
ACHIEVING
HIGH
AVAILABILITY
IN
UC

Contents

Introduction 3
UC & High Availability Defined 4
The Many Forms of UC 4
Why High Availability? 6
The True Cost of UC Downtime 7
Cost Categories 7
Calculating Costs 8
Outage Costs on the Rise 10
Availability vs. Reliability 11
Creating a Fault Tolerant UC Data Center 12
What Typically Fails? 13
The Active/Passive Cluster Server Problem and the Fault Tolerant Solution 14
Fault Tolerant Data Center Classification 15
JITC Certification: A UC Benchmark 16
Certification by the Joint Interoperability Test Command (JITC) 17
JITC Certification for Mission Critical UC Applications 17
UC Approaches 18
The UC Cloud Service Choice 18
The UC Hybrid Premises Cloud Approach 18
UC On-Premises Only Choice 19
Evaluating the Best Solution 20
Solution Comparison 21
Product Highlight 22



Introduction

This paper will provide insight into:

- Types of UC
- Reasons for High Availability
- True Data Center outage costs
- The differences between Availability and Reliability
- Creating a Fault Tolerant UC Data Center
- JITC UC Certification
- UC approaches
- Solution comparisons

Modernization has resulted in an increased reliance on Information Technology and Software systems. In most industries, performing daily activities require tools that are IT functional and data-centric. The ability of a given employee to do their job appropriately, i.e. access the system and submit or alter their work depends on the system's availability.

If the employee as a user cannot access the system, it is, from the user's point of view, unavailable, and unavailability (a.k.a. downtime) generally means loss of revenue. Learning the value and operation of High Availability in a UC environment is now critical to anyone working in the field of communications and collaboration.

You'll find many resources that discuss high-availability concepts and strategies, as well as the engineering details of high-availability solution components. What you're not likely to find, is a resource that shows you how to put those components together in a UC-enabled communications environment.

High availability delivers the maximum form of dependability a product or service can achieve. The mark of a highly available system means that it's accessible whenever it's needed with little or no interruption. Unified Communications (UC) is a capability that is becoming so integral to the success of SMBs and Enterprises, that its loss, even for a few minutes, can have a negative impact on daily operation—crippling employee productivity that's not only valuable to sales but also customer loyalty.

While having a backup system in place for your hardware and software systems is prudent, high availability for UC cannot be achieved through those means alone.

UC & High Availability Defined

"The essence of communication is breaking down barriers. In its simplest form, telephones, instant messaging, and conferencing break the barriers of distance and time so people can communicate in real time when they aren't together."

Since modern communications systems reside on networks consisting of many parts, all of which usually need to be present in order for the whole to be operational, much planning for high availability is needed. Availability can be measured relative to ***"100% operational"*** *or* ***"never failing"****. A widely held but difficult-to-achieve standard of availability for a system or product is known as* ***"five 9's"*** *(99.999 percent) availability.*

The Many Forms of UC

Unified communications is an industry term used to describe all forms of call and multimedia/cross-media message-management functions controlled by an individual user for both business and social purposes.

UC can be a single product or service or a combination of products and services that provide a consistent unified user-experience and interface across multiple devices and media-types.

UC encompasses many media and formats for exchanging information as well as storing, retrieving, forwarding, and consuming it.

What makes UC so unique, is that it allows a user to send a message over one medium and receive the responding communication on another. For example, a UC user can receive a voicemail message by accessing it directly from their desk phone, or through their email or cell phone. The receiver can then check the sender's presence—determine whether or not the sender is taking calls and can immediately send a response via text chat or voice/video call.

In information technology, high availability refers to a system or component that is continuously operational for a desirably long length of time.

If you think of UC as a communications and IT network, and the individual UC tools as connectors, then the Top 10 UC Connectors would be:

1. Traditional Voice Calls
2. Voice/Audio Conferencing
3. Video Calls & Conferencing
4. Instant Messaging (IM) & Chat
5. Presence Information
6. Interactive Whiteboards
7. Email
8. Fax
9. Data/Screen Sharing
10. Unified Messaging

UC & High Availability Defined

Why High Availability?

High availability refers to systems, devices, or components that are operational without interruption for an extended period. As mentioned previously on page 4, availability can be measured relative to "100% operational" or "never failing." Most legacy PBXs have achieved the "five 9's" and have an availability of 99.999%. 99.999% availability means that your business should only experience 5 minutes and 15 seconds of downtime in one year of continuous operation.

