Aligning disaster recovery to company technical direction and objectives

Received (in revised form): 1st August, 2023

Andrea Houtkin

Disaster Recovery Specialist, Houtkin Consulting, USA

Andrea Houtkin

***Andrea Houtkin** has been a technical project manager and disaster recovery specialist for 30 years, working in various global technical and business environments. Her knowledge of 'what could happen' is informed by feet-on-the-ground experience from over 15 disasters and is central to the business continuity and disaster recovery processes and procedures that she shares with the companies she works with. She dedicates her work to those technical colleagues who lost their lives in the 9/11 attacks.*

Houtkin Consulting,
101 East 23rd Street, #165,
New York, NY 10011,
USA

Tel: +1 212 627 8314;
E-mail: houtkin@houtkinconsulting.com

Journal of Business Continuity
& Emergency Planning
Vol. 17, No. 3, pp. 206–219
© Henry Stewart Publications,
1749–9216

Abstract

One of the many concerns of disaster recovery specialists is how to create disaster recovery scenarios, strategies and related solutions that meet the vision of management while building solutions for the critical business process within budget, with refined technical resources and operational and maintenance processes and procedures similar to those utilised in production. Rather than consider disaster recovery as a separate environment from production, this paper suggests that there are areas where the disaster recovery solution can map more closely to production solutions to better manifest the critical business process, avoiding the decreased sales forecasts and reputational impacts resulting from an outage. There is no magic here — just ideas for designing a solution and enhancements to the disaster recovery programme that may help to meet business expectations. A disaster recovery site based on similar production technical solutions and overall corporate IT vision can provide such benefits as: faster recovery time objective; faster availability of the data while maintaining data integrity; fewer manual procedures during switch/failover; ability to utilise similar resources to work both environments resulting in a smaller training programme; similar operational and maintenance processes and procedures; ability to switchover components rather than declaring disaster recovery; and an environment that supports production by running critical business process while production suffers an outage or requires maintenance. This paper provides readers with ideas to take back to their disaster recovery solution and how it manifests the critical business process during an outage.

Keywords: disaster recovery, business continuity, recovery time objective, recovery point objective, risk assessment, risk mitigation, business impact analysis, disaster recovery scenario, disaster recovery solutions

EXECUTIVE SUMMARY

The chief executive officer (CEO) defines the annual business objectives and key performance indicators (KPIs) for the company based on his or her overall vision for the business. This vision cascades down through the corporate organisational design including the technology division.

The chief information officer (CIO) in turn establishes the vision for technology in support of the CEO's vision.

With the help of their technical management team, the chief technology officer (CTO) manifests this vision into consistent strategies and technical solutions, weighing how/when to upgrade current technology; how to introduce new technology and deal with end-of-life hardware/software — all while ensuring as little interruption to the business as possible — permitting an open path for the technology teams to define how they can follow the CEO's vision and directives — in both the production and alternate/disaster recovery environments.

One of the many concerns of disaster recovery specialists is how to create disaster recovery scenarios, strategies and related solutions that meet management vision while building solutions for the critical business process within budget, and that can ensure, based on corporate disaster recovery policy, that the business will not suffer reputational loss and still be able to meet sales forecasts as agreed with the business.

How close to production can we build? Do we create standalone solutions, hybrid or cloud-based solutions? Should we consider third-party application, database service hosting? Can we ensure that all critical systems will switch/failover from one environment to another without issue or additional manual work? How do we ensure that disaster recovery maps technically to that in production while avoiding performance issues and maintaining business KPI and sales or revenue-generating levels? Can we design/build a solution that meets the recovery time objective (RTO) and recovery point objective (RPO)? Are we facilitating training and cross-training by installing the same technical solution? Are the business and technical staff willing to spend time away from day-to-day project and operational priorities to perform disaster recovery technical and process testing and walkthroughs in an environment that is different from production?

