The cloud after tomorrow

ARE YOUR BUSINESS CONTINUITY AND DISASTER RECOVERY PLANS READY?

Prepare for data disaster

Cloud vendors inherently provide some support for the availability of their customers' data and applications, but this level of availability is often not enough.

Should a cloud infrastructure failure occur — caused by utility issues, natural disasters, human action or other events — the data and applications hosted in the cloud may not be available to users for an extended period of time. Even worse, they may be lost altogether.

For many organizations, this is simply unacceptable. Data loss, service interruption and lost sales are particularly damaging to a business. Organizations that use cloud services should prepare for potential failures and plan accordingly to mitigate the associated risks.

Fortunately, proactive cloud providers are increasingly offering one or more data and application replication services to their customers. These services support business continuity and disaster recovery in the cloud by ensuring that cloud-based data and applications are replicated to another location to minimize the impact of a failure at any particular location.

Some services can automatically failover when a problem occurs, making the transition nearly seamless to users. The most capable services, such as hypervisor-based solutions, can limit data loss and downtime to seconds or no time at all. In contrast, other services can only limit data loss and/or downtime in terms of minutes, hours or even days.

Every organization that stores data and applications in a cloud environment should give serious consideration to deploying replication services from their cloud provider to mitigate the risks posed by a failure within the cloud.

> "Every organization that uses cloud services should prepare for potential failures and plan accordingly to mitigate the associated risks."

Be prepared for the unknown

It's human nature for people to assume that nothing disastrous is going to happen to them. Many organizations treat their infrastructures, data and applications in the same manner. They never seriously consider the "what if" possibilities, let alone address them.

This lack of strategic planning may work in the short term, but in the long run it's likely the organization will have some sort of interruption to its operations. This can be caused by power outages, air conditioning failures, burst pipes, Internet service outages and other utility issues.

Then there's Mother Nature. Floods, earthquakes and other natural disasters may cause havoc to organizations and their ability to function at a normal capacity.

Operational interruptions may also be triggered by equipment theft, failure of a critical service vendor (e.g., bankruptcy), pandemics or epidemics, or targeted cyber attacks.

Business continuity and disaster recovery (BC/DR) are the services that plan for mitigating these interruptions to reduce their impact to acceptable levels. BC/DR is particularly important for preserving data, which could be lost during a disaster. This may also be known as business continuity management (BCM).

"Over time, recovery, hosting, application and storage cloud providers may offer more robust service availability and data protection alternatives compared to inhouse IT," said Gartner analyst John P. Morency in a July 2014 report.[1]

Unfortunately, some organizations incorrectly assume that if they move their data and applications to the cloud, they will be provided with much more redundancy and resiliency because those are the characteristics of clouds.

While there is some truth to that — cloud workloads being seamlessly migrated from server to server for example to avoid maintenance downtime — it is generally not a cloud provider's responsibility to perform backups of customer data and applications.

It's also rare that they provide advanced data replication services more rigorous than weekly or daily backups unless specifically requested (and paid for) by the customer.

> "Over time, recovery, hosting, application and storage cloud providers may offer more robust service availability and data protection lternatives compared to in-house IT."
>
> GARTNER
> "HYPE CYCLE FOR BUSINESS CONTINUITY MANAGEMENT & IT DISASTER RECOVERY MANAGEMENT"

[1] "Hype Cycle for Business Continuity Management and IT Disaster Recovery Management, 2014," John P. Morency & Roberta J. Witty, Gartner, Inc., July 22, 2014.

Gartner does not endorse any vendor, product or service depicted in its research publications, and does not advise technology users to select only those vendors with the highest ratings or other designation. Gartner research publications consist of the opinions of Gartner's research organization and should not be construed as statements of fact. Gartner disclaims all warranties, expressed or implied, with respect to this research, including any warranties of merchantability or fitness for a particular purpose.

4 critical disaster recovery assessments

Each cloud customer must conduct its own BC/DR planning then enact those plans for data and applications in the cloud. Each organization has a unique set of requirements for BC/DR, as well as different needs for each data set and application.

