10th International Conference of the International Institute for Infrastructure Resilience and Reconstruction (I3R2)
20–22 May 2014
Purdue University, West Lafayette, Indiana, USA

Data Center Site Redundancy

H. M. Brotherton and J. Eric Dietz
Computer Information Technology, Purdue University

ABSTRACT

Commonly, disaster contingency calls for separation of location for redundant locations to maintain the needed redundancy. This document addresses issues for the data center redundancy, including limits to the distribution, distance and location that may impact on the efficiency or energy.

1. INTRODUCTION

Disaster contingency and business continuity planning are important elements of data center management. Business continuity plans and disaster mitigation are an absolute must for business critical systems. Tier 4 data centers must meet many requirements to be certified as Tier 4. One of those requirements is less than half an hour of down time per year. The minimum Tier 4 requirements alone may not be enough to maintain 99.995% availability. Failure to do so may not only result in financial losses but in longer term financial damage due to loss of operational credibility. It is for this reason that alternate processing sites are recommended for data center business continuity planning. The rational for this level of redundancy, recommendations, and pitfalls are presented in this document.

2. REDUNDANCY VERSUS GEOREDUNDANCY

Canonical fully redundant data centers are “like an ark where everything goes two by two” (High Scalability, 2013). Redundancy in the context of the data center typically involves backup equipment to cover failures. The servers and storage would be configured in a clustered fashion to provide seamless failover in the event of hardware, software, or other equipment or service failure.

Redundancy alone does not equate to a substantial increase in reliability. This is because if the redundancy is colocated there is no mitigation in place for site failures such as common power outages. Colocated redundant equipment provides only increased equipment failure resiliency generally at the cost of increased energy load for equipment that may just be on standby.

Georedundancy solves the vulnerabilities of colocated redundant equipment by geographically separating the backup equipment to decrease the likelihood that occurrences, such as power outages, will make compute resources unavailable. Table 1 presents Disaster Recovery Tiers as defined by IBM in 2003. Figure 1 shows the relationship between time and recovery cost for the Disaster Recovery Tiers. This graphic was created in the 1980s and does not reflect technological advances reflected in the updated table. However, the overall relationship is the same.

3. SITE REDUANDANCY

Unfortunately, redundancy can be expensive and is often seen as wasteful. Cold sites can be thought of as very expensive insurance. This is a valid concern.

Table 1. Disaster recovery tiers

Tier 0	No off-site data	Have no Business Continuity Plan or equipment
Tier 1	Backup with no hot site	Back up their data and send these backups to an off-site storage facility
Tier 2	Data backup with a hot site	Regular backups Physical transport to off-site recovery facility and infrastructure
Tier 3	Electronic vaulting transport	Some mission critical data is electronically vaulted
Tier 4	Point-in-time copies	Two active sites with application software mirroring
Tier 5	Transaction integrity	Two-site, two-phase commit
Tier 6	Zero or near-Zero data loss	Disk and tape storage subsystem mirroring
Tier 7	Continuous operations	Automated failover and site recovery

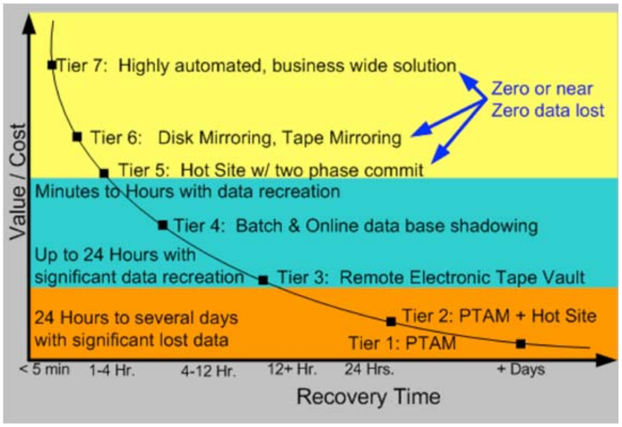


Figure 1. Seven tiers of disaster recovery. Source: http://recoveryspecialties.com/7-tiers.html

In addition, it has been shown that bringing up a cold site in the midst of a disaster recovery situation can be a process fraught with unanticipated incompatibilities resulting in excessive recovery time and expenses.

