# What is Your Business Continuity Plan for the Inevitable?

**A Gravic, Inc. White Paper**

# What is Your Business Continuity Plan for the Inevitable?

## Executive Summary

The many events that can lead to application outages are not rare; it is a case of when, not if.

> *"96% of global IT decision-makers have experienced at least one outage in the past three years."*[1]

This problem is compounded by critical business IT systems' costs of prolonged downtime; these costs are often *significantly more* than companies perceive (and can even shutter some companies).

> *"Companies with frequent outages and brownouts experience 16 times higher costs than companies with fewer [outages and brownouts]."*[1]

This problem is mitigated with a complete, documented, and well-tested business continuity plan. Maybe you are under the impression that you already have one. Unfortunately, the data does not support this thinking, resulting in many companies operating with the mistaken belief that their business continuity plans will save them from an outage.

The first step is to make sure that this way of thinking does not apply to your company! You need to ensure that your business continuity plan completely documents all facets of the recovery process, and then (of course) regularly exercise it in practice.

In order to complete a thorough business continuity plan, the correct IT architecture must be selected to maintain services during a planned or unplanned outage.[2] This paper describes several architectures, from classic active/passive, through sizzling-hot-takeover (SZT), to active/active, using HPE Shadowbase[3] asynchronous and synchronous data replication technologies.

Many teams never evolve beyond a basic active/passive architecture. Unfortunately, this architecture has many issues that can prevent a successful and timely failover. It is difficult to test and suffers from failover faults. Consequently, this model is reactive, risky, and provides a false sense of security. While adding marginal complexity and expense, the more sophisticated business continuity architectures (SZT and active/active) are actually far more cost-effective when considering the Total Cost of Ownership (TCO). The chances are that if you use an active/passive architecture, one outage during peak processing hours will teach you this lesson the hard way.

Asynchronous SZT and active/active configurations are far superior to active/passive, and should be the minimum solution chosen. Why? Because we believe all business environments can run at least at SZT if not fully active/active. However, when implemented with an asynchronous technology, these configurations can also have issues, including some data loss (in both cases, although active/active is at least 50% less than SZT), and possible data collisions (active/active only). Perhaps for your applications, one of these architectures is a perfect fit. Nevertheless, sometimes even SZT and active/active architectures do not provide enough data protection. For example, applications where no data loss is acceptable (or applications which must be 100% available) cannot employ an active/active architecture because data collisions cannot be tolerated or avoided. For these applications, HPE Shadowbase synchronous replication offers a solution, eliminating both data loss and data collisions.[4]

The solution is in your hands. The attention-grabbing headlines never need to apply to your company. Make sure that your business continuity plan is based on more than just faith, and that you know for a fact that it works because you have exercised it regularly.

---

[1] *Research Note: NonStop TCO Comparison*: LogicMonitor survey
[2] For more information, please see the Gravic Shadowbase white paper, *Choosing a Business Continuity Solution to Match Your Business Availability Requirements*.
[3] Please visit ShadowbaseSoftware.com.
[4] Please contact Gravic for the future availability of synchronous replication on various platforms and database environments.

## Table of Contents

**Executive Summary** ..... **2**

**The Costs and Risks of Extended Downtime** ..... **5**

**The Fallacy: "HPE NonStop Servers are too Expensive"** ..... **9**

*Do Not Confuse Price with Value* ..... *9*

*How Does this Example Relate to Enterprise-Grade Servers?* ..... *10*

*Maintenance* ..... *10*

**Is *Your* Business Paying Enough Attention?** ..... **10**

*Fault Tolerance* ..... *11*

**Business Continuity Architectures: Pros and Cons** ..... **11**

*The Public Cloud is not Fault Tolerant Enough* ..... *12*

*Too good to be true?* ..... *12*

*Asynchronous Active/Passive – Classic Disaster Recovery* ..... *12*

*Asynchronous Active/Almost-Active – Sizzling-Hot-Takeover (SZT)* ..... *13*

*Asynchronous Active/Active – Partitioned* ..... *14*

*Asynchronous Active/Active – Route Anywhere* ..... *14*

**New Business Continuity Technologies – Synchronous Replication** ..... **19**

**Scalability** ..... **21**

**Summary of Business Continuity Architectures and Technologies** ..... **22**

**In Summary – Match Your Business Continuity Solution to Your Business Risk** ..... **22**

**International Partner Information** ..... **24**

**Gravic, Inc. Contact Information** ..... **24**

## Table of Figures

Figure 1 – Frequency and Causes of IT Outages[4] ..... 5
Figure 2 – Have You Experienced an Unplanned Datacenter Outage in the Past 3 Years? ..... 6
Figure 3 – Percentages of Complete Datacenter Outages by Industry Segment During 2023 ..... 6
Figure 4 – Duration of Complete Datacenter Outages by Hours ..... 7
Figure 5 – Average Cost per Hour of Downtime for Various Industries ..... 7
Figure 6 –Total Outage Cost is Only Increasing ..... 8
Figure 7 – Who Says There Is No Such Thing as Bad Publicity? ..... 9
Figure 8 – Perception vs Reality Regarding IT Failure Preparedness ..... 11
Figure 9 – How Prepared Are You, Really?[8] ..... 11
Figure 10 – RPO and RTO for the Various Business Continuity Replication Architectures ..... 15
Figure 11 – Estimated Service Unavailability Costs for a Financial Application ..... 15
Figure 12 – Estimated Service Unavailability Costs for a Financial Application ..... 16
Figure 13 – Average Cost per Transaction of Lost Data for Various Industries ..... 17
Figure 14 – TCO vs Complexity/Cost of Implementation of Various Business Continuity Architectures ..... 18
Figure 15 – Estimated Outage Data Loss Costs for Various Industries by Business Continuity Technology .20
Figure 16 – Estimated Costs of Lost Data for Average Stock Trade Transaction ..... 20
Figure 17 – RPO and RTO for Asynchronous vs Synchronous Replication ..... 21
Figure 18 – Business Continuity Architecture and Technology – Pros and Cons ..... 22

# What is Your Business Continuity Plan for the Inevitable?

## The Costs and Risks of Extended Downtime

Like it or not, things happen. The occurrences of events that threaten the continued availability of IT services are not infrequent. Figure 1 shows the types of events that have caused extended IT outages in the past three previous years.

For the unprepared, such incidents will take down systems, and the online business processes of the company, both internal and external, will come to a halt. You may think that such a circumstance could never happen to you because you are well prepared, but the statistics tell a different story.

