RL Consulting

People • Process • Technology
Organization • Integration

IT Service Management

Service Continuity Methods
(Disaster Recovery Planning)

White Paper

Prepared by:
Rick Leopoldi
May 25, 2002

SERVICE CONTINUITY METHODS

This white paper describes the major elements of the approach to Service Continuity that includes the following topics:

- Business Impact Analysis
- Disaster Recovery Planning
- Information Technology Infrastructure for Service Continuity

It should be noted that the section on business impact analysis is included as the main purpose of any disaster recovery effort since a business impact analysis is necessary to provide input to a subsequent disaster recovery planning effort. In addition, disaster recovery is a main and critical component of Service Continuity.

Business Impact Analysis

Through interviews with users or their representatives, a business impact analysis will verify the major systems that must be recovered, help determine the effects of an interruption of those systems, and verify the priority and sequence of functions and application systems which should be recovered if an emergency occurs.

The results of the business impact analysis will help the Customer determine the true effects of a disaster to the computer resources and in turn help to justify a recovery strategy.

Without conducting a proper business impact analysis, an organization may run the risk of making arbitrary and possibly incorrect decisions about the importance and recovery sequence of their business functions and supporting application systems. As a result, they may either arrange for inadequate recovery resources or pay more for recovery than is required.

Business impact analysis requires extensive fact-finding and analytical activities. The benefits of a business impact analysis are:

- To verify the major systems that must be recovered in the event of an outage
- To determine the operating, financial, legal, and regulatory effects of an interruption of these systems
- To determine the capability of each department to respond to an outage
- To verify the priority and sequence of functions and application systems which should be recovered if an outage occurs
- To assess the need to develop a disaster recovery plan and to arrange for the availability of a recovery facility and its required size

- To present to management an assessment of the organization's reliance upon a centralized data center or server cluster, its ability to carry on business during an outage, and its preparedness to restore its original computing capabilities.

Disaster Recovery Planning

The disaster recovery planning methodology consists of many disaster recovery planning processes. The contemplated disaster recovery planning effort, which will be performed after the disaster recovery business impact analysis, will be accomplished utilizing the disaster recovery planning processes, which are summarized below.

Please note that the sections on *Application Inventory and Criticality Ratings*, *Critical Application System Resource Requirements*, and *Alternate Processing Strategies* are partly covered in a business impact analysis, but are included here for completeness.

1) Overview

The overview is the first step addressed in the disaster recovery planning process. The project boundaries (purpose, scope, objectives and assumptions) to be used in developing the disaster recovery plan should be defined, documented, and approved by management. The types of disasters addressed by the disaster recovery plan will depend upon the physical location of the company and its information processing facilities.

2) Application Inventory and Criticality Ratings

In a business impact analysis, all application systems are identified and categorized as to their criticality relative to the continuation of the business operations.

To define the criticality and recovery sequence of an application system, its major functions, size and complexity, impact on other systems, and processing schedule should be taken into consideration.

Processing of critical systems should be resumed as quickly as possible after a disaster. Less critical systems should be processed as soon as possible after data processing operations are restored or if time and resources are available, after restoration of critical systems. Non-critical systems should be restored when data processing services are available.

3) Disaster Recovery Teams

Establishment of functional teams and identification of individuals is necessary to assure preparedness in the event of a disaster. Designating personnel to be responsible for specific functions before, during and after an emergency situation is essential to assure that the proper actions will be taken expeditiously.

Team members must participate in the disaster recovery planning process to assure that responsibilities and roles are defined, documented, assigned to specific individuals, understood, executable and accurately represent what is required for a successful recovery.

4) Off-Site Backups

The required off-site movement of software, associated data, documentation, and any additional resources needed in the event of a disaster should be identified, documented and implemented.

The backup scheme used should respond effectively to the most likely problem situations while still providing a tolerable solution in major recovery situations. Criticality and execution frequency of each application system will determine the frequency of backup and the safeguards designed into the backup scheme.

5) Critical Application System Resource Requirements

Once critical systems have been identified and restoration sequences defined, the resource requirements needed to process each critical system must be determined.

