SITE SELECTION IS THE FOUNDATION FOR YOUR BUSINESS CONTINUITY STRATEGY

April 2020

The topic of Disaster Recovery (DR), and its big brother, Business Continuity (BC) is one of the most misunderstood in the world of business. It is not misunderstood because it is necessarily complicated; the ability to replicate, store and then access data is not a new one. Typically, the confusion comes when trying to comprehend the massive scope and amount of resources these projects require. Most organizations are simply not prepared for the many months or even years of pre-work necessary for just one bit to be replicated and "protected". In addition to the daunting scope, DR/BC takes ongoing care. True DR/BC touches every aspect of the business, spanning IT, Human Resources, Executives, Board and shareholders.

No one argues that data is one of the most valuable assets any company owns. It is a well published fact that over 50% of companies who lose data are out of business within 24 months. Imagine a large company losing their accounts receivable, or customer account information, or stock transactions worth billions! Ransomware adds a whole new dimension to the concept of declaring a "disaster".

The challenge with Disaster Recovery is the fact that it means so many things to so many different people. Interdependencies between disparate data sources and the applications using that data must be identified and documented prior to any "real work" beginning. This confusion gives way to sheer terror when the C-suite realizes the ongoing budget implications. DR/BC planning is the very definition of "Multi Factor Decision Making". There are 1000s of critical decisions to be made during the journey, any one of which can completely blow the project timeline or —even worse— derail the project.

We're not going to try and boil the ocean with this paper, but we are going to propose a methodology to start the process. There's one decision you make which will impact all the others, and that is what we are going to address here.

Before we get into the specific criteria, we need to build a common dictionary of Business Continuity terms. As with most of the IT world, DR/BC is alphabet soup when it comes to the acronyms. Even more confusing is the tendency for each element to start very broad and then splinter into ever smaller branches, influenced by business type, budget and governance/SLA requirements. Establishing a baseline knowledge prior to engaging vendors and internal sponsors will cut months from the project timeline. Understanding the many interdependencies and then applying that to the site selection criteria will mitigate risk, reduce costs and potentially avoid a disaster after the disaster! Let's get started understanding the basic concepts and definitions of DR/BC:

Business Continuance (BC)

This is an over-arching concept considering all aspects of the business, not just the data and IT infrastructure. BC includes the human resources planning, access controls, alternative office space, etc. Disaster Recovery defined below is a sub-set of your BC planning.

Disaster Recover (DR)

Typically defined as a set of policies, tools and procedures to enable the recovery or continuation of vital technology infrastructure and systems following a natural or human-made disaster. This makes it sound easy, it is not. DR refers to all of the hardware, software, people, documentation, etc., you need to resume IT operations. If you are managing 1000s of servers and petabytes of data, it quickly becomes clear you are in the deep end of the pool.

Recovery Point Objective (RPO)

How much data can you afford to lose? In a large organization, there may be 1000s or even 10s of thousands of transactions per hour. The answer to this defines your RPO. This is the most impactful element of DR; it will drive all of the other architectural and budget decisions. Example: "Acme Corp. has defined acceptable data loss of no more than 15 minutes". In order to achieve this requirement, you must implement hardware, software, telephony and facilities to replicate all "critical" data.

Recovery Time Objective (RTO)

Defined as the targeted duration of time and a service level within which a business process must be restored after a disaster, in order to avoid unacceptable consequences associated with a break in continuity. RTO and RPO are typically considered complimentary. RTO is essentially how much time can you afford to wait (lose) for critical business functions to resume. As an example, restarting and redirecting applications to new storage volumes may take many hours. So, the RPO is 15 minutes, the RTO may be 8-hours+. What if less time is desired? You guessed it; more architectural decisions must be made to mitigate the RTO impact.

Source

Where the data originated. Companies typically call this "production".

Target

Where you are sending the data to protect or use it for other purposes.