The solution is to integrate the alternate site or sites into daily workload processing. This configuration is referred to as a hot site. Hot sites are failover sites that are configured using active-active clustering. In hot site or active-active configurations, “each site is active for certain applications and acts as a standby for applications which are not active on that site” (Cisco, n.d. b). This configuration creates resilience at the site level, allowing failover of the data center as a whole. This provides “a key opportunity for reducing the cost of cloud service data centers” by eliminating “expensive infrastructure, such as generators and UPS [uninterrupted power supply] systems, by allowing entire data centers to fail” (Greenberg, Hamilton, Maltz, & Patel, 2009).

4. DATA REPLICATION

4.1. Synchronous

Synchronous data replication guarantees that data at the target location is the same as the data at the source location. The cost of achieving this level of data fidelity is often degraded application performance speed because new write operations cannot be initiated until confirmation that the current write operation has been completed is received. This often means that it is impractical to perform synchronous replication over long distances. “Synchronous replication is best for multisite clusters that are using high bandwidth, low-latency connections” (Microsoft, 2012).

4.2. Asynchronous

Asynchronous replication can provide replication over longer distances at increased speed and reduction in negative impact to application performance in comparison to synchronous replication over the same distance. This is because it is possible to initiate multiple write operations before receiving confirmation of successful completion of preexisting write operations. The drawback of this replication method is that “if failover to the secondary site is necessary, some of the most recent user operations might not be reflected in the data after failover” (Microsoft, 2012). Data loss can occur in synchronous replication as well but is easier to identify incomplete replication transitions.

4.3. Recovery Objectives

Choosing the appropriate type of data replication and effective disaster recovery and business continuity planning must begin with business requirements and establishing the Maximum Tolerable Period of Disruption (MTPOD), Recovery Time Objectives (RTO), and Recovery Point Objectives (RPO). MTPOD relates to how long your business can be “down” before damaging the organization’s viability. Using established tolerances and objectives, as illustrated in Table 2, based on organizational characteristics, direction would be provided in terms of what mitigation techniques to implement.

5. GEOCLUSTERING

Geoclustering georedundant data center resources can solve the problem of unused computing resources by integrating the otherwise standby equipment into the daily workload while providing increased availability and flexibility of workload management. Geoclusters are classified differently and configured to replicate differently based on distance.

5.1. Metro Cluster

Clustered data centers that are approximately 62 miles apart or less are considered metro clusters. Synchronous data replication can be used in metro clusters. According to Cisco (n.d. a), substantially increased reduction in performance will be experienced at this distance for certain configurations.

5.2. Regional Cluster

Regional data center clusters are at distances of more than 62 miles apart. Regional clusters less than 93 miles apart may use synchronous data replication if a performance penalty of 50% is acceptable. Assisted disk failover technologies are highly recommended for this scenario (Cisco, n.d. a).

Table 2. Replication (Gowans, 2008)

	Asynchronous Replication	Synchronous Replication
Resilience	Two failures are required for there to be loss of service	A single failure could lead to the loss of the service
	Failures which lead to data corruption will not be replicated to the second copy of the data	Failures which lead to data corruption are faithfully replicated to the second copy of the data
Cost	Asynchronous replication solutions are generally more cost effective	Synchronous replication tends to be considerably more expensive to buy and manage
Performance	Less dependent on very low latency, high bandwidth network links between units of storage	Dependent on very low latency, high bandwidth network links between units of storage
Distance	Global	Up to 150 miles
Recovery Point Objective	Some data loss acceptable	Zero data loss(Some solutions guarantee no data loss)
Recovery Time Objective	Hours	Zero down time