The resource and processing requirements identified for all critical systems are analyzed in the recovery and alternate processing strategies process to determine the minimum resource requirements and configuration (hardware, software, communications, forms, supplies, special equipment, etc.) required to process all critical systems in the event of a disaster.

6) Alternate Processing Strategies and Inventories Required

The overall recovery strategy and listings of the hardware, software, telecommunications, support equipment, forms, supplies, etc. needed in the event of a disaster must be identified and documented.

The recovery strategies to be used in the event of a disaster should be based upon a summary of the requirements and resources identified for each critical system.

The strategy decision must be unique to the particular company and situation. Seldom will a single strategy be the optimum solution. For instance, a single strategy for disaster critical applications might be considered, with reduction or withdrawal of services for less critical or non-critical applications.

7) Operating Procedures

The capability to recover from both major and minor disasters necessitates the availability of a comprehensive set of operating procedures describing the steps required to re-establish data processing in the event of a disaster. Therefore, the documentation required should consider the various levels of disasters that could occur and support the objectives, assumptions and strategies of the Disaster Recovery Plan.

Operating procedures can be incorporated either in the disaster recovery plan, or separately with a reference note in the disaster recovery plan. The procedures must be available to technical support and operations personnel with a current copy placed in the off-site vault.

The procedures needed to recover from a disaster will be specific for the actual restorations and, in most instances, for normal processing will be the same as those used in daily operations. Existing procedures should be used where possible to satisfy disaster planning documentation requirements. This will minimize the effort necessary to produce procedures and the training of personnel in the recovery process.

Disaster recovery procedures should be tested periodically and also modified as changes in the data processing facility and environment occur.

8) Critical Application Systems Recovery

The procedures and strategies required for recovery of each critical application system needed in the event of a disaster should be documented and tested. Application system recovery is the process of restoring application system data and software to a point of known integrity after the original integrity of the processing, or data, has been lost. Once the total recovery process has been completed, the system can be restarted and normal operations resumed.

Recovery techniques will vary depending upon the nature of the processing; whether it is batch, on-line, or a combination of both. The criticality of the system and the data involved dictates the degree of recovery measures implemented.

Recovery can involve the use of techniques to reconstruct the status of the system as it was immediately prior to failure. If the system cannot be recovered to where it was immediately prior to the disaster, the recovery process also involves the reconstruction and/or correction of any data that was lost or damaged.

Systems recovery should be closely coordinated with off-site backup efforts to assure the proper and most current data is available.

9) Vendor Contact Information

Documentation of vendor information is needed to provide for expeditious recovery. Designation of individuals at each vendor company will provide a focal point when emergencies occur and allow a vendor to be familiar with the requirements for successful recovery.

Vendor agreements should be copied and maintained with the disaster recovery plan. This will allow easy reference and provide the information vendors require to service the account. Special procurement, supply and shipment considerations from vendors, in the event of a disaster, should be negotiated, documented and included in the agreements.

10) Support Considerations

In the event of a disaster, a timely and orderly recovery of data processing capabilities is dependent on the availability and coordination of personnel and resources throughout the entire company.

Therefore, as the disaster recovery strategy is formulated and external tasks and resource requirements are identified, involvement from external organizations (e.g. finance, facilities, personnel, public relations, etc.) is mandatory to produce a workable plan.

11) Facility, Office and Reconstruction

Facility specifications and office space requirements needed in the event of a disaster should be documented. This forecast of facilities and office space must consider the recovery priorities assigned to each system, the amount of time during which processing will be performed without full capability and the hardware, software and support needs of existing systems.

One approach to defining facility and office requirements in the event of a disaster is to document existing facilities, office space, physical and environmental specifications and any other information pertaining to the current environment. This specification of the current environment will serve a two-fold purpose. It will provide a reference for defining and documenting minimum facility and office requirements in a disaster situation and will assure the availability of documentation needed for reconstruction if a total disaster occurs.

12) Training

Personnel should be knowledgeable of their responsibilities, roles and specific tasks required to execute the disaster recovery plan. Each supervisor or manager who has disaster recovery responsibilities or whose work force members are expected to participate in disaster recovery operations should insure that assigned personnel receive proper instructions. Well-trained personnel can often prevent a disaster or, at the very least, limit the resulting damage.

