

Data Recovery

Outside the Core Data Center

Information availability is critical to an enterprise. Most enterprise IT departments have procedures in place to help ensure data availability, but if a move to a remote site becomes necessary, significant downtime may be incurred. To enhance disaster recovery capabilities and mitigate downtime, enterprises can build dual data centers. This article discusses secondary-site strategies and provides guidance on their deployment.

BY MICHAEL KIMBLE

Related Categories:

Business continuity

Disaster recovery

Visit www.dell.com/powersolutions for the complete category index.

High availability of data is vital to enterprises. Without constant access to electronic information, services can be disrupted and enterprises can suffer far-reaching business and financial consequences. Minimizing this possibility is the goal of most enterprise IT departments. Although IT departments typically have procedures in place to meet availability requirements for localized problems such as data corruption or system failures, they can be unprepared to maintain information availability in the event of a disaster or catastrophic systems failure. Reasons for this may include the following:

- No single disaster recovery implementation meets every requirement.
- Business expectations are not aligned with IT processes.
- Thorough documentation and details for recovery are not up-to-date.

Recent events, including terrorist attacks and natural disasters, have demonstrated that an enterprise's need for adequate disaster recovery planning is greater than ever before. Although the odds of needing to transfer IT operations to a secondary data center are small, the consequences of not being prepared to do so are enormous.

Knowing this, many enterprises are choosing to build secondary data center sites as a way of mitigating downtime risk. However, the time and expense required to do this means that they must consider all available options. This article reviews strategies for setting up secondary sites and provides guidance for estimating recoverability requirements using two criteria: allowable data loss, which is known as the recovery point objective (RPO); and the time required to restore operations, which is known as the recovery time objective (RTO).

Selecting the level of remote recoverability

For the past few decades, one of the most common data recovery methods has been to back up data to tape and send it off-site to be retrieved if necessary. In some cases, tapes are sent to secondary locations where the recovery process begins. This still holds true for the majority of today's enterprises—those that regard an RTO of one day or longer as acceptable—and probably will for the foreseeable future.

However, some enterprises prefer to have dual data centers that can provide transparent failover of applications with little interruption to end users. This has resulted largely from advances in data replication software and the reduced costs of reliable hardware. Although this is

Guideline	Class A	Class B	Class C	Class D
RPO	No data loss	Less than 10 minutes	4 hours	Best effort
RTO (internal)	Less than 30 minutes	Less than 4 hours	Less than 24 hours	Less than 48 hours
RTO (external)	Less than 24 hours	Less than 48 hours	3 days	1 week
Planned downtime	Less than 12 hours per year	Less than 24 hours per year	Less than 24 hours per year	Best effort
Remote technology used	Synchronous mirror	Asynchronous mirror	Host-based replication	Tape only
Local technology used	Point-in-time copy	Point-in-time copy	Backup to disk or tape	Backup to tape

Figure 1. Example recovery guidelines for various data classes

a legitimate goal, many enterprises lack the resources, budget, or justification for the expense and labor required for such an undertaking. For those reasons, enterprises should set realizable goals when planning a secondary site for disaster recovery.

Aligning business requirements with strategies

Just as not all data is created equal, the resources and requirements that must be committed to that data are not equal. A clear understanding of which applications and data affect an organization the most is the first step to defining a disaster recovery strategy. This understanding comes from performing a thorough business-impact analysis of the environment and documenting the findings in a data-classification report. The business-impact analysis estimates potential costs in lost revenue and productivity as well as damage to an enterprise's reputation and business relationships should a lengthy downtime occur.

Data classification involves many facets of an organization—from IT to business units to executive management. Each group often views the importance of an application and its data very differently. For example, e-mail might be the most important application to end users, while the enterprise's leaders might view an order management application as the most important because the enterprise cannot generate revenue and process orders without it. These differing views demonstrate the need to classify applications and data. Figure 1 shows an example classification matrix wherein each application or data pool is assigned to a defined class.

Examining secondary-site deployment options

To fully realize its goals for remote recoverability, an enterprise must examine its secondary-site data recovery options. Commercial hot sites and internally funded dual data centers are among the most popular choices, but other available options may be of interest.

Determining the most suitable option should be based on the guidelines set forth in each enterprise's data classification and recovery decisions. Figure 2 summarizes secondary-site data recovery options.

Budget considerations

One of the primary factors in choosing a secondary-site strategy is the cost required to implement and maintain it compared to the return for potential use.

Hot site. With commercial hot sites, enterprises pay recurring fees to maintain a facility and keep equipment and resources available should they need them. These fees can vary widely based on the established RTO and can range from US\$100,000 for recovery over several days to upwards of US\$1 million for recovery in hours. An advantage of this option is that administrators can test preparations regularly and help ensure that processes and procedures are in order.

Cold site. With cold-site facilities, an enterprise can own or hold a lease on the space it requires and can then purchase hardware, software, and other items as needed. Although this can be a significant expense, an enterprise has to spend money to furnish and operate the secondary data center only if a disastrous failure occurs. This can be an appealing alternative to both hot-site and dual data center approaches. However, data recoverability using a cold site is largely based on the ability of hardware and software providers to deliver their products in a timely fashion.