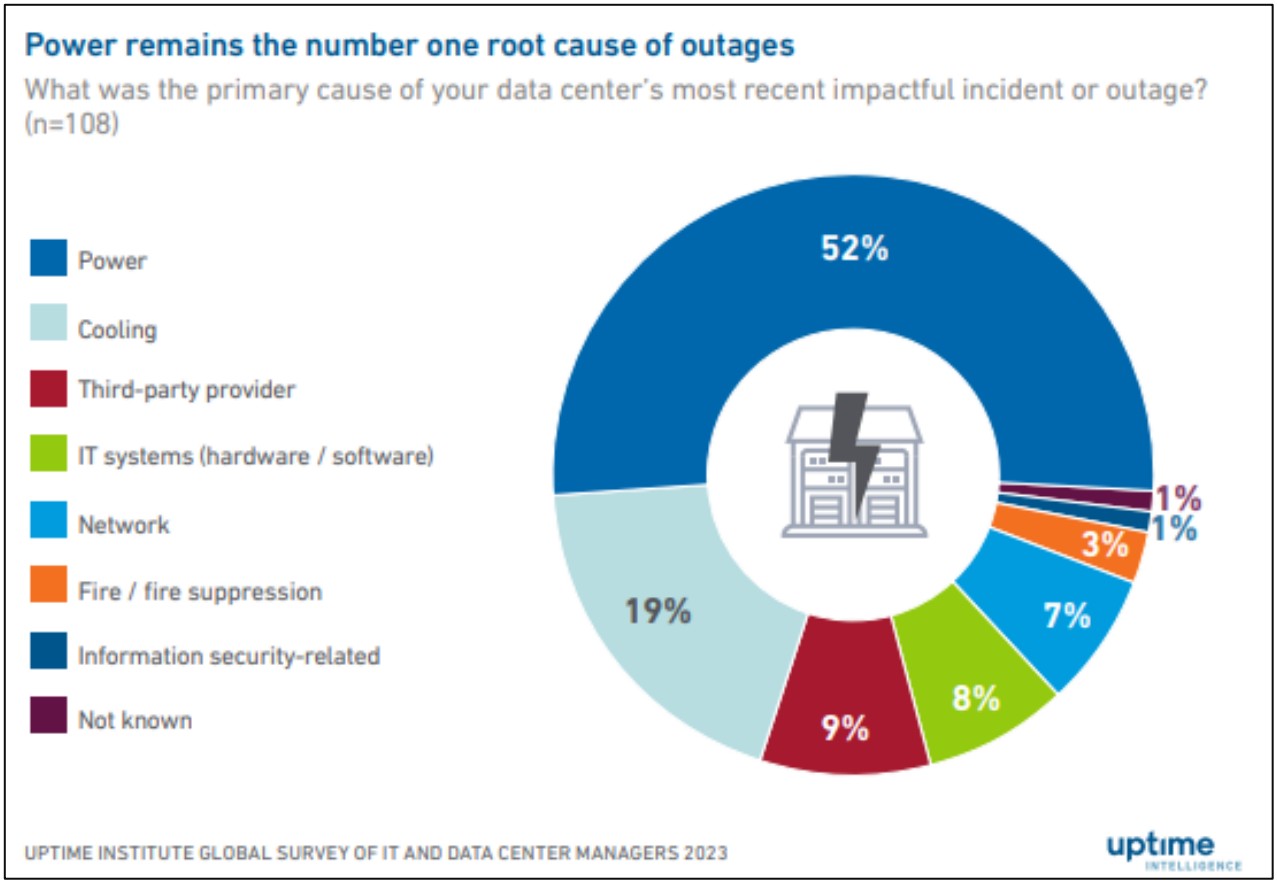



**Figure 1 – Frequency and Causes of IT Outages**[4]

Figure 2 shows the incidence of unplanned datacenter outages across a range of industries during the three previous years: 55% of respondents indicated they had experienced an unplanned datacenter outage. These outages range from negligible to severe, with 27% of reported incidents being significant, serious, or severe. While cases of cyber incidents are rising, four out of five of those issues were preventable.

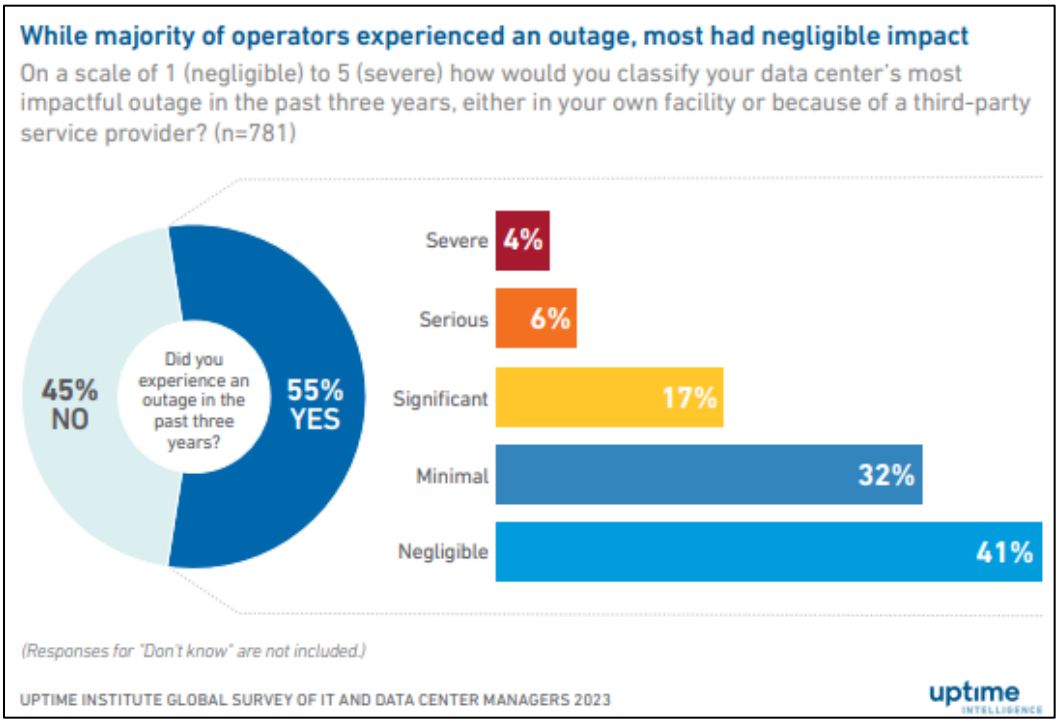



**Figure 2 – Have You Experienced an Unplanned Datacenter Outage in the Past 3 Years?**

Note, these outages are not limited to non-mission critical types of applications, where they might be less impactful. Figure 3 and Figure 4 show the number of complete datacenter outages and their duration by industry segment[4] over a 24-month period. Industries such as Financial and Healthcare, where one would presume that continuous availability of IT systems is absolutely critical, show an outage incidence of almost two and three, with average periods of 73 and 122 minutes, respectively. Note, these outages are *complete datacenter outages*, not just individual systems!