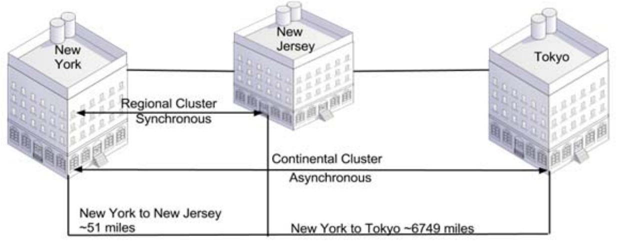

New York
New Jersey
Tokyo
Regional Cluster Synchronous
Continental Cluster Asynchronous
New York to New Jersey ~51 miles
New York to Tokyo ~6749 miles

Figure 2. Three node geocluster

5.3. Continental Cluster

Data center clusters separated by 621 miles or more are considered continental clusters. Synchronous data replication is not generally considered acceptable for continental clusters because the performance impact is too great to avoid degraded processing. Asynchronous replication is recommended to take “full advantage of the bandwidth between the data centers” (Cisco, n.d.a).

6. COMBINING GEOCLUSERS

Depending on an organization’s budget and needs, a combination of geoclusters may be ideal. If there are at least three site nodes in the geocluster, it is possible to have both synchronous and asynchronous replication being performed on different data center nodes as seen in Figure 2. This allows the data center to take advantage of having both real-time replication and processing with reduced impact to application speed.

The combination shown is excellent for recovery scenarios. For example, as illustrated in Figure 2, if the New York data center went out of service and it was necessary to move operations, the New Jersey site already has all the data and is processing jobs because it shared the processing load with New York and is synchronized. Essential employees from the New York data center can report to the New Jersey data center in less than two hours to perform any manual action necessary after the automatic failover. The New Jersey site will still be performing replication with the Tokyo data center site, keeping both remaining sites protected, if degraded.

For larger organizations, particularly those that are global with high computing needs, combining geoclusters can provide some interesting new data center management possibilities in addition to operational resilience.

7. LATENCY

Network latency is an issue that must be addressed when considering georedundancy. The issue cannot be addressed by purchasing higher network bandwidth, though not having adequate bandwidth will certainly slow down data transmission. The issue here is physics—the speed of light.

7.1. Fiber Optic Data Transmission

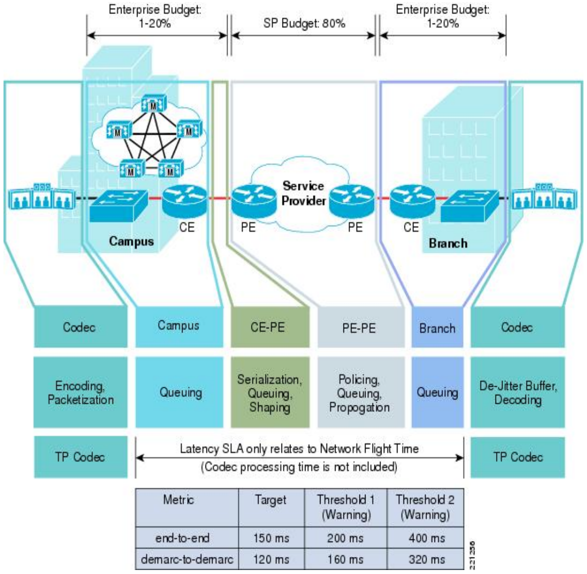
The speed of light is 186,000 miles per second (NASA LTP, n.d.). The speed of light though fiber optic cables is 33% less, or 122,000 miles per second (Hamilton, 2012a). The distance around the equator is 24,901 miles (National Geographic Society, n.d.). This means, in theory, data could be transmitted around the earth’s equator in just over one-fifth of a second.

Unfortunately, that is not good enough. Delays of one-half of a second have shown 20% in revenue loses, and delays of one-tenth of a second have shown 1% revenue decreases (Greenberg et al., 2009). In reality, it is unlikely that data is transmitted at the maximum speed possible through fiber optic cable. The reason for this is, typically, there will be sections where the data is routed though copper, and the customer may not have high bandwidth, therefore, further slowing transmission.