Dual data centers. This approach to data recoverability requires that enterprises construct secondary facilities to house their emergency processing needs. Although this option provides the fastest recoverability, it can cost up to three times as much as

Type of recovery site	Description	Typical recovery time
Commercial hot site	Commercial hot-site vendors offer fully staffed facilities ready to support an enterprise's processing needs should the need arise. They offer not only processing hardware, but also workspace, telephones, and network connectivity.	24 to 48 hours
Cold site (shell site) and replacement systems	A cold-site facility is prepared to receive replacement hardware and technologies to resume processing. Hardware (and software, if not vaulted with tapes) is procured from a primary vendor or third-party contract providers.	3 to 7 days
Mobile shell site	A mobile shell site is a mobile processing facility that can be delivered to a destination of choice. The facility contains hardware and limited workspace.	1 to 3 days
Reciprocal backup agreement	A reciprocal backup agreement is an agreement between two noncompeting enterprises to share hardware resources in the event of failure.	12 to 36 hours
Dual data center	A dual data center is an internally funded data center equipped with the same type of hardware as the primary data center. It may be ready for processing immediately, depending on the vaulting strategy.	Instantaneous to 12 hours

Figure 2. Types of secondary sites for data recovery

the primary data center—and it may never be used. This additional cost can be attributed to obtaining the facility; purchasing hardware and software; and retrofitting the infrastructure to meet HVAC, telecommunications, security, and workspace requirements.

Because executive management often has a hard time justifying such a large expense for something to sit idle, organizations often try to use the secondary site for some primary processing. However, a thorough review of the impact of a catastrophic failure on an enterprise can often justify the need for a dual data center. For example, if an enterprise stands to lose US\$1 million for every hour that its primary applications are down, the cost of a standby data center can be acceptable.

A further consideration is that although the primary- and secondary-site configurations are initially sound and the data is mirrored, the two sites often lose congruity as time progresses. This can occur when an enterprise adds new hardware and functionality to its primary site but overlooks updating the secondary site because of cost, staffing shortages, or other reasons. This problem has led to the adoption of virtualization technologies at the remote site to keep physical hardware costs low. Most disaster recovery situations do not require all operations to function at full capacity, so virtualization technology can be used to consolidate operations onto fewer physical servers than exist at the primary site—which can lead to savings in hardware costs.

Implementing a dual data center

After investigating all the available options, an enterprise may decide that a dual data center approach is the best choice. In that event, the enterprise should carefully choose an appropriate location and create a comprehensive plan for hardware, staff, and data connectivity.

Location and facility considerations

Choosing a location for a secondary data center can be difficult. Consideration should be given to distance, need for site renovation, availability of utilities, and staff proximity, among other factors. Distance from the primary site typically depends on where the facility is located and the type of threats to which it is exposed. For example, in Florida and the Gulf Coast of the United States, the wide path of hurricanes dictates the need for distance separations of 100 miles or more, whereas in a tornado zone, a much smaller distance—for instance, seven miles—would suffice.

These factors also affect the enterprise's recovery objectives and the type of replication it employs. An RPO of zero data loss, for example, requires the use of synchronous replication software, which is often limited to approximately 60 miles. Going beyond that distance may require asynchronous tools that increase RPO exposure.

Power grid considerations

After a geographic location has been identified, the task of selecting a specific site begins. The most important factor is determining which power grid the facility operates on and which other agencies

use that grid. Typically, first-responder facilities such as hospitals, police stations, fire stations, and city functions are the first to be restored in a power grid. After power is restored to a data center, data communications services and connectivity soon follow. Although an enterprise's data center may have power and data available, it provides no advantage unless users can access it both locally and remotely. Therefore, enterprises should ensure they have processes in place for users to gain connectivity to the secondary data center.


Another important factor to consider in choosing a site is the nature and extent of the resources that are required to make the facility an acceptable data center. Enterprises should determine whether a costly retrofit—such as a raised floor, HVAC, or security—is necessary and whether the site can provide adequate workspace for staff operations.

Understanding factors that can inhibit disaster recovery

Personnel and data are two key components in any successful disaster recovery. However, other factors can greatly affect an enterprise's success and its ability to meet recovery objectives:

- **Applications:** Although data may be backed up and protected, current copies of an enterprise's applications also should be stored off-site.
- **Legacy hardware:** The need for legacy hardware such as specialized cards and readers should not be overlooked.
- **Documents and forms:** Preprinted forms such as checks and statement forms should be stored off-site or arrangements should be made to procure them quickly.
- **Local user data:** A significant amount of data may be stored on local users' hard drives, and this data must be backed up.

Finding the right fit

Selecting a disaster recovery strategy requires a significant amount of research and time. However, the time and research invested in studying an enterprise's environment and defining realizable recovery goals can help enterprises determine their choices and reach sound decisions. If enterprises choose to implement a secondary site for disaster recovery, they can benefit from the business continuity and high availability that this disaster recovery option provides. 

Michael Kimble is an enterprise technologist in the Advanced Systems Group at Dell. Working with Dell consultants and customers, he helps design storage implementations for disaster recovery and business continuity. Michael has a B.S. in Finance and Economics from the University of Central Florida.

FOR MORE INFORMATION

Dell business continuity:
www.dell.com/businesscontinuity