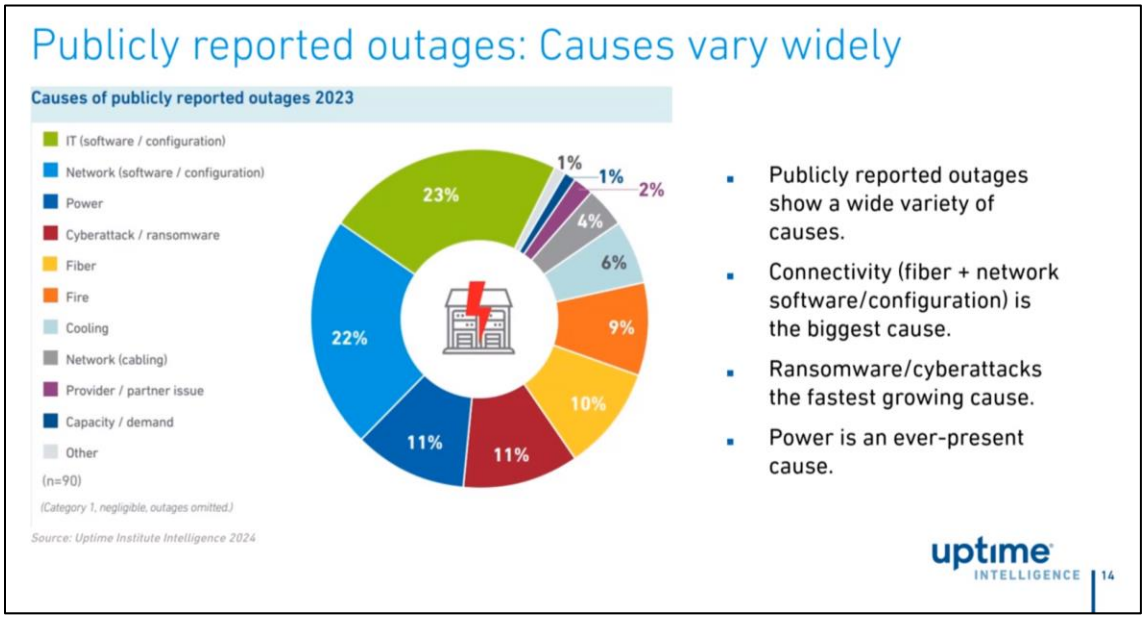



**Figure 3 – Percentages of Complete Datacenter Outages by Industry Segment During 2023**

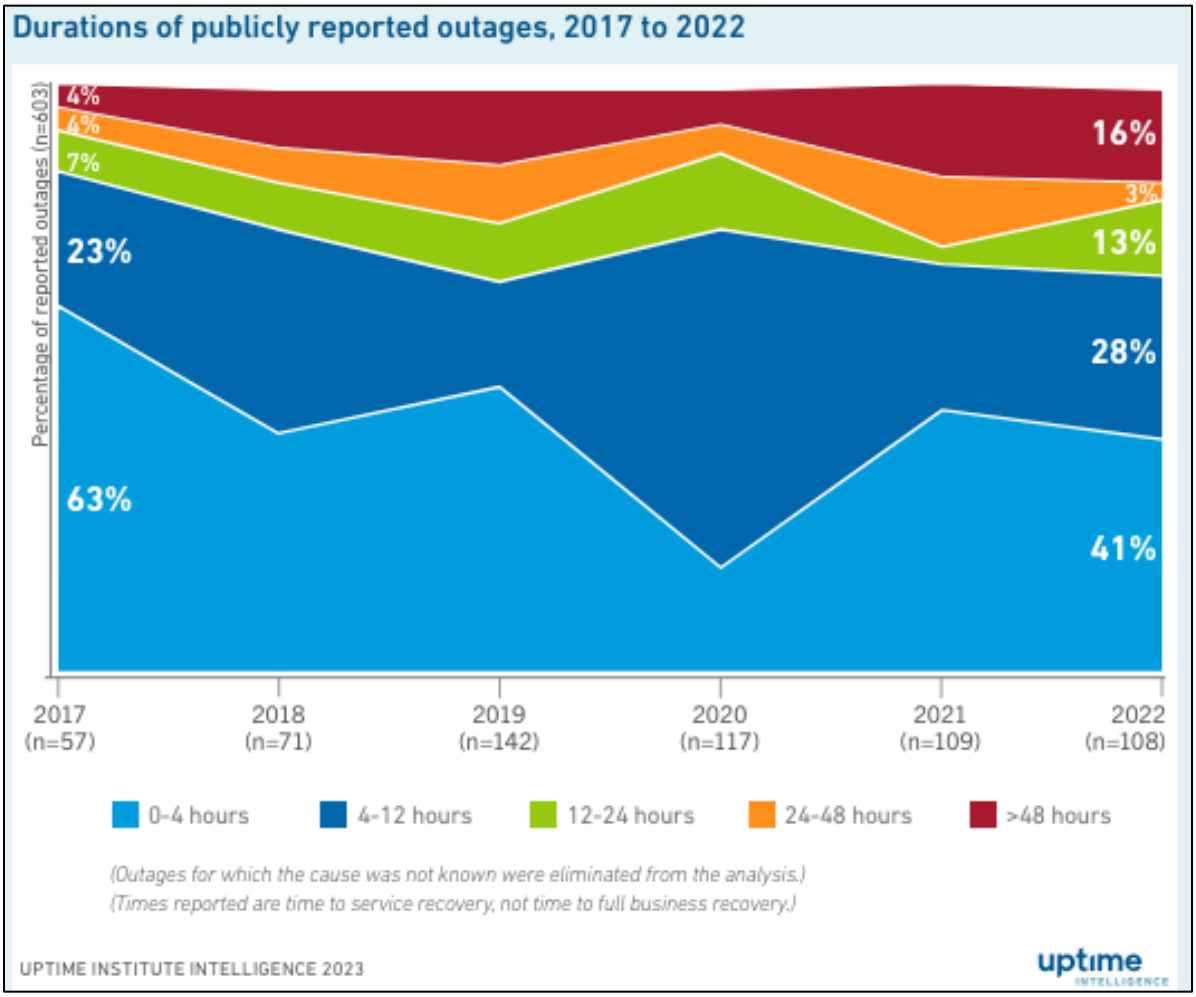



**Figure 4 – Duration of Complete Datacenter Outages by Hours**

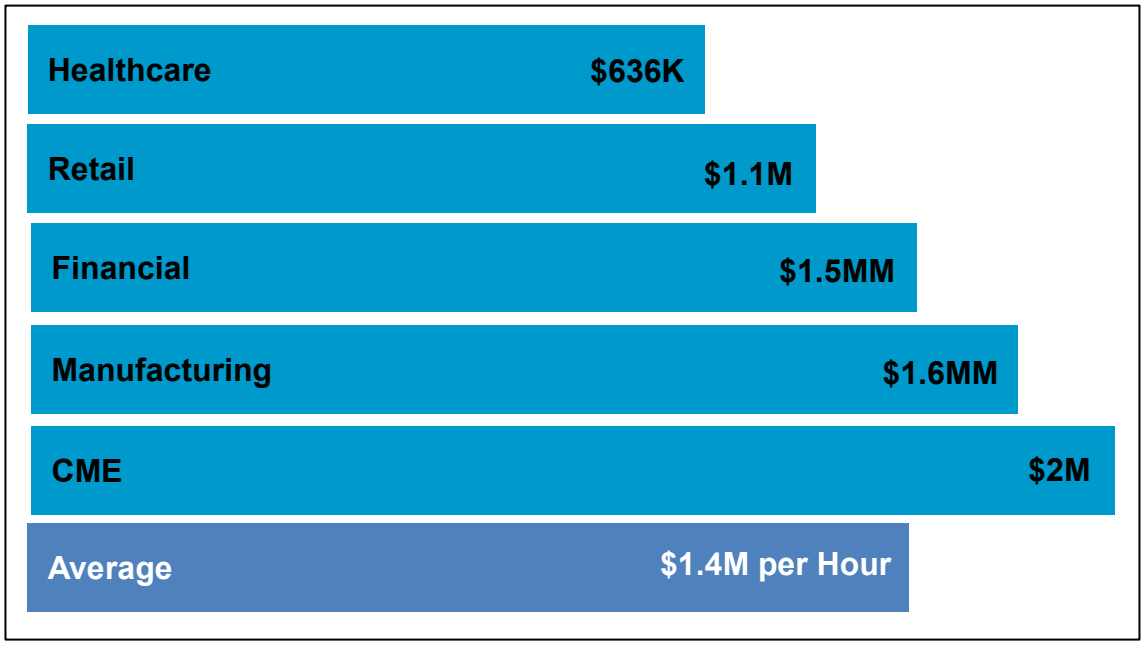



**Figure 5 – Average Cost per Hour of Downtime for Various Industries**

(Sources: Network Computing, the Meta Group, Contingency Planning Research)

So, you have an outage, so what? For businesses that depend upon IT to provide services to their customers, or to perform in-house processes (which today is almost every business), unavailability of IT systems will incur a cost. That cost will vary depending upon the nature of the business and the duration of the outage. Some average costs per hour of downtime across various industries are shown in Figure 5, and total unplanned outage costs across a range of industries are shown in Figure 6.

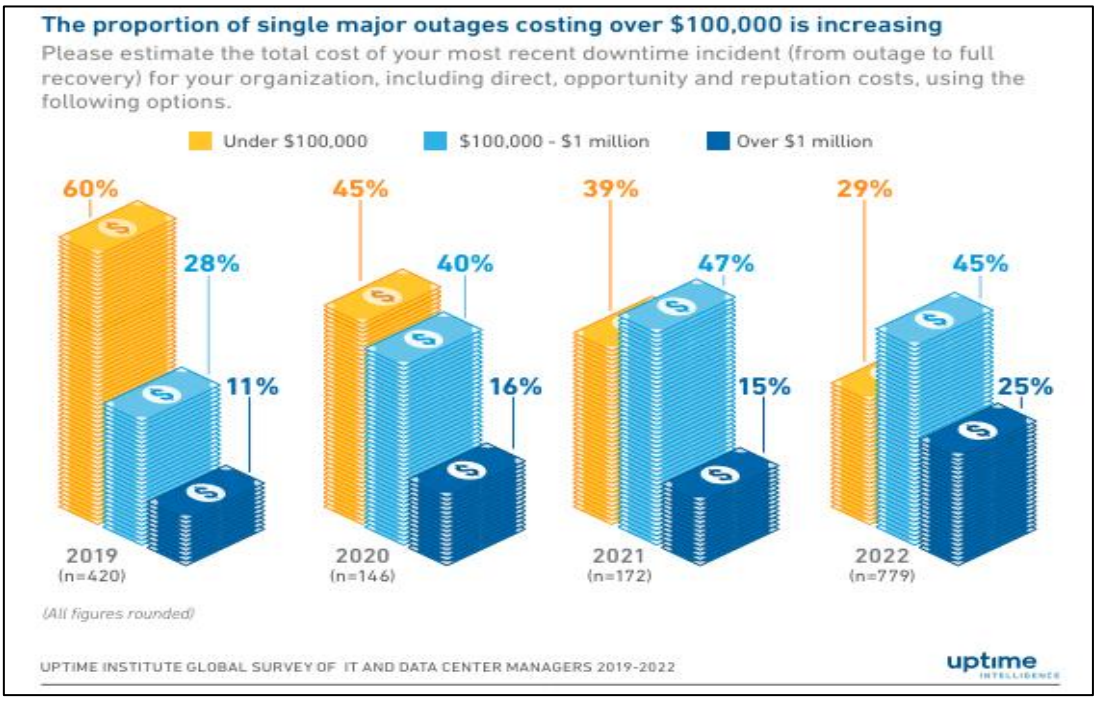



**Figure 6 –Total Outage Cost is Only Increasing**

Obviously, these monetary costs are non-trivial, to say the least. In fact, 46% of these outages cost more than 100k, 38% up to one million, and 16% over one million. The 73-minute outage for the financial sector costs an average of $1.8M. But, beyond the immediate monetary costs, as bad as they are, there are further hidden ramifications which could have even worse consequences for the company:

- Threats to human health and safety (e.g., for healthcare providers and emergency services)
- Legal, regulatory, and contractual compliance exposure
- Loss of customer/partner/supplier confidence and loyalty
- Reduced employee productivity
- Negative publicity impacts brand integrity and corporate reputation
- Missed commitments and associated liabilities and penalties

Unfortunately, 15-20% of high-profile organizations will suffer a major outage with financial, reputational and other consequences. And no CEO wants to see these kinds of headlines about their company all over the news (Figure 7).


