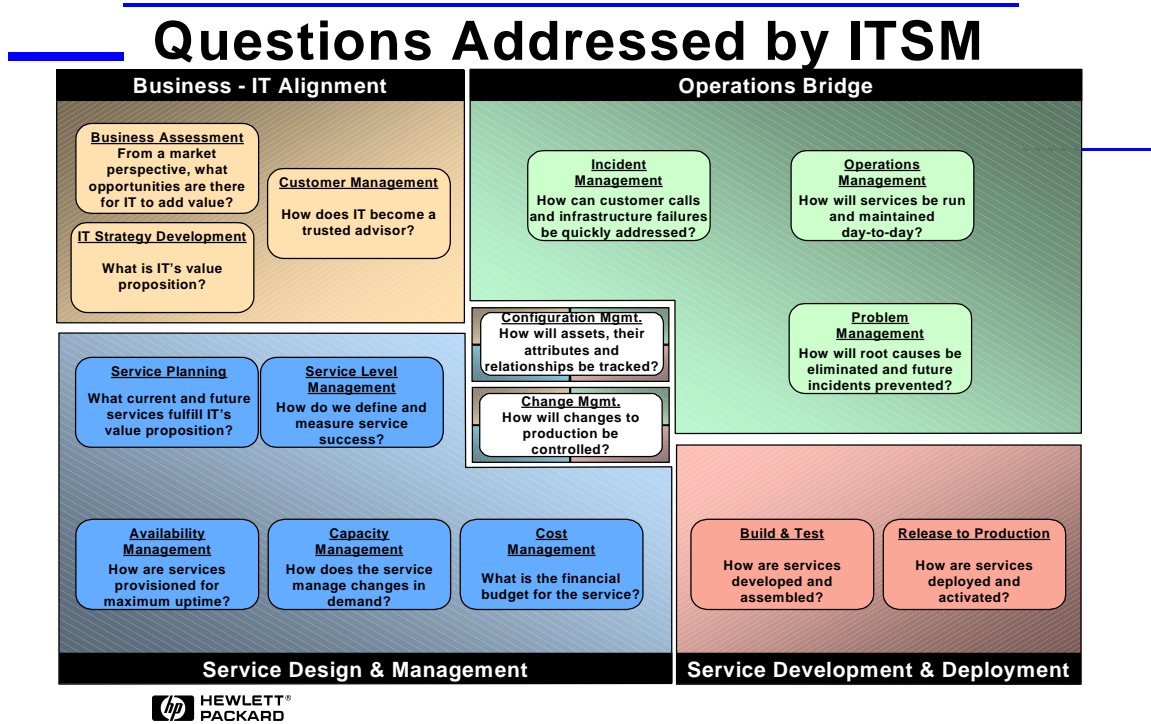


Figure 5: Questions Addressed by ITSM

Availability Processes

Let's take a closer look at the model for some processes that impact availability. While all of these areas are important and there are many processes within each that impact availability, let's concentrate on a few major ones.

Change management is more critical in a high availability environment than elsewhere. Any change to the environment, even patches to software, must undergo a rigorous change review before being permitted. Changes should include a risk analysis, a detailed schedule, and have back-out plans.

In the Business/IT Alignment area, the most important processes are working with your end user customers to determine their needs and then understanding how those needs map to the overall needs of the business. You need to use business priorities to determine how to make availability investments.

In Service Design and Management, you need processes to design the infrastructure to meet the availability needs, to establish supplier agreements, to manage the costs of availability, to determine the service level requirements, to ensure these service levels are being met, and to ensure you have enough capacity to meet the needs. You also need to address business continuity planning, or disaster recovery, as part of this area.

One of the often-overlooked areas is in Service Development & Deployment. It is critical to have a test environment in which to test new services and changes and/or fixes to existing system components. A formalized “release to production” process will ensure that approved changes do not cause disruption to availability when they are put into your production environment.

The Operations Bridge area must include processes for problem management, escalation, and problem analysis. The procedures for recovery of all parts of the environment must also be detailed here. This area needs to include processes for the management and monitoring of your high availability systems. In addition, all the procedures used by the Operations Bridge need to be documented and tested.

Value of ITSM Reference Model

Since the processes you have in place in IT are critical to achieving availability goals, it is important these processes be clear and well defined. This will ensure you can **Keep it Running**.

Use the HP ITSM Reference model to identify processes you currently have in place, their status, value, and relationships with other key IT processes. The model is a quick reference tool that demonstrates the “desired state” to be achieved by IT. With the assessment of the current state and this “best” state model, you can easily identify the “gaps,” what needs to be accomplished to get to the availability level you desire.