Active/Active

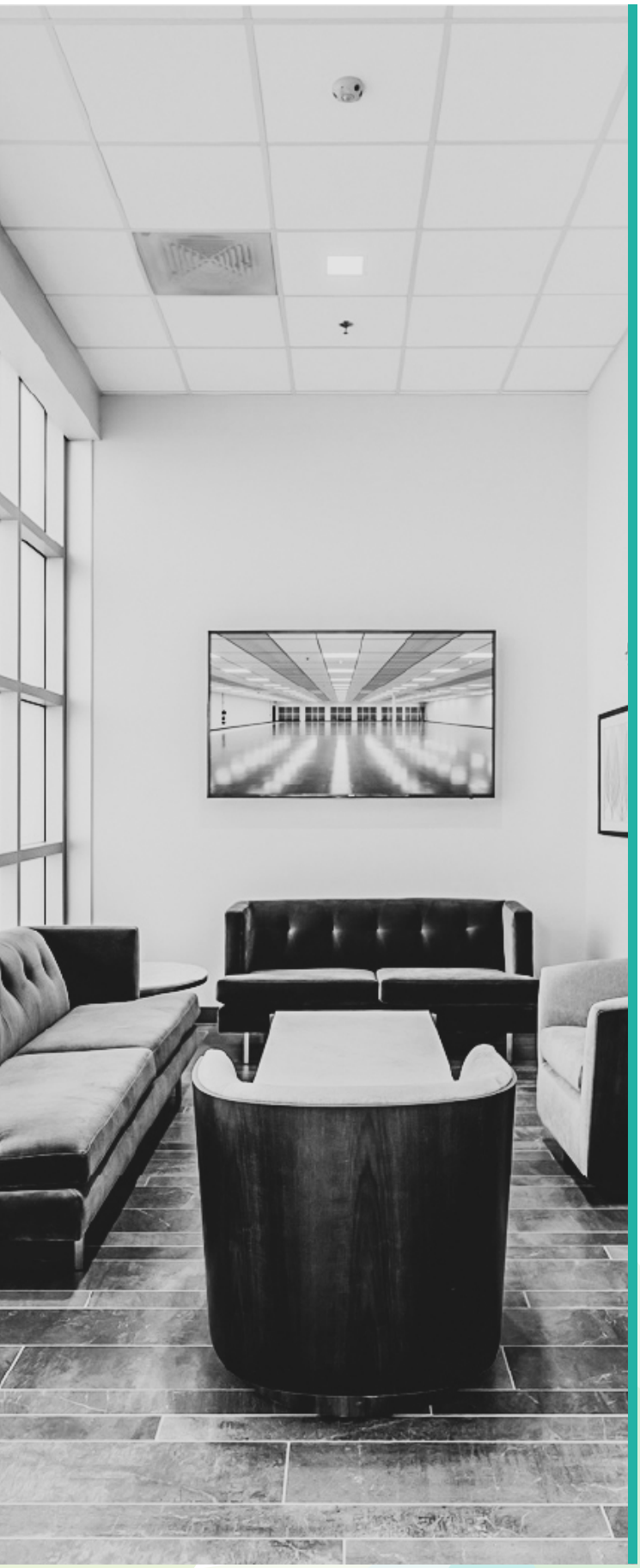
An architecture where all data is replicated in real-time between the source and the target, either one of which can become the source or "primary" in the event of a disaster declaration. This is typically a scenario where the target is also leveraged for IT operations such as test/dev and backup.

Mirror

A full copy of the source with no data out of synchronization. This essentially provides a RPO of zero.

Okay, that's enough as far as definitions. Are there more? Of course, hundreds more. However, these are the big picture items which drive the scope and therefore the budget of Business Continuity. We suggest adding each of these criteria to your decision process and assigning a weight based upon your company's RPO and RTO tolerances. It may be difficult to find a perfect location that meets all of the criteria below, so a scoring process based on weighting is typically employed.