**Figure 7 – Who Says There Is No Such Thing as Bad Publicity?**

*"In an updated report, IDC provided the following: The cost of downtime is increasing as businesses become more and more dependent on their infrastructure for daily operations. For 20.7% of organizations, the cost of downtime is $5,000-$10,000 per hour; for 18.4%, it is $10,000-$25,000 per hour; for 17%, it is $25,000-$100,000 per hour; and for some businesses (1.4%), it is $500,000.[8]"*

But wait, it gets worse! The results from extended outages can be severe enough to actually put a company out of business:

- 93% of companies that suffer a significant data loss are out of business within five years.[5]
- 60% of companies that suffer a disaster and have no recovery plan are out of business within three years.[6]
- 40% of companies without access to data for 24 hours go out of business.[7]

## The Fallacy: "HPE NonStop Servers are too Expensive"

***Do Not Confuse Price with Value***

*A local hardware store that claims it is the "lowest price provider" displays a large sign: "If you buy your tools anywhere else, you're throwing away your money." Any tool that my friends or I have ever bought at the store worked well – until it broke after the first use or turned out to be defective. This is ironic. Warren Buffett once said, "Don't confuse price with value." Unfortunately, this mantra applies here. I bought a trailer kit for $200 one year. "What a great deal! I'll save so much money!" After years of difficulty and conflict that included 20 hours of assembly, $700 worth of materials, 10 hours of documentation, waiting on the phone with the Department of Transportation, and working with the local notary (a small local business that helps individuals apply for state titles), I finally received the title, and with it the ability to legally use it on the road.*

*The cost of a brand new, pre-assembled steel trailer? $550. This story is a sad example of where I confused price with value. In a way, I got what I paid for, and perhaps, even what I deserved. I implore you, do not be like me.*

[5] Source: U.S. Bureau of Labor
[6] Source: Univ. of Minnesota
[7] Source: Eagle Rock Ltd. Continuous Planning and Mgt. Survey

***How Does this Example Relate to Enterprise-Grade Servers?***

It is too easy to run numbers and confuse a price tag with the Total Cost of Ownership (TCO). What goes into calculating the TCO for a server? For the majority of models:

- Initial purchase price,
- license renewal fees,
- personnel costs and time required to perform maintenance,
- cost of downtime,
- support,
- datacenter and upkeep costs,
- and software costs.[8]

***Maintenance***

> *"The cost of maintaining a well-running HPE NonStop server is minimal, if at all."*
>
> --Former AOL technologist (25+ years of experience with HPE NonStop servers)

How many employees does it take to run a typical Linux server with an Oracle database? On average, five-six personnel. How many employees does it take to run an HPE NonStop server with comparable throughput? One employee.[8]

## Is *Your* Business Paying Enough Attention?

These statistics are all sobering. Perhaps you think that your business is protected against outages by a thoroughly resourced and tested business continuity plan and redundant IT infrastructure, so industry-grabbing headlines of an extended outage involving your company are impossible, or at least very unlikely. Again, this belief is not borne out by a recent Ponemon Institute survey[4], which found that:

- Only 36% believe they utilize all best practices in datacenter design and redundancy to maximize availability
- Less than half (44%) believe datacenter availability is their highest priority
- Only 38% agree there are ample resources to bring their datacenter up and running if there is an unplanned outage
- Over two-thirds (68%) agree that availability has been sacrificed to improve efficiency or reduce costs
- Less than half (41%) believe senior management fully supports their efforts to prevent and manage unplanned outages
- Most telling of all, 52% believe all or most unplanned outages could have been prevented

These results tell us that the management support and resources necessary to prevent prolonged outages are not being applied in practice. Hence outages are still occurring, many of which could have been prevented.

This issue of perception not equating to reality is also seen in the results of another study (Figure 8). While 82% of respondents are confident they are protected against outages, only 65% have sufficient 24x7 technical support coverage. How do the other 35% think they are actually going to execute their business continuity plan when needed if qualified and trained support personnel are not available? In addition, 87% of respondents are confident they are protected against data loss, but only 54% actually test that assertion including the folks that actually use and update that data.

[8] Source: *Research Note: NonStop TCO Comparison*

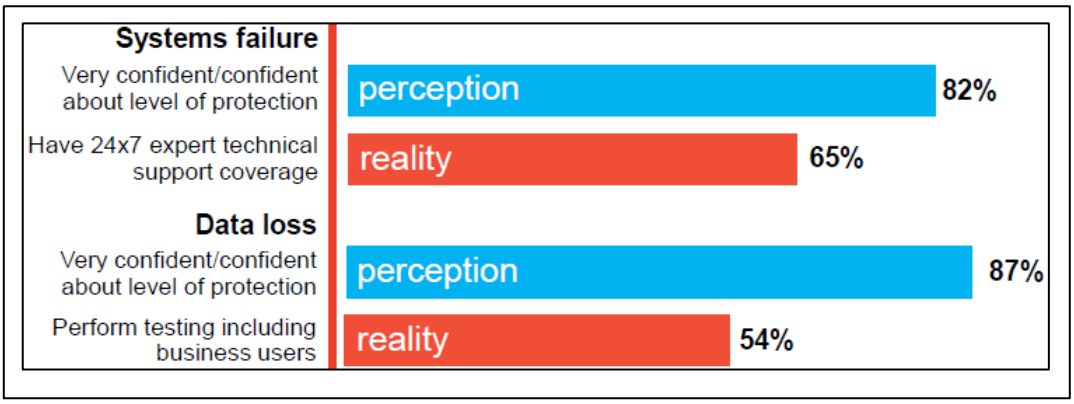



**Figure 8 – Perception vs Reality Regarding IT Failure Preparedness[9]**

***Fault Tolerance***

There are further telling results which highlight the disconnect between belief that sufficient IT procedures are in place should failures occur, and the reality (Figure 9). This disconnect exists even in the banking industry, which is among the most demanding in terms of system availability, service level agreements, and audit compliance requirements:

- 11% do not perform backup/restore testing
- 18% do not have a fully documented DR plan
- 20% perform no internal audit of their procedures

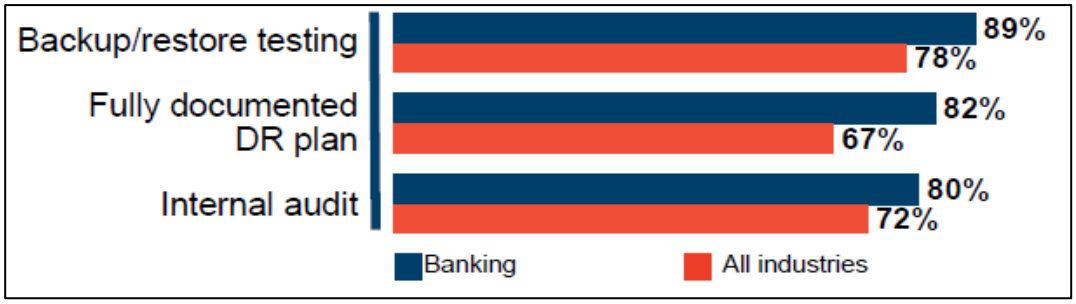



**Figure 9 – How Prepared Are You, Really?[8]**

How can 18% of banks have no fully documented DR plan? Additionally, the banking industry is actually in better shape compared to the average across all industries, where one third do not have a documented DR plan!