For example, one application and its data may need to be available at all times for an organization, with no loss of data. Meanwhile, it may be perfectly fine for another application to be unavailable for a day or two and/or to have multiple days of data lost.

Assessing an organization's requirements for BC/DR for its cloud-based data and applications should be executed via a formal process. Common steps for such processes include the following:

Define the potential business impact of downtime and/or data loss

Many factors are potentially used to measure this impact, such as loss of revenue and reputation. Once an organization has determined the impact that can be tolerated, it then needs to translate it into actionable information.

The terms most often used to express this impact are RTO and RPO. RTO, which stands for recovery time objective, measures how long an application and/or its data being unavailable can be tolerated by the organization.

Similarly, the recovery point objective (RPO) quantifies how much data loss is acceptable — none, seconds,minutes, hours, days, etc.

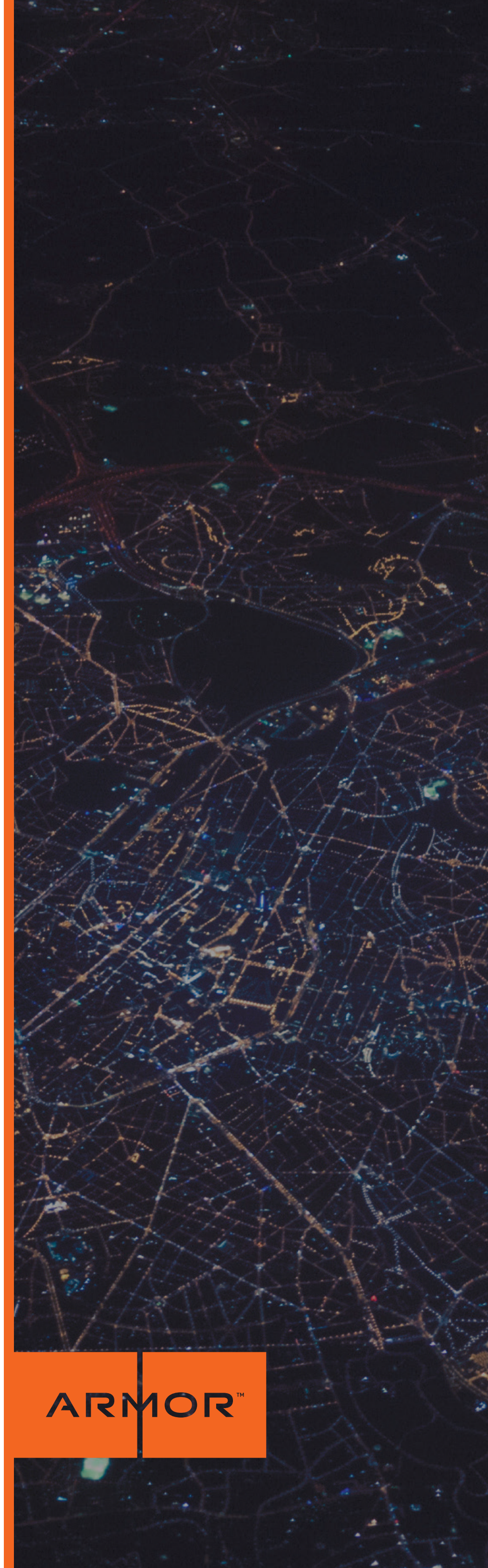
Examine existing requirements

Many organizations are subject to one or more sets of existing requirements that may affect BC/ DR planning.

For example, there are various laws and regulations for particular types of data (e.g., health data, financial records) that may mandate what data must be maintained and for how long. This could affect how often data is backed up or otherwise replicated, and how long those backups are preserved.

An organization may also need to meet requirements that they have previously agreed to with their own customers, such as data availability clauses in contracts and service-level agreements (SLA).

The need to meet existing requirements may lead to a more rigorous solution for business continuity and disaster recovery than would otherwise be selected.



Identify infrastructure needs

Replicating data, particularly in real-time or near-real-time, can require significant increases in IT infrastructure because of bandwidth consumption and performance needs.