To prepare for High Availability, you must include planning for backup and failover processing, data storage, and access.

UC justification is typically centered on UC's main benefits, which include: better revenue management and increased profitability and customer satisfaction. If you eliminate communications and collaboration, all of these values disappear. The loss of UC, even for a few hours, can be a substantial hit monetarily speaking to an enterprise. Besides the costs involved, there is also a reputation aspect. Downtime means customer unrest, possible customer loss, and that leads to reputation damage. Regaining reputation, once harmed, is a long and costly process. Therefore ensuring the high availability of UC becomes a paramount enterprise goal.

AVAILABILITY	ANNUAL DOWNTIME
99.9999%	32 sec
99.999%	5 min, 15 sec
99.99%	52 min, 34 sec
99.9%	8 hrs, 46 min
99%	3 days, 15 hrs, 36 min

Most legacy PBXs have achieved the "five 9's" and have an availability of 99.999%.

99.999% availability means that your business should only experience 5 minutes and 15 seconds of downtime in one year of continuous operation.

> **The loss of UC, even for a few hours, can be a substantial hit monetarily speaking to an enterprise.**

1 Ponemon Institute, LLC. *http://www.ponemon.org/.*

The True Cost of UC Downtime

Enterprises are in business to generate revenue and profit. An outage will almost always impact revenue. This includes revenue lost during the outage as well as lost potential revenue because customers avoid doing business with an enterprise in midst of a crisis or failure.

Imagine that you have lost your data center. All of your applications, including Unified Communications, are down. The cost to respond to and resolve the outage is high.

Your customers may not return, and you will have to rebuild your enterprise's reputation. In some cases, the enterprise will have to invest considerably in costly marketing efforts to regain lost customers.

Cost Categories

The costs incurred due to an outage typically fall into three categories:

1 Direct

Direct costs can be determined by the outlay of cash to the organizations who fix the outage—such as vendors, providers, and consultants. These costs are typically much easier to calculate, because you have all the bills available to prove them.

2 Indirect

The second source of costs is more difficult to calculate. These are the ***indirect costs***, which are comprised of the time or overtime IT staff and others spend focusing on the outage and its resolution, rather than their normal work. There is always a cost involved when other work that IT should be performing is delayed.

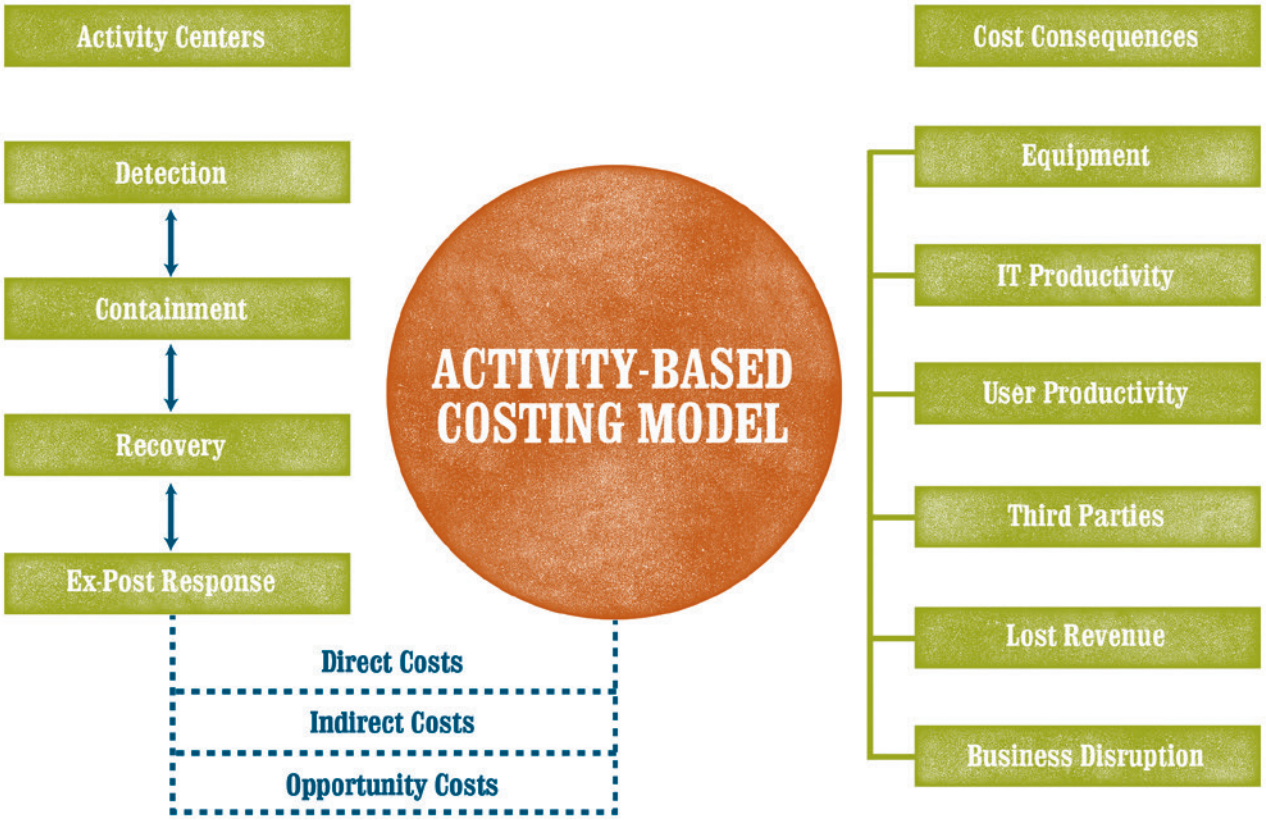
Be careful to not underestimate indirect costs. Productivity loss is expensive. The greater the costs incurred during the outage, the more likely it will be that IT can justify investments that will mitigate or prevent another outage in the future.

