System Construction Technology and Solutions to Support Disaster Recovery

TAKAYA Masahiro, TOMI Mitsuhiro

Abstract

This paper is intended to consider the range of recoveries from disaster, describe the systems architectures of disaster recovery (DR) systems and indicate some key points of relevance to effective solutions. It also summarizes the data replication systems offered by NEC from the viewpoint of systems layers, and introduces individual products together with their features. Finally, it also explains the concepts that are underpinning the development of network circuits.

Keywords

Disaster Recovery (DR), system construction technology, data protection, remote backup, data replication

1. Introduction

This paper introduces system design and implementation procedures for the disaster recovery (DR).

In order to continue business using a substitute system when the production system becomes unavailable due to unforeseen circumstances, it becomes necessary to create overall system architectures from the viewpoint of business continuity (BC).

The substitute system will not be adequate if it simply consists of installations of substitute equipment to replace the existing systems. Network circuits for enabling access to the substitute systems and two-site systems operation are also important.

The key to business continuity using a substitute system is how to maintain the data. One of the most well known means of achieving this is by data replication using storage functions. However, this paper will also introduce the more recent developments.

2. Recovery Ranges of BC and DR

Recovery after disaster can roughly be divided into two levels. One is the recovery level targeting BC (Business Continuity), and the other is the recovery level targeting DR (Disaster Recovery), or "protection of the IT systems and data."

The need for BC has recently been increasing because it has become evident that DR measures are not by themselves enough to ensure effective BC. However, as the approach to BC necessitates a large amount of time and costs, many cases begin by applying DR measures as a realistic solution (**Fig. 1**).

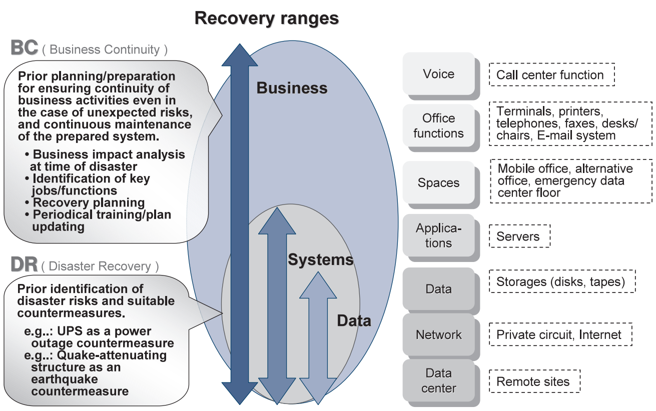

Fig. 1 Recovery ranges of BC and DR.

The BC/DR system is a copy of the production system, and can be regarded as a customization of the production system to meet the requirements set by considering the costs and functions (**Fig. 2**). These are the requisite indices for studying the BC/DR systems; 1) RLO (Recovery Level Objective); 2) RTO (Recovery Time Objective); 3) RPO (Recovery Point Objective). Defining these three objectives also determines the direction of each BC/DR system at the same time as defining the continuity capability of each business.

Like the design of IT systems, an explicit definition of system requirements is essential in the design of BC/DR systems. It is due to these requirements that the system architecture is decided and specific system designs are conducted.

