



The three levels of high availability – Balancing priorities and cost

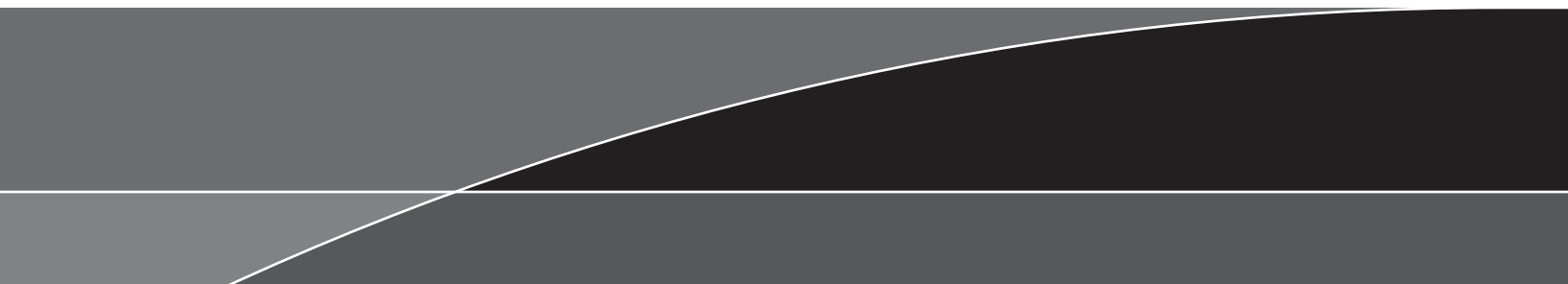


Table of contents

Executive summary

Availability – Goals and metrics

Evaluating downtime	3
Availability levels	4

Availability options in XenServer environments

Level 1 – Core failover, recovery and restart	6
Level 2 – High availability with component-level fault tolerance	7
Level 3 – Continuous availability with system-level fault tolerance	7
Availability level comparison.....	8
Resource requirements and TCO.....	8

Summary

Executive summary

Virtualization technologies make it much easier and less expensive to keep key applications available through failures, outages and even regional disasters. By carefully balancing downtime costs and availability options, organizations can extend the benefits of high availability (HA) to even more applications, locations and users.

Citrix Systems, Inc. and Marathon Technologies have collaborated on a range of HA alternatives based on Citrix XenServer™ virtualization. Their solutions, ranging from core failover/recover/restart through high and continuous availability, offer capabilities appropriate for every application, business and budget.

Availability – Goals and metrics

Information technology exists to deliver business processes and information to employees, customers and partners, so availability is a strategic IT imperative. But not all applications have the same priority, so it's important to distinguish among different availability goals and standards:

High availability means applications remain accessible a very high percentage of the time and recover quickly from localized failures so users experience little or no interruption.

Continuous availability is availability's upper limit—nonstop operation through failures and outages, eliminating downtime and associated data and transaction loss.

Disaster recovery is restoration of IT services after a widespread disruption: floods, regional power interruptions or other major disasters.

New technologies are driving down IT costs, but better availability always consumes more resources—of hardware, software, time, attention and money. Continuous availability for every application in every conceivable catastrophe is unrealistic and wasteful. A far better approach is an application-by-application review that matches availability objectives with costs, using three metrics:

Recovery Time Objective (RTO) – The time by which an application must recover from a failure, or maximum acceptable loss of service

Recovery Point Objective (RPO) – The point in time at which an application must recover, or maximum acceptable loss of data and transactions

Total Cost of Ownership (TCO) – The direct and indirect costs of achieving the specified RTO and RPO over an application's lifespan

Evaluating downtime

It's easy to evaluate direct losses of current revenue when customer-facing transaction-processing applications go down. But the indirect costs may be even greater, and may extend even to applications like e-mail and Web services that are not directly in the revenue stream. An end-to-end assessment of downtime costs will include losses of:

Productivity	Employee hours lost, multiplied by their fully-burdened hourly rate.
Reputation	Net present value of lost business when customers and partners take their business elsewhere.
Revenue and margin	Penalties and other compensation, billing delays and investment losses.
Other financial	Revenue recognition errors and delays, discounts, payment guarantees, credit ratings and stock price.