With the advent of the cloud and advanced infrastructure-related technology, disaster recovery services the business most when it aligns to the production environment — its technical strategies and solutions and operational/maintenance processes and procedures are the same in both environments and many software developers and vendors have designed resilient technical solutions to keep within the same service/product scope. The disaster recovery specialist faces a dilemma, though, because we are blocked from considering other solutions should the resilient application result in high costs. If cost is an issue, we could potentially render the application useless to the business, especially if we choose an alternative architecture or method of implementation that is not supported by the vendor. The more we step back from the production solution, the more we could impact application performance and the end-user experience. However, costs can be decreased over time if we use the same resources in both production and disaster recovery and consider ways to bundle the solution to lower the initial purchase and installation costs — or annual support and licensing.

What elements of our design would we map to production? Will the disaster recovery solution drive a change in the production approach? We would first look at business requirements and see how the critical process that the business needs in disaster recovery is manifested in production. Once we understand what disaster scenarios we are building to and what we need to glean from the production design, we can determine the network, infrastructure and application technology and how robust it needs to be based on the organisation's tolerance for financial loss. If the business does not need to perform certain critical business functions until the third day of an incident, then the solution will be different to ensure that only critical and regulated business functions can be performed as part of the phase 1 switch/failover. Which site design does the business need and what can it tolerate: hot, warm or cold? Will we build in the cloud or as a hybrid solution? Should we consider an active-passive or active-active data centre architecture or do we want to migrate disaster recovery to the cloud and various applications/databases to hosted providers so that we can refine what is truly disaster recovery and what is production high availability (switchover within the same region). What do we need to implement in advance to ensure that we can meet the RTO (less work at invocation) and where can we 'wait' to see what is actually required based on the issues resulting from an incident? The closer we build to production, the main benefit that we will see is that the time to invoke is decreased, the applications and software are just as robust as in production, and the infrastructure/application/software can be simultaneously maintained in both sites as long as there is policy to support how closely the disaster recovery solution is maintained in relation to production.

The disaster recovery mission, policies and procedures can be integrated into those already created and implemented for the production environment, and staff can be more readily trained and cross-trained as they use the same technology with the same policies and procedures as production. Disaster recovery becomes part of the solution, not an afterthought, and also supports the CEO's vision through good times and bad. This can hold true whether disaster recovery is hosted internally or via service providers.

If we do anything, we must meet the RTO and RPO. If that means we build up-front and minimise post-invocation work, then we have designed a workable solution. In addition, if the disaster recovery environment is built as close to production as possible, then we have provided an added benefit to the business. We can isolate the site to improve testing of new technology; we can support production by running from disaster recovery during production site maintenance; and we can switchover components from one data centre to another, intra-day, to avoid taking down more technology than required and perform other break-fix work without exceeding acceptable downtime. There is nothing magical here — many companies do this — but it is also important to note that maintenance in one site requires a fast turnaround in the other to ensure that these benefits are viable.

As always, we may have to juggle as building a full or mini-production environment for disaster recovery may be costly, so our presentation to the business and our management may require several options and possibly, be in receipt of some 'no's'.

It is time to stop thinking of disaster recovery strategies, solutions and technical environments as extraneous to production. We now have the technology and service providers to design and implement resilient solutions in both production and disaster recovery. This paper recommends opportunities to integrate the disaster recovery solutions as one with production — a change in mindset.

Why? Because the business process may be narrowed in scope during the first few days of a disaster, yet the business may not be willing to give up current gains or lower its tolerance for financial loss to run critical business processes from 3–5 years ago.

WHERE DO WE START: THE BUSINESS DEFINES THE TECHNICAL SOLUTION

Technology exists in companies to serve

the needs of the business. Our solutions should ensure that we are aligning to the CEO's vision for the business and how the business manifests that vision.

As disaster recovery is being owned more and more by the technical teams within the enterprise, we are beginning to see that sometimes disaster recovery concepts and the business requirements are not included in the design. The environment may not be designed to support switch/failover, invocation and switch/fallback, or policies and guidelines are missing that ensures that the environments exhibit some level of compatibility with current technical capabilities built in production: Our daily practice should keep in touch with the following:

- Maintenance of the current while understanding the future trajectory of the technology that we have implemented and continue to implement in our data centres;
- Maintenance of the fundamental data centre architecture and design;
- Awareness of how data centre, infrastructure, application and software architectures support business requirements, especially when advanced or upgraded technology that gives the business the edge, is implemented;
- Consideration of how to support disaster recovery while the production technology profile is being advanced before a long-term solution is implemented; and
- A new set of questions for infrastructure, development and other third parties that includes whether their solution is sufficiently flexible to be built as high availability between data centres (within the same region) or for switch/failover from one data centre to another (out-of-region).