Participation in plan testing and documentation of test results will insure that training takes place, that disaster recovery awareness exists and that responsibilities are understood and executable.

13) Testing

In order to evaluate the disaster recovery plan and verify the accuracy of the data and strategies developed, the plan must be tested on a reoccurring basis. Testing serves to:

- Demonstrate the ability to recover from a disaster
- Verify that the disaster recovery plan documentation works
- Test the feasibility and compatibility of backup facilities
- Identify and correct weaknesses in the plan
- Train members of the disaster recovery teams
- Increase confidence in the ability to recover
- Evaluate the adequacy of external support
- Enforce continual maintenance of the disaster recovery plan

The overall testing of the disaster recovery plan usually proceeds in phases. The first phase tests the contents of the off-site storage facility. The next phase is a "paper" test involving each disaster recovery team. This is followed by testing recovery at the alternate processing facility with involvement of team members. A large "paper" test involves all teams working together to solve an exercise in responding to a disaster. It is best to wait until the plan has been tested many times before calling a mock disaster where most participants do not know when a test will be run.

It is usually practical to start with testing components of the plan and, as problems are resolved, to build up to more comprehensive tests.

Remember, an un-tested, un-maintained disaster recovery plan can be worse than no plan at all since management believes the business is protected, when in reality it is not.

14) Maintenance and Review

The primary objective of maintaining disaster recovery capabilities is to ensure that planned and tested procedures will work properly regardless of personnel, hardware or system changes.

Maintenance requirements must be incorporated into normal business functions and accomplished on a continuing basis. Maintenance instructions should be a part of the standard documentation that is used in the normal course of business operations.

Changes in software, hardware, facilities, telecommunications or personnel require maintenance procedures to keep the disaster recovery plan up to date. This ensures that procedures are accurate and reflect the current environment.

I/T Infrastructure and Service Continuity

There are many conflicting disaster recovery design objectives when considering an information technology infrastructure for Service Continuity. Consequently, an individual solution is always a compromise, and there is no single solution or recommendation that fits all.

The decision criteria are the cost of the disaster recovery solution, disaster coverage and residual risk, speed of recovery, and completeness of recovery and data integrity. Between any two of these criteria, there is a certain conflict and therefore some trade-off must be found.

Within the scope of the involvement of the Service Continuity effort, the existing information technology infrastructure as well as current available and proven technology are reviewed and evaluated.

Recommendations will evolve from this analysis to arrive at the desired solution, which will implement and support the selected disaster recovery strategy in a cost-effective manner.

This section provides some background related to the disaster recovery design and information technology infrastructure.

1) Data Backup and Recovery

Information is critical to the survival of any business. Of all I/T resources, data is the most important. Other resources, such as processing power, software, DASD storage, and building facilities are all ultimately replaceable, but much data is not. Data is also the most volatile and complex of all I/T resources.

This complexity and volatility of data makes it the most difficult resource to manage during recovery. Whereas the relatively static nature of hardware and building facilities enables sites to be ready before recovery is necessary, the volatile nature of data means it must be managed as an ongoing process. Either data is kept current at the recovery site or it must be made current as part of the recovery process.

Data can be divided into data managed by a database management system (DBMS) and other data that is not. DBMS data is more complex in its structures and requires special consideration for backup. Consequently, most database management systems provide utilities for this purpose. Non-DBMS managed data can be even more challenging for disaster recovery.

Any disaster recovery plan should assume some loss of data. This data loss can be planned or unplanned. Planned data loss is inherent in the chosen disaster recovery strategy. There is also the possibility that data will be lost inadvertently through error, either in the design of the recovery process or its execution.

Generally, data is recovered by restoring a copy of the data taken at some previous time and then applying any necessary updates to it. This approach assumes that the updates made to the prime copy of the data are repeatable such that the equivalent updates can be made to the remote copy at some later time. In a database management system this is achieved by logging all updates to the log data set. This method protects the data up to the point of the latest safe log data set.

When a database management system is not used, other methods of repeating these updates must be employed. In some cases it may be possible to rerun the original transactions, or to capture information on the updates in another way, such as at the workstation or on data entry forms. It may be possible to rebuild the actual data from other sources.