Although the model represents what processes IT organizations must have in place to deliver service management, in reality, every corporation differs in its immediate needs. The organization, therefore, must consider a variety of implementation priorities for different processes. The reference model expedites this effort because of its emphasis on inter-process relationships and linkages, thereby helping your IT organization to judge the impact and value of one implementation approach versus another. The highest priority can be given to processes affecting availability.

While the model can be used to assess the current state of IT with regard to IT processes, it can also be used to take the problems uncovered by an assessment and map them to the appropriate processes so you can see where the fix is required. If you determine that updates to your software are causing downtime, you need to look at Change Management, Build & Test, and Release to Production processes for resolution.

Given the uniqueness of each IT organization, no single organizational model can be applied within each corporation. However, the model can be used as a reference for restructuring IT organizations along process and service lines. This will help to break down the silos that often exist today, where systems are managed in one group, networks in another, and applications in yet another. If your organization can be structured so that one group is responsible for providing an end-to-end service, the cases of finger pointing between groups should disappear, enhancing your ability to provide availability.

Drilling down into the model and analyzing process hand-offs and integration points provides you with the ability to target potential areas for implementation of process-enabling technologies. These technologies can be vital in helping you monitor and manage your environment and know quickly about potential availability issues.

Finally, implementing the processes defined by the reference model will help you make cost-justified decisions regarding which services make business sense to keep in-house, and which should be considered opportunities for selective outsourcing.

Support Partnerships

Now, let's look at the last pillar, Support Partnerships. Support partnerships refer to all of the people required to maintain and support your IT environment. These are the people who have the ability to help **Keep it Running** and **Fix it Fast** if it fails. Support partnerships include external organizations you use for outsourcing, as well as those with whom you contract for hardware and software support.

The highly skilled and trained people in your own organization need to be backed up by the skills and training of your vendors. You need to be able to count on your vendors to be there when you need them, with the right parts and the right answers.

Vendors can provide assistance to **Keep it Running** by providing the reliable hardware and software we have already discussed. They can also, through contracts, provide proactive monitoring of your existing systems and networks and participation in your change management processes. Vendor consultants can also assist you in designing your infrastructure for the highest availability, in assessing your current capabilities, and in planning and carrying out improvements.

A.R.I.S.E.

Support partnerships can help you achieve availability by providing technology and services built on the principles of avoidance, regeneration, inclusion, simplicity, and expertise.

Support Partnerships Provide Availability through:

- **A**voidance - prevention of downtime
- **R**egeneration - transparent repair
- **I**nclusion - extension of availability beyond systems
- **S**implicity - ease of installation and use
- **E**xpertise - demonstrated knowledge of IT needs



Figure 6: Support Partnerships Provide Availability

- **Avoidance** - To prevent downtime, you need partners who provide advanced diagnostics to protect against, predict, and prevent component failures. Partners should provide advice and assistance in ways to eliminate planned downtime, and should help to design an infrastructure and processes that prevent and control both planned and unplanned downtime.
- **Regeneration** - Look for partners who can provide regeneration within systems and networks, such as is possible with mirrored or RAID disks and clustering products. Also important is the ability to quickly and easily repair systems and restore them to full operation.
- **Inclusion** - Since the availability of a system no longer guarantees the availability of your application, you need partners who can extend availability to include the entire application environment, including networks, databases, applications, and your end-user workstations.
- **Simplicity** - Simplicity is an important element in availability. A system or application that is easy to install, easy to use and easy to customize will reduce human errors that can cause downtime. Partners who provide a single point of administering, monitoring, and managing the environment can help you keep downtime to a minimum.
- **Expertise** - Finally, look for partners who take the time to know your business and your IT needs, and who are leaders in creating and integrating new technologies.

Repair Contracts

Having the proper level of support in place is critical when something fails. Vendor support options must provide repair and restore commitments to ensure that systems can be fixed fast. Twenty-four hour per day seven day per week support is usually necessary for mission critical services. You may also require direct or immediate access to the vendor's highly skilled technicians at all times during the year.

Some types of predictive support can allow potential problems to be reported and fixed before they cause an outage, and can even monitor your environment for failures of redundant components, that might otherwise go unnoticed. Repair contracts that provide for proactive maintenance such as this can be critical to keeping downtime to a minimum.

Sustaining a Complex Environment

Prevention is clearly the preferred approach in a mission critical environment. In clustered environments, the risks from poor processes or change management increase significantly (when compared to stand-alone systems). Using consultants to assist you in designing and implementing robust processes can really improve the level of availability you are able to reach. Involving support vendors in your change management processes can also be a significant help.

Sustaining a cluster is much more complex than a stand-alone system. Routine administrative tasks for stand-alone systems (e.g. adding a piece of hardware) take on much greater significance, and difficulty, in a highly available cluster.

In a clustered environment, the failure of one node in the cluster could result in degraded performance of the remaining nodes, so it is important that someone notice the failure immediately.