In addition to technology's future paths, we also need to consider the application of these industry-standard risks identified within the company, the various providers and each critical business application or software: climate/weather; the company's position in the marketplace, the company's financial stability, history, viability of the business' operational processes and procedures as well as continued risks that may result from internal or external malicious intent.

Many in IT consider disaster recovery as a purely technical pursuit. It is not. We are stewards of the critical business process and our designs must ensure its ability to continue or recover during an outage.

The business defines the critical business process, approves the technical approach, costs to manifest these critical processes and the RTO — their tolerance for financial loss. Not every business will support a full mapping of disaster recovery to production; however, they may approve the buildout of just critical services and functions in the alternate data centre. Note, however, each component that is not deployed could potentially inhibit other business and technical functions that may not be performed during a data centre outage.

The business owns the data that we so carefully house, administer, replicate, backup and restore. Only the business can define the required integrity of the data before the data are no longer viable for use during recovery; eg recovery point objective (RPO). What are the switch/failover and switch/fallback solutions for the database and what infrastructure do they depend on? Is it resident in the main data centre or as an outlier, hosted by a database solution provider? How we switch/failover and fallback needs to be considered — do our plans cover scenarios where replication endures during an outage or stops as a result of the incident? What if, in the worst-case scenario, replication stops and the database needs to be re-created once technology switches back to the primary site — is that considered in our design?

Initially, the disaster recovery solution will most likely not be able to fully support the CEO's vision and the CTO's technical strategies and solutions if it does not keep up with the same technology, policy and operational procedures that support production. Furthermore, if the CEO's annual or five-year vision does not include disaster recovery, we may be left with a 'less than' solution, without the support we require, and the need to have that uncomfortable discussion with the business to explain why the solutions do not support their and their customer's expectation that they worked so hard to create and achieve. So, while considering options for approval, also consider a phased-in approach to a robust disaster recovery environment if the company prefers that approach.

How do we justify a disaster recovery solution that maps to the production environment and the critical business process? Start with the CEO's/CIO's and CTO's view of business continuity and disaster recovery within their annual and 3–5-year vision statements. This gives some indication of where they may align funding. Is the goal to be out of the data centre business? Migrate fully to the cloud or third-party hosting facilities? What functions do they want the business to perform during an outage? Look at the business directly and define what is critical, the timeframes for availability in times of a disaster, ease of switch/failover and switch/fallback or site invocation at time of disaster, and how the solution can benefit production and facilitate the business during day-to-day operations. What changes would be required to define a solution that supports the vision, the business-critical process, a site that can support component-level switch/failover (if this is a requirement) and support the production environment by permitting critical processes to continue while the production environment is being maintained?

Some companies may want only to recover (suggesting a gap between impact and hand-off of the invoked site). Some want the business to continue — that requires a more technically robust set of solutions — as here, we need to ensure there is as little a gap as possible. Some businesses may be willing to pay more up-front if they can see the added value of disaster recovery in support of production — during the day-to-day — and not just used as a solution to a data centre outage. In addition, stakeholders may choose this solution if they can feel in control of the technology during an outage by having a solution that is prepared, ready and supportive of the business process, and available in a timely manner — without worrying how long it is going to take to switch over.

THE BUSINESS GUIDELINES FOR DISASTER RECOVERY

The recommended path for this pre-design discovery could include:

- Meetings with the CIO and CEO and critical business heads;
- Analysis of the business-critical processes and procedures, their RTO and RPO;
- The priority for each critical business process, whether they are in or out-of-scope for both continuity and recovery and when they need to be available;
- The business and technical risk assessment and resulting business impact analysis;
- Review the business data classification requirements and tiering that defines the timelines for availability; and

- History of the company, health and welfare, regulatory constraints and any risks that are open and have not yet been mitigated.

