UNIT-4
Chapter 11
Introduction to Business Continuity

Continuous access to information is a must for the smooth functioning of business operations today, as the cost of business disruption could be catastrophic. There are many threats to information availability, such as natural disasters (e.g., flood, fire, earthquake), unplanned occurrences (e.g., cybercrime, human error, network and computer failure), and planned occurrences (e.g., upgrades, backup, restore) that result in the inaccessibility of information. It is critical for businesses to define appropriate plans that can help them overcome these crises. Business continuity is an important process to define and implement these plans.

Business continuity (BC) is an integrated and enterprise-wide process that includes all activities (internal and external to IT) that a business must perform to mitigate the impact of planned and unplanned downtime. BC entails preparing for, responding to, and recovering from a system outage that adversely affects business operations. It involves proactive measures, such as business impact analysis and risk assessments, data protection, and security, and reactive countermeasures, such as disaster recovery and restart, to be invoked in the event of a failure. The goal of a business continuity solution is to ensure the “information availability” required to conduct vital business operations.
Section III: Business Continuity

11.1 Measuring Information Availability

11.1.2 Causes of Information Unavailability

Various planned and unplanned incidents result in data unavailability. Planned outages include installation/integration/maintenance of new hardware, software upgrades or patches, taking backups, application and data restores, facility operations (renovation and construction), and refresh/migration of the testing to the production environment. Unplanned outages include failure caused by database corruption, component failure, and human errors.

Another type of incident that may cause data unavailability is natural or man-made disasters such as flood, fire, earthquake, and contamination. As illustrated in Figure 11-1, the majority of outages are planned. Planned outages are expected and scheduled, but still cause data to be unavailable. Statistically, less than 1 percent is likely to be the result of an unforeseen disaster.
Section III - Business Continuity

at the component level, which may result in data unavailability. MTTR includes the time required to do the following: detect the fault, mobilize the maintenance team, diagnose the fault, obtain the spare parts, repair, test, and resume normal operations.

IA is the fraction of a time period that a system is in a condition to perform its intended function upon demand. It can be expressed in terms of system uptime and downtime and measured as the amount or percentage of system uptime:

\[ IA = \frac{\text{system uptime}}{\text{system uptime} + \text{system downtime}} \]

In terms of MTBF and MTTR, IA could also be expressed as

\[ IA = \frac{MTBF}{MTBF + MTTR} \]

Uptime per year is based on the exact timeliness requirements of the service, this calculation leads to the number of "9s" representation for availability metrics. Table 11-1 lists the approximate amount of downtime allowed for a service to achieve certain levels of 9s availability.

For example, a service that is said to be "five 9s available" is available for 99.999 percent of the scheduled time in a year (24 x 7 x 365).

Table 11-1: Availability Percentage and Allowable Downtime

<table>
<thead>
<tr>
<th>UPTIME (%)</th>
<th>DOWNTIME (%)</th>
<th>DOWNTIME PER YEAR</th>
<th>DOWNTIME PER WEEK</th>
</tr>
</thead>
<tbody>
<tr>
<td>98</td>
<td>2</td>
<td>7.3 days</td>
<td>3 hr 22 minutes</td>
</tr>
<tr>
<td>99</td>
<td>1</td>
<td>3.65 days</td>
<td>1 hr 41 minutes</td>
</tr>
<tr>
<td>99.8</td>
<td>0.2</td>
<td>17 hr 31 minutes</td>
<td>20 minutes 10 sec</td>
</tr>
<tr>
<td>99.9</td>
<td>0.1</td>
<td>8 hr 45 minutes</td>
<td>10 minutes 5 sec</td>
</tr>
<tr>
<td>99.99</td>
<td>0.01</td>
<td>52.5 minutes</td>
<td>1 minute</td>
</tr>
<tr>
<td>99.999</td>
<td>0.001</td>
<td>5.25 minutes</td>
<td>6 sec</td>
</tr>
<tr>
<td>99.9999</td>
<td>0.0001</td>
<td>31.5 sec</td>
<td>0.6 sec</td>
</tr>
</tbody>
</table>

11.1.3 Consequences of Downtime

Data unavailability, or downtime, results in loss of productivity, loss of revenue, poor financial performance, and damages to reputation. Loss of productivity reduces the output per unit of labor, equipment, and capital. Loss of revenue includes direct loss, compensation payments, future revenue losses, billing losses, and investment losses. Poor financial performance affects revenue
Chapter 11 - Introduction to Business Continuity

recognition, cash flow, discounts, payment guarantees, credit rating, and stock price. Damages to reputation may result in a loss of confidence or credibility with customers, suppliers, financial markets, banks, and business partners. Other possible consequences of downtime include the cost of additional equipment rental, overtime, and extra shipping.