Availability levels

Once you have established downtime costs for your most important applications, you are ready to match them to availability options. Availability metrics vary: probability of unplanned downtime and nines of percentage uptime (e.g., 99.99 percent = four nines) are two scales in wide use. Organizations may even use different metrics for different applications. But all availability measurements span the levels illustrated in Figure 1:

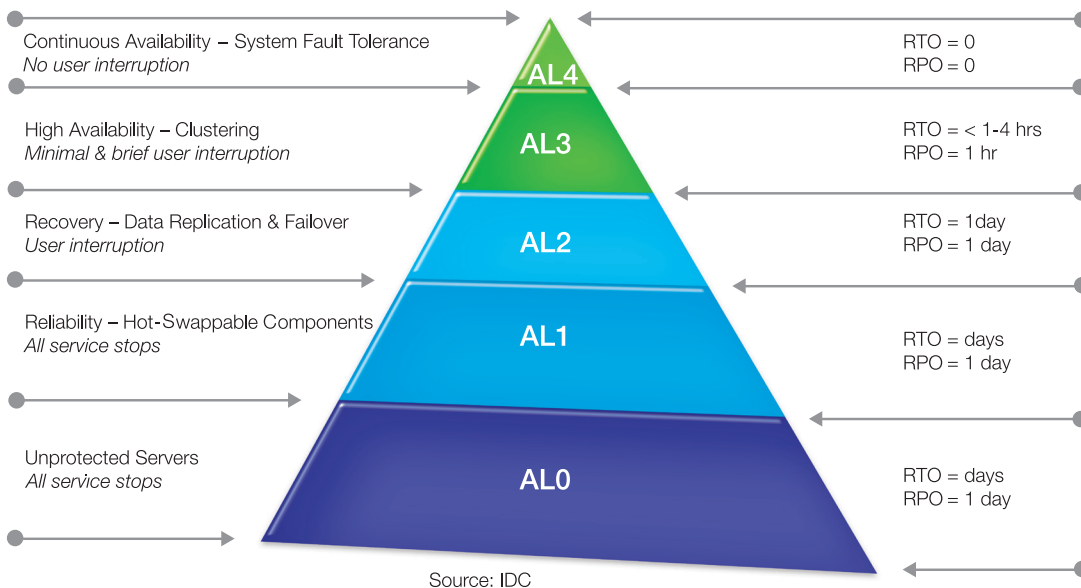


Figure 1 - Application Availability Pyramid

Reliable	Many—often most—applications require no protection or hot-swappable hardware. You will restore them eventually after a major failure, but the business does not depend on them.
Recoverable	Web services, DNS and Active Directory servers, and other redundant infrastructure components should recover automatically, but some downtime is acceptable.
Highly available	CRM, ERP, database and e-mail servers run the business. Service interruption and data loss are very expensive, so RTO and RPO are demanding.
Continuously available	For a few applications such as process control, security, trading and banking, no downtime is acceptable. Some businesses may have no applications in this category.

Availability options in XenServer environments

Citrix XenServer is open, powerful server virtualization that radically reduces datacenter costs by transforming static and complex datacenter environments into dynamic, easy-to-manage server workload delivery centers. Based on the open-source Xen® hypervisor, XenServer delivers a secure, mature server virtualization platform with near bare-metal performance.

XenServer virtualization solutions dramatically reduce the costs of availability, bringing high and continuous availability, and prompt disaster recovery to a broad range of business applications. XenServer allows IT organizations to consolidate multiple server workloads into a shared pool of physical servers and storage, for simple, powerful, cost-effective high availability.

Citrix has worked closely with Marathon Technologies to create options that span business availability requirements at a far lower TCO than previously possible.