To summarize, IT outages are not uncommon, they *will* happen, and when they do, the consequences can be dire for the business. In addition, when they do happen, you may well find that your company is not as well protected to handle the event, as you would like to think it is.

However, it does not have to be so. There *are* ways to ensure that your company is well protected against these circumstances, you *can* continue operations with little if any disruption, provided you put the necessary IT infrastructure, resources and procedures in place to meet your availability goals, and test them regularly. It takes focus, investment, and planning.

## Business Continuity Architectures: Pros and Cons

We will look at what is necessary to ensure your company does not end up in the headlines with a sensationalistic anti-availability storyline. Ultimately, an understanding of the total cost of ownership of the various business continuity architectures is required, so you can make an informed decision as to the best solution to meet your availability goals. To that end, we must first introduce the various business continuity technologies, and the differences between them.

[9]Source: IBM Global Reputational Risk and IT Study. Global survey of senior executives across all industries.

A couple of terms must be understood to help illustrate the differences between business continuity architectures:

1. Recovery Point Objective (RPO). This is the maximum acceptable amount of data loss arising from an outage of an active system. In practice, it is the data updated in the period between the last time the data was saved to (remote) recoverable media, and the point of failure.
2. Recovery Time Objective (RTO). This is the maximum acceptable time for recovery from an outage. In practice, it is the period between the time of failure and the point at which services are restored to an acceptable level.

The different business continuity architectures offer different capabilities with respect to RPO and RTO. We will briefly look at each of these major architectures.[10] [11]

***The Public Cloud is not Fault Tolerant Enough***

> *"Even as availability and security in the public cloud have greatly improved, true fault tolerance continues to be seen as an on-premises or hybrid cloud capability, not as a public cloud capability. IDC research shows that 38.5% of businesses host the highest availability tier on on-premises infrastructure, whereas only 2% of businesses host this tier in a public cloud."*[12]

Make no mistake: every technology has advantages, disadvantages, and a solution it attempts to solve. That being said, those implementing the public cloud while placing full faith in it as a fault tolerant, 100% secure solution may want to reconsider their assumptions.

Google, the "free"[13] provider of Google Search, YouTube, and Google Drive, and also considered one of the top technology companies in the world, experienced an average of 23.27 reported outages per day with a maximum of 469 over the past 90 days, since this article's posting. (Note, these are only *reported* outages.)[14] Additionally, Google recently experienced an hour of downtime for YouTube, Google Docs, Gmail, and Google Classroom.[15] Apple iCloud has faced serious security fallbacks. In one such hack, 40 million accounts were remotely compromised.[16]

In summary, the public cloud claims to give the best of both worlds: outsourcing storage and maintenance tasks to another company, consumption-based usage (only pay for what you use), and top-tier security ("we take security practices very seriously"). Unfortunately, many companies using Amazon Web Server (AWS) are surprised when they get their bills (sometimes millions of dollars more than they expected).[17] Microsoft Azure[18] and Google users experience enough downtime that they cannot be labeled as fault-tolerant, and 40 million iCloud accounts have been hacked.

***Too good to be true?***

Unfortunately, yes.

***Asynchronous Active/Passive – Classic Disaster Recovery***

In this architecture, all transactions are executed on a single system (the active node), and the database updates are replicated asynchronously to a backup system (the passive node). In the event of a failure of the active node a failover to the backup node is executed, users are switched to the backup node, the applications are brought up with the local (synchronized) database open for read/write access, and processing resumes.

---

[10] For a much more detailed description of the various business continuity architectures, see the Gravic white paper, *Choosing a Business Continuity Solution to Match Your Business Availability Requirements*.

[11] Note that each of these architectures use *asynchronous* replication, where there is a slight delay between when the data is updated on one system, and is safely stored on another system.

[12] Source: *Research Note* by Pyalla Technologies, LLC: NonStop TCO Comparison (Worldwide AL4 Server Market Shares, 2019: Fault-Tolerant systems become Digital Transformation (DX) platforms. Paul Maguranis Peter Rutten)

[13] "Free" is in quotes because Google mines, archives, and sells data to advertisers and third-party companies. A popular quote about "free" online products and services is that "if it's free, you are the product."

[14] Source: Outage Report

[15] Source: Google was hit with massive outage, including YouTube, Gmail, and Google Classroom

[16] Source: 40 Million iCloud Accounts Hacked? Hackers Hold iOS Devices To Ransom

[17] Source: As AWS Use Soars, Companies Surprised by Cloud Bills

[18] See Outage Map

The key issues with this architecture are:

- All users see the outage and need switched to the backup node. All in-flight transactions will fail and require re-submission.
- There is more data loss (higher RPO) than the other architectures we will consider.
- It is very difficult to test the backup node and failover procedures. This testing requires an outage of the primary node and may take a long time, and so failover testing is very often not performed at all or not to completion if it is attempted (because it takes longer than the available or scheduled outage window). It is also possible that restarting the production system after the test has completed may not work, another reason why testing may not be performed fully. Hence, this architecture is risky, the state of the backup system is not really known, and failover faults may occur and the failover may be unsuccessful, or at least take a long time. Because of this uncertainty, management is often slow to initiate a failover in the first place, further delaying recovery after the primary node fails. Consequently, this architecture has the possibility (probability) of a high RTO, several hours or even days.
- Testing costs are high (the process of halting the primary system, failing over to the backup, and then failing back to the primary, is time-consuming and resource intensive). Per test, average costs were cited of $30K-$40K, or even as much as $100K.[19] Again, this fact results in limited testing and increased risk.
- The capacity of the backup system is under-utilized.
- Depending on the data replication product used during replication, the backup database may be inconsistent with the primary and hence, using it even for read-only access during this time may not be feasible.[20]

While a basic active/passive architecture offers some protection, it is by no means the best solution. It should really only be considered as a starting point, or used for non-mission critical applications.

***Asynchronous Active/Almost-Active – Sizzling-Hot-Takeover (SZT)***

While looking almost the same as a classic active/passive architecture, SZT has one difference that makes it a much better solution. That difference is that while all transactions are still routed to a single active node, the backup node has the applications already up-and-running, with the local database open for read/write access.[21]

The key benefit that this architecture confers vs classic active/passive is to obviate the uncertainty around the state of the backup system. Since the applications are up and running on the backup node with the local database open read/write, it is easy to send test transactions to validate the backup system at any time, with no impact to the active system. Hence, the backup system can be regularly validated, and becomes a *known-working system* (it is, to all intents and purposes, a fully active system, with the exception that it is not processing online transactions). If (and when) an outage occurs of the primary node, the decision to fail over can be made immediately, with confidence that failover faults will not arise, and the failover will succeed quickly. Therefore, this architecture gives a much better and repeatable RTO for SZT vs classic active/passive architectures.

One other difference of SZT vs basic active/passive architectures is that HPE Shadowbase bi-directional data replication is configured between the active and passive nodes. With this setup, on a failover, all changes to the database on the backup node will be queued. As soon as the down node is recovered, HPE Shadowbase replication will automatically replay the queued updates to bring the two databases back into synchronization. This replay re-establishes backup protection quickly, and allows for faster fail back to the recovered node if required.

SZT does still suffer some of the same issues as classic active/passive however (all users see the outage and are affected; more data loss; under-utilization of backup capacity). However, it is an excellent solution when a fully active/active configuration is not possible, and much, much better than classic active/passive. (There is really no reason why everyone should not be using an SZT architecture as opposed to classic active/passive.)

[19]Gartner Infrastructure Summit.
[20]This inconsistency is not the case with HPE Shadowbase data replication, which maintains transactional consistency between the active and backup databases during normal operations.
[21]Not all data replication products allow the backup database to be open for application read/write access during replication, but Shadowbase solutions have no such restriction.

***Asynchronous Active/Active – Partitioned***

In a partitioned active/active architecture, the applications are active on all nodes, transactions are routed to all nodes, and each node has a copy of the database that is kept synchronized by HPE Shadowbase bi-directional data replication. However, the data is partitioned such that transactions are routed to a specific node based on some key in the data. For example, the database may be split by customer name, and all transactions for customers A-M are executed on one node, and customers N-Z on the other. This architecture brings most of the benefits of active/active, but avoids one of the biggest potential issues, data collisions (where the same record is updated simultaneously on multiple nodes, resulting in data inconsistency that must ultimately be identified and resolved).

