1 Downtime and Availability

CHAPTER OBJECTIVES

In this chapter, you will learn to:

- ✓ Describe and differentiate the terms *downtime* and *availability*
- ✓ Explain the business impact of system downtime
- ✓ Describe and calculate the *Mean Time between Failures* (MTBF) and estimate its impact on the IT environment
- ✓ Describe the *Mean Time to Repair* (MTTR) and explain its importance in the overall recovery
- ✓ List and describe technologies that address availability, such as *Redundant Array of Inexpensive* (or *Independent*) *Disks* (RAID), snapshots, and replication, and contrast them with backup
- ✓ Explain and use *Recovery Point Objective* (RPO) and *Recovery Time Objective* (RTO) as parameters for backup planning

Introduction

When it comes to downtime, availability, fault tolerance, backup, recovery, and data archiving, several topics come to mind. Almost everyone knows something about them, usually firsthand. Who has not faced a system failure, a loss of data, or some type of a disaster that affected normal operations of a network, the Internet, a point-of-sale system, or even your own computer?

All these terms deal with protecting business or personal assets—either systems, applications, processes, data, or the ability to conduct business. Each varies, however, in its approach and methodology.

This chapter lays the foundation for this course and the subsequent chapters by defining and describing the key concepts behind the quest for protecting IT resources. It begins by describing and differentiating between downtime and availability. It then explains how a component or a system failure, which leads to downtime, is measured and prevented. Lastly, it takes the opposite view, one of system availability, and how it can be achieved. Within the context of both downtime and availability, it positions backup and recovery.

Downtime

Let us begin by focusing on what prevents a system from operating normally and what the

consequences might be.

Defining downtime

Downtime is a period of time during which a system, an application, a process, a service, or data is unavailable. During this time, a business entity cannot conduct sales operations, support a customer, provide a service, or conduct transactions, that is, if this business entity solely relies upon what just became unavailable.

There are many kinds of downtime, but all cause some type of disruption to normal business operations. These events cause loss of productivity, inability to communicate with clients, or even loss of sensitive or important information. Downtime of an IT system can cause loss of productivity of a single user, a workgroup, or even the entire company. Whenever downtime impairs business, the outage carries serious consequences.

To prevent or minimize the impact of downtime, you must understand what causes downtime in the first place.

Describing causes of downtime

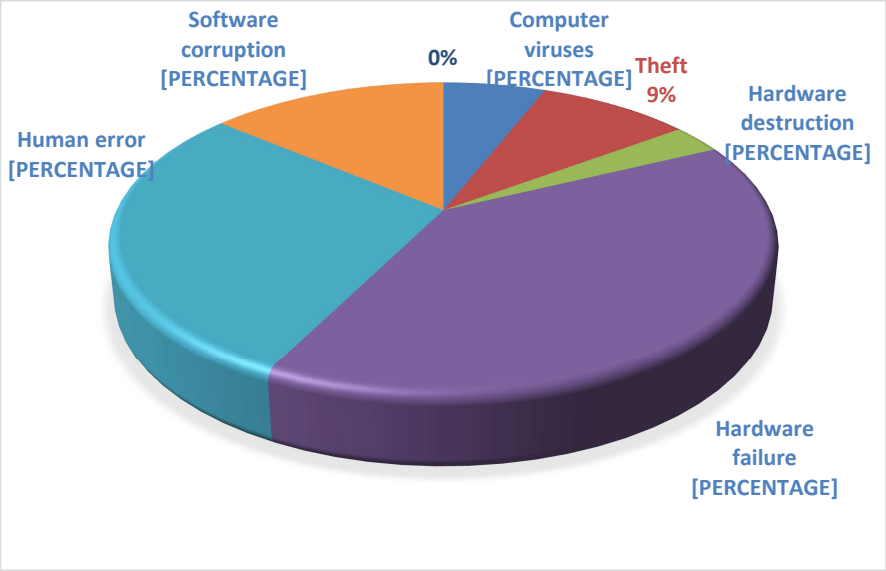

Figure 1-1 Causes of downtime and data loss or unavailability

Figure 1-1 lists the most typical causes of downtime and data loss or unavailability. Note that the largest cause of downtime is due to hardware failures and then to human error. Other causes, not

shown in this chart, include the following:

- Power outages
- Data center cooling issues
- Software bugs
- Cyber attacks
- Database inconsistencies or design flaws
- _____ (add your own)
- _____ (add your own)
- _____ (add your own)

Disasters, without a doubt, affect computer operations and data unavailability. A disaster in the IT world is any event that disrupts a company's computer operations and causes downtime. Disasters are not controllable, but precautions against them make the recovery much faster and easier.