3 Opportunity

The third cost area to consider is ***opportunity costs***. These costs represent the lost business, lost reputation and lost market share the company experiences during the outage. The marketing department would be the best source for finding out what these costs are. These costs are also difficult to determine and take time calculating. It may be months before the outage's opportunity costs become apparent.

The True Cost of UC Downtime

Figure 1: Activity-based cost account framework



Calculating Costs

The Cost of Data Center Outages[2] report discusses in depth how to calculate outage costs. Basically, the total cost of an outage is the sum of the activity the outage causes and the resulting consequences of the outage/those activities (see Figure 1).

2 Ponemon Institute, LLC. *2013 Cost of Data Center Outages*.

The True Cost of UC Downtime

The following list comprises the specific cost categories that can be factored into the total outage cost. Sometimes IT departments can miss/forget to factor these in and thereby underestimate the total outage cost.

1 Detection

The cost of monitoring and discovering if, when, and how much of an outage has occurred. This also includes costs incurred when fully researching the extent of the outage and its effect on the enterprise.

2 Containment

Outages can have a domino effect. One outage may start another outage or cause the existing outage to worsen. There is a cost to limiting the outage influence.

3 Recovery

The effort to restore the systems and networks back into acceptable operation.

4 Post Recovery

There may be other incidental costs such as informing users and customers about the outage, as well as a plan/strategy to prevent the outage in the future.

5 New or Replacement Hardware/Software

Costs related to hardware failure or software upgrades/replacement.

6 IT staff

Considerable IT labor may be expended during and after the outage. This labor expenditure also interferes with ongoing projects. Delayed projects may have a cost impact as well.

7 User productivity loss

End user downtime and the overtime needed to make up for the downtime should be calculated.

8 Third party expenses

These expenses are the result of engaging outside services to review the outage and recommend resolutions. There may be reports for regulated industries that need to be produced by independent third parties.