Also, there is a risk during a failure that a second failure could cause application downtime. Therefore, it is critical to detect and repair failures in a cluster before another failure occurs.

Clustering software is extremely sensitive to things like adding new hardware that it is not configured for, adding new technology it is not certified for, making script changes in application software, and applying operating system upgrades and patches. Implementing any of the tasks listed above without fully understanding the implications for the cluster can result in the entire cluster failing. For example, if an application script has a hard-coded address included in it, the application may not run if it is failed-over to another server. Similarly, if a new piece of hardware which is not certified is added, the cluster could fail.

In a clustered environment, the value of expertise and appropriate process cannot be underestimated. Similarly, due to the dynamic nature of the support environment, where new problems are found and fixed every day, access to a vendor's knowledge database can be crucial in avoiding "surprise" outages. Tools that identify and track changes can also add great value.

In a complex environment, normal reactive support from a vendor is not sufficient. You need proactive support from someone with knowledge of your environment. You also need your support partner to have ready access to dedicated experts who are committed to resolving your problem as fast as possible. This is the **Fix it Fast** commitment you need from a vendor.

Conclusion

As we have seen, availability is not just a product you can buy. It really is a continuum, offering different levels of availability at different costs. In order to **Build it Right**, **Keep it Running**, and **Fix it Fast**, you must make investments in all three areas of: technology, IT processes, and support partnerships. The three pillars supporting your availability infrastructure are all equally important to being able to achieve the level of availability necessary to support your mission critical applications.

HP can provide not only the technology that is required for mission critical solutions, but also the services required to complete the mission critical foundation of support partnerships and IT processes.

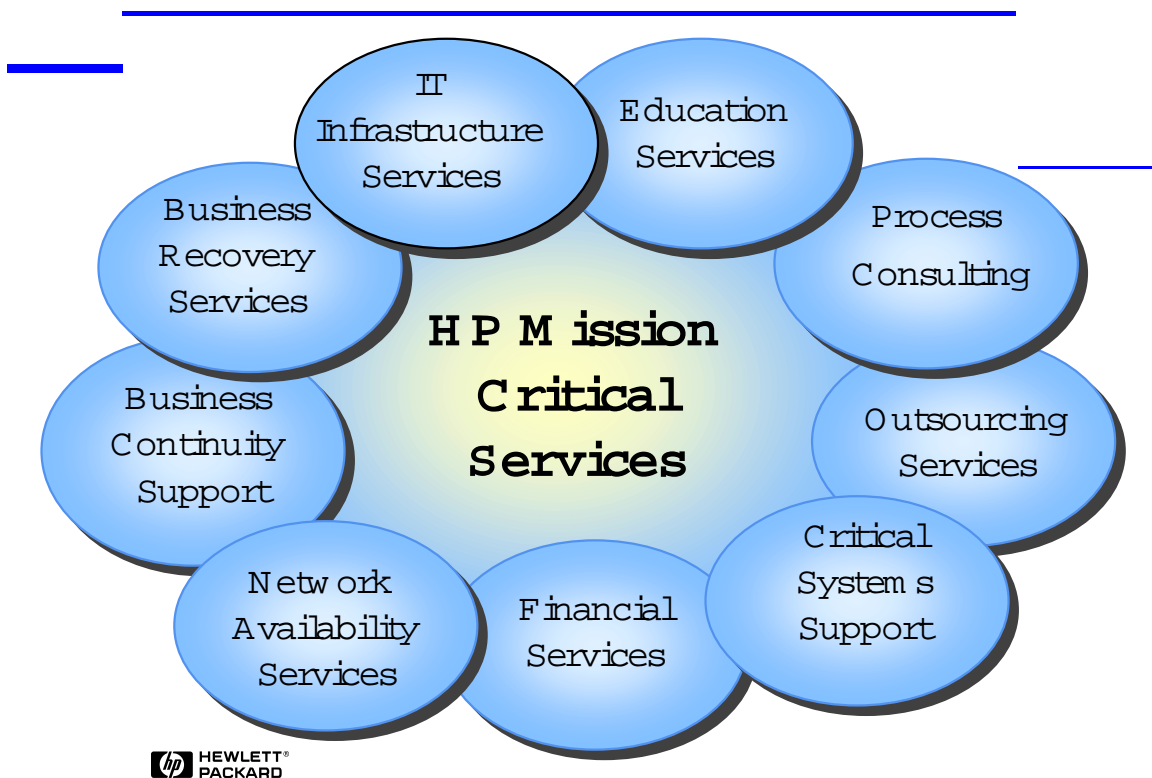


Figure 7: HP Mission Critical Services

Each of these ovals represents a service or group of services provided by HP. Through these services, HP can assist you in building your high availability infrastructure.