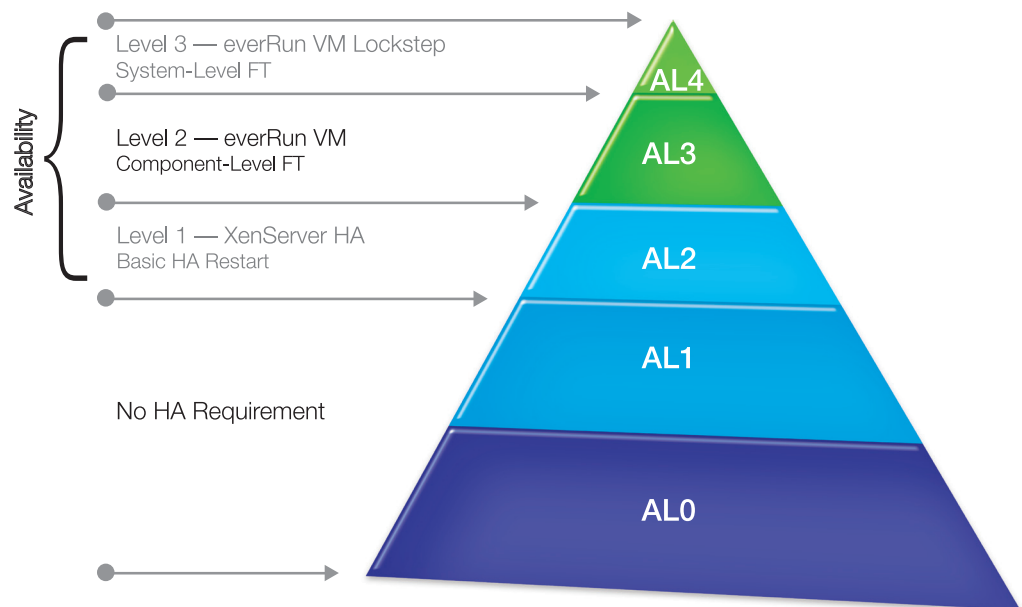


Figure 2 - Protecting Virtual Environments

As illustrated in Figure 2, the combination of XenServer and Marathon everRun™ virtualization solutions offers three levels of recoverability and availability:

Level 1	XenServer delivers out-of-the-box high availability, including cost-effective core failover, recovery and restart capabilities for applications running in the virtual environment.
Level 2	Marathon everRun VM delivers high availability from component-level fault tolerance, eliminating downtime caused by I/O component failures and guaranteeing recovery from system failures.
Level 3	Marathon everRun VM Lockstep (available 1H 2009) delivers continuous availability from system-level fault tolerance, eliminating data loss, downtime and transaction loss.

Tight integration between XenServer and Marathon everRun technologies balances availability benefits and costs for every requirement. In any combination, these solutions protect every virtual machine through a single management interface that spans the entire virtual environment.

Level 1 – Core failover, recovery and restart

Level 1 protection is appropriate for reliable and recoverable applications—those that are not absolutely critical, and where some delay can be tolerated while virtual machines are automatically restarted. Most organizations will apply this level of protection to knowledge databases, file-transfer utilities, project-management tools and HR applications. XenServer provides continuous assessment and monitoring of virtual server failover environments, with smooth failover, recovery and restart ideal for routine or infrastructure applications, or test and development systems.

To deliver Level 1 protection, XenServer monitors hosts in a resource pool to detect any host failure. When it finds one, it directs the virtual machines on that host to restart on a different host in the pool. Manual intervention is required only when a virtual machine cannot be restarted because the recovery host has also failed, or lacks resources.

XenServer provides basic high availability:

- Advance calculation of a XenServer pool's resources to determine whether they can handle a preset number of simultaneous host failures
- Continuous health monitoring for hosts within the pool
- Smooth, automated failover and restart using another host in the pool
- Prioritized restart of virtual machines so the correct workloads are running after failover

Level 1 availability requires only XenServer and a shared storage configuration to provide storage and data protection. In multilevel high-availability environments, XenServer can be managed along with other availability levels using the everRun Availability Center interface.

Level 2 – High availability with component-level fault tolerance

For business-critical applications, where data loss and downtime carry extremely high costs, Level 2 provided by everRun VM is the optimal choice. It is ideal for e-mail, customer relationship management, enterprise resource planning, back-office databases, financial software and other critical applications. The solution provides component-level fault tolerance, and sustains the loss of individual networks or storage components with no service interruption or data loss.

Running within the XenServer environment, everRun VM uses Marathon's proprietary ComputeThru® technology to monitor all system components; it also maintains an exact, synchronous duplicate virtual machine on a secondary host within the XenServer pool for full redundancy. The solution identifies faulty I/O pathways before they become a problem and responds to a wide range of I/O and component failures. Active validation of all components on primary and secondary hosts ensures smooth recovery following any primary host component failure.

Level 2 availability delivers component-level fault tolerance:

- Automated setup, runtime and fault management with policies to handle system, network and disk I/O failures
- Assured recovery of virtual machines with zero downtime from I/O failure and zero data loss
- Disaster tolerance from synchronous data mirroring between geographically separate hosts with no requirement for shared storage
- Continuous active validation of all production and standby system components for complete redundancy at all times
- System, network and data availability in a single integrated solution

Level 2 protection requires XenServer, Marathon everRun VM and dedicated secondary hosts to maintain redundancy within XenServer resource pools.

Level 3 – Continuous availability with system-level fault tolerance

For mission-critical and life-sustaining applications where any service interruption is unacceptable, Marathon everRun VM with Lockstep provides system-level fault tolerance and continuous availability in the face of component or system-wide failures. Applications that control life-sustaining processes in power, healthcare, air and rail transportation, and military and civilian security applications are just a few of the cases for which the extreme cost of downtime justifies investment in continuous availability.