9 Lost revenues

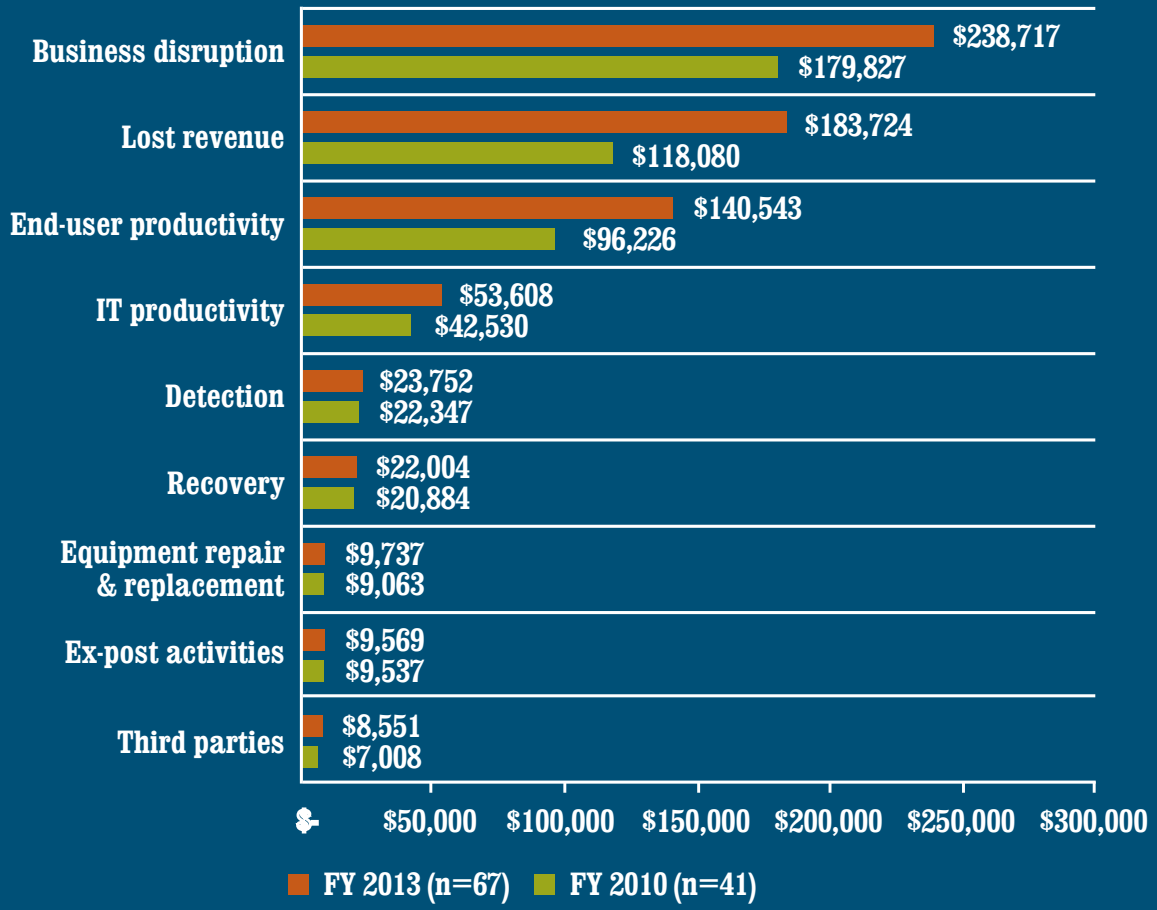
The total revenue loss from customers and potential customers because of their inability to access core systems during the outage period.

10 Business disruption

The total economic loss of the outage including reputational damages, customer churn, and lost business opportunities.

The True Cost of UC Downtime

Figure 2. Comparison of FY 2010 & FY 2013 activity cost categories
from the Ponemon report



Outage Costs on the Rise

The Cost of Data Center Outages report also provides a wealth of information about the real nature of several outage costs that occurred 2010 to 2013.

Based on nine of the 10 specific cost categories listed previously, the report infers the cost of an unplanned data center outage. The costs of data center outages have increased steadily across every category from 2010 to 2013.

According to the report, there is a "significant variation across nine cost categories for FY 2010 and FY 2013. The cost associated with business disruption, which includes reputation damages and customer churn, represents the most expensive cost category."

Availability vs. Reliability

When we talk about "five 9's," we're talking more about availability than reliability, although the latter is integral to the former.