Critical business process

Take the time to review and understand the critical business processes, the up/down dependencies and how the process works and then research how they are manifested in production. What infrastructure, operating system or rev level supports the process? Is there a project in place to upgrade the solution in production? Is it a third-party solution or in-house developed? What is the priority of the process within the full set of critical processes and what is the order of switchover and how long does it take to cumulatively fail the full technology stack? Is there a pricing profile that supports either two distinct installations or high availability between data centres? Does the application require certificates and user or site licensing? Is there separate pricing for two data centres? Does the upgrade process require special requirements? Are there defined service-level agreements that we need to maintain with our customers?

Risk assessment and risk profile

Performing a business and separate data centre risk assessment balances the business risk assessment and business impact analysis. Basically, the business-critical process helps to define the disaster recovery scenario and the business impact analysis tells us what we are building and how it is to be built to mitigate the risk points. For example, you may find that what was thought to be an easy solution may be marred by previous outages or by telecommunications providers that may not be able to provide the required circuits and bandwidth. Ask the following questions:

- What grade is the data centre?;
- Has the provider and/or data centre experienced outages within the past year? What was the root cause and how did they inform their customers and communicate with customers during problem resolution?;
- Is the application or software developed in-house or by a third party?;
- Does the development team build security into their applications and database?;
- Does the software/application third party offer a disaster recovery solution or are you left creating two instances of the same application? Does the application support a highly available architecture between data centres?;
- If the business would like to use disaster recovery as a production break-fix solution, what dependencies and risks need to be considered, how will the data centres communicate; do public-facing load balancers play a part in the production data centre and do you have the expertise to configure and use global site load-balancing or software-defined networks?;
- What other utilities, software or databases are the software/application dependent on? Are they listed in the 'in-scope' list even if they are only a priority 2 or priority 3 application?;
- How costly is the disaster recovery solution? Is it worth the price if the disaster recovery process is only tested twice a year?

Recovery time objective

The RTO is our agreement with the business to build solutions that support their tolerance for financial loss. When defining the RTO, both the technical and business RTOs must be added to create a single increment of time. If the business RTO is two hours for critical applications or services, it needs to be two hours — meaning that the design must permit the full technical stack to switch/failover and switch/fallback in two hours, not three or four.

Most companies look at how long it may take to switch/fail a single application without consideration of the time to switch/failover the full technical stack that supports it. How long does it take to switch/fail the database that the application is dependent on or how long does it take to propagate DNS changes through the network?

This is a good reason to create a switch/failover and switch/fallback timeline that defines the reality of how long it really takes to invoke the site and in what technical order. If you have 200 critical services and applications, that adds up. The timeline helps us to choreograph this function and keep our agreement with the business. Remember that many companies still have legacy infrastructure, applications and technology that are still deemed as critical to the business. They may take longer to switch/failover. It is necessary to include these legacy resources in the overall switch/failover timeline — as this will help you to understand whether you can meet the requested RTO and RPO.

The business sees the RTO as a pause in the company's ability to generate revenue or perform critical processes/services for their customers. Every hour has a financial value. Create a chart that includes the financial loss thresholds by RTO and hourly cost for the prioritised list of critical processes. This is the truth that must be shared with the business and can be an important criterion in determining how the solution is architected and how much of the solution will be built. The current technical solution may not be able to meet the organisation's RTO. How do we respond to that concern? There is no magic here: it may result in the need to revisit the design process perhaps two or three more times or search for a new service provider.

Not all business processes require a 30-second RTO on the same day as the impact. Some processes do not need to be engaged until day three after an outage, for example. Some set the thresholds between data centres based on seconds or the loss of a data centre heartbeat. The business provides availability timelines for the critical business process and lets us know what they are willing to accept. It is our job to design a solution that can ensure a handshake between business process and technology, between incident and time.