7.2. Latency Budget

The amount of data transmitted by data centers is huge. Many transactions require a series of communications between servers and clients. This increases latency beyond the simple figures presented in the previous paragraph. In order to avoid financial losses due to latency, business determine their latency tolerance. For example, financial trading firms have a latency budget of 50 microseconds for messaging and less than one microsecond for trading (Heires, 2010). To put this into perspective, it takes 74,000 microseconds for data to travel the 2,913 miles from New York to California (Hamilton, 2012b). This yields a data transfer rate of approximately 0.039 miles per microsecond. Therefore, a data center for trading would need to be located within 1.9 miles of the trading market to be within the bounds of the latency budget.

Few other industries are likely to have the low tolerance for latency that trading firms have based on business requirements. The tolerance for Voice Over Internet Protocol (VoIP) is 150 milliseconds end to end (Szigeti & Hattingh, 2004). As illustrated in Figure 3 Cisco TelePresence has a latency budget of 150 milliseconds for what is refers to as “network flight time” and generic video conferencing has a tolerance of 400–450 milliseconds. Each organization must determine its latency budget and determine data center locations based on this budget. Ecommerce has a latency budget of less than 100 milliseconds. Table 3 illustrates the maximum recommended data center distance based on the previously stated latency budgets.



Metric	Target	Threshold 1 (Warning)	Threshold 2 (Warning)
end-to-end	150 ms	200 ms	400 ms
demarc-to-demarc	120 ms	160 ms	320 ms

Figure 3. Cisco TelePresence Latency budget. Source: http://www.cisco.com/en/US/docs/solutions/Enterprise/Video/tpqos.html

Table 3. Latency budget

Maximum Data center Distance		
Industry	**Budget (in milliseconds)**	**Distance (in miles)**
Financial trading	0.05	1.97
Ecommerce	100	3884
Voice Over Internet Protocol	150	5826
Video conferencing	450	17478

Table 4. Data center sizing

Data center Type	Number of Servers	Power requirements
Mega Data center	Tens of thousands	Mega-Watts
Micro Data center	Thousands of servers	100s of kilowatts.

8. FOLLOW THE MOON

Power and cooling are the largest expenses in data center operation (Langborg-Hansen, 2013). The follow-the-moon data center operation strategy allows organizations with data centers around the globe to optimize operations to take advantage of lower power cost, cooling requirements, and temperatures by processing loads at night. This strategy is in use by hyper-scale data center operators such as Amazon and Google. Google "automatically shift(s) its data center operations from the chiller-less data center if the temperatures get too high for the gear (Higginbotham, 2009).

This is achieved by shifting virtualized computing loads to be processed by data center infrastructure where it is most cost effective based on power cost. This allows data center owners to shift workloads to save money and operate more efficiently. The operating policies used to implement this strategy also allow data centers to define policies that make it feasible to operate using renewable energy resources and to increase reliability and availability though the use of virtualized workloads (Higginbotham, 2009).

9. RIGHT SIZING REDUNDANCY

Organizations not competing on a hyper-scale should not despair. There is more than one way to get the benefits of geoclustering. Small- and medium-sized organizations can reap many of these benefits by contracting with cloud service providers. This could mean that the primary control data center is maintained locally by the organization and that a virtual hot site is maintained by a service such as Amazon Web Services.

Very few organizations require multiple megadata centers. These would generally be operated by cloud service providers such as Google, Amazon, or Microsoft. Data center sizing classifications, shown in Table 4, can help describe scale during planning. Microdata centers may also be operated by cloud providers to reduce latency, but these data centers are also a cost-effective option for implementation of georedundancy.