Availability is a function of two basic factors: the Mean Time Between Failures (MTBF) and Mean Time to Repair (MTTR). Both are usually measured in hours. Availability is described by the following equation:

$$\text{Availability} = \frac{\text{MTBF}}{\text{MTBF} + \text{MTTR}} \times 100 = 9x.xxxxx\%$$

It turns out, however, that MTTR is really the mean time to restore rather than to repair, and it includes the following five activities (note that all five can be significantly reduced or eliminated by redundant components and automatic reconfiguration):

- *Failure Detection*
- *Failure Notification*
- *Vendor/User Response*
- *Repair/Replacement*
- *Recovery/Restart/Reboot*

MTBF provides a measure of a system's reliability. But over the course of a system's life, the metric doesn't necessarily denote physical reliability. Your system could have 99 percent availability and still suffer a disaster—one huge outage—or hundreds of short outages. But while the metrics are indifferent to the impact of an outage(s), they still provide a useful function. They provide a frame of reference. It does not accurately predict whether there is one long outage or multiple short outages. The outage length can usually be determined by experience with the hardware and software configurations in use. The outage behavior generally improves as the IT staff learns how to respond to outages.

Fault Tolerant systems offer the "end of the rainbow" solution to data center outages.



Your system could have 99 percent availability and still suffer a disaster.



Creating a Fault Tolerant UC Data Center

Fault-tolerant describes a computer system or component that is designed so that, in the event in which a component fails, a backup component or procedure can immediately take its place without interrupting the service.

A Fault Tolerant UC system is one that can continue to operate while sustaining failures of individual components without disrupting the users. Fault Tolerance can be provided with the appropriate hardware, software, or some configuration of both.

Fault Tolerance hardware is produced by designing two of each element into the system. Take Disk Mirroring as an example. Disk mirroring replicates data onto two or more disks. So in a UC environment, multiple processors are synchronized (in lockstep) and are processing the same data simultaneously. Typically the results are compared for accuracy. But when a problem occurs, the faulty element can be determined and removed from service, while the mirrored element takes on the full workload. As a result, the system continues to function as usual.

The goals of fault tolerant systems include:

Non-stop operation

Employing redundant hardware for continuous operation helps protect against component failures.

Non-disruptive maintenance

Hot-swappable components help enable replacement modules without interrupting operation.

General operating systems

General operating systems (including Windows® / Linux® / VMware®) help deliver the same operability as the most widely-used servers.

Creating a Fault Tolerant UC Data Center

What Typically Fails?

While any server component can malfunction during system operation, there are three candidates that typically cause most failures. These are in order of most likely to least likely:

Disk Drives

These types of mechanical systems tend to malfunction as a result of becoming "worn out."

Power Supplies

These electrical components tend to age and are often prone to heat-induced failures.

Memory

An electrical component that can sometimes be forgotten. A single memory error can remove the unit's entire memory from operation.

In Fault Tolerant equipment, use of "dual" components can help minimize and mitigate the failure of standalone components. Solving the disk drive failure issue is relatively simple. Use of dual disk drives with synchronized copies of data, can offset the failure of one single disk drive.

There are multiple ways to include Dual Disk drives in a server environment. One solution is to employ a Redundant Array of Independent Disks (RAID). Another good approach is to use a storage area network (SAN).

The same can be said for the power supply issue. Implementing dual power supplies that both operate simultaneously can solve failure problems, so long as both independent power supplies are rated to carry the entire power load.

Preventing memory failure proves to be a bit different. When there is memory failure, the memory component needs to be physically replaced to continue operation.

The common condition that causes the most failures across all three components is fluctuation in the electrical power used to run the hardware. Surge protectors can help forestall the fluctuation issue, but isolating the hardware through the use of an Uninterruptable Power System (UPS) is usually the best solution.