Recovery point objective

The RPO is a time-based measurement that defines the maximum amount of tolerable data loss. So, if the data required for an application to work effectively can be no more than six hours old, the RPO is six hours. This requirement can define how we replicate data between data centres or availability zones (AZs), the replication schedule, and the backup and retrieval solution. Data classification analysis defines the priority of the data based on its capture or use as part of a critical business process. Our job is to map data availability and integrity to the business process and ensure that it is stabilised by database solutions and infrastructure that keeps the data available and where corruption can be kept at bay. This solution includes the database, the infrastructure where it resides, the storage area network and the backup/restore solution and processes. If our RPO is three hours, it cannot take six hours to retrieve that last committed piece of data.

At issue here is the location of the database — and the time it takes for switch/failover. If implemented within the same network as the applications, it could be faster and easier to switch/failover. If located within another cloud or network, there are firewall requirements as well as bandwidth concerns that could potentially impact the time to switch/failover and in worst case, failover, the reconstruction of the database in production during fallback.

Disaster recovery scenarios

Disaster recovery scenarios require some level of thought. They must be tied to the business-critical process yet not be so detailed that they become difficult to create, build to or explain. These are our guidelines not for only business continuity planning but for determining what disaster recovery solution is implemented.

Start with the easiest. For disaster recovery, it can be loss of a critical production data centre where the impacted data centre is dark or perhaps partially lit, adding additional process to avoid having two instances of the same application/script, etc. run simultaneously.

Be simple yet ensure that your individual scenarios (if you have more than one) can cover the full set of impacts to the critical business process that would result in a need to switch/failover to disaster recovery.

DEFINING THE TECHNOLOGY SOLUTION TO MEET THE BUSINESS REQUIREMENTS

Map the critical business process to the production technology. When a solution is defined, consider both production and disaster recovery requirements simultaneously to determine whether all layers can work in both data centres. The devil is in the details, but each detail can help determine what is being architected and designed as well as the total cost — in financial terms (the spend) and loss because of a slow or laboured switch/failover. In this activity, we are not talking about infrastructure/application/software device configuration — we are also looking to understand how that technology works in the midst of the technical stack within the data centre. Consider the following:

- Is the overall design to recover or to continue the critical business process?;
- Is the network design resilient or redundant?;
- Are changes required to supporting infrastructure and operating systems?;
- Are changes required to support advanced technology or additionally robust applications that exist in production?;
- How will legacy applications/software be treated?;
- How will active-active or active-passive be designed? Will you need to build out a complete copy of the active data centre in a passive or active-active design or just critical infrastructure/applications, etc.?;
- What policies are required in support of the final design? For example, it may be prudent to have a policy that states 'all critical applications are to be designed as two separate instances in two remote data centres or AZs or as HA (high availability) between data centres and AZs, and the design must support the switch/failure process based on the scenario' or 'all infrastructure, systems and applications must be designed to run in production and be readily available in disaster recovery in case of a component-level outage in production';
- Delineate all internal and external up/down-stream dependencies for each technology, how they connect and whether they can connect as easily to and from the alternate data centre;
- Review the database solution and how the application/server connects to the database and define a consistent strategy for all database connectivity;
- Review the operational and maintenance processes and procedures for

the production solution and whether changes to production and disaster recovery can be performed through the same work order;
- Define the technology-related switch/failover and switch/fallback strategy and processes/procedures. Determine whether the implementation in production needs to be refined to support the solution in both environments;
- Research known issues that have plagued the production solution or provider;
- Research open technical issues.

Delineate the assets and related components for disaster recovery

Many companies use end of service life (EOSL) or older models for their disaster recovery environment. Remember that the business bases their metrics and KPIs and forecasts on the current technology and taking a step back may result in infrastructure that cannot support the current critical business. It will be difficult for the business to have to factor in 'loss' because of an incident when the technology can ensure that the business can maintain its KPIs and forecasts.

Migrating to the cloud may result in a 'no more hardware' policy from management. How will you determine which older infrastructure to use, if required?

Perform an asset inventory of your critical production systems and technologies that support the business-critical applications. Take note of the infrastructure: server configurations, operating system, bios and idrac versions, code, release process and validation. Determine whether an actual replacement or upgrade is required, or whether the technology requires additional, manual switch/failover procedures. Switching or failing to an alternative data centre with older technology could result in performance issues and more support calls. If the alternative data centre is implemented using the same technology, operational and maintenance processes and procedures, we can continue or recover critical business processes to a business-as-usual standard, as well as support a longer outage.

