Data Center
Site Redundancy

I3R2

Heather Brotherton
PHD Candidate

May 20, 2014

PURDUE
UNIVERSITY

Data Center Site Redundancy

Introduction

Data centers are vital to supporting business continuity particularly as pertains to business critical applications.

Examples of low tolerance applications:

- Emergency 911 systems
- Banking
- Electronic medical records

Data Center Site Redundancy

Redundancy

Canonical fully redundant data centers are *"like an ark where everything goes two by two."* (Possibility Outpost, 2013)

- Redundancy can be expensive and is often seen as wasteful.
- Cold sites can be thought of as very expensive insurance.
- 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 expense.

Data Center Site Redundancy

Redundancy vs. Geo redundancy

- Co-located hardware redundancy
  - provides no mitigation for site failures such as common power outages
  - Provides equipment failure at the cost of increased energy load

- Geo redundancy
  - solves the vulnerabilities of co-located redundant equipment by geographically separating the backup equipment
  - Can also result in increased energy costs

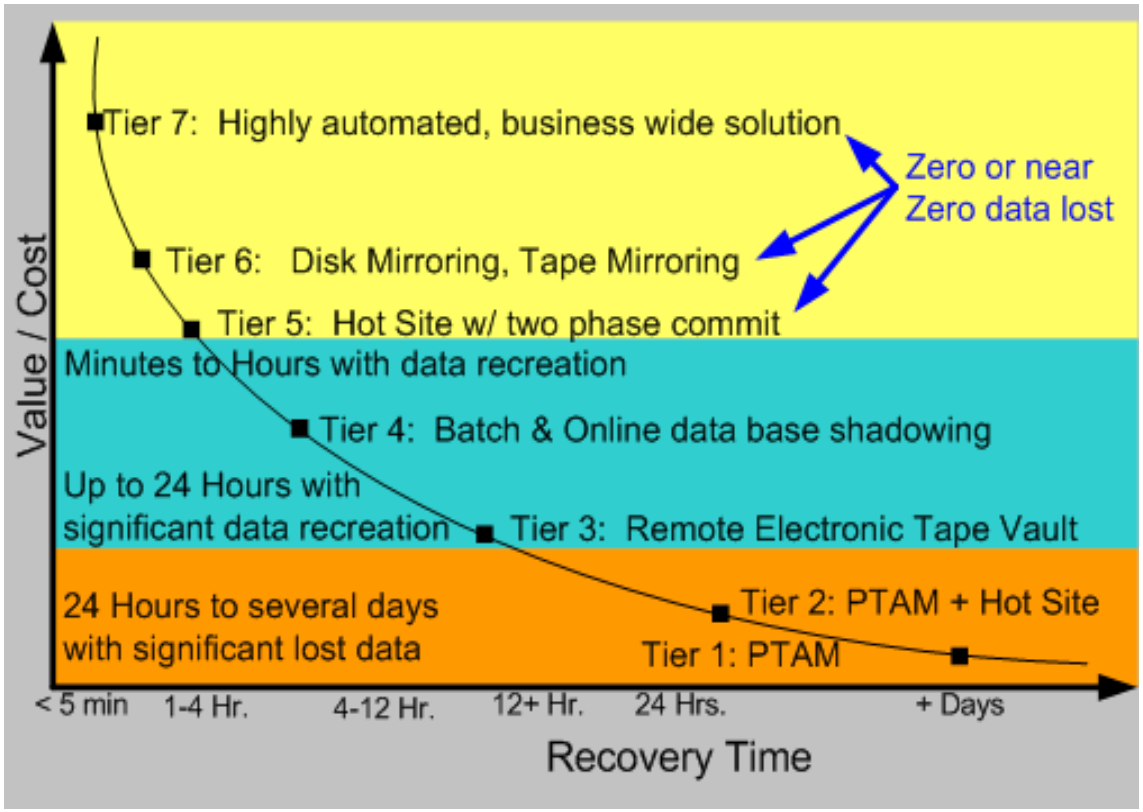
Data Center Site Redundancy

Data center 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

Data Center Site Redundancy

Seven Tiers of Disaster Recovery



PURDUE UNIVERSITY

Data Center Site Redundancy

Data Replication

Synchronous data replication

- Guarantees that data at the target location is the same as the data at the source location
- Cost is often degraded application performance speed
- Expensive to buy and operate
- Practical over distances of less than 150 miles