The benefits of this architecture compared to classic active/passive and SZT for a two-node configuration are:

- On outage, only half of the users are affected and need to be switched. The other half of users see no outage at all, i.e., better RTO.
- Only half the in-flight transactions will fail and require re-submission.
- There is about half as much data loss (in a two-node configuration). i.e., better RPO, because only the updates in the replication stream on the failed node are lost. The updates in the replication stream on the remaining node(s) are unaffected, and will be replayed once the down node is recovered.
- There are no testing costs/issues, and no failover faults. All systems in the configuration are known to be working at all times (also true for SZT).
- The capacity of the backup system is better utilized.

There are still some issues however:

- Not all applications/data can be partitioned
- Because transactions must be routed to specific nodes, imbalanced load distribution is possible
- As for any active/active solution, it is more complex to implement and manage

***Asynchronous Active/Active – Route Anywhere***

There is always a price to pay of course, and in this case, it is the possibility of data collisions. For some applications data collisions may be practically impossible (for example, it is highly unlikely the same credit/debit/ATM card would be used simultaneously for multiple transactions). However, if collisions are possible, they must be dealt with. HPE Shadowbase data replication includes functionality to automatically detect, report, and resolve data collisions. User exits are also provided to enable more sophisticated processing of data collisions if necessary. Figure 10 helps to visualize the difference between these various architectures with respect to the parameters of RPO and RTO.[22]

[22]Note that with respect to RTO and RPO, there is no difference between Asynchronous Active/Active – Partitioned and Asynchronous Active/Active – Route Anywhere.

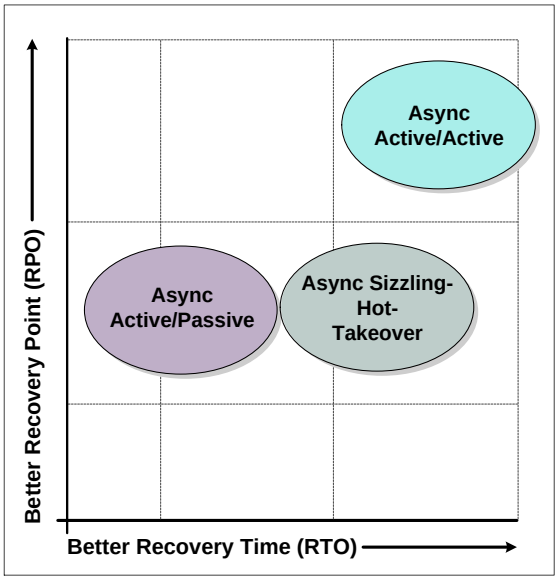



**Figure 10 – RPO and RTO for the Various Business Continuity Replication Architectures**

There is another way of looking at the various business continuity architectures which provides a more striking view of the differences between them and hence, their relative benefits, and that is to look at the total cost of ownership (TCO). It is true that active/active configurations are more expensive and complex to implement, but when looked at through the lens of TCO, these issues pale into insignificance.

Using the average cost per hour of downtime for a financial application of $1.5M/hour (Figure 5), and making some informed estimates about typical periods of recovery time (RTO), we can estimate actual outage costs for the various business continuity architectures (Figure 11).

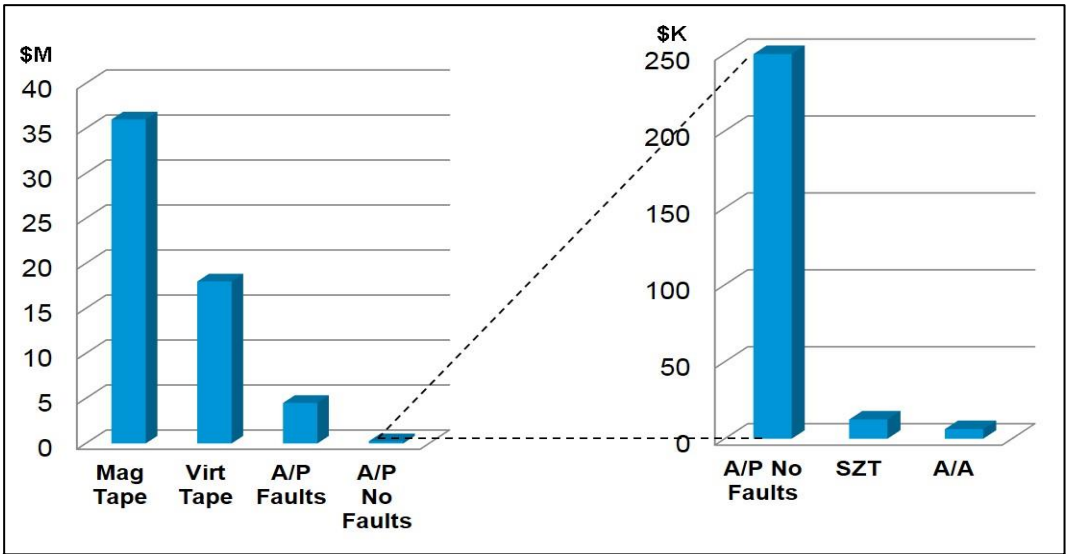
| Architecture | RTO | Outage Cost |
|---|---|---|
| Active/Passive[1] | ~ 3 hours (if at all) | ~ $4.5M |
| Active/Passive[2] | ~ 10 minutes | ~ $250K |
| Sizzling-Hot | ~ 30 seconds[3] | ~ $12.5K |
| Active/Active | ~ 30 seconds | ~ $6.25K[4] |

**Figure 11 – Estimated Service Unavailability Costs for a Financial Application**

What is glaringly obvious is that basic active/passive architectures are very expensive when looked at in terms of TCO. They may be easier and cheaper to implement, but when outages do occur, they are likely to cost you much, much, more in the long run. Even in the best case for a well-tested system and a trouble-free failover, basic active/passive is still going to be ~20 times higher in outage costs compared to an SZT configuration. For a worst-case scenario (much more likely given the difficulties of testing and probability of failover faults as previously discussed), it is ~36 times costlier at ~$4.5M *per outage*.

**The cost differences become even more apparent when viewed graphically (Figure 12). Given the marginal incremental cost and complexity, coupled with the significant decrease in potential outage costs, there is really no reason why anybody should run that risk and not move immediately from an active/passive to an SZT architecture. Figure 12 and**

Figure 13 also illustrate just how good an SZT setup is even compared to an active/active implementation, when viewed solely in terms of RTO performance. There are of course other benefits of active/active vs SZT, as discussed above.




[1] Worst case: with failover faults, management indecision, etc.

[2] Best case: with no failover faults, prompt management action, etc.

[3] Possibly slightly longer depending on network switching.

[4] Half of users see no outage at all (less than half if > 2 replicated nodes)

**Figure 12 – Estimated Service Unavailability Costs for a Financial Application**

**As well as the cost of downtime, there is also the cost of lost data, as shown in**

Figure 13. Even though data loss (RPO) goals based on the average value of a transaction may appear acceptable, some transactions (data) are much more valuable than others and cannot be lost, period:

- Healthcare – lost dosage records result in patient overdosed on medication
- Manufacturing – car manufacturer can tolerate short production line outage, but cannot lose data regarding bolt torque settings, etc., for fear of lawsuits in case of accidents
- EFT – some transactions are worth $M, even if the average transaction is much lower
- Stock Trades – like EFT, some transactions are worth $M, and stock price is based on previous trades (none can be lost)

Therefore, RPO goals must be set based not on the value of an average transaction, but on the value of the most expensive/critical transaction. If the cost of losing the most valuable/critical data is very high, then an active/active architecture is the best solution, since it has the best RPO characteristics (least data loss).