Similarly, performing regular backups can take up substantial storage in the cloud because of the need to preserve backups for a period of time.

There are other needs to be considered as well, such as data replication/backup software and management interfaces for accessing that software. To meet these infrastructure needs, it may be necessary to acquire a variety of additional resources in the cloud.

Develop a remediation plan

The remediation plan should take into account all of the above steps — the potential business impact of downtime and data loss, any needs to meet existing requirements, and any necessary changes to cloud or organization infrastructure.

The organization should use this information to identify any shortcomings in its current disaster mitigations and develop a contingency strategy to address those shortcomings. This strategy should ensure that the right technology, people and processes are in place in the event of a disaster.

Choosing your contingency

There are many different types of BC/DR services for cloud-based data and applications. All of these services synchronize data and/or applications between servers — preferably at different physical locations — but each type of service uses a different method for performing this synchronization.

Many characteristics differentiate these types of services, including RTO, RPO, failover and failback capabilities, hardware/software dependencies, management, scalability and performance impact.

Traditional data backup

The BC/DR service periodically makes backups of the data within the workload. Traditional data backups often occur daily or weekly.

Snapshot-based

The BC/DR service periodically takes a snapshot of the workload (e.g., every four hours) and saves that snapshot elsewhere. The snapshot serves as a backup for the entire workload (including its data, applications and operating system).

Array-based

The BC/DR service duplicates the contents of one storage array on another storage array. The two storage arrays must be using identical hardware.

Guest-based

The BC/DR service runs as an agent within the workload's guest operating system. It periodically backs up data and applications running on top of that guest operating system.

Orchestration-based

The BC/DR service provides cloud orchestration capabilities that support application availability, typically in conjunction with a separate data replication service.

Hypervisor-based

The BC/DR service automatically replicates the workload (virtual machine) from one server to another, synchronizing both the data and applications. Replication typically occurs in near-real-time (e.g., every few seconds) or continuously as changes are made.

Recovery time objective (RTO)

As discussed, RTO refers to the acceptable period of time for applications and their data to be unavailable for use. Downtime is a fundamental consideration for nearly every BC/DR scenario.

That said, some older BC/DR solutions only preserve data and not applications. These solutions assume that the organization is already maintaining duplicate copies of the application or that the organization is prepared to stand up new instances of the application whenever needed.

"Not only is the market need for more predictable operations recovery increasing, but the required recovery times for the most important mission-critical applications continue to be measured in the order of minutes or hours versus in days," said Morency.

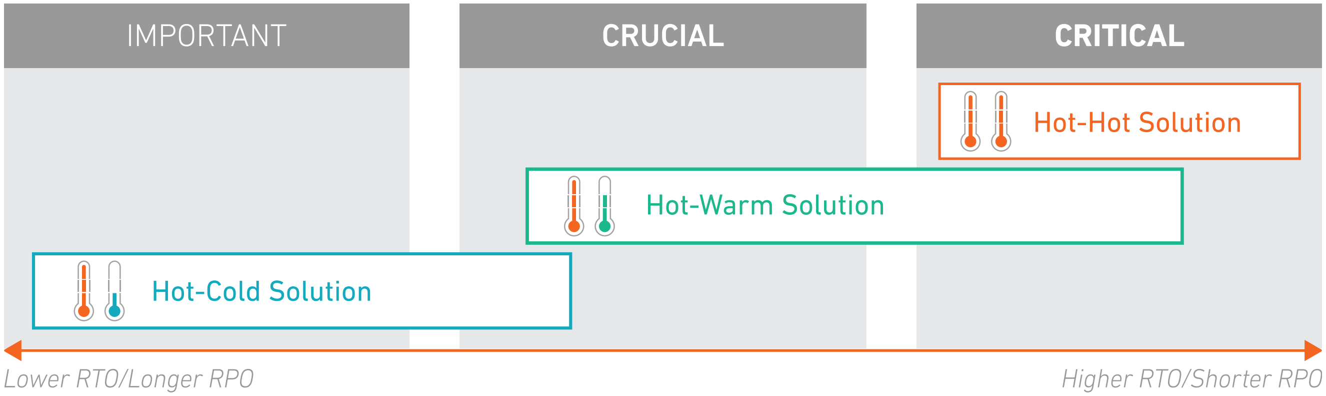
> "Not only is the market need for more predictable operations recovery increasing, but the required recovery times for the most important mission-critical applications continue to be measured in the border of minutes or hours versus in days."