Disasters can be categorized according to the affected area:

- Building-level incidents
- Metropolitan area disasters
- Regional events

Furthermore, downtime can be planned as well as unplanned.

Building-level incidents

Figure 1-2 Disasters affecting a building

Disasters affecting a building (Figure 1-2) usually impact computer operations in that building as well. There may not be a direct damage to the IT systems, but these incidents may prevent access to them or to the data they host, or they may interrupt operations.

Metropolitan area disasters

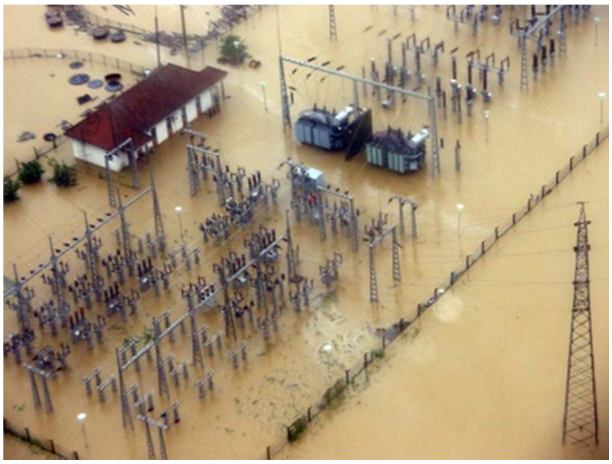

Figure 1-3 Metropolitan area disasters

Usually floods, fires, large chemical incidents, moderate earthquakes, severe winter storms, or blackouts affect entire cities, impacting their infrastructure and disrupting IT systems (Figure 1-3).

Regional events

Figure 1-4 Natural disasters

Computer operations may be interrupted by natural disasters that affect an entire region within a radius of hundreds to tens of thousands of miles/kilometers. These disasters include large floods, hurricanes, earthquakes, political instability, and wars.

Planned vs. unplanned downtime

Planned downtime occurs when system administrators intentionally restrict or stop the system operations to implement upgrades, updates, repairs, or other changes. During planned downtime, a particular time period is set aside for these operations, which are carefully planned, prepared, executed, and validated. On the contrary, unplanned downtime is when an unintentional intervention restricts or stops the system availability.

Planned downtime means downtime as well.

Return, for a moment, to Figure 1-1, which shows the major causes of downtime and data loss/unavailability. Are these causes intentional/planned or unintentional/unplanned? If you answered *unintentional/unplanned*, you answered correctly.

Overall, the majority of system and data unavailability is because of planned downtime due to required maintenance and upgrades. In fact, unplanned downtime accounts for only about 10% of all downtime, but its unexpected nature means that any single downtime incident may be more damaging to the enterprise, both physically and financially, than many occurrences of planned downtime.

Thus, understanding the cost of downtime is critical in either case.

Estimating the cost and impact of downtime

Quantifying downtime is not an easy task because the impact of downtime varies from one case to another. Losing a second of time in an air traffic control system or in a hospital life-support environment can have dire consequences, while losing hours in a billing system may not have a significant impact at all if these billing transactions were queued and committed only when the system became available again.

Before you can calculate downtime, you must know its root cause. And not all root causes are strictly IT issues. To begin with, it is important to identify and understand both internal and external downtime threats—you need to know *what* and *who* has the potential to take your business down. You also need to know *how*.

IT-related outages, planned or unplanned, can unleash a procession of costs and consequences that are direct and indirect, tangible and intangible, short term and long term, and immediate and far-reaching.

Tangible and direct costs[i]

Tangible and direct costs refer to expenses that can easily be measured and documented, are incurred up front, and are tracked in the business general ledger.

These costs can be “touched” or “felt” or “easily determined.”