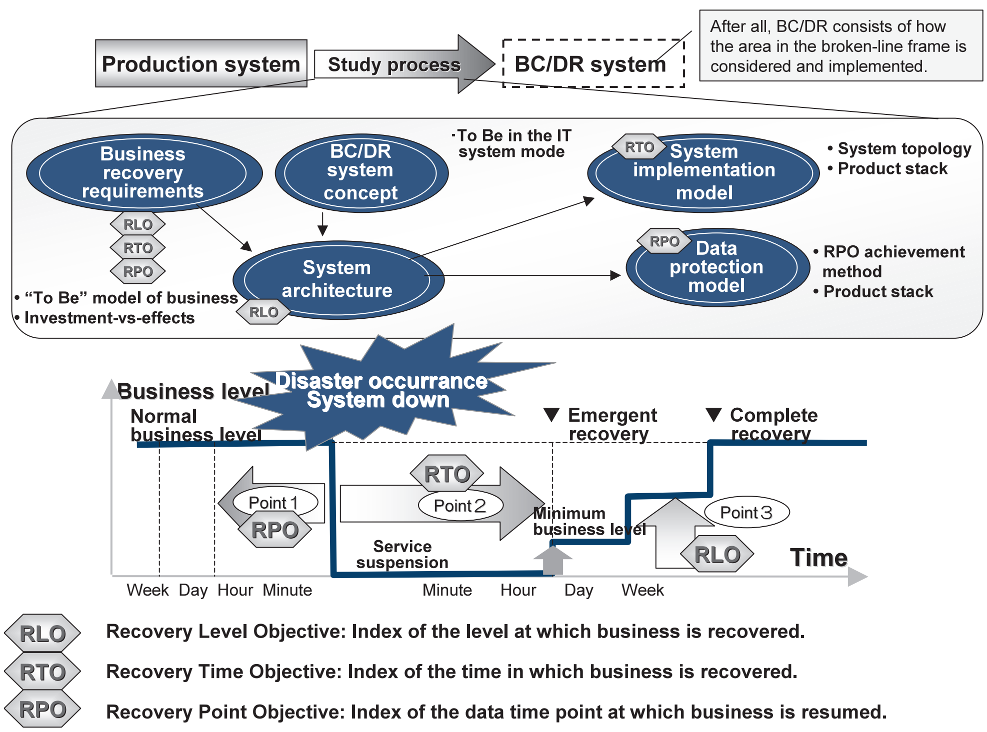

Fig. 2 Concept of BC/DR system.

3. Study of Systems Architectures

As discussed in Section 2 above, the study of BC/DR systems architectures is not enough if it is focused only on the IT systems, but it is essential that it also considers other business aspects. In other words, the introduction of BC/DR systems is almost the same as creating a completely new system. This means that a wide variety of items should be studied and the design will not be advanced unless there is a clear policy with respect to the system. Therefore, it is strongly desirable to define a clear policy by conducting business impact analysis before proceeding to the study of the IT system.

(1) RLO (Recovery Level Objectives)

One of the indices determining the system architecture at time of disaster is RLO. It can be described as the SLA (Service Level Agreement), and is a critical index for business execution.

Like the SLA of production systems, RLO does not express the SLA of a specific part but is a general SLA aimed at continuing the business, i.e., the SLA for the entire IT system. Consequently, it should be studied based on the recognition of the IT system components (referred to in this paper as the system configuration stack). **Fig. 3** shows the system configuration stack.

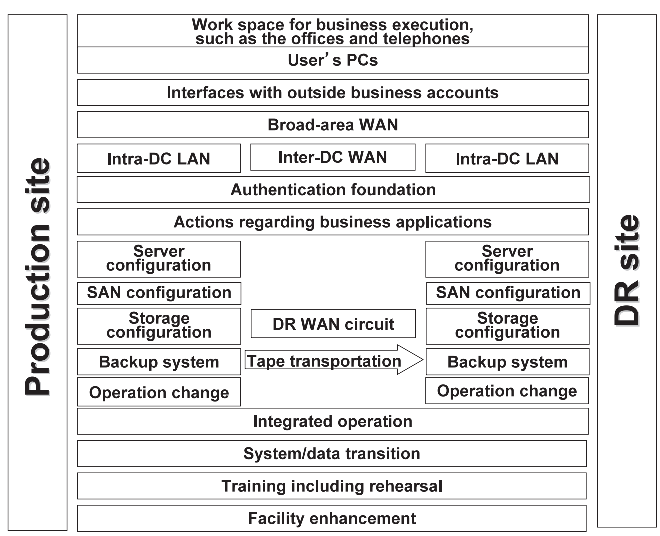

Fig. 3 System configuration stack.

(2) Costs

The equipment costs for servers and storages of the IT system are often regarded as a critical factor in the BC/DR system design. However, since system implementation closely resembles the addition of a completely new system, the items that should be regarded as being the most important are "integrated operations" and "networking (broad-area WAN)." The important system factor is that any study of these two items requires that the two sites to be considered are available for comparison, this is particularly important for the smooth operation of the BC/DR system.