GARTNER

"HYPE CYCLE FOR BUSINESS CONTINUITY MANAGEMENT & IT DISASTER RECOVERY MANAGEMENT"

Downtime can be limited in several ways

	Hot-Hot	Hot-Warm	Hot-Cold
Protection	Multi-Site	Multi-Site	Single Site
Cost	$$$$	$$	$
Outcome	Fully replicates an environment and allows a company to recover from complete datacenter disaster with minimal RTO/RPO effort.	Allows for partial or total data restores which can be targeted to any datacenter by the user.	Allows a company to restore a server with a longer RTO and restores are only available in the same datacenter as the source server.
Intended Use	All application updates or configuration changes that are made to the first instance are automatically replicated in the second instance so that it is kept current at all times. Data is nearly continuously synchronized from the first instance to the second instance.	This model is similar to the Hot-Hot model, except that interruptions to the first instance are not automatically detected, nor is control automatically transferred to the second instance. Rather, a human administrator must intervene to perform these recovery actions.	The Hot-Cold model only has one active instance of the application that is not loaded and running on a production server. If a failure of the active instance occurs, a human administrator must intervene to transfer data from the active instance's backups or other locations, and then make the application available to users post re-configuration.
RTO	Seconds	1-3 Hours	>3hrs
RPO	30 minutes	30 mins-1hr (As per customer policy)	Daily



Implementing these models in pre-cloud days used to be much more costly. The organization had to acquire, in advance, the hardware to support them; lease space at recovery sites; deploy spare servers and networking equipment; and provision the necessary software. Then the organization had to maintain that equipment on a regular basis.

Now, with the ability to leverage cloud resources, it is much easier for organizations that were previously limited to Hot-Cold architectures to be able to afford Hot-Warm or even Hot-Hot architectures. This evolution has helped greatly decrease RTO.

Still, even with these lowered costs, there remain significant cost differences between Hot-Hot, Hot-Warm and Hot-Cold architectures.

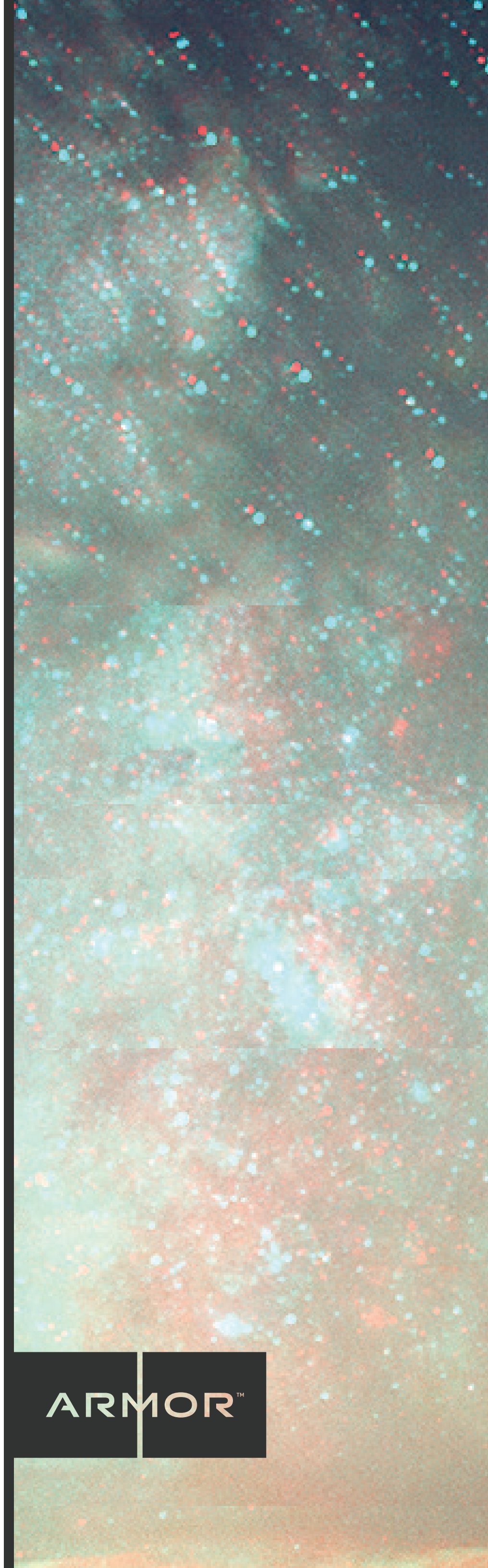
Generally speaking, the lower the RTO the higher the cost of the service. So, Hot-Hot is more expensive to achieve than Hot-Warm, which in turn is more expensive than Hot-Cold.