Stand-alone; active-active; active-passive: hot, warm and cold

Keep in mind that the further your design moves away from 'hot' or 'active', the more time it may take to invoke the alternative site because of the additional switch/failover steps that are required to be performed.

If you ultimately want to have an active-active relationship between data centres or AZs, consider a phased-in approach by building the alternative site as passive and focus on the enterprise-supporting technology to switch/failover on the datacentre layer, not just on the application layer. The enterprise-supporting technology may need to be the same in both environments. This permits a more controllable and standardised switch/failover.

No matter what is being implemented — include testing or shakeout with every technology that is implemented. It is better to know where there are gaps as you build — not after the stack is complete or worse yet, during an incident.

Data centre fundamental requirements

Perform a review of the production technical requirements and based on the business requirements in the business continuity plan, consider the relationship between production and disaster recovery:

- Will the number of application transactions increase or will the payload increase? If so, can the bandwidth between data centres support this?;
- Has the traffic flow between data centres been tested? Are there any impediments?;
- Perform load-testing on the network on

a regular basis — both production and disaster recovery;

- Determine circuit and bandwidth requirements between production and disaster recovery and between the disaster recovery and external third parties;
- Can the wide area network (WAN) and local area network (LAN) design handle traffic for two sites? At points during switchover and switch/fallback, there may be additional traffic on the wire;
- How much bandwidth would be required if there is a failover of the database and it needs to be recreated in the production data centre as part of the fallback process? How long will it take to recreate the database?;
- What other services are required to support the critical business process for the planned phases of an outage? Can the network support?;
- What WAN design is supported in both environments: resilient (two circuits, two carriers) or redundant (two circuits, same carrier) because of cost; what WAN circuit type and protocols are utilised? *NB: For a resilient design, additional processes and components (autonomous numbers) may be required to keep two carriers playing nicely in the sandbox*;
- Remember to include edge security technologies (eg intrusion protection, detection, firewalls, security information and event management, application programming interfaces) used to protect production. For example, do not allocate ports on firewalls dedicated to disaster recovery to respond to production requirements. This is why it is better to implement disaster recovery changes as part of work orders for production changes so firewalls, for example, already have the ports for disaster recovery allocated to disaster recovery before an outage;
- Connections to third parties should be consistent between environments.

Operating system services

How will production traffic be redirected in light of a production outage? What is the strategy and solution for the redirects? Via DNS push, for example? There must be a clear understanding of when to perform a DNS push on a company-wide level and how long it takes to propagate through the network.

Active Directory should be replicated from controller to controller via the company's Active Directory replication solution.

If using DNS for traffic redirects, has a DNS push ever been performed and timed? How long does a DNS push take to propagate through the network? Are there areas in the network where some level of reconfiguration or redesign can support a faster propagation speed?

The database

Special care is required when designing the database solution. How will systems access the database? How will duplicate requests be handled? What is the replication methodology?

A database switchover suggests that the database can still replicate between production and disaster recovery in this case the disaster recovery database becomes the primary repository and the production database becomes secondary.

Failover is defined when the replication methodology has been impacted. There are two concerns:

- Unless preparation for a safety repository or local backup is made, there may be only one database running during the disaster recovery event;
- To fallback, a new database may need to be recreated in the production site — and based on the bandwidth between production/alternative sites, can take more time than considered — impacting the ability to normalise back to the

production environment. Consider both switch/failover and switch/fallback in your design and processes/procedures.

Determine whether the current storage solution is robust enough to support the full disaster recovery response. How quickly can the data be back online? How long does it take to recycle the application servers to connect to the alternative data repository?

Look at your current backup/restore solution, schedule and process. How quickly can data be restored? Test this solution every year — and depending on identified gaps, test again. Determine the time to find the correct data store and time to restore.

Implement a ransomware solution for the database such as air-gap (may be difficult in the cloud) or immutable storage.