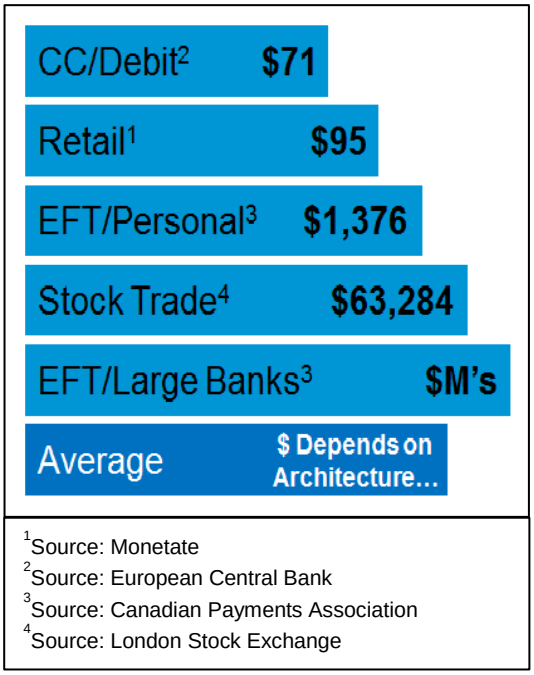



**Figure 13 – Average Cost per Transaction of Lost Data for Various Industries**

**As well as the cost of downtime, there is also the cost of lost data, as shown in**

Figure 13. Even though data loss (RPO) goals based on the average value of a transaction may appear acceptable, some transactions (data) are much more valuable than others are and cannot be lost, period:

- Healthcare – lost dosage records result in patient overdosed on medication
- Manufacturing – car manufacturer can tolerate short production line outage, but cannot lose data regarding bolt torque settings, etc., for fear of lawsuits in case of accidents
- EFT – some transactions are worth $M, even if the average transaction is much lower
- Stock Trades – like EFT, some transactions are worth $M, and stock price is based on previous trades (none can be lost)

Therefore, RPO goals must be set based not on the value of an average transaction, but on the value of the most expensive/critical transaction. If the cost of losing the most valuable/critical data is very high, then an active/active architecture is the best solution, since it has the best RPO characteristics (least data loss).

To summarize, TCO decreases by orders of magnitude more than the cost that the business continuity solution increases, as illustrated by Figure 14:

- The better the availability, the greater the complexity and implementation cost
- The better the availability, the lower the outage cost
- Net result, as implementation cost increases, overall TCO decreases, but at a much faster rate

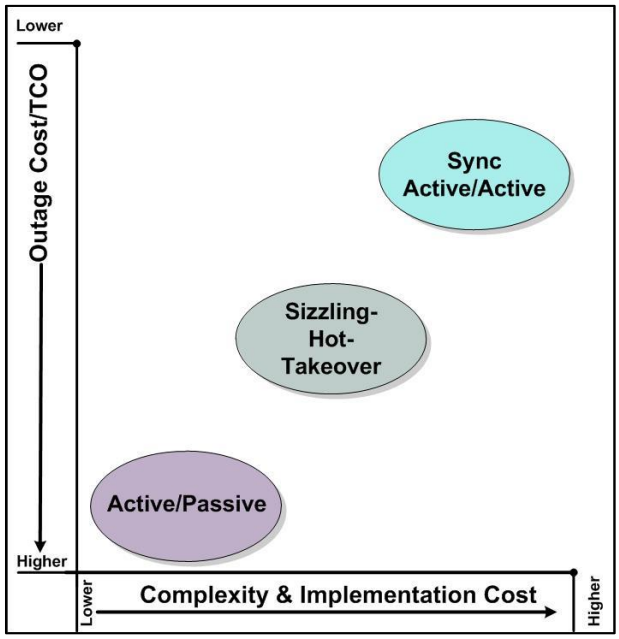



**Figure 14 – TCO vs Complexity/Cost of Implementation of Various Business Continuity Architectures**

By this measure, the cost and complexity of an active/active solution is clearly more than outweighed by its superior overall TCO. It likewise illustrates how much better SZT is in terms of TCO compared with basic active/passive, and for only a marginal increase in cost and complexity. Everyone should at least be running an SZT configuration!

## New Business Continuity Technologies – Synchronous Replication

This paper would be incomplete without a brief discussion of upcoming new features in the HPE Shadowbase suite of products to address some of the inherent shortcomings of asynchronous replication, to enable you to make the most informed choice as to which business continuity architecture is right for you.

Asynchronous replication represents state-of-the-art technology, and offers excellent levels of protection against outages (especially in SZT and active/active configurations), and is more than enough for most applications.

However, it does have some limitations:

- Data loss (in active/passive, SZT, and active/active modes)
- May require application/data partitioning (in active/active mode), which may result in imbalanced load across systems
- May incur data collisions (in active/active mode)

**If we take the average costs of lost data per transaction for various industries (as shown in**

| Technology | RPO[1] | Retail[2] | CC/Debit[3] | EFT[4] | Stock Trade[5] |
|---|---|---|---|---|---|
| A/P + S/H[6] | ~ 1 sec | $47.5K | $35.6K | $688K | $31.6M |
| A/A[6] | ~ 0.5 sec | $23.8K | $17.8K | $344K | $15.8M |

[1] Example assumes rate of 500 transactions per second
[2] Retail average transaction ~ $95 (US online) (Source: Monetate)
[3] CC/Debit average transaction ~ $71 (UK) (Source: European Central Bank)
[4] EFT average transaction ~ $1,376 (Source: Canadian Payments Association)
[5] Stock trade average transaction ~ $63,284 (Source: London Stock Exchange)
[6] Asynchronous replication

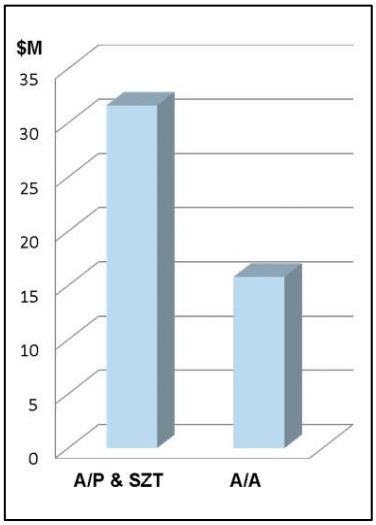
Figure 13), we can then estimate the actual data lost costs arising from an outage depending upon which business continuity technology is employed (Figure 15 and Figure 16). We can see that all of the asynchronous replication architectures are subject to the possibility of high costs associated with data loss (even active/active, although it is considerably better than either active/passive or SZT).

**Figure 15 – Estimated Outage Data Loss Costs for Various Industries by Business Continuity Technology**

Therefore, for the most critical applications, those for which *any* lost data or downtime will incur unacceptable levels of business cost, asynchronous replication may be insufficient. For such applications, a new Shadowbase technology, synchronous replication, will soon be available, which resolves all of these issues.

With asynchronous replication, the source and target databases are updated independently and consequently, and there is a separation between when the data is updated on the source system and safe-stored on the target system. (It creates a window where a failure of the source system can result in updates not replicated to the target system – and the data would be lost.) With Shadowbase synchronous replication the data is updated simultaneously on both systems (updates are only committed on the source system if and when the data has also been safe-stored on the target system).[23]

**Figure 16 – Estimated Costs of Lost Data for Average Stock Trade Transaction**




[23]For more details on HPE Shadowbase synchronous replication technology, see the Gravic white paper, *Choosing a Business Continuity Solution to Match Your Business Availability Requirements*.

Synchronous replication provides the following benefits compared to asynchronous replication:

- Zero data loss (RPO = 0).[24] This benefit is graphically illustrated in Figure 15, where the data loss costs for synchronous replication in any configuration (active/passive, SZT, or active/active), is $0 across the board.
- No possibility of data collisions (in active/active mode). During the source transaction update, the data records are locked on all replicant systems, so it is not possible for another transaction to update the same data simultaneously.[25]
- No need for application/data partitioning (in active/active mode), because data collisions are avoided.