Another consideration when selecting a BC/DR solution is the synchronization method it uses. Some synchronization methods support lower RTO than others. For example, hypervisor-based solutions typically are capable of providing the lowest RTOs (seconds or minutes), while snapshot-based solutions have higher RTOs (several hours). Traditional data backups don't even synchronize applications. The other methods fall somewhere in between, typically an RTO of minutes or hours.

Complement your strategy with data archiving

Data archiving is a process that complements data replication and may significantly reduce BC/DR solution resource usage. These services take data that is no longer needed for daily use, but must be kept for months or years for data retention purposes, and move it to a different storage media.

This may make it a bit more time-consuming and difficult to access the data, but there are several benefits from doing this. Data archiving can greatly reduce the amount of data being stored in a cloud workload, and thus correspondingly reduce the processing, bandwidth and storage resources necessary to replicate that data.



Recovery point objective (RPO)

For many situations, RPO is as important or even more important than RTO. For example, it might be acceptable for an application to be unavailable for a period of hours, but loss of any data whatsoever cannot be tolerated.

A common instance of this is an application involving financial transactions. The integrity of the related bank records, credit card charges or other financial data might be seriously compromised if recent transaction data were lost.

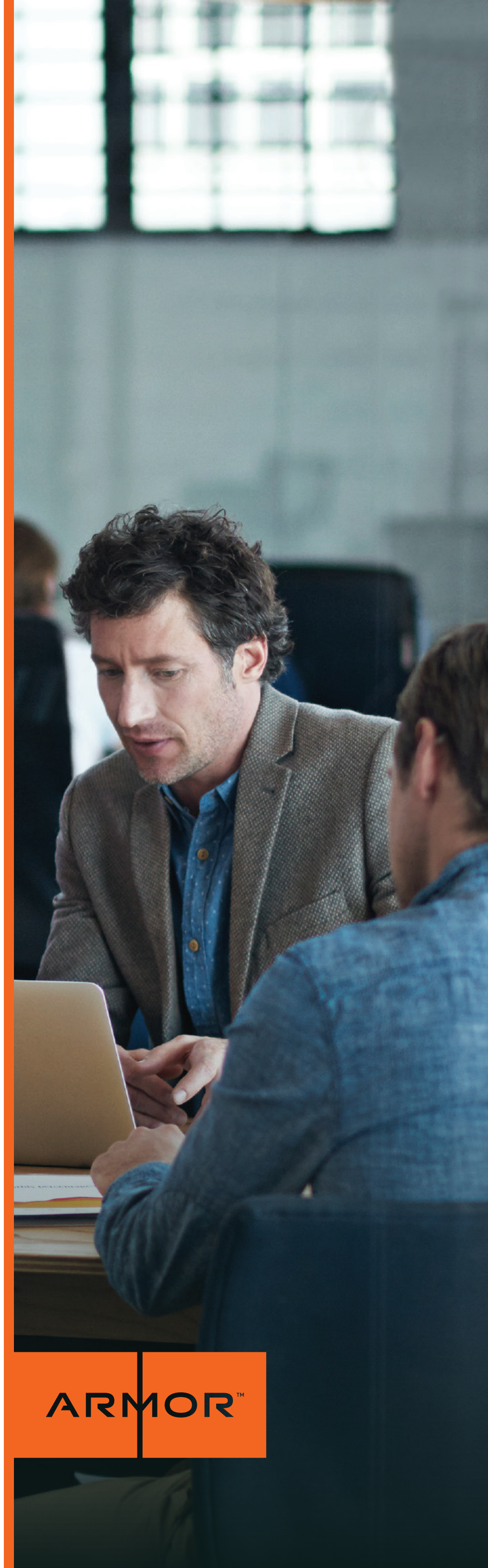
When designing, deploying and maintaining Hot-Hot, Hot-Warm and Hot-Cold architectures, organizations have a great deal of control over how frequently data is replicated from one location to another.