Application-layer standards and guidelines

The application is a layer within the full technical stack. We cannot lose sight of that because if we do, then there could be gaps regarding up/down-stream dependencies, hooks into tools, infrastructure, utilities and connectivity to third parties. Think end-to-end process to understand the critical process flow — overlaid onto the disaster recovery topology. When analysing the applications that are required in disaster recovery ask the following questions:

- Are there corporate development standards and guidelines? Are the Open Worldwide Application Security Project (OWASP) and National Institute of Standards and Technology (NIST) guidelines for application security included in the standards and guidelines?;
- Is the architecture and implementation consistent from one system/application to another and can it support expansion to another AZ or data centre?;
- If using third-party software, does the vendor offer a disaster recovery solution that maps to the business requirement and the data centre strategy?;
- Have applications/technologies that are considered non-critical yet support a critical business process been included in the 'in-scope' list and implemented at the alternate/disaster recovery site?;
- How will the application integrate into the overall design (eg active-passive solution) and what might it take to create an active-active data centre relationship if required to map to the RTO? Is this a goal worth pursuing?;
- How will you determine whether the impacted site is still lit if the strategy is not to isolate the site at the WAN. This could result in duplication of jobs/scripts during switch/failover or invocation and additional invocation steps. Plan for this in your timeline and whether it impacts RTO;
- Look at how applications connect with their external partners and services and the traffic flow of data.

Operations and maintenance

Here is where creating a disaster recovery solution as close to production as possible can pay off. How many resources are required to perform daily operational/maintenance processes and procedures; how long do they take and what is their schedule in terms of day/time (daily, monthly, etc.)?

Check your support, licences, certificates and service-level agreements for all engaged services, hardware and software supporting the critical business process. Understand the full cost of using them in an alternative data centre. Identify the best methods of keeping the disaster recovery environment up-to-date and in sync with production.

The full data centre switch/failover

The design is not complete until you understand how each technical solution will switch/failover or switch/fallback. Are there manual processes required or can it be automated? Are there any switch/failover processes that can be further refined? Does the full stack need to be invoked on '0' hours? How long will the full stack take to switch/failover and can you meet the RTO/RPO. Look at those areas where strategies/solutions can be refined and determine how this can be achieved.

Create a detailed switch/failover timeline of all the technology in the correct order. Practise this before an actual test as a walkthrough. Look at possible technical changes from year to year and always add the total time to fully understand what is required for a full switch/failover.

Many fail to consider switch/fallback. It is crucial that your design can support both directions.

How exactly will you switch/failover? Thresholds on a public-facing load balancer? Use of the global site load-balancing protocol? Software-defined WAN? Site isolation with a DNS push? If your data centre switch/failover strategies are defined at the load-balancer, remember that this mechanism is incredibly fast. There may be critical legacy or external applications that may not move as quickly. The technical RTO may need to include when legacy technology can be available as well as the current, faster technology.

Technical runbooks

Not all technical shops include disaster recovery switch/failover and switch/fallback processes and procedures in the production runbooks, and some do not have production runbooks at all. Create them for each critical technology/application and include production and disaster recovery processes and procedures within this single document. This helps focus on normalising disaster recovery with production and ensures that procedures used in both environments are only written once. A consistent format is required to ensure that anyone on-call can support invocation. People can often be stressed during an incident, and having the same format facilitates the invocation process by obviating the need to fumble around looking for procedures. Should a data centre be included in the seating facility, and if there is an incident, you may need to bring in third parties to perform the invocation while you account for staff.

Testing

Create a robust testing programme of walkthroughs and actual tests. If active-active, some companies switch/failover and run individual components or the complete production environment from disaster recovery for two weeks to several months to look for gaps and maintenance issues or give them a chance for maintenance at the production site. As always, be very clear about the scope and methodology for the test. We may need to revert to walkthroughs if there is any risk to production or customer services.