Marathon everRun VM with Lockstep offers all of the benefits of everRun VM (Level 2) and adds two important attributes:

- Zero downtime, even with complete host failure
- Preservation of application and memory states during failures

Availability level comparison

This table summarizes the features of Levels 1 through 3:

Feature	Level 1	Level 2	Level 3
Protection from host failure	Yes – VM restarts on surviving host	Yes – VM restarts in redundant environment	Yes – VM continues without interruption
Protection from VM failure	No	Yes	Yes
Protection from I/O failure	External, using NIC bonding or multi-path Host Bus Adapters	Yes	Yes
Storage	Requires shared storage	Mirrors local or shared storage between hosts	Mirrors local or shared storage between hosts
Monitoring and validation	Resource calculation and health monitoring	Active component validation	Active component validation
Supports geographic separation for disaster tolerance	No	Yes	Yes
Resource requirements	Reserved memory for VM recovery	Reserved redundant memory, network and storage resources	Reserved and running redundant CPU, memory, network and storage resources

Resource requirements and TCO

Risk management decisions balance the marginal costs of solutions against the marginal costs of the risks they mitigate. Choosing an availability solution is no different. Managers calculate the costs to raise availability a few percentage points (or even fractions of a point), then determine whether the expected reduction in downtime justifies them.

Realism about both benefits and costs is the key. The TCO for any level of availability includes initial direct infrastructure and software costs, direct and indirect costs of implementation, operating overheads, staffing and consulting costs, and more. For most businesses, modest improvements in availability of noncritical applications seldom justify Level 3 availability. Although a chemical plant or financial exchange may consider Level 3 a bargain, a retailer or warehouse may find that Level 2 provides all the protection they need, even for their most critical applications.

Virtualization technology has changed the cost structure of high availability dramatically in just a few years, in many cases replacing expensive server clusters and exotic remote mirroring strategies with innovative combinations of routine hardware and high-performance software. Still, careful consideration of the total costs of owning and operating a virtualization solution can have dramatic effects on the scale and performance of your high availability initiative.

Citrix pioneered the delivery of applications to users, and its Citrix XenApp™ application virtualization technology continues to lead the industry. Affordable, high-performance desktop and server virtualization are natural extensions of the company's vision—and Citrix now offers the lowest TCO among application, desktop and server virtualization solutions. The company's focus on cost, performance, commitment to open-source development of critical hypervisor technology and effective partner network bring unparalleled value to every customer virtualization and high availability initiative.

Summary

Citrix XenServer and Marathon everRun have enhanced the XenServer environment to run applications at availability levels that range from routine recovery and restart operations to complete fault tolerance. Together, XenServer and Marathon everRun offer availability solutions at attractive price points to match every business or application need, without adding complexity.

The flexible everRun management framework allows you to specify the availability level appropriate for each application you add to your virtualized environment. You may start with XenServer for infrastructure applications, for example, and add everRun VM to protect more business-critical applications as you add or virtualize them. With easy-to-use availability solutions integrated into your virtual computing environment, you can enjoy the benefits of highly available virtual servers for a broad range of applications.

Citrix Worldwide

Worldwide headquarters

Citrix Systems, Inc.
851 West Cypress Creek Road
Fort Lauderdale, FL 33309
USA
T +1 800 393 1888
T +1 954 267 3000

Regional headquarters

Americas
Citrix Silicon Valley
4988 Great America Parkway
Santa Clara, CA 95054
USA
T +1 408 790 8000

Europe
Citrix Systems International GmbH
Rheinweg 9
8200 Schaffhausen
Switzerland
T +41 52 635 7700

Asia Pacific
Citrix Systems Hong Kong Ltd.
Suite 3201, 32nd Floor
One International Finance Centre
1 Harbour View Street
Central
Hong Kong
T +852 2100 5000

Citrix Online division
6500 Hollister Avenue
Goleta, CA 93117
USA
T +1 805 690 6400

www.citrix.com

About Citrix

Citrix Systems, Inc. (Nasdaq:CTXS) is the global leader and the most trusted name in application delivery. More than 215,000 organizations worldwide rely on Citrix to deliver any application to users anywhere with the best performance, highest security and lowest cost. Citrix customers include 100 percent of the *Fortune* 100 companies and 99 percent of the *Fortune* Global 500, as well as hundreds of thousands of small businesses and prosumers. Citrix has approximately 8,000 partners in more than 100 countries. Annual revenue in 2007 was \$1.4 billion.

©2009 Citrix Systems, Inc. All rights reserved. Citrix®, Citrix XenApp™, Citrix XenServer™ and Xen® are trademarks of Citrix Systems, Inc. and/or one or more of its subsidiaries, and may be registered in the U.S. Patent and Trademark Office and in other countries. All other trademarks and registered trademarks are property of their respective owners.

0209/PDF