Organizations can replicate data as often as they would like, even continuously, if they have the sufficient resources and finances available to do so.

Similar to RTO, the lower the desired RPO, the higher the cost of the service. Factors to consider include the bandwidth needed to frequently transfer data and applications between locations, and the storage and processing needs for those transferred components.

In terms of synchronization methods and their impact on data loss, hypervisor-based solutions can achieve an RPO of seconds or, for the most sophisticated services, zero (no data loss).

> "The ability to manage recovery service levels in an automated, repeatable and timely manner is becoming increasingly critical for many organizations. As Web-based applications support more business-critical processes, managed recovery service levels will become an important basis for improving business resiliency."
>
> GARTNER
> "HYPE CYCLE FOR BUSINESS CONTINUITY MANAGEMENT & IT DISASTER RECOVERY MANAGEMENT"



Failover & failback capabilities

Many BC/DR services offer automated failover and/or failback capabilities. Failover is the transfer of ongoing operations from the primary workload instance to the secondary workload instance.

Failback refers to transferring the operations back from the secondary to the primary. Failover and failback features, at their most sophisticated, can quickly transfer transactions that are currently in progress from one workload to another so that there is no visible interruption to user availability.

At the other end of the spectrum are services that lack failover and failback capabilities, requiring administrators to manually transfer operations from one server to another by redirecting network traffic destined for the first server or otherwise adjusting application access.

Without failover and failback capabilities, administrators may also have to manually synchronize applications and data, which can be time-consuming, slow and error-prone.

Support for failover and failback capabilities varies widely among synchronization methods. For example, orchestration-based solutions offer failover capabilities but not failback. Array-based solutions and traditional backup solutions simply copy storage from one place to another and have no awareness of virtualization, so they cannot automate the failover and failback processes. On the other hand, hypervisor-based solutions are fully virtualization-aware, and they provide both failover and failback support.

Hardware & software dependencies

Some BC/DR services have hardware and/or software dependencies that need to be taken into consideration. For example, the guest-based solution requires the installation and configuration of an agent within each virtual machine, and the subsequent maintenance of each agent.

Other methods, namely array-based solutions, are necessarily dependent on the availability of duplicate storage hardware, while other methods, including hypervisor-based and snapshot-based methods, typically do not have hardware or software dependencies.

Management & testing

A BC/DR service has to be managed from time to time, such as monitoring the service for errors and altering the frequency of data synchronization operations.

In some cases, BC/DR management can be performed through existing cloud management tools; in other cases, it has to be handled by yet another separate management interface or console. The first option is preferred by most organizations because it helps to streamline operations.

“The ability to manage recovery service levels in an automated, repeatable and timely manner is becoming increasingly critical for many organizations,” said Morency. “As Web-based applications support more business-critical processes, managed recovery service levels will become an important basis for improving business resiliency.”

One of the most important facets of BC/DR service management is testing. It’s generally necessary to test a service at least once a year to ensure that it is operating optimally. Consider it a fire drill for data and applications. Some BC/DR services inherently provide support for testing, which can greatly simplify the testing process.