Tangible and direct costs related to downtime include:

- Loss of transaction revenue
- Loss of wages due to employees' idle time
- Loss of inventory
- Remedial labor costs
- Marketing costs
- Bank fees
- Legal penalties from inability to deliver on service-level agreements (SLAs)

Intangible and indirect costs

Intangible and indirect costs refer to business impact that is more difficult to measure, is often felt or incurred at a later date, and is not related to a physical substance or intrinsic productive value.

These costs are nonetheless real and important to a business' success or failure. They can be more important and greater than tangible costs.

Intangible and indirect costs related to downtime include the following:

- Loss of business opportunities
- Loss of employees and/or employee morale
- Loss of goodwill in the community
- Decrease in stock value
- Loss of customers and/or departure of business partners
- Brand damage
- Shift of market share to competitors
- Bad publicity and press

Outage impact to business

The cost that can be assigned to a measurable period of downtime varies widely depending upon the nature of the business, the size of the company, and the criticality of the IT system related to the primary revenue-generating processes. For example, a global financial services firm may lose millions of dollars for every hour of downtime, whereas a small manufacturer using IT as an administrative tool would lose only a margin of productivity.

According to a Gartner document titled *How Much Does an Hour of Downtime Cost?*, for a

conventional brick-and-mortar business, estimating the cost of an outage is relatively simple, compared to, let us say, a global financial services firm. In either case, such estimations are never trivial.

Consider this example:

Assume that a firm conducts business in Western Europe and North America during regular business hours. This firm needs its systems and services to be available 40 hours per week, or 2000 hours per year (accounting for holidays, vacations, and weekends).

Therefore, the first order of approximation for the cost of an outage would be to distribute the firm's revenue uniformly across those 2000 hours. Thus, if the firm's annual revenue is $100 million, an average hour would represent $50,000 of its revenue. Consequently, one-hour outage would cost this firm $50,000.

Two immediate objections arise to this assessment and both lead to important refinements. First, revenue is almost never distributed uniformly across all working hours. Second, most businesses experience seasonal fluctuations. Many retail organizations make 40% of their revenue and 100% of their profits in the last eight weeks of the year. One-hour outage on December 23 will have a much greater impact on the firm's financial performance as compared to the same outage in late June, for instance. Therefore, the cost of an outage must reflect its potential impact at a particular time in the business cycle.

Table 1-1 shows the cost of downtime, estimated in 1998 by the Gartner Group. You can see that the cost of downtime is company-specific and related intangible costs are very difficult to estimate.

Table 1-1 The cost of downtime, estimated in 1998 by the Gartner Group

Industry	Application	Average cost (per hour of downtime)
Financial	Brokerage operations	$6,500,000
Financial	Credit card sales	$2,600,000
Media	Pay-per-view	$1,150,000
Retail	Home shopping (TV)	$113,000
Retail	Catalog sales	$90,000
Transportation	Airline reservations	$89,500

Consequences of downtime

What happens to an organization when its system goes down?

Before you can predict downtime and plan its prevention, you must have a clear understanding of what happens when a system goes down. Specifically, *what* and *who* is impacted and *how*.

You should consider the following:

- **Processes:** Vital business processes such as order management, inventories, financial reporting, transactions, manufacturing, and human resources may be interrupted, corrupted, or even lost.
- **Programs:** Revenue can be affected and a key employee or customer activities might be missed or lost.
- **Business:** If customers cannot access a website, they might purchase from someone else. You might lose a customer now and forever.
- **People:** Salaries might not be paid, and even lives could be lost due to downtime of life-sustaining medical systems.
- **Projects:** Thousands of person-hours of work can be lost and deadlines can be missed, resulting in failure-to-perform fees and noncompliance penalties.
- **Operations:** Those who manage daily activities of an organization may find themselves without the data they need to make informed decisions.

The Gartner Group recommends estimating the cost of an outage to your firm by calculating lost revenue, lost profit, and staff cost for an average hour—and for a worst-case hour—of downtime for each critical business process. Not-for-profit organizations cannot calculate revenues or profits, so they should focus on staff productivity and qualitative assessment of the outage to their user community. An outage can weaken customer perception of the firm, harm the wider community, and derail a firm's strategic initiatives, but these impacts may be difficult to quantify, and should, in most cases, be left unquantified. Any outage assessment based on raw, generic industry averages alone is misleading.