If none of the above techniques are available periodic backups alone will have to be used. This may be acceptable if the data is very static, if the updates occur at regular and predictable intervals, or if some level of data loss is acceptable. In all other cases, the result will be an unacceptable data loss.

When reviewing the backup and logging needs of data, it is important to consider whether the data can be recreated, whether it is volatile, whether the updates are predictable, whether the data is important, what the acceptable window is for recovery, and how much data loss is acceptable before deciding upon the method of backup, logging, and recovery.

2) Disaster Recovery Concepts

It is desirable that the time between a disaster and the beginning of production at the recovery site is short. This is possible if the second site is ready at all times, with all required hardware installed, and with all data at a very current level. The degree to which this goal can be achieved depends on the recovery concept implemented at the recovery site:

- Cold backup

  This is a recovery site that is equipped with an I/T infrastructure, such as a raised floor, air conditioning, and network connections, but without I/T equipment installed.

- Warm backup

  This is a recovery site that is operational and available for takeover after some delay. The delay may be caused by the time it takes to restore data.

- Hot backup

  This is a recovery site that is operational. Some or all application data is on-line, such that the time required to prepare the recovery site for takeover is much reduced.

To become effective, the above definitions should apply to the disaster recovery strategy adopted by the organization. As such these definitions do not accurately describe a given disaster recovery strategy. A precise description of the readiness of a recovery site has to consider the application systems. Main aspects are the backup method and the techniques used to manage backup data.

The readiness level of an application system may be:

1) No provision is made for disaster recovery.
2) Periodic backup

   The installation will at certain times take a consistent copy, which allows recovery to that point, and send it to an off-site location.

3) Semi-Roll-Forward

   In addition to periodic backups, update logs are also sent to the off-site location. Transport may be physical or electronic. Recovery will be to the last log data set received.

4) Roll-Forward

   A shadow copy of the data is maintained at the recovery site. Update logs are periodically applied to the shadow copy through recovery utilities. Transmission may be physical or electronic.

5) Real-Time-Roll-Forward

   Similar to roll-forward, except that updates are transmitted and applied at the same time they are being logged in the production site. This near real time transmission and application of log data would not impact transaction response time at the production site.

6) Real-Time-Remote-Update

   This is the capability to update both the primary and shadow copy of data, prior to sending transaction response or completing a transaction.

3) Interconnection Technology

The primary and recovery site need not necessarily be interconnected for disaster recovery. Data on tape, for instance, may be regularly carried to the recovery site. Still, in order to provide the best possible data protection as well as the quickest possible takeover in the case of a disaster, data must be transmitted off-site as fast and frequently as possible. This is best achieved by using high-bandwidth interconnection technology. Apart from cost, distance is a limiting factor here.

Interconnection technology available today has currently limited distance capability. Emerging technology may extend this capability in the future, but the highest bandwidth interconnection technology will always imply a certain distance limitation.

With the recovery site not interconnected, and to facilitate disaster recovery, backup data must be physically transported to the recovery site. This method of data transportation is well suited to very large amounts of data. It is also suitable for any information stored on paper, as there will always be a need to regularly ship documents to the recovery site.

Disadvantages are that there is a great potential for data loss in case of a disaster; the data is "unsafe" as long as the media is queued for transport. Also, third parties may be involved and this introduces potential management and security issues. At some point tape volumes must be tracked and recycled.

With a network connection to the recovery site, the network allows some data transmission and user access to the recovery site. In most cases this will not eliminate the need for physical data transport. The network connection provides a means for moving data to the recovery site as soon as it is made transferable and avoids the disadvantages of physical transportation.

Remote tape has the advantage of placing backup data in a safe state without an intermediate period of being transferable. This can be done with a transmission bandwidth in excess of a network solution. It facilitates the integration of in-house backup and disaster recovery backup.

Remote DASD may be used to maintain backup data in a safe and usable state. It allows remote copies of critical data to be maintained at a very current level, providing warm or even hot backup.

4) Distance from Primary Site

If for example a fire is the only kind of disaster to worry about, the recovery site can be as close as two machine rooms in the same building, isolated from each other by a fireproof wall. However, disasters may affect more than a single machine room or building.

