Chapter 4:

Cloud Service Supports

Cloud Service Supports

4.1 Learning Objectives

After completing this chapter, learners will be able to:

1. Distinguish between service implementation and support in the cloud.
2. Describe the testing and validation requirements for post-cloud implementation.
3. Articulate the special role that configuration management plays in cloud computing.
4. Identify resource management challenges with cloud computing implementation.

4.2 Overview

Section 1.4 discussed cloud implementation from a high-level view and covered topics such as establishing teams, migrating data and conducting post-implementation reviews—all of which could apply to any "typical" in-house IT implementation. In some ways, it is like traditional infrastructure; cloud is, in fact, hardware and software and is, therefore, subject to many, if not most, of the same concerns as the enterprise's in-house systems.

But cloud is not exactly like an in-house implementation. While the technology may be the same, the business model is different. Its very essence—the fact that the enterprise lacks authority over its implementation, controls, testing, business practices, compliance activities, etc., and is dependent on the provider to perform up to the enterprise's expectations—causes it to present unique challenges that require special attention.

4.3 Testing and Validation

Enterprises are accustomed to testing and validating their own systems. They test resilience, security, controls, compatibility and a host of other elements of the technology, the programming and the users who operate it, with the purpose of checking that a system meets specifications and fulfills its intended purpose. In validating their own systems, enterprises freeze the systems, and any intermittent changes requested are implemented under a strict change control process.

This sort of testing is unavailable to enterprises in the cloud. Enterprises cannot freeze and control the changes happening to the system. Changes and improvements are continuous, necessitating a process of continuous validation to be sure that those changes are not negatively impacting the state of the system. Enterprises must be assured that cloud-hosted applications are successfully qualified and in a validated state for the entire duration of their life cycle, and they remain in compliance with the regulating bodies.

The testing and validation process must be negotiated through the SLA and tracked through monitoring. And, while CSPs are generally on board with testing, they may be reluctant to share all aspects of it due to the proprietary nature of some products or the security concerns that may arise from revealing processes.

The enterprise must fully understand the nature of its data being placed in the cloud to ensure proper validation:

- Is it PII?
- Does it have specific compliance requirements, such as those associated with Payment Card Industry Data Security Standard (PCI DSS), the US Health Insurance Portability and Accountability Act (HIPAA), the US Sarbanes-Oxley Act (SOX) or the European Union's General Data Protection Regulation (GDPR)?

The answers to these questions can inform the enterprise's discussions with the CSP and inspire other questions the CSP should address, such as:

- Can a test period be provided for new application candidates to identify possible issues early in the deployment?
- If data are demanded by authorities, can the data be provided without compromising other information?
- What sort of compliance testing is done? How often?
- When change plans are presented, does the CSP perform compatibility testing?
- Is the backend secure and does the enterprise have authorization to run tests on it? If so, how?
- Are business continuity and disaster recovery plans well documented and tested?
- Is segregation of duties between provider employees maintained? How is different customers' information segregated? What controls are in place to prevent, detect and react to breaches?
- How often is backup testing done? Are the procedures for doing the testing consistent with the enterprise's security policy?

Validation processes conducted by the CSP should be supported by a validation plan, current versions of all requirements documents and any qualification testing documents. Validation testing will ensure that all predefined requirements and specifications are successfully challenged and met, and testing is properly documented.

The basic principles of validation, which must be considered when tackling the validation of a cloud-based system, are:[14]

1. Fitness for the intended use of a computerized system is verified and documented.
2. Risk associated with using the system is mitigated.
3. The validated state is maintained through effective change control mechanisms.

Enterprises will be especially interested in the supplier assessment, which determines the provider's ability to meet the enterprise's expectations surrounding controls to ensure security, data integrity, compliance with the enterprise's procedures and regulatory compliance. Audits—both offsite and onsite—can provide this information.

Ultimately, the validation of cloud-based systems should align with the risk level of the applications.

4.4 Configuration Accessibility

Configuration is an important element of cloud computing because mistakes made in configurations can offer opportunities for attackers to access data. The types of misconfigurations that can result in attacks can be categorized as follows:[15]

- **Storage access**—Enterprises may have false confidence in the security surrounding their cloud storage because they believe access is afforded to "authenticated users" only, a term they believe applies only to those who are authenticated with their organization or the relevant application. However, when speaking of Amazon Web Services (AWS), the term "authenticated users" refers to anyone with an AWS authentication: effectively, any AWS customer. Controls must be established to ensure that storage access is limited to only those within the enterprise who need access.
- **"Secrets" management**—This refers to secrets such as passwords, API keys, administrator credentials and encryption keys. Misconfiguring secure storage of this information can reveal it to those who would use it with negative intent. One solution is to maintain an inventory of all secrets the enterprise uses in the cloud and checking it regularly to see how each is secured. A secrets management system can be helpful as well.
- **Disabled logging and monitoring**—Public clouds provide logs and telemetry data that can be useful to an enterprise, but many enterprises fail to enable, configure or use them. Enterprises should assign responsibility to someone on the cloud team to regularly review the information and flag security-related events.
- **Overly permissive access to hosts, containers and virtual machines**—Enterprises must secure important ports and disable—or at least lock down—older, insecure protocols in the cloud, just as they would in their on-premises data center.
- **Lack of validation**—Whether an internal resource or an outside auditor is used, enterprises must assign someone responsibility for regularly (on a set schedule) verifying that services and permissions are properly configured and applied. Cloud configurations should also be audited periodically.

In addition to these potential misconfigurations, enterprises must give attention to configuration accessibility. Many cloud solutions offer multiple services but not every enterprise needs or wishes to use them all. For some enterprises, taking the entire configuration as offered may be the right decision; others may decide they can address certain needs (e.g., a firewall function) on their own or with a third party.

Effective configuration management requires that the management and provisioning procedures are segregated from the CSP, are limited to a security operations function within the enterprise and provide audit trails to document all activities.

4.5 Management of Resources

Cloud services are rented, not owned, and are paid for on a sliding scale. The compensation corresponds to the number/amount of resources (CPU, memory, energy consumption, etc.) used. Therefore, it is

cloud services are rented, not owned, and are paid for on a sliding scale. The compensation corresponds to the number/amount of resources (CPU, memory, energy consumption, etc.) used. Therefore, it is in the enterprise's best interest to take a keen interest in managing the use of those resources at every phase of their life cycle (discovery, allocation, scheduling and monitoring). Inefficient resource management has a direct negative effect on performance and cost, while also indirectly affecting system functionality.

Resource management, which is a core function in any cloud system, is not entirely about cost containment, however. Because of the nature and structure of the cloud, there are other concerns related to resource management, such as identity management, security, availability and privacy.

The policies for resource management can be grouped into a few general categories:[16]

- **Admission control**—To prevent the system from accepting workload that violates high-level system policies and prevent the system from completing work already in progress or finalized
- **Capacity allocation**—To allocate resources for individual instances, a feat that is more difficult when the state of individual servers changes rapidly
- **Load balancing and energy optimization**—To