10. IMPORTANCE OF TESTING

Contrary to popular belief, the primary barrier to resiliency is not lack of investment in mitigation or planning. The major cause of failure in execution of recovery plans is lack of testing or drilling. Testing cannot be overlooked in recovery planning. It is arguable that a data center service provider who has not successfully performed site failover and recovery should not have Tier 4 certification. As studies of data center failure have shown, the lack of adequate testing with documented successful results is the hallmark of failure in actual disaster recovery scenarios.

Testing is the sign of a mature, well documented plan. Testing means that the employees charged with data center recovery are well versed in the procedures. Successful testing results indicate that "unforeseen" circumstances during recovery have been uncovered during testing and successfully overcome.

11. CONCLUSION

Resiliency planning is very complex, and many decisions need to be made at a business-requirements level before implementation planning can begin. Poor planning can result in increased cost. Careful planning and implementation could result in an exponential increase in resilience at a relatively low cost. Georedundancy provides some opportunities for implementing cost saving techniques through the reduction of UPS generators. Other cost savings opportunities include shared processing and the ability to use the follow-the-moon technique for global georedundant implementations.

REFERENCES

Cisco. (n.d. a). Geoclusters. In *Data center high availability clusters design guide* (Chapter 3). Retrieved from http://www.cisco.com/en/US/docs/solutions/Enterprise/Data_Center/HA_Clusters/HAGeoC_3.html#wp1082661

Cisco. (n.d. b). *Data center—Site selection for business continuance.* Retrieved from http://www.cisco.com/en/US/docs/solutions/Enterprise/Data_Center/dcstslt.html

Gowans, D. (2008, December 12). Asynchronous or synchronous replication? [Web log post]. Retrieved from http://blogs.msdn.com/b/douggowans/archive/2008/12/12/asynchronous-or-synchronous-replication.aspx

Greenberg, A., Hamilton, J., Maltz, D. A., & Patel, P. (2009, January). The cost of a cloud: Research problems in data center networks. *ACM SIGCOMM Computer Communication Review, 39*(1), 68–73. http://dx.doi.org/10.1145/1496091.1496103

Hamilton, J. (2012a, February 21). Communicating data beyond the speed of light [Web log post]. Retrieved from http://perspectives.mvdirona.com/2012/02/21/CommunicatingDataBeyondTheSpeedOfLight.aspx

Hamilton, J. (2012b, July 13). Why there are data centers in NY, Hong Kong, and Tokyo? [Web log post]. Retrieved from http://perspectives.mvdirona.com/default,date,2012-07-13.aspx

Heires, K. (2010, January 11). *Budgeting for latency: If I shave a microsecond, will I see a 10x profit?* Retrieved from http://www.securitiestechnologymonitor.com/issues/22_1/-24481-1.html

Higginbotham, S. (2009, July 16). *Google gets shifty with its data center operations.* Retrieved from http://gigaom.com/2009/07/16/google-gets-shifty-with-its-data-center-operations/

High Scalability. (2013, January 23). *Building redundant data center networks is not for sissies: use an outside WAN backbone.* Retrieved from http://highscalability.com/blog/2013/1/23/building-redundant-data center-networks-is-not-for-sissies-us.html

Langborg-Hansen, K. (2013, February 12). *Following the moon*. Retrieved from http://blog.schneider-electric.com/data center/2013/02/12/following-the-moon/

Microsoft. (2012, August 29). *Requirements and recommendations for a multi-site failover cluster.* Retrieved from http://technet.microsoft.com/en-us/library/dd197575(v=ws.10).aspx

NASA LTP. (n.d.). *How “fast” is the speed of light?* Retrieved from http://www.grc.nasa.gov/WWW/k-12/Numbers/Math/Mathematical_Thinking/how_fast_is_the_speed.htm

National Geographic Society. (n.d.). *Equator.* Retrieved from http://education.nationalgeographic.com/education/encyclopedia/equator/

Szigeti, T., & Hattingh, C. (2004, December 17). QoS requirements of VoIP. In *Quality of Service Design Overiew.* Retrieved from http://www.ciscopress.com/articles/article.asp?p=357102