In connection with the "integrated operations" care is required regarding the fact that the addition of a completely new system results in an increased number of operations management targets. In addition, in order to prepare for an emergency, the operation should also guarantee the normal functioning of system components that are not used in the operation of the production system at normal times without any unexpected event occurrences. It is no exaggeration to say that these tasks almost double the burden of operation.

With regard to "networking (broad-area WAN)," connections at time of disaster should be guaranteed by ensuring switching to the DR site. It is also necessary to consider if the network circuit has an alternative circuit that can be connected in the event of an anticipated broad-area disaster, should the circuit be run by a different carrier, etc.

(3) DR Site Environment Configurations

The environment of the DR site may be provided in one of the following three configurations.

① Standby System in an Environment Degrading the Production Site Environment

This configuration is usually selected. The DR site for an existing environment is built in a degraded environment that can meet the lowest required minimum RLO in order to reduce additional investment to as low a level as possible.

② Standby System in a Similar Environment to the Production Site

This configuration is selected when degrading the production system environment is not allowable. The DR site environment is completely identical to the production site environment so a reduction in the burden of the operation can be expected.

③ Active Systems at Both Sites

This configuration can be regarded as the optimum from the viewpoint of running costs as there is no standby system environment that does not contribute productions. However, introduction of such an environment is rather difficult because complicated and transitional design and operation as a DR system is necessary to the existing environment. As a result, this configuration could be considered most often when an entire system is newly constructed.

(4) Concept of Production Site Redundancy

Consideration of the SLA for the production site is required before studying the RLO.

It may seem that the scale of the production site redundancy can be reduced when a DR site is constructed. However, just as with cluster systems, the redundancy and reliability of the production site should be secured before studying the equipment configuration of the DR site. A frequent mistake is to design the DR site by regarding it as a clustering standby system in a remote site. This concept is not so erroneous when one views it as a single system from a microscopic viewpoint. However, when the entire system is considered from a macroscopic view, it will be noticed that there is usually almost no business that is running on one system only and that, to run a system on the DR site, other intact associated systems should also be run on the DR site in order to continue the business. Therefore, the redundancy of the production site should be secured in the same way as before when considering the configuration of the DR site.

(5) RTO (Recovery Time Objective)

RTO is the index of the time limit until all the businesses become completely available on the DR site. Note that it is not the time until the IT system infrastructures start operation.

Measures related to the infrastructures are not the sole factors determining RTO, but the measures related to the business system and operations are also important. These measures should be taken into consideration from the following points of view.

- What kinds of IT system devices can be prepared at the DR site in advance? Are they available immediately?
- When will the data be available?
- Are the job checks procedures prepared and functional?

(6) RPO (Recovery Point Objective)

RPO is dependent on the data protection model, and its most important factor is the method of data replication from the production site to the DR site. As the data protection model also affects RTO, the selection of the data replication method eventually determines the character of the DR system.

Fig. 4 shows the system of NEC's DR solutions for RTO and RPO.

4. Data Replication Methods

The method for synchronizing data should be selected by

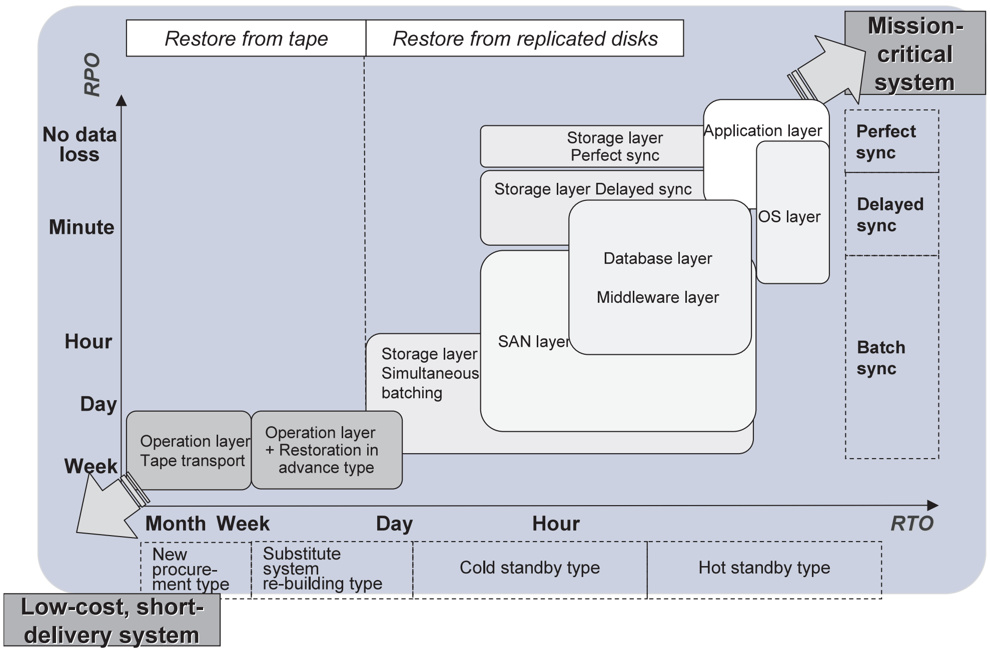

Fig. 4 NEC's DR solution system.

examining RTO/RPO as well as by considering the equipment configuration of the overall system and the cost efficiency.

The method of data replication can be classified roughly into the following three types. The method of layer division that will be described below can also be classified into these three methods. As each of these types involves different resources and risks, full studies are necessary on the system configurations and each of these methods.

① Perfect sync type: A perfect match is maintained between data held at both the production and DR sites. The necessity of updating the DR site at the remote location affects the response of applications.

② Delayed sync type: Updating of the production sites via applications processing is reflected at the DR site asynchronously (with a certain delay).

③ Batch sync type: The stationary point of the data is determined and data is replicated instantaneously.

The data replication method making use of the storage function is the most widely recognized method. However, this method is not universal. To deal with various requirements and system configurations, several methods have been developed for each system as shown in **Fig. 5**. The following sections will describe these methods.

With the perfect sync and delayed sync types, the consistency in the system is not guaranteed and the data at the DR site may become inconsistent. As a result, these types necessitate thorough diagnoses and restoration of the file systems as well as of the database recovery before resuming business. These operations can recover the consistency of a file system but until what process of business is reflected completely in the database remains unknown. As this means that the significance of data is not guaranteed from the viewpoint of business, confirmation is required in the actual business environment.

4.1 Application Layer Method

This refers to the method that handles all processes at the business application level.

This method can offer the best accuracy but is the most difficult to apply. With this method, the data consistency between sites is controlled by the applications, by combining inter-program communications and file transfers. Consequently, there is no fixed solution for this method and the system should be built depending on each business operation. Since the method exerts an important impact on the basic designs of the applications, it cannot be applied for the extension of an existing system. It can be studied only at the time of a complete system replacement or upgrade.

・Representative solutions

1) Robocopy (Windows resource kit)

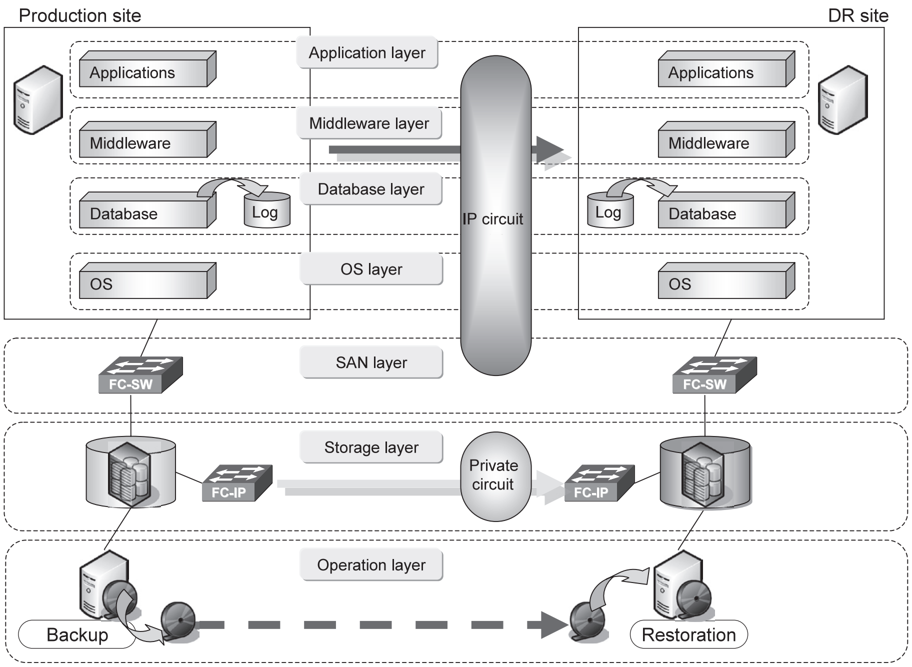

Fig. 5 System layers.

4.2 Middleware Layer Method

This method performs data replication between sites at the middleware level.

The basic processing is that the middleware acquires the I/O processing operations on the production site and restores them on the DR site. Consequently, the middleware is greatly dependent on the file system of the server OS. As it is also greatly dependent on the system environment including the OS version and applied batch level, its impact on the existing environment is large. In addition, as work such as assessment and confirmation after batch application is increased in actual operation, operational considerations become important. The middleware overheads do not need to be considered with recent systems, but care is still required on the inconsistency with other middleware because this method is associated closely with the system environment. In addition, care is also required on the characteristics of the applied files and it is also necessary to be careful with the middleware functions, for example whether the unit of data replication is per file or per block.

・Representative solutions

2) VVR (VERITAS Volume Replicator)

3) VRE (VERITAS Replication Exec)

4) ISV products such as Double Take

4.3 Database Layer Method

This method transfers the updated journal data of the production site database to the DR site to reflect the data in the database of the DR site.

This method requires that the databases are run in the journal output mode, so it may be accompanied with a system review.

This method cannot replicate data of files outside the database. This poses a problem for example with a configuration in which the database stores only the pathnames of flat files and the flat files are managed as ordinary files. With this configuration, inconsistency may be produced in the DR processes because the database is protected by the journal while the flat files are not. This problem should be solved by preparing the consistency securing function for the business applications.

In addition, care is also required that the DR site also runs a database and that the license for this database is required separately.

Fig. 6 shows an example of this method with ORACLE.

・Representative solutions

1) Oracle (Data Guard)

This is the function incorporated in Oracle.

2) Microsoft SQL Server (log distribution)

3) Data Coordinator/RA

This is the software implementing “data linkage between dissimilar databases.” Data replication is possible by linking

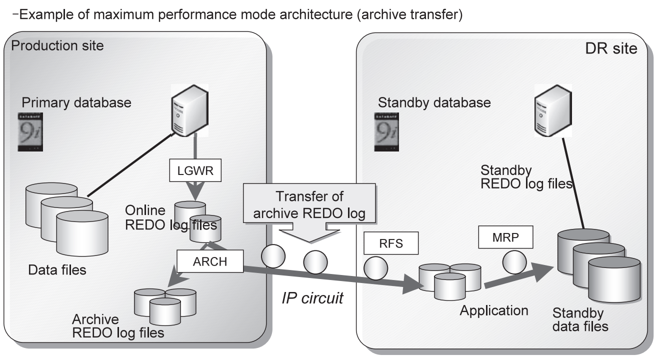

Fig. 6 ORACLE data guard.

data with a remote location using this function.
4) Other: A system using this method can be constructed by using RDBMS with a journal operation capability.

4.4 OS Layer Method

This method handles data replication on the extension of the clustering function.

Ordinary clustering technology necessitates a shared disk, but CLUSTERPRO can implement a cluster configuration without using a shared disk. This facility can be used to replicate data in the standby cluster node in a remote location. Since the basis is the clustering technology, it is relatively easy to apply this method to a system that has already installed the CLUSTERPRO environment. The standby node receiving data at the DR site can accept data from multiple active nodes so the equipment configuration can be reduced.

The special characteristic of this method is the possibility of failover in the standby node at the DR site. But care is required in the system operation because failover is possible only with systems with cluster configurations. All of the associated systems are not failed over at once.

Fig. 7 shows an example of configuration using CLUSTERPRO.

・Representative solutions
1) CLUSTERPRO
2) SystemGlobe RemoteCluster with MC/ServiceGuard

This function implements a cluster configuration in a remote location. However, it is indispensable to replicate data in the storage layer.

4.5 SAN Layer Method

This method replicates data using the FCSW (Fiber Channel Switch) of the SAN Island or an equivalent function.

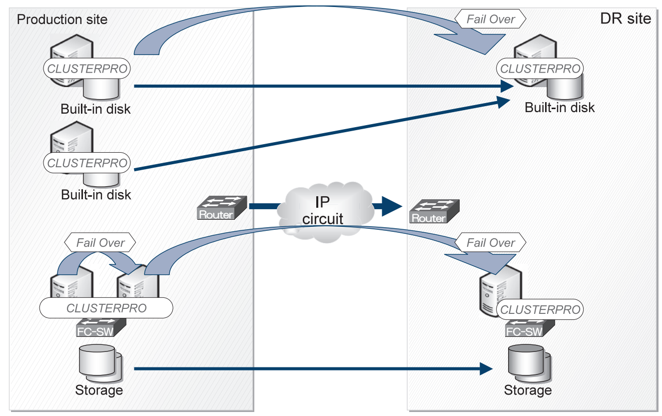

Fig. 7 Example of configuration using CLUSTERPRO.

Similarly to the middleware layer method, this method achieves data replication by acquiring the I/O data passing through the SAN at the production site and restoring the data at the SAN in the DR site. This technology can be regarded as an extension of the virtualization technology of SAN and, consequently, can be applied in an environment composed of storages of various coexisting models or of storages without a data replication capability.

However, the applicable systems are limited because the system must have the SAN configuration. In addition, the use of SAN configuration causes the problem of connectivity of SAN with the system, like other SAN configuration systems. This means that sufficient verification of connectivity is required before introducing this method.

Fig. 8 shows an example using SANsymphony.

・Representative solutions
1) MelodiousStor
2) System Globe SANsymphony
3) EMC/Invista

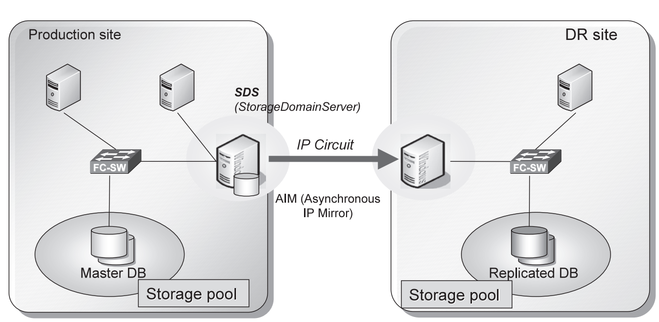

Fig. 8 SANsymphony.

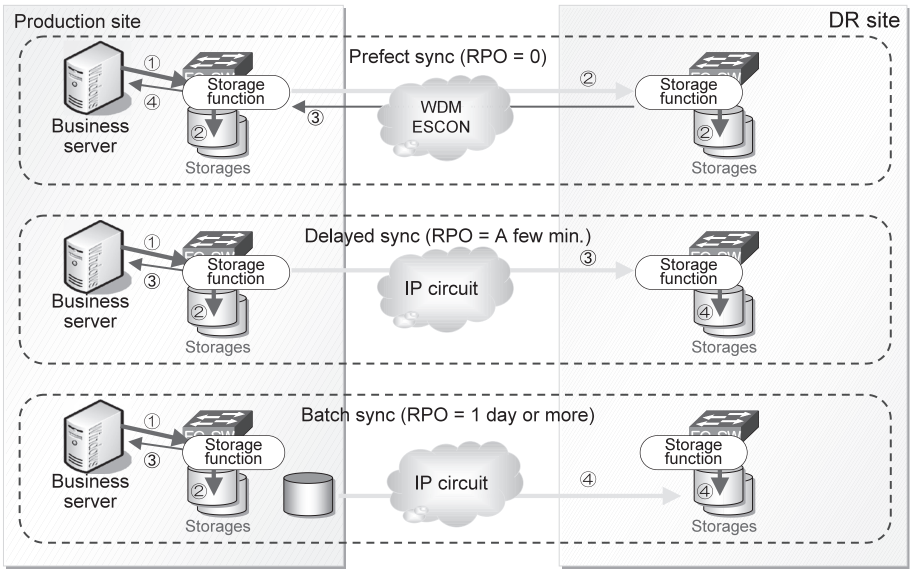

Fig. 9 Data replication in the storage layer.

4.6 Storage Layer Method

This method replicates data between storage cabinets.

As it replicates disk data physically using the function of the storage, so duplication is in principle possible only between storages of the same model (**Fig. 9**).

・Cascade method

With the perfect sync type, the applicable distance between sites is restricted due to the property of the processing method, and the distance limit is usually regarded as being about 60 kilometers. The cascade method is designed to avoid this restriction, and is composed of the combination of perfect sync type + delayed or batch sync type. **Fig. 10** shows an example of this method.

・Representative solutions

NEC iStorage

1) SystemGlove RDR (Remote Data Replication)

EMC Symmetrix

2) SRDF

3) Open Replicator (for use between different storage models)

FMC CLARiX

4) Mirror View family

5) SAN Copy (for use between different storage models)

4.7 Operation Layer Method

This method selects the data to be replicated, backs it up on

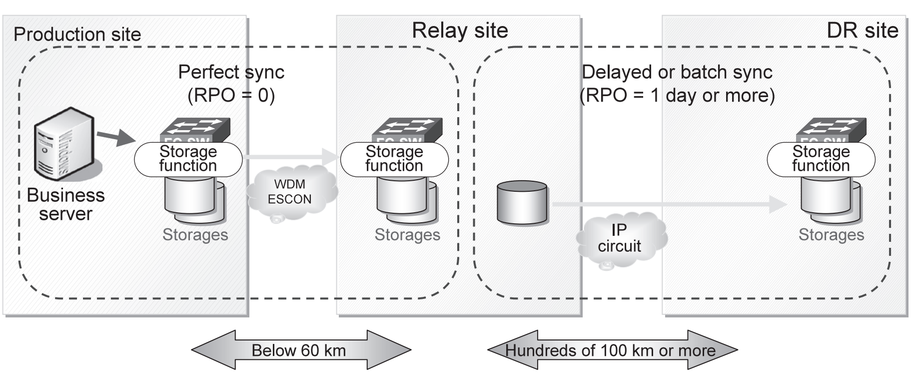

Fig. 10 Cascade method.

tape and sends the tapes to the DR site.

(1) Tape Transport Operation

This method is of the lowest cost but the operation load is significant. This method can be classified further into the following two methods according to the operation of the DR site receiving the tapes.

①Tape Storage Only

The DR sites simply store the tapes. In case of a disaster, data is recovered by restoration from the tapes.

② Pre-Restoration Operation

After the DR site receives the tapes, the data is restored in the DR site system in daily operation. This can reduce the RTO after a disaster by the same period as the restoration period.

Since the tape operation is accompanied by the outside transport of tapes, additional security measures should be applied when creating the tapes to be transported or by encrypting the transported tapes.

(2) Tape Image Transmission

This method can be regarded as an extension of the tape transport method. Instead of transporting the tape media physically, this method creates the data of the tape images in disks and transmits the disks to the DR site. Other operations details are identical to the above method.

(3) Remote Backup

This method does not transmit the tape image to the DR site but collects the data backup across the network. However, due to the problems in the WAN circuit bandwidth and product support, this method may not be widely applicable except in very few cases.

Fig. 11 shows an example of tape transport method operation.

・Representative Solutions

1) Tape Transport Service

2) Virtual Tape Device

iStorage T5100VT, Data Domain, etc.

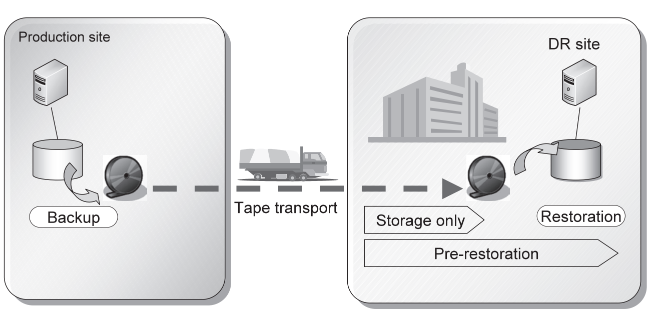

Fig. 11 Data replication in the storage layer (Tape transport method).

4.8 DR of ACOS-4

There are mainly three methods for replicating the data of ACOS-4, NEC proprietary mainframe computer.

(1) Data Synchronization in the Storage Layer

Since all of the ACOS-4 disks are composed of iStorage, the data replication method of iStorage can be applied. This method features a function that is not available with other open systems, which is the incorporation of an OS boot area in iStorage that makes it possible to synchronize data in the system area in addition to the ordinary synchronization of business data. This frees the DR site from the necessity of a system environment change and thereby decreases the operation load. As this method also synchronizes the system files that are changed in real-time, even the execution history of each job can be replicated in the DR site. This can greatly improve the ease of identification of the business status at the moment of a disaster.

(2) VSAS Journal Transfer Method, as an Application of "Data Synchronization in Database Layer"

The standard file system of ACOS-4 (VSAS) allows the journals to be collected from any file. As a result, data can be replicated by applying the journal operation to the flat files as well as the database.

(3) Method Using ACOS-Specific Tool

The disk area of iStorage used by ACOS can be accessed from an open system. This tool makes it possible to replicate data using an open system in place of the ACOS function. However, the same care on the code system and filenames as that in the case of ACOS ⇒ open system transition.

This method can be applied partially to ACOS-2.

・Representative solutions

1) DISOA/PX21 delayed transfer function

2) DISKSHARE & refam

4.9 DR of NAS

When a file server is built using an ordinary server OS, data can be replicated using the methods described above. In this section, we discuss the NAS-specific method of the NAS appliances.

The NAS appliances consist of a proprietary, OS and file system, so data replication is difficult to achieve using the above mentioned methods. Therefore, the NAS appliances are equipped with a specific data replication mentioned method. However, with Windows Powered NAS based on a Windows OS, a method similar to that used with some Windows servers can be selected.

In the data replication operation, it is required to understand the following characteristics that are specific to the file servers.

① Existence of a large number of users makes the usage authority setting important and complex. Linkage with an authentication function such as a Windows Active Directory is indispensable.

② The file environment contains “a huge number of very small files.” This property affects the response of replication.

・Representative solutions

iStorage NV Series

1) MVDSync

EMC Celerra Series

2) Celerra Replicator

Windows only

3) Windows 2003/R2 DPS Replication

5. Network Circuit

While achievement of DR is possible with network realization of data replication between sites, achievement of BC necessitates on considerations of the entire network associated with the system.

① Data Replication Network

The network for use in data replication is required to have high quality and a guaranteed bandwidth. It is the critical network of any mission-critical system.

② User Network

The network used by the system users to access the business system needs optimally to be capable of switching the site to the DR site in case of a disaster. Considerations for the security and switching of the authentication system are also important at such a time.

③ Outside Connection Network

The network for use in linkage with customer’s system located outside NEC should be prepared after concluding an agreement with the customer.

④ Development Network

To prepare for unforeseen circumstances, it may also be necessary to consider the possibility of DR of the development environment, though there may be a question of whether the development environment is to be used in the case of a disaster.

⑤ Monitoring Network

The network for use in operation management should be renewed by assuming a 2-site operation.

⑥ Remote Maintenance Network

Coordination with the service providers/vendors is required.

6. Conclusion

This paper has described the concepts supporting the implementation of BC/DR systems and the methods of data replication that support such systems.

It is expected that in the future the demand for BC/DR will increase following realization of the predicted increase in data capacity and due to further growth in dependency on IT systems. In this context it is anticipated that the diffusion of BC/DR systems will be driven by WAN cost down, etc. It is also expected that in the future BC/DR systems will advance by applying the emerging business grid technology.

We are determined to continue to provide stable systems for our customers by providing optimum combinations of NEC solutions.

Authors' Profiles

TAKAYA Masahiro
Group Manager,
IT Platform Systems Development Division,
Systems Software Operations Unit,
NEC Corporation

TOMI Mitsuhiro
Manager,
IT Platform Division,
Platform Integrated Operations Unit,
NEC Systems Technologies, Ltd.