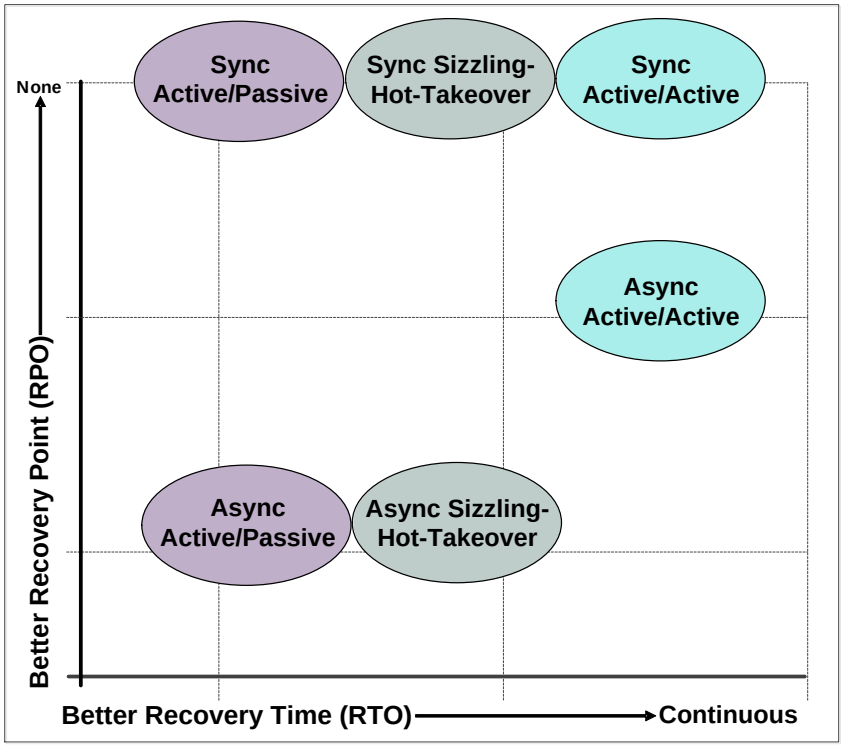



**Figure 17 – RPO and RTO for Asynchronous vs Synchronous Replication**

HPE Shadowbase synchronous replication can be run in any of the same three modes as discussed above for asynchronous replication. In active/passive and SZT modes, it has the same attributes as asynchronous replication, with the big exception that there is zero data loss on failover. In active/active mode, as well as zero data loss on failover, it also avoids data collisions with no application/data partitioning required, enabling the active/active route anywhere model for any application, ensuring balanced load distribution across systems.

There is now a business continuity replication solution for those applications where any data loss cannot be tolerated, and for those applications that previously could not run in an active/active configuration (because partitioning was not possible and data collisions could not be tolerated or resolved). Therefore, Shadowbase synchronous replication enables the minimum (best) possible values for RPO and RTO, for the widest possible range of applications (Figure 17).

## Scalability

Scalability is critical to avoid “brownouts.” Does anyone remember in 2018 when Bitcoin’s USD value skyrocketed and the transaction rate slowed to a crawl? This example is a modern-day brown-out. It can occur in the enterprise world when demand peaks, taxing server workloads, ultimately lagging transaction throughput, and subsequently slowing down service to the end users. Architectures that can scale will automatically resolve this issue.[26]

[24]This feature is available with the new product, Shadowbase ZDL. Contact Gravic for more details.
[25]This feature will be available with the follow-on product, Shadowbase ZDL+. Contact Gravic for more details.
[26] For more information, please see: *The Benefits of Switching from Scale-up to Scale-out Architectures*

"A small two-processor system, to a large 16-processor system can be clustered into large systems of 4080 processors for a total of 24,480 cores with the same system image and application environment."[27]

Summary of Business Continuity Architectures and Technologies

Figure 18 summarizes the characteristics and differences between the various business continuity architectures and technologies discussed above. The horizontal axis lists each of the architectures (in both asynchronous and synchronous modes), and the vertical axis lists the main parameters which describe the attributes of each architecture. Notice that moving from left to right, from active/passive to SZT to active/active, the characteristics improve, and in each case, the synchronous mode is better than the asynchronous mode. Which is the best solution for your applications depends upon the specific business continuity requirements of each application, but given those, this table helps determine which architecture and technology would best satisfy your requirements. In any case, there is no better solution than synchronous active/active!

| Replication Mode / Attribute | Asynchronous Active/Passive | Synchronous Active/Passive | Asynchronous Sizzling-Hot-Takeover | Synchronous Sizzling-Hot-Takeover | Asynchronous Active/Active | Synchronous Active/Active |
|---|---|---|---|---|---|---|
| Failover Faults | Yes | Yes | No | No | No | No |
| Application Outage | Yes | Yes | Minimal[1] | Minimal[1] | No | No |
| Data Loss | Yes | None | Yes | None | Yes | None |
| Data/Request Partitioning | Not required[2] | Not required[2] | Not required | Not required | May be required | Not required |
| Data Collisions | Not possible | Not possible | Not possible | Not possible | Possible | Not possible |
| Backup Utilized | No[3] | No[3] | No | No | Yes | Yes |

[1] All users affected, but takeover time same as for Active/Active modes
[2] "Required" if run in Reciprocal mode
[3] "Yes" if run in Reciprocal mode

**Figure 18 – Business Continuity Architecture and Technology – Pros and Cons**

## In Summary – Match Your Business Continuity Solution to Your Business Risk

It is clear that the costs of prolonged downtime of critical business IT systems are significant (potentially to the point of shuttering the company). The significance is compounded by the fact that the many events that can lead to such outages are not rare; it is a case of when, not if. This reality is only acceptable if you have a complete, documented, and well-tested business continuity plan in place. Maybe you think that you do, but the data does not support this way of thinking. Many companies are operating with the mistaken belief that their business continuity plan will work when the time comes, only to find out the hard way that it does not. The first step is to make sure that this misplaced confidence does not apply to your company; completely document your business continuity plan, and regularly exercise it in practice.

In order to complete a thorough business continuity plan, the IT architecture to be employed in order to maintain services in the event of an outage (planned or unplanned) must be selected. This paper has described several such business continuity architectures, from classic active/passive, through SZT, to active/active, using HPE Shadowbase asynchronous and synchronous data replication. Many users never get beyond basic active/passive, but as has been described, this architecture has many issues, which can prevent a successful and timely failover. It is difficult to test and is risky because it can suffer from failover faults. Consequently, this model is reactive and provides a false sense of security. The more sophisticated business continuity solutions (SZT and active/active), while marginally more complex and expensive to implement, when looked at through

[27] *Research Note: NonStop TCO Comparison*

the lens of TCO are in fact far more cost-effective. If you are running an active/passive architecture, it will probably only take one outage during peak processing hours and the costs incurred for you to realize that you need to move to an SZT or active/active architecture (that is if your business survives at all!).

Asynchronous SZT and active/active configurations are far superior to active/passive, and should be the minimum solution chosen. However, these configurations also have issues: some data loss (in both cases, although active/active is 50% less than SZT), and possible data collisions (active/active only).

It may be that for your applications, one or other of these solutions is perfectly good enough (match the attributes of the solution to the potential costs of a prolonged outage). But, for some, even SZT or active/active is not good enough (for example, those applications where no data loss is acceptable, or which must be 100% available, but cannot employ an active/active architecture, because data collisions cannot be tolerated or avoided). For these applications, HPE Shadowbase synchronous replication offers a solution, eliminating both data loss and data collisions when running active/active.

The solution is in your hands, the attention-grabbing headlines need never apply to your company. Make sure that your business continuity plan is based on more than just faith, and that you know for a fact that it will work when you need it because you have exercised it regularly. Choose an appropriate business continuity architecture and data replication solution with the necessary attributes to eliminate the risk of large costs due to an IT outage (which should be at least an SZT configuration). You do not just have to cross your fingers and hope!

## International Partner Information

### Global

**Hewlett Packard Enterprise**
6280 America Center Drive
San Jose, CA 95002
USA
Tel: +1.800.607.3567
www.hpe.com

### Japan

**High Availability Systems Co. Ltd**
MS Shibaura Bldg.
4-13-23 Shibaura
Minato-ku, Tokyo 108-0023
Japan
Tel: +81 3 5730 8870
Fax: +81 3 5730 8629
www.ha-sys.co.jp

## Gravic, Inc. Contact Information

17 General Warren Blvd.
Malvern, PA 19355-1245
USA
Tel: +1.610.647.6250
Fax: +1.610.647.7958
www.shadowbasesoftware.com
Email Sales: shadowbase@gravic.com
Email Support: sbsupport@gravic.com

Hewlett Packard Enterprise Business Partner Information

Hewlett Packard Enterprise directly sells and supports Shadowbase Solutions under the name ***HPE Shadowbase***. For more information, please contact your local HPE account team or visit our website.

Copyright and Trademark Information

This document is Copyright © 2015, 2017, 2020, 2022, 2024 by Gravic, Inc. Gravic, Shadowbase and Total Replication Solutions are registered trademarks of Gravic, Inc. All other brand and product names are the trademarks or registered trademarks of their respective owners. Specifications subject to change without notice.