Predicting downtime

Can you predict downtime and events that lead to it? If you could, would it mean that you could then better prepare for such events, or even prevent them from happening?

There are causes of downtime which cannot be predicted, such as natural disasters or fire. You just have to determine the best prevention or recovery mechanism if they occur. On the other hand, failure rates of computer system components can be predicted with a level of certainly. To determine and express failure rates of computer components, you can calculate the MTBF.

Defining MTBF

MTBF is the measure of expected failure rates. To understand MTBF, it is best to start with something else—something for which it is easier to develop an intuitive feel.

Let us take a look at a generalized MTBF measure for a computer component, such as a hard drive,

a memory DIMM, or a cooling fan. This component has an MTBF of 200,000 hours. Since there are approximately 9000 hours in a year, 200,000 hours is about 21 years. In other words, if the MTBF of this component is 200,000 hours, it is expected that the component fails every 21 years.

Now, take a sample of 10,000 units of this particular component and determine how many units fail every day, over a test period of 12 months. You may determine that

- In the first 24 hours, two components fail.
- In the second 24 hours, zero components fail.
- In the third 24 hours, one component fails, and so on.

You then ask

- If five units fail in 12 months, how long would it take for all 10,000 units to fail at this rate?
- If all units fail in regular intervals over this period, how long is this interval?

If a failure is the termination of the component's ability to perform its intended function, what is then MTBF? MTBF is an interval of time used to express the expected failure rate of a given component. It does not indicate the expected lifetime of that component and says nothing about the failure likelihood of a single unit.

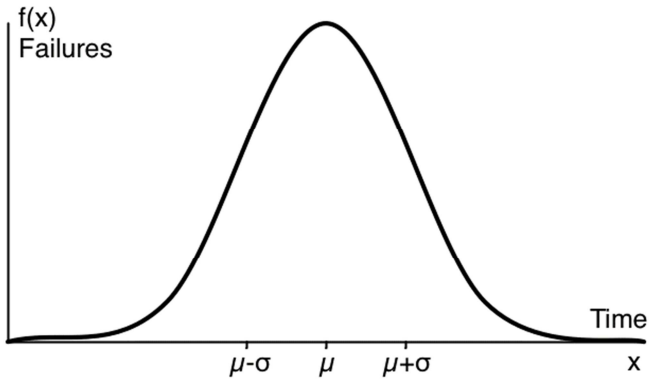

Figure 1-5 Reliability bell curve: number of failures vs. time distribution

This reliability bell curve (Figure 1-5) is also known as a normal distribution, where the highest point of the curve (or the top of the bell) represents the most probable event (at time μ). All possible events are then distributed around the most probable event, which creates a downward slope on each side of the peak.

Given the reliability bell curve and a sample of components, most of them will fail around time μ. Some components will fail early in the cycle, whereas others will last much longer. But all of them

will fail at some point on this curve, with a high probability that the majority of failures will be distributed according to the bell curve.

INTERNET:
Trent Hamm has written a good explanation of the reliability bell curve in August 2014, using the analogy of low-end and more reliable washing machines. You can find this article at: http://www.thesimpledollar.com/the-reliability-bell-curve-what-does-more-reliability-actually-mean/.

Calculating MTBF

How does one calculate an MTBF for an environment which consists of a number of such components? What if all these components were identical? What if they all were different?

Once again, let us use an example:

If you have 2000 identical units in your environment and each unit has the MTBF of 200,000 hours, what is the associated total MTBF?

The formula for the total MTBF is as follows:

$$\text{MTBF}_{\text{total}} = \frac{MTBF \text{ of the } unit}{\#\ of\ units}$$

Therefore,

$$\text{MTBF}_{\text{total}} = \frac{MTBF \text{ of the } unit}{\#\ of\ units} = \frac{200{,}000\ hours}{2000\ units} = 100 \text{ hours} = \sim 4 \text{ days}$$

Within your particular environment of 2,000 identical units, you can expect a failure approximately every 4 days.

It may be easy to understand the MTBF of a single unit, but in reality, systems are complex and consist of a number of different components. Within a complex environment, you can expect a failure of any particular component, a failure between components, or a failure of the entire system.