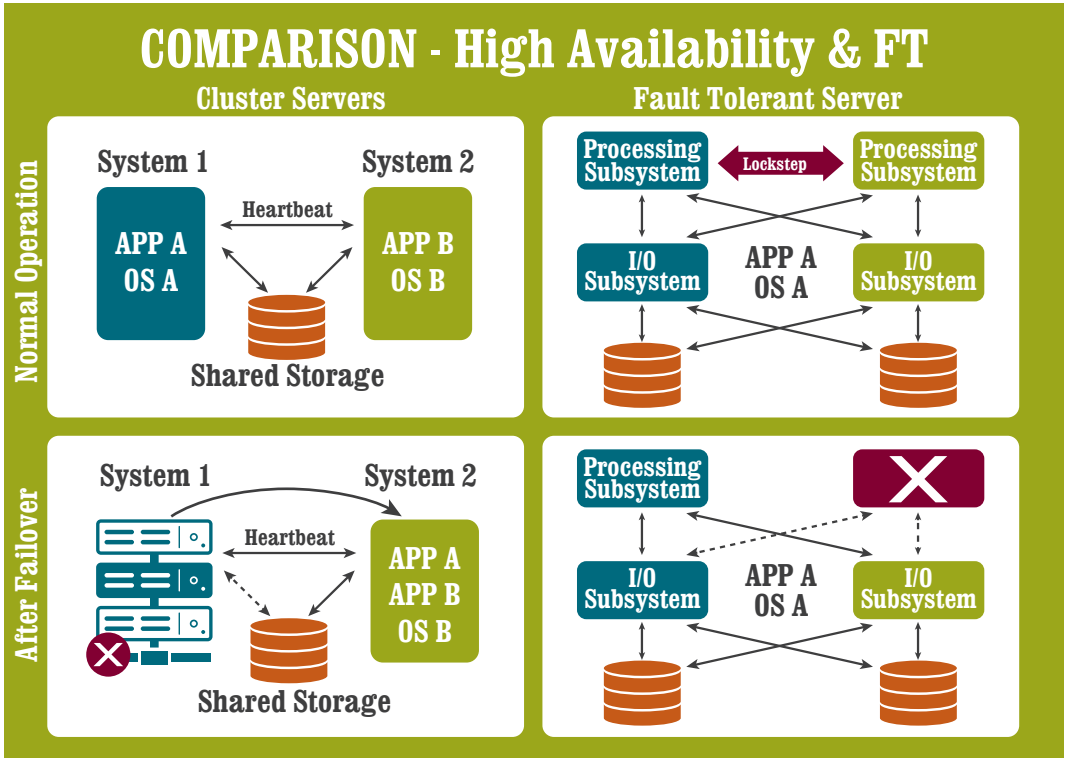
Creating a Fault Tolerant UC Data Center

The Active/Passive Cluster Server Problem and the Fault Tolerant Solution

In an Active/Passive Load Balancing configuration, there are two hardware systems that support the operating load. The backup system does not overshadow the operating system, but rather operates in listening mode, monitoring the performance of the primary load balancer.

There is a heartbeat (keep alive) function operating between the primary and backup systems. When a failure occurs in the primary system, it is taken off line and the backup takes over to support the UC service. This causes an interruption to service that can last minutes to hours. Data may be lost and the stored records may be inaccurate, which is a headache for systems that drive data specific business processes—i.e. financial or database/record-based programs.

The Fault Tolerant version has a storage area network between the primary and backup systems (SAN). In this configuration, there will still be an interruption of service while the backup system takes over operation, but in this instance, the backup server will have access to all the data stored by the primary system so there should be no data inaccuracies.



Creating a Fault Tolerant UC Data Center

Fault Tolerant Data Center Classification

The Uptime Institute created the standard Tier Classification System[3] as a means to effectively evaluate and provide business requirements for data center infrastructure system availability. The Tier Classification System provides the data center industry with a consistent method to compare unique, customized facilities based on expected site infrastructure, performance, and/or uptime. Classifications range from the least robust at Tier I to the most robust at Tier IV.

Organizations selecting a Tier III infrastructure typically have high-availability requirements for ongoing business or have identified a significant cost disruption due to a planned data center shutdown. Tier IV site infrastructure builds on Tier III, adding the concept of Fault Tolerance to the site infrastructure topology. With Fault Tolerance built into the infrastructure, if/when individual equipment fails or a distribution path interruption occurs, the effects of these events are stopped short of interrupting IT operations.

Organizations that have high-availability requirements for ongoing business (or mission) imperatives, or that experience a profound impact of disruption due to a data center shutdown select Tier IV site infrastructure. Tier IV is justified most often for organizations delivering "24xForever" services. As UC continues to become a critical business/mission component, it will need to be supported by Tier IV data centers.

"As UC continues to become a critical business/mission component, it will need to be supported by Tier IV data centers."

3 The Uptime Institute. *Tier Certification*.

JITC Certification: A UC Benchmark

Hardware is usually referred to when discussing availability. But a case can be made to test and evaluate software for reliability and availability as well. Software testing requires that an independent third party evaluate the software solution. There are organizations that can perform software testing, but they usually do so at the behest of and to acquire payment from a software vendor. Vendor sponsorship does not automatically make the test results questionable, but it should, however, prompt you to question the criteria used, and to ask what the parameters of testing were. And, of course, the test results should adhere to known and accepted standards for your industry.