- Most importantly, it is important to understand how each technical solution is validated;
- Is component testing performed for each deployed solution before end-to-end processing?;
- Can all critical applications work from the alternative site without dependencies on the primary site?;
- Is end-to-end testing of each critical business process performed?;
- Can the end-to-end process work without dependency on any technology in the primary data centre?;
- Is a walkthrough and test performed to

validate the data centre switch/failover processes?;

- Do you perform an actual switch/failover and fallback testing at least twice a year?;
- Is the backup/restore process and solution tested at least twice a year?;
- Is circuit bouncing between primary/secondary circuits tested with your carrier(s)?;
- Do you test with critical service providers?

SUMMARY

Disaster recovery is changing by virtue of the technical opportunities that are now before us. For some companies, disaster recovery is a separate environment that contains the baseline services to support the business for a short period of time and is considered only while planning and performing the annual or biannual test. Some companies do not even test switch/failover or switch/fallback, relying only on walkthroughs to test the invocation timeline. Others will see the value in a peer-to-peer data centre design, facilitating switch/failover and switch/fallback, that supports production data centre maintenance and opportunities for production break/fix during day-to-day operations.

If the business looks at disaster recovery as an insurance policy, it may simply require the disaster recovery solution to be available only in the event of a disaster, through a hosting service rather than paired with production, and built using older infrastructure (EOSL) and earlier versions of software. However, if the production environment utilises new technology and more robust infrastructure, it behooves the business to consider the additional costs in upgrading the disaster recovery environment to map to that of production. Otherwise, it can be difficult to continue business because the solution requires recovery that is dependent on a longer, more manual switch/failover process.

Do we need to build all of production technology into the disaster recovery or paired data centre at the get-go? No, but whatever technology that supports the critical business process, if built, should map to production solutions and technology — this is the only way to ensure that the business can maintain its KPIs and forecasted sales during an incident.

A fully active-active data centre pair could be a goal — whether via standalone or cloud-based data centres/AZs but it is best considered as a long-term goal and implemented in a phased in approach (starting with an active-passive architecture) if money and resources are an issue.

We now have the technical capability to consider the alternate AZ within the same region as a solution for a production intraday outage or outage of a complete application. In these cases, we do not need to declare a disaster for a single component outage. It is difficult to do that with just two standalone data centres, but it can be implemented using load-balancing technology, for example, and a good understanding of the critical business process, the RTO and whether the technical solution permits. Beware, however: kludged solutions are not the answer. Every kludge requires someone to remember, to document and to ensure that it can integrate into the standards and guidelines defined by the corporation.

The technology also allows for a faster switch/failover — however, we need to keep focused on the data centre as a technical stack, not just a solution for the application layer. We need to understand each layer of the enterprise-supporting technology, the order of switch/failover and the technical order of invocation. Ultimately to reach this goal, the design would include the same enterprise-supporting technology that permits this flexibility and removes additional switch/failover steps or reduces the time to perform them.

There is work to be done to come to the final decision of whether to move towards peered environment. However, if we take too many steps back from production, the application, designed to run on more robust infrastructure may not perform as expected. This results in additional work for the technical teams and possibly additional calls for the customer support teams. We may not be able to recover production in the time requested by the business and the current data may not be properly synched or may be unavailable due to an inadequate back/retrieval process or solution.

When a CEO defines a vision for the business, that business vision is manifested in technical solutions and the technical solutions built in disaster recovery should rightly support the business as it does in production. Technology is faster now, more robust and any incompatibility between production and disaster recovery will be easily identified.

Look to production and how production responds to the business need. Consider a change in thought from disaster recovery as a point-in-time solution, an insurance policy, the outlier. The CEO's vision for the business requires the technology team to design, build and maintain solutions in both the production and disaster recovery environments whether during the day-to-day or an incident.

Remember also that a disaster can occur at any time — and it most likely will not be defined in your set of disaster recovery scenarios. Being prepared and maintaining the disaster recovery solution as production can ensure that we are better prepared to manoeuvre following such events.

If budget is a problem, consider the bare minimum — what you need to really recover the critical business process in times of an outage, or look at your opportunities to plan and design a solution that may be more appropriate at a later time, implemented using a phased-in approach. But try not to go backwards. If you have the support, try to keep disaster recovery as close to production as possible. It will save your company's reputation from lost revenue and possibly loss of market share.