Criteria	Description	Critical Target Site Considerations
Natural and Human-Made Disaster	An appropriate DR site should minimize the influences of natural disasters. These are categorized as weather related and geological related. Typically, an "acceptable distance" is 50-100 miles, but there are other considerations as noted at right and below.	The target should be separated from as many common geologic and weather criteria as possible. Consider hurricanes on the East Coast. 500 miles between sites may not matter in the event of a massive storm-you could lose both source and target. Distance is one consideration but not the only. Using the risk of earthquakes in San Francisco as an example: 50 miles to the south is still considered a very high quake and flood risk. 80-100 miles east is a completely different seismic profile, far from quakes, tidal flood risk and human threats. Bottom line? Distance does not always equal safety.
Utilities, Electricity and Cooling	The DR site should have stable power and cooling to prevent power outages and system shutdowns.	As with most of the site criteria, you should ensure no commonality between the source and target as it relates to utilities. The target site should be supplied by a different utility provider and should have diverse sub-station feeds into the target site (two at a minimum). Consider the ongoing issues with one of the largest providers in the Western U.S., Pacific Gas & Electric. In 2019 and 2020 the grid was intentionally shut off multiple times for many days to avoid fire threat. In addition, major financial woes are placing the company at risk of a public sector takeover. Why place both source and target at risk by relying on the same provider for both? Criteria to consider: Separate grid from source, separate supplier from source, onsite sub-station, diverse feeds to that sub-station and financial stability of the target electricity supplier.
Distance from Source	The distance between the source and target site(s) depends on the Risk Assessment Process. Multiple criteria are used, but in general it must have adequate distance to insulate against multiple threat types.	The concept that distance alone is protection from disaster is simply not correct, there are again multiple criteria to consider outside of physical distance. For instance, a data center located in Florida with replication 100 miles away, means nothing if a large hurricane hits both sites. The selection process for a target site should consider all the following criteria: a different weather profile, a different seismic profile, a different utility provider and multiple transportation option for critical employees. There are other critical criteria such as latency and connectivity, but the list above is a good start towards eliminating the higher-risk sites.
Transportation and urban infrastructure	The transport network between sites includes the availability of major roads, airports, port and railway options. In addition, accommodations such as hotels and restaurants should be considered in the event IT staff must remain onsite.	An often-overlooked criteria is the ability for critical employees to travel to the target site within a reasonable amount of time. It may be tempting to choose a target literally in the middle of nowhere, thinking this provides great protection. However, what if critical staff must manually restart systems, enter passwords and restore some data from backups? Booking and taking a flight to remote areas (assuming the airports are open) can take many hours, even days. A disaster may render one or more transport methods inoperable, so look for urban infrastructure with multiple options nearby. Ideally the site is a reasonable driving distance with other options such as train, bus or port. Typically, a 1-3-hour drive is perfect. Consider accommodations such as hotels and restaurants as well. Recovering a large IT environment may take weeks. This element of criteria has the largest potential for blowing past your RTO agreements.
Cost	Cost cannot be ignored during this scoring process. It should have a weight and score like any other criteria	It seems "dangerous" to allow something like cost to influence this decision, but in any real-world business decision cost matters. You may identify the perfect location, but if the cost is 2-3x your source site you will have a difficult time convincing anyone this is smart. Ideally, you will identify a site with lower hard costs. Costs such as utilities, real estate, personnel, sales tax, etc. should be factored in. In some situations you may even be able to mitigate or eliminate the cost of DR based on the target site selection (more on that below). It is an unfortunate reality that the RPO and RTO decisions (the most critical you will make in this process) are driven by cost. For example, an RPO of "zero data loss" will force the proximity within ~30 miles at most (typically a Mirror in active/active configuration), to avoid major application impact. That proximity may be a site in Manhattan or other extremely expensive location.



All of the criteria in the table above is a good start towards empirically proving your target site is ideal to protect the business. It should be clear by this point however that “your mileage will vary”. There may be 10s or even 100s of legacy applications to consider. Some may be latency sensitive; some may take hours to boot and map volumes and some may be “feeders” to other applications, so they are “Tier-I by association”. All Business Continuity elements are intertwined. RPO and RTO are impacted by distance due to latency, the safety score is impacted by distance + target weather/ seismic profile, RTO is impacted by distance + transport, etc.

We alluded above to the fact that this decision is an example of a multi-criteria decision process. Step 1 is to understand that location will impact all the other criteria related to BC and it will take expert resources and executive sponsorship to execute. We are hopeful this paper has given you the insight needed to start the journey towards full Business Continuity.