The Federal government is one of the third-parties that has a vested interest in evaluating a wide-range of products and services. Looking to the federal government for help with hardware and software evaluations might not be a bad idea. The Federal government has some of the most stringent requirements and criterion; many of which non-governmental testing organizations do not typically cover.

The government wants to ensure that the software testing is applicable to all its agencies. The testing is easily applicable to non-government organizations with similar requirements.

“Looking to the federal government for help with hardware and software evaluations might not be a bad idea.”

JITC Certification: A UC Benchmark

Certification by the Joint Interoperability Test Command (JITC)

The Federal government has created the Joint Interoperability Test Command (JITC)[4]. The Joint Interoperability Test Command (JITC) is a U.S. military organization that tests technology that pertains to multiple branches of the armed services and federal government. JITC's mission is to test and evaluate products and services that advance global net-centric testing in support of warfighting capabilities. To do this, JITC must provide a full range of rapid, standardized, and customized test, evaluation, and certification services to support global net-centric warfighting capabilities under all conditions of peace and war.

The JITC has produced a valuable set of benchmarks for computer and communications platforms which can be used to evaluate existing products and services for both the government and non-government organizations. Any enterprise should consider JITC certification as an unbiased and independent evaluation for comparing high availability UC solutions.

JITC Certification for Mission Critical UC Applications

JITC certification extends beyond the Department of Defense (DoD) into a wide array of enterprise opportunities in regulated industries, publicly funded organizations, and Federal, State, local governments/agencies, and any other organization demanding security, reliability, high availability, UC functionality, and standards-based interoperability.

How to spend money on an IT solution can be a contentious subject. Chief Financial Officers (CFOs) have to consider multiple factors to determine expenditures every year, i.e. how market conditions and the costs of implementing a new IT solution might bring a host of changes, both good or bad.

- **UNIVERGE 3C - Certified added to the Approved Products List (APL) for US Department of Defense Sales**
- **Unified Capabilities Requirements (UCR) - "Gold Standard" for Public Sector and Defense Organizations across the Globe**
- **The UCR defines a Scalable, Interoperable & Global all-IP & IT-centric UC&C Solution**
- **Verified by Extensive Tests on Security, Reliability, High Availability, UC Functionality and Standards Based Interoperability**

4 *Joint Interoperability Test Command*.

UC Approaches

FUNCTION	OWN ON PREMISES	UC as a SERVICE (UCaaS)
Customer provided physical facilities	YES	NOT OFFERED
Customer provided equipment & software	YES	NOT OFFERED
On-premises networking	YES	NOT OFFERED
WAN connection to data center	YES	MAY BE OFFERED
Storage	YES	YES
Servers	YES	YES
Virtualization	YES	YES
Operating system	YES	YES
Middleware	YES	YES
Customer program execution control	YES	YES
Data	YES	YES
Applications	YES	YES

Who is Responsible?
(YES = it is Offered)

The UC Cloud Service Choice

IT and communications budget restraints can make an on-premises UC solution too "capital" intensive. Many enterprise CFOs are shifting IT investments from a capital expenditures based model to an operating expenditures based model. A cloud solution that is expensed (OPEX) with little or no capital expenditure (CAPEX) may be desired. A cloud solution offers a fixed cost per month based on the number of users and the features used.

Cloud UC can be subscribed to by feature group allowing greater flexibility for the enterprise when determining what UC features should be offered to what users. Most enterprises have implemented a few UC features and are observing and monitoring their use to determine what are the feature benefits and their ROI.

The cloud service approach requires less IT staff. It does require, however, more performance management and monitoring of the cloud service provider.

The UC Hybrid Premises Cloud Approach

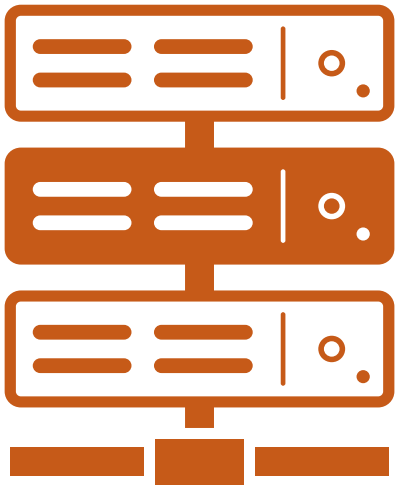
A hybrid solution is a computing service that is composed of some combination of on-premises and cloud solution. A hybrid cloud solution breaks down the isolation and provider borders.