5) Distance and Disaster Scope

The disaster recovery scope may be such that a whole geographical area is affected. Therefore, a greater distance between I/T sites results in greater security against any widespread disaster. Greater distance however has its price in terms of interconnection, relocation effort, etc. With current available technology, it is neither easy nor cheap to interconnect two sites over hundreds of kilometers at a bandwidth high enough to keep large amounts of data at both sites fully synchronized.

A moderate distance may give less protection against wide-spread disasters, but has advantages such as:

- High interconnection bandwidth

  In this context, high bandwidth provides links in the megabit or gigabit range for remote attachment of tape drives and DASD.

- Low interconnection cost

  The cost of interconnections is a function of the distance covered. the cost of connections may be prohibitive over a long distance.

- Relocation in case of a disaster

  From a network connection viewpoint the recovery site may be relocated anywhere. If there is considerable physical interaction with the primary I/T center, such as central printing and diskette or magnetic tape exchange, operations might not function at a remote recovery site.

- Real-time-remote-update using interconnection technology

  Hot backup strategies are not easy to implement over long distances. They require that a similar system is active at the recovery site. The primary site transmits database changes to the recovery system which applies them in real time to a shadow database. Most of this technology must be implemented within application systems and requires that both systems are available at any one time. Functions like error recovery, synchronization, and takeover, are usually extremely difficult to implement.

It is much easier to attach the shadow database to the primary system using interconnection technology, where. both the primary and the mirror database are controlled from the primary system. This depends on the ability of the application and/or system software to keep the shadow copy updated.

- In-house disaster backup

  Two separate sets of backup copies normally have to be maintained. One is required for recovery at the primary site to rebuild a database in the case of an error. The second is required for disaster recovery and should be taken off-site as soon as it is created. In many cases, the database system must be taken down during the backup process. The time available to do this can be very limited and maintaining two redundant backup processes may be a major problem.

  A solution to this problem may be to use remote tape for backup. Tape drives could be physically located at the recovery site, and connected to the primary site using interconnection technology links. These backups are immediately available for in-house recovery as well as for remote recovery in the event of a disaster.

  The distance between the primary and recovery sites being a given, technology that is currently available and proven must be reviewed and adopted to implement and support the disaster recovery strategy.

6) Network Infrastructure

A network capable of functioning after a disaster at the primary site is a major requirement in disaster recovery design.

In addition to handling the workload generated by user access to the primary site, a network designed for disaster recovery must provide the following:

- Data shipment between sites

  There will be a requirement for a constant data exchange between both sites. This data may include remote operation, system and application changes, systems management, database logs, and database backup copies.

  If complete files or database backup copies are to be sent across the network, it will require extremely high capacity transmission links. Current available interconnection technology does not provide a bandwidth of that range.

– Capacity required in the event of a disaster

  In the event of disaster, the resulting workload takeover can cause major shifts in network traffic. Users may change from local to remote attachment, and remote users may require significantly different transmission routes through the network.

7) Network Topology

A network topology that provides a disaster recovery capability must to have the following properties:

– Network access for both sites

  This is required both for workload takeover, as well as for periodic testing.

– Uses separate gateways, one for each site

  In the context of disaster recovery, a network gateway machine may be a single point of failure. Disaster recovery requires that a second gateway exists, and that these gateways are not located close to each other. The goal is that the recovery site does not lose its network access when a disaster occurs at the location of the primary site's gateway.

– Both sites must be able to control the network

  Large networks are typically controlled from one location. In the event of a disaster at the controlling location, a provision must exist to move the network control function to another site.

– Isolated external paths

  Despite the fact that two sites exist that each have their own network gateway, a single point of failure may exist in the external connection between the gateways and a public telephone interchange facility. Ideally, both gateways should be connected to separate public telephone interchange facilities.

– Alternate paths for all locations

  The network topology should provide alternate paths between host locations and all remote user locations.

– Automatic path switching

  Wherever path switching is required in the event of disaster recovery, it should take place automatically, or under central control. Remote users are typically not trained or sufficiently experienced to perform this task reliably.