BCP/DR Primer – Part 2 – BCP Tiers and recovery requirements

In Part1 of our BCP/DR Primer series we defined some key terminology that we will use as we build up our plans and document our systems as they map against our ability to recovery them.

The BCP Tiers are common among many organizations although you may seem some deviation in the specific time-frames that they use to define them. I have used the same definitions with multiple organizations and these are what I will be showing you here.

Each BCP Tier is defined by RPO and RTO. These are both required for the system(s) to be included in the tier. The reason that we have to use both RPO and RTO is that we cannot consider a system as being tier 1 by RTO, yet the RPO is 5 days. What we mean by this is that there is no value in being able to bring a system online very quickly, yet the data is far out of date. When we are defining the BCP tier we use the lowest common denominator of the system recovery requirements. This will make sense with as we walk through the BCP tier definitions.

An important note is that these are IT recovery definitions. Many businesses may refer to having "tier 1" systems that have a requirement to be recovered within 48 hours. Quite often these terms and definitions come from regulatory requirements where a company must have certain capabilities within 48 hours of a Significant Business Disruption. It is important to differentiate between business definitions and IT definitions.

Tier 1

- **RPO**: <15 minutes
- **RTO**: < 1 hour

This is your highest availability and highest recoverability category. The key definition for a system to fall into this environment is that it requires near-zero or real-time replication of data and automated failover to the alternate site. If the data is replicated, but the recovery requires manual intervention you are immediately pushed into Tier 2. The reason for automation being a requirement is that we cannot assume availability of people resources when a disaster event takes place so a manual recovery may take longer than 1 hour. A majority of your Oxygen Services should be in this category.

Tier 2

- **RPO**: < 4 hours
- **RTO**: <24 hours

Tier 2 is the sweet spot for many systems. The operational costs to maintain a true Tier 1 environment is often beyond the reach for many application environments. Requirements for Tier 2 include replicated data and hot or warm standby environments in the alternate site. Any environments that will require recovery from tape/archive data are automatically moved into Tier 3 or even into Tier 4. This Tier will include the remainder of your Oxygen Services which ensures that the other IT application environments and your business application environments have the required infrastructure to be brought online at the alternate site in Tier 2 and below.

Tier 3

- **RPO**: <24 hours
- **RTO**: <48 hours

Tier 3 is going to be a heavily populated recovery Tier also. Requirements for Tier 3 recovery include a minimum of cold standby systems (may also include warm) and data should be in place, or recovery from archive is possible provided that archive content is available at the alternate site within the RTO window.

Tier 4

- **RPO**: <24 hours
- **RTO**: <5 days

The final category is Tier 4 which will encompass all systems remaining, including those with manual recovery processes, data recovery from archive for medium to large data sets and anything which is deemed as “less important” than systems which are categorized as Tiers 1-3. Remember that importance is defined by the business sponsor, and cost of recovery may impact how systems become Tier 4.

Where is Tier 5?

There are few environments which have a Tier 5 simply because any system which would fall under this category is a very manual process or system and may even be just paper data which is stored off site in archival storage. While we may have some systems that could live in the Tier 5 space, usually through the BIA process you will find that the business will want these systems classed as Tier 4.

You may find that Tier 5 exists for your organization so it should not be assumed to always be unnecessary. I have just found that in practice that most organizations require recover in one way or another within the 5 day window. If you define Tier 5 it is typically an RPO of 24 hours and RTO of 5-14 days. The Tier 5 category is what we would call “everything else”.

Why the Lowest Common Denominator?

LCD isn’t just for mathematics. When we define the BCP Tier for an application or system, we have to use the lowest common denominator for the RPO and RTO to decide where it belongs in our recovery plan. What we mean by this is that you should not have a system that has near-zero replication of data, yet the core application requires 48 hours to be recovered. While this may happen, we have to define this application as Tier 3 because regardless of the currency of the data (<15 minutes) we still require the 48 hours to recover the whole application environment.

What is zero hour? The Declaration of Disaster

This is a very good question that is hot topic among the BCP community. When does the clock start? There are 2 events which will occur in a Significant Business Disruption or Disaster Event. First we have the actual event itself, followed by the “declaration” by the business sponsors on the BCP team who will officially initiate BCP recovery to an alternate site.