Scalability

Scalability is another important consideration when selecting a BC/DR service. This relates to some of the previous discussions, such as hardware and software dependencies. These can reduce the scalability of a service by increasing costs and delaying deployment (because of necessary purchases, installation, configuration, testing, etc).

Another limiting factor for scalability is latency. Some BC/DR models, particularly snapshot-based services, simply cannot keep up with higher replication loads because there is too much latency. Generally speaking, hypervisor-based services are the most readily scalable.

Performance impact

The final consideration for BC/DR service selection is its impact on performance. Organizations often underestimate how much BC/DR services can negatively affect performance.

The most common criteria include the performance of the applications being replicated; servers that those applications are running on; and networks and storage devices involved in transferring and storing data and applications on behalf of the BC/DR service.

Each service type (synchronization method) has its own implications for performance, and no service type is clearly superior because of the multitude of factors involved.

For example, most services will have a rather minimal performance impact if the application and its data rarely change. Of course, that tends to be rather uncommon. Instead, most applications are frequently updated, and data in particular is often changing.

For such cases, it may be best in terms of performance to use a hypervisor-based solution because it can synchronize based on only the data and application elements that have changed since the previous synchronization.

Other solutions, such as snapshot-based services, have to transfer the entire virtual machine's contents every time they perform a synchronization, and the process of capturing this information may significantly impact server processing capabilities. The transfer of this information across networks to another server may negatively affect bandwidth as well.

Performance under pressure

How your infrastructure performs before, during and after a disaster or critical event is crucial. Keep track of the performance of:

- Applications being replicated
- Servers that those applications are running on
- Networks and storage devices involved in transferring and storing data and applications on behalf of the BC/DR service

Assess & invest

By default, most cloud environments do not emphasize the availability of customer data and applications. They do automatically migrate workloads among servers if possible when server failures occur or are imminent, but they do not make special efforts to ensure that customer data and applications remain available.

It is up to cloud customers to assess their own BC/DR needs and, at the outcome of this assessment, develop and implement a remediation plan that will reduce the impact of cloud downtime and data loss to acceptable levels.

Odds are that this plan will call for the use of different BC/DR services and configurations for different data sets and applications.

For example, the most business-critical data and applications may necessitate very short RTO and RPO, while it may be acceptable for other data and applications to have a much longer RTO and RPO. Making the remediation plan granular in this way optimizes the value of the BC/DR services.

Fortunately, the top cloud providers offer one or more BC/DR services for their customers. There is a wide range of services, from the most basic (traditional data backups) to the most sophisticated (hypervisor-based replication), with other services between these extremes, including snapshot-based, array-based, guest-based and orchestration-based.

> "There is a wide range of services, from the most basic (traditional data backups) to the most sophisticated (hypervisor-based replication), with other services between these extremes ... "

These service types — and the individual service offerings within each type — can be differentiated by several characteristics, such as downtime, data loss, failover and failback capabilities, hardware/software dependencies, management, scalability and performance impact.

No BC/DR service type is the best for every situation. Each situation has its own unique needs. However, some BC/DR service types are clearly much more capable than others.

Hypervisor-based solutions generally offer the lowest RTO and RPO, the most automated failover and failback capabilities, and a minimal performance impact when compared to other solutions. In addition, hypervisor-based solutions can often have their management handled by existing cloud management solutions, and there are no hardware or software dependencies to take into consideration.

Finally, hypervisor-based solutions are highly scalable, and they can readily support any cloud BC/DR needs, from the smallest to the largest. Organizations considering the adoption of cloud-based BC/DR should give serious consideration to hypervisor-based solutions.

> “It is up to cloud customers to assess their own BC/DR needs and, as the outcome of this assessment, develop and implement a remediation plan … ”
>
> GARTNER
> “HYPE CYCLE FOR BUSINESS CONTINUITY MANAGEMENT & IT DISASTER RECOVERY MANAGEMENT”

US 2360 Campbell Creek Boulevard, Suite 525, Richardson, Texas 75082 | Phone: +1 877 262 3473
UK 268 Bath Road, Slough, Berkshire SL1 4AX | Phone: +44 800 500 3167