Data Center Site Redundancy

Data Replication

Asynchronous replication

- Can provide replication over longer distances at increased speed
- Data after failover may not include most recent operations

Data Center Site Redundancy

Choosing Data Replication

First establish:

- Maximum Tolerable Period of Disruption (MTPOD)
- Recovery Time Objectives (RTO)
- Recovery Point Objectives (RPO)

Data Center Site Redundancy

Data Replication

	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

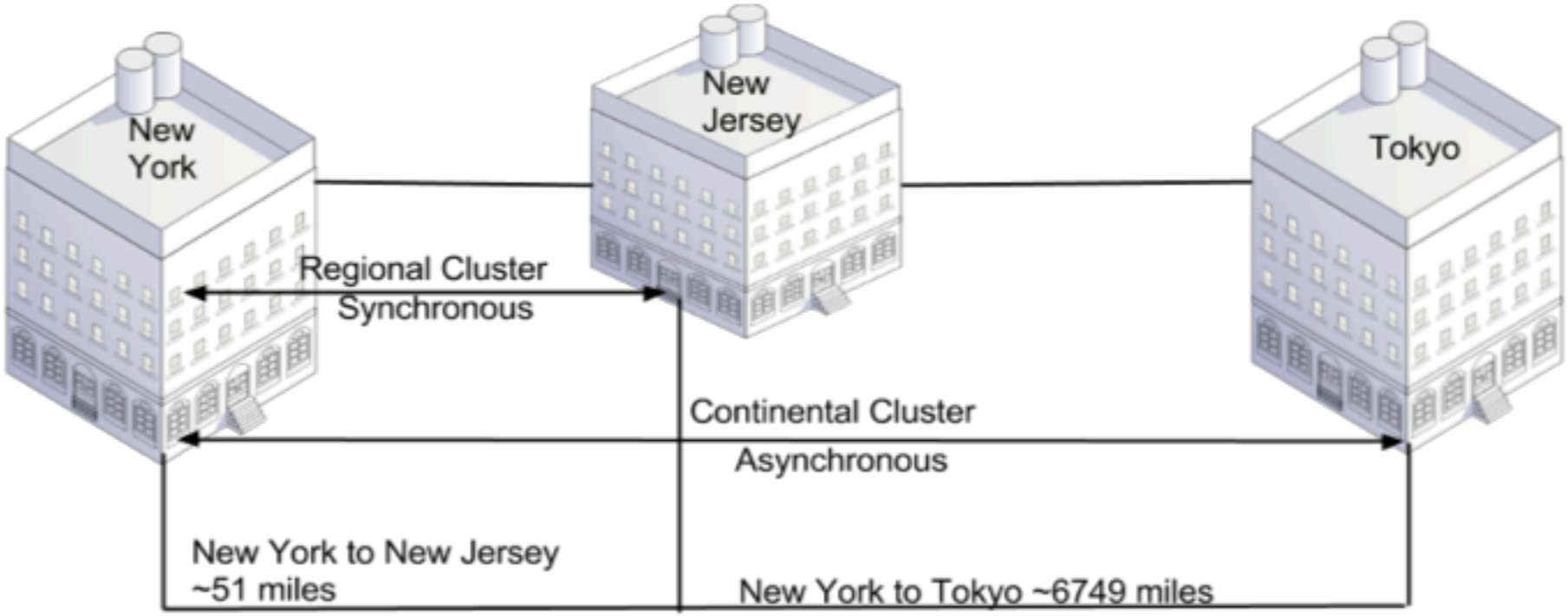
Geoclusters

Georedundancy

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.

Combining Geoclusters

Three Node Geocluster



PURDUE UNIVERSITY

Data Center Site Redundancy

Latency

Maximum Datacenter 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

Follow the Moon

Power and cooling are large expenses in data center operation. 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.

Follow the Moon

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."

Follow the Moon

This is achieved by shifting virtualized computing loads to be processed by datacenter 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.

Follow the Moon

Data center efficiency Foundations

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.

Data Center Site Redundancy

Right sizing redundancy

Practical options for small to medium sized organizations:

- Contract with cloud service providers
- Work together through mutual aid agreements

Data Center Site Redundancy

Testing

The major cause of failure in execution of recovery plans is lack of testing or drilling.

- Testing results in employees well versed in recovery procedures.
- Successful testing results
  - Identify weakness
  - Implement plan improvements

Questions?
Comments?
Suggestions?

PURDUE
UNIVERSITY