As an example, let’s use a localized power loss to your data center. Assume for our example that you

have 6 hours of UPS available to withstand an outage. We must also assume that our network provider(s) have similar protection to provide you with connectivity during the time when you are operating on UPS.

While we have an immediate issue with the loss of power, we have been able to maintain our Tier 1 systems. During the course of the first hours (less than 6 in this case) we must prepare to fail-over environments if we continue to suffer the loss of power. We may also find out that we will get our power back after 8 hours. In this case the BCP team may declare the start time as of 4 hours into the loss of power once we come to a agreement that we will initiate some fail-over procedures.

So for this example, the “zero hour” is actually 4 hours beyond the actual event. This is just an example to illustrate that the event may not always be considered the start point because we consider the declaration of disaster to be the true starting point where we enact the recovery plans.

Other Restrictions which affect the BCP Tier

There are a number of things which can negatively affect the recoverability of a system. Most importantly, we have to be aware of many issues that will arise that can, and will affect the BCP Tier definition of an application or system.

If we look at the distribution of applications among the BCP Tiers we will see that many will land in Tier 2. The challenge that we have with that is that if there are too many in Tier 2, the manual intervention and time required to complete recovery of many of those systems may exceed the RTO of Tier 2. In other words we may have packed 2 days worth of work into a 24 hour recovery window.

For systems that require data restore from archive, you have to carefully measure the recovery time. While some tape recovery can still qualify as Tier 3, if the data size is too great this will also push the application into Tier 4.

It is important that we evaluate the overall recovery plan along with the individual itemized plans to be sure that we correctly categorize application recovery. If we find that there are too many in the Tier 2 range then we must either modify the recovery process to move it to Tier 1 or we must work with the business to prioritize it accordingly and move some applications into Tier 3.

Mapping the Application to a BCP Tier

I will use 2 examples to show how to map the application into the appropriate BCP Tier. It will become simpler to do this process as you spend more time with it, so the key is choosing applications with which you have familiarity.

System 1: Microsoft Active Directory, DNS and WINS

> It’s really 3 systems, but because I co-locate them on a single server (two or more in each network location). This is an easy one for BCP Tier definition because by design it is a geographically dispersed cluster with data transmitted every 5 minutes and the longest wait duration is 15 minutes across the WAN.
>
> **DESTINATION**: **Tier 1**

System 2: PHP Web server on Microsoft Windows with MySQL

> The web application is protected by nightly Robocopy tasks and the database is backed up to

disk using mysqldump and the content is Robocopied to the alternate server along with the web application code. The destination server automatically imports the nightly database dump. All that is required for recovery is to modify the CNAME record to point to the new target server.

DESTINATION: **Tier 2**

It's just that easy Of course you will have to try this out with your systems and through evaluating each one you will get much more familiar with the recovery processes and the mapping of applications to BCP Tiers should become easier as you go.

What's Next?

In our next post we will build the BCP Recoverability Matrix

BCP/DR Primer – An introduction to planning your IT recovery strategy – Part 1

You see the acronyms all over the place, and it can be very overwhelming. There are so many articles about cloud, SPOF, RPO, RTO, five 9 up-time and in the end it is really all meaningless unless you understand your own environment.

A common misconception is that by putting your data "in the cloud" that you have a Disaster Recovery (more often caled BCP, or Business Continuity Planning) strategy. Wrong! There is much, much more to protecting your data, applications and business than simply copying your data to somewhere that boasts (and I stress the word boasts) a "99.999%" uptime.

This series will cover the fundamentals of IT BCP strategy with a split focus on technology as well as the business requirements to define your overall BCP program. Because BCP is just that, a **Business Continuity Planning** strategy, you have to begin by understanding the nature of your business and what is required to meet the needs of your staff and your customers.

BCP Speak

Much like what I call "project speak" there is "BCP speak". These are the basic, and much used acronyms and phrases when defining your BCP strategy. You've most likely seen these before but it is key to understand these key phrases and features of a BCP program.

RPO

Recovery Point Objective. This is the point in time that your data will be recovered to in the event of a disaster event. The most common RPO is 24 hours because we assume that you have a backup from within a 24 hour period which you can restore from. This does not mean that it takes 24 hours to get your system online, but it means that whenever your system is recovered that the data will be

up to 24 hours old as of the disaster event.