The measure of a failure rate for a system consisting of one to *n* components, which may not necessarily be identical, is expressed as the *MTBF for a complex system*. Its formula is as follows:

$$\text{MTBF}_{\text{complex system}} = \frac{1}{\frac{1}{MTBF1} + \frac{1}{MTBF2} + \cdots + \frac{1}{MTBFn}}$$

Even though individual MTBF rates for a single component have been improving, the entire environment is still vulnerable to component failures due to the interdependent nature of complex and multivendor solutions. Consider these examples:

- **Hardware:** Storage systems may fail to provide adequate service due to MTBF problems or due to corrupted data (resulting from viruses or data integrity issues). Computing systems may lose power or can have various other problems, such as memory or processor failures. One corner stone for a highly available environment is a good network. High-speed connections for transaction processing and robust connections for client/server backup are critical.
- **Software:** The basis of any system is a stable operating system because a failure at the OS level may lead to vulnerability and data loss in everything that is at higher levels of the stack, such as applications, databases, and processes.

Defining and calculating MTTR

You are looking at a component or a system failure rate in order to predict, prevent, and minimize downtime. There is another variable which has not yet been discussed, one of recovery. *As soon as a unit fails, how long does it take to repair or replace it?* To answer this question, another term is used, called MTTR (Mean Time to Repair or Mean Time to Recovery).

MTTR represents the average time it takes to repair or replace the defective component and return the system to its full operation. It includes the time needed to identify the failure, to diagnose it and determine its root cause, and to rectify it.

MTTR can be a major component of downtime, especially if you are unprepared.

Planning for and preventing downtime

While planning for downtime, you have to be aware of the level of problem you are planning for. While you cannot influence metropolitan or regional incidents because they are usually naturally occurring disasters, you can create an appropriate disaster recovery plan and implement a disaster recovery solution once you know what it is you are planning for (also called *local problem area*). Many times, you can also put in place a solution that may reduce the impact of a disaster or even prevent it altogether.

Lastly, regardless of the cause, location, or the level of downtime, some type of administrative intervention is usually required to quickly recover from the problem. This intervention always depends on the exact cause of the problem and differs from one case to another. To increase your chances of success, make sure that you have

- Adequate problem detection in place
- A recovery plan that is validated and tested
- Protection technologies that ensure desired levels of availability
- Monitoring and diagnostic tools and processes

- Skilled staff educated in the adopted technologies and recovery processes
- _____ (add your own)
- _____ (add your own)
- _____ (add your own)

Learning check

Reinforce your knowledge and understanding of the topics just covered by completing this learning check:

1. What measures the time needed to repair a component after it fails?
    a. MTBF
    b. MTTR
    c. RTO
    d. RPO
2. What measures the MTBFs of a complex system?
    a. MTBF
    b. MTTR
    c. RTO
    d. RPO
3. Floods, wildfires, tornados, and severe snowstorms are examples of what type of disasters?
    a. Building-level incidents
    b. Metropolitan area disasters
    c. Regional events
    d. Localized incidents
4. What is the major cause of system and data unavailability?
    a. User errors
    b. Metropolitan area disasters
    c. Planned downtime
    d. Inadequate backup strategy

Learning check answers

This section contains answers to the learning check questions.

1. What measures the time needed to repair a component after it fails?
   a) MTBF
   b) MTTR
   c) RTO
   d) RPO
2. What measures the MTBFs of a complex system?
   a) MTBF
   b) MTTR
   c) RTO
   d) RPO
3. Floods, wildfires, tornados, and severe snowstorms are examples of what type of disasters?
   a) Building-level incidents
   b) Metropolitan area disasters
   c) Regional events
   d) Localized incidents
4. What is the major cause of system and data unavailability?
   a) User errors
   b) Metropolitan area disasters
   c) Planned downtime
   d) Inadequate backup strategy

Availability

Up to now, this module covered the effects of downtime of a computer system or unavailability of data due to a failure or a disaster. The opposite of downtime is *availability* (or *uptime*).

This section discusses the terminology and methods of expressing and calculating the availability times and requirements.

Learner activity

Before continuing with this section, see if you can define the terms below and answer the questions. Search the Internet if you need to.

- In your own words, define these terms:
    - High availability:

_____

_____

_____

_____

    - Fault tolerance:

_____

_____

_____

_____

    - RPO:

_____

_____

_____

_____

    - RTO:

_____

_____

_____