The business impact of downtime is the sum of all losses sustained as a result of a given disruption. An important metric, average cost of downtime per hour, provides a key estimate in determining the appropriate BC solutions. It is calculated as follows:

Average cost of downtime per hour = average productivity loss per hour + average revenue loss per hour

Where:

Productivity loss per hour = (total salaries and benefits of all employees per week) / (average number of working hours per week)

Average revenue loss per hour = (total revenue of an organization per week) / (average number of hours per week that an organization is open for business)

The average downtime cost per hour may also include estimates of projected revenue loss due to other consequences such as damaged reputations and the additional cost of repairing the system.

11.2 BC Terminology

This section introduces and defines common terms related to BC operations and are used in the next few chapters to explain advanced concepts:

- **Disaster recovery**: This is the coordinated process of restoring systems, data, and the infrastructure required to support key ongoing business operations in the event of a disaster. It is the process of restoring a previous copy of the data and applying logs or other necessary processes to that copy to bring it to a known point of consistency. Once all recoveries are completed, the data is validated to ensure that it is correct.

- **Disaster restart**: This is the process of restarting business operations with mirrored consistent copies of data and applications.

- **Recovery-Point Objective (RPO)**: This is the point in time to which systems and data must be recovered after an outage. It defines the amount of data loss that a business can endure. A large RPO signifies high tolerance to information loss in a business. Based on the RPO, organizations plan for the minimum frequency with which a backup or replica must be made. For
example, if the RPO is six hours, backups or replicas must be made at least once in 6 hours. Figure 11-2 shows various RPOs and their corresponding ideal recovery strategies. An organization can plan for an appropriate BC technology solution on the basis of the RPO it sets. For example:

- **RPO of 24 hours**: This ensures that backups are created on an offsite tape drive every midnight. The corresponding recovery strategy is to restore data from the set of last backup tapes.
- **RPO of 1 hour**: This ships database logs to the remote site every hour. The corresponding recovery strategy is to recover the database at the point of the last log shipment.
- **RPO of zero**: This mirrors mission-critical data synchronously to a remote site.

![Diagram showing recovery point and time objectives](image)

**Figure 11-2: Strategies to meet RPO and RTO targets**

- **Recovery-Time Objective (RTO)**: The time within which systems, applications, or functions must be recovered after an outage. It defines the amount of downtime that a business can endure and survive. Businesses can optimize disaster recovery plans after defining the RTO for a given data center or network. For example, if the RTO is two hours, then use a disk backup because it enables a faster restore than a tape backup. However, for an RTO of one week, tape backup will likely meet requirements. Some examples of RTOs and the recovery strategies to ensure data availability are listed below (refer to Figure 11-2):
  - **RTO of 72 hours**: Restore from backup tapes at a cold site.
  - **RTO of 12 hours**: Restore from tapes at a hot site.
  - **RTO of 4 hours**: Use a data vault to a hot site.
Chapter 11 • Introduction to Business Continuity

- **RTO of 1 hour**: Cluster production servers with controller-based disk mirroring.
- **RTO of a few seconds**: Cluster production servers with bidirectional mirroring, enabling the applications to run at both sites simultaneously.

- **Data vault**: A repository at a remote site where data can be periodically or continuously copied (either to tape drives or disks), so that there is always a copy at another site.

- **Hot site**: A site where an enterprise's operations can be moved in the event of disaster. It is a site with the required hardware, operating system, application, and network support to perform business operations, where the equipment is available and running at all times.

- **Cold site**: A site where an enterprise's operations can be moved in the event of disaster, with minimum IT infrastructure and environmental facilities in place, but not activated.

- **Cluster**: A group of servers and other necessary resources, coupled to operate as a single system. Clusters can ensure high availability and load balancing. Typically, in failover clusters, one server runs an application and updates the data, and another server is kept redundant to take over completely, as required. In more sophisticated clusters, multiple servers may access data, and typically one server performs coordination.