RTO

Recovery Time Objective. This is the duration of time to recovery your system(s) after a disaster event. While you may have an RPO of 24 hours because you have a 24 hour old backup tape, it may take 3 days to recover that system or data thus resulting in a 72 hour RTO with a 24 hour RPO.

SBD

Significant Business Disruption. This term is used because we are not always faced with a “disaster” but perhaps we have situation like your primary data center loses power for 8 hours and you only have a 2 hour battery resulting in a 6 hour outage. This is not a “disaster” but a “disruption”. It may be as simple as the transit being shut down for some reason which impedes your staff from getting to their place of work for a day or more.

BIA

Business Impact Analysis. This is a planning and documenting process where IT and the business coordinate to document and define the needs for a business process. This will include people, processes and technology requirements and from the BIA we will be able to define the RPO and RTO for the technology components.

Alternate Site

This term simply refers to a secondary site which is usually more than 100 km (60 miles) from the primary site. This is usually referring to the data center where servers and infrastructure are held.

Synchronous versus Asychronous

Data synchronization is done in one of two ways which is either Synchronously, or Asynchronously. Synchronous replication means that the data is current in both locations and when data is written to the primary location, it is simultaneously written to the secondary location. In database terms this is referred to as “dual commit” where the transaction is not considered to be completed until the second write is confirmed.

Asynchronous replication means that the data is written to primary location and then sent to the secondary location to be written as soon as is technically possible. The advantage to asynchronous is that while there may be a slight delay in the writing of the data to the secondary location, there is less latency because the transaction is completed as soon as the primary data is written.

When your primary and secondary site are separated by more than 100 km you will find there are technical limitations to providing sychronous replication. This is a matter of physics and at this time is not up for debate unfortunately.

Near-Zero

With asynchronous replication we may still have “near” to synchronous speeds but because it is not a guarantee that we have that speed, and that we know for a fact that the transaction is not synchronous, we often refer to this as Near-Zero because it may be under a minute, or under 5 minutes. If your system has limited data change you may be able to be comfortable with under 15 minutes as the threshold for data synchronization.

Real-Time

Much like Near-Zero, the term Real-Time is used to describe the currency of data in the secondary site. Data transfer which is referred to as Real-Time is usually synchronous data replication.

Hot, Warm and Cold Standby

The terms hot, warm and cold when referring to standby systems are used to refer to the currency of the data and availability of the system in the alternate site.

Hot standby is defined as an online recovery system with synchronous data, or possibly asynchronous, but near-zero replication. A hot standby system is immediately available in the event of an outage in the primary site.

Warm standby is defined as an online recovery system with asynchronous data replication. While the data transfer may be as close as near-zero, there is some recovery required to bring the system online to recover that service. This may be automatic or manual.

Cold standby is defined by hardware or virtual systems available in the alternate site which can be used for the recovery of a system. There is manual intervention in restoring data and bringing the system online, but it can be done without the purchase and installation of hardware or software.

SPOF

Single Point of Failure. This is where a system has one or more component, which if removed, renders the system unusable. Even when we build redundancy into systems, we often have a single point of failure. A simple example is a web application which has a database connection, yet there is only a single database server. If the database system were to go offline then the web application would also become unavailable.

BCP Tiers

The tiers in BCP are defined by RPO and RTO ranges. Using these tiers we map the business requirements against the BCP tier and this will define the technnology and people factors involved to maintain a recovery strategy within a specific timeframe.

High Recoverability versus High Availability

A system may be fully redundant by having multiple paths to a database, or mutliple servers in a site servicing the delivery to the customer, but redundancy is often done locally with low latency technology. This is what we refer to as **High Availability** because it has the ability to be available despite a number of local distruptions or events.

High Recoverability is the capability to recover the business or technical function in an alternate location. It is entirely possible to have a system which is higly available, but with limited recoverability and the reverse is possible also. The example of High Recoverability could be a simple standalone application server with no local redundancy, but a warm standby in an alternate site.

Oxygen Services

The term Oxygen Services refers to core IT systems which are required such as network, name resolution, directory services. This name was chosen because without these services we cannot do

anything else. These are the essential systems and services requiredbeforewe are able to recover our business systems.

What’s Next?

Our next post in the BCP/DR Primer series will take the key things we’ve spoken about here and we will look at how we define the BCP Tiers for our environment with our initial focus on Oxygen Services.