The hybrid solution allows the enterprise to maximize the features that are mandatory for the entire organization, while using cloud based features for those in the organization that need them. With the combination of your current on-premises technology and cloud, the expense of procuring new on-premises features for departments that do not need them is eliminated. And, the enterprise can always move functions from the cloud to a full premises-based system if the cost of cloud service begins to exceed the cost of on-site implementation. The hybrid approach can also deliver business continuity failover services at a much lower cost. The enterprise does not have to have a new or expanded on-premises backup system—rather all of these functions can be moved to the cloud.

UC Approaches

UC On-Premises Only Choice

The traditional approach to communications for most businesses has been to purchase and operate a system on their premises. This continues to be true even with the advent of competing cloud services. The business that wants to have an on-premises system usually has one or more reasons for this decision.



- *The business is capital intensive so an OPEX approach does fit into its financial schemes.*
- *The business already owns an on-premises system and is satisfied that it meets their requirements.*
- *Signing a cloud contract can produce provider lock-in which the business wants to avoid.*
- *There is customization required that is not offered by the cloud service.*
- *The security and control requirements by the business do not allow outsourcing to a cloud provider.*
- *The cloud Service Level Agreement (SLA) does not satisfy the business requirements nor do the financial credits offered by the cloud provider offer any substantial monetary payment to offset the cost of an outage.*
- *The business has significant seasonal traffic fluctuations where the cloud contract cannot be easily reduced without financial penalties.*

Evaluating the Best Solution



There are multiple reasons to adopt any one of the UC approaches listed on the previous 2 pages. One of the most common reasons for not adopting a cloud service is security. This is especially true for regulated vertical markets like healthcare, financial services, and some government operations. If you own your IT equipment, you typically feel more secure than if the responsibility for security was under the control of a third party. It should be mentioned though, that regardless of approach, the enterprise is still the final responsible party when it comes to security.

Another common decision-making factor when choosing IT equipment is control. If the enterprise looks at UCaaS as the best communications solution, then it means that they are willing to place most of the control into the hands of the service provider.

A third common concern is provider lock-in. Once the enterprise changes its IT staff to support cloud services, it is a considerable effort to return those cloud-based services back into an on-premises solution/network. Often, hybrid solutions allow the enterprise to select what functions to implement in the cloud and what to retain on-premises balancing cost, control, and security.

All enterprises should look at cloud services though, even if the final decision is not to use them. If a cloud service looks attractive, then the enterprise should consider the least mission critical functions as candidates for the cloud. If the cloud is not successful, the impact of moving out of the cloud will be minimal.

Solution Comparison

The on-premises solution has potentially better security and does not require an Internet connection. The cloud service solution can be cheaper, and require less IT staff time, but will require greater Internet bandwidth and have greater security issues. Both solutions (on-premises and cloud) are equal when considering endpoints like phones and video devices and on-site LAN operation.

What is important are the financial, technical, facility, and staff support all of which can force the implementation decision in a particular direction. The balance of all the decision factors and how to weight their values is up to the enterprise.

Unified Communications is all about multi-media communications and collaboration. There is no single right or wrong approach to implementing UC for any particular organization size or market. What is important is that the UC function delivers high availability.

> **UC is the key to staying competitive and agile in a global economy no matter what size the business.**

Product Highlight

NEC offers High Availability and Fault Tolerant Solutions that work together to produce a fully comprehensive UC&C communications system that can be deployed on-premises, hybrid, or through the cloud. When NEC's UNIVERGE 3C unified communications platform is coupled with their Express5800 Series Fault Tolerant servers, businesses can be assured of continuous high availability and disaster recovery with unmatched security, scalability, interoperability and five nines (99.999%) availability. To learn more NEC's solutions, please visit:

www.necam.com.

About the Author

Gary Audin *has more than 40+ years of computer, communications and security consulting and implementation experience. He has planned, designed, specified, implemented, and operated data, LAN, and telephone networks. These have included local area, national and international networks as well as VoIP and IP convergent networks in the U.S., Canada, Europe, Australia, Caribbean, and Asia. He has also produced over 1000 blogs, articles, podcasts, white papers and webinars.*