### 11.3 BC Planning Lifecycle

BC planning must follow a disciplined approach like any other planning process. Organizations today dedicate specialized resources to develop and maintain BC plans. From the conceptualization to the realization of the BC plan, a lifecycle of activities can be defined for the BC process. The BC planning lifecycle includes five stages (see Figure 11-3):

1. Establishing objectives
2. Analyzing
3. Designing and developing
4. Implementing
5. Training, testing, assessing, and maintaining
Figure 11-3: BC planning lifecycle

Several activities are performed at each stage of the BC planning lifecycle, including the following key activities:

1. Establishing objectives
   - Determine BC requirements.
   - Estimate the scope and budget to achieve requirements.
   - Select a BC team by considering subject matter experts from all areas of the business, whether internal or external.
   - Create BC policies.

2. Analyzing
   - Collect information on data profiles, business processes, infrastructure support, dependencies, and frequency of using business infrastructure.
   - Identify critical business needs and assign recovery priorities.
   - Create a risk analysis for critical areas and mitigation strategies.
   - Conduct a Business Impact Analysis (BIA).
   - Create a cost and benefit analysis based on the consequences of data unavailability.
   - Evaluate options.
3. Designing and developing
   - Define the team structure and assign individual roles and responsibilities. For example, different teams are formed for activities such as emergency response, damage assessment, and infrastructure and application recovery.
   - Design data protection strategies and develop infrastructure.
   - Develop contingency scenarios.
   - Develop emergency response procedures.
   - Detail recovery and restart procedures.

4. Implementing
   - Implement risk management and mitigation procedures that include backup, replication, and management of resources.
   - Prepare the disaster recovery sites that can be utilized if a disaster affects the primary data center.
   - Implement redundancy for every resource in a data center to avoid single points of failure.

5. Training, testing, assessing, and maintaining
   - Train the employees who are responsible for backup and replication of business-critical data on a regular basis or whenever there is a modification in the BC plan.
   - Train employees on emergency response procedures when disasters are declared.
   - Train the recovery team on recovery procedures based on contingency scenarios.
   - Perform damage assessment processes and review recovery plans.
   - Test the BC plan regularly to evaluate its performance and identify its limitations.
   - Assess the performance reports and identify limitations.
   - Update the BC plans and recovery/restart procedures to reflect regular changes within the data center.
11.4 Failure Analysis

Failure analysis involves analyzing the data center to identify systems that are susceptible to a single point of failure and implementing fault-tolerance mechanisms such as redundancy.

11.4.1 Single Point of Failure

A *single point of failure* refers to the failure of a component that can terminate the availability of the entire system or IT service. Figure 11-4 illustrates the possibility of a single point of failure in a system with various components: server, network, switch, and storage array. The figure depicts a system setup in which an application running on the server provides an interface to the client and performs I/O operations. The client is connected to the server through an IP network, the server is connected to the storage array through an FC connection, an HBA installed at the server sends or receives data to and from a storage array, and an FC switch connects the HBA to the storage port.

![Figure 11-4: Single point of failure](image)

In a setup where each component must function as required to ensure data availability, the failure of a single component causes the failure of the entire data center or an application, resulting in disruption of business operations. In this example, several single points of failure can be identified. The single HBA on the server, the server itself, the IP network, the FC switch, the storage array ports, or even the storage array could become potential single points of failure. To avoid single points of failure, it is essential to implement a fault-tolerant mechanism.

11.4.2 Fault Tolerance

To mitigate a single point of failure, systems are designed with redundancy, such that the system will fail only if all the components in the redundancy group fail. This ensures that the failure of a single component does not affect
data availability. Figure 11-5 illustrates the fault-tolerant implementation of the system just described (and shown in Figure 11-4).

Data centers follow stringent guidelines to implement fault tolerance. Careful analysis is performed to eliminate every single point of failure. In the example shown in Figure 11-5, all enhancements in the infrastructure to mitigate single points of failures are emphasized:

- Configuration of multiple HBAs to mitigate single HBA failure.
- Configuration of multiple fabrics to account for a switch failure.
- Configuration of multiple storage array ports to enhance the storage array's availability.
- RAID configuration to ensure continuous operation in the event of disk failure.
- Implementing a storage array at a remote site to mitigate local site failure.
- Implementing server (host) clustering, a fault-tolerance mechanism whereby two or more servers in a cluster access the same set of volumes. Clustered servers exchange heartbeats to inform each other about their health. If one of the servers fails, the other server takes up the complete workload.

**Figure 11-5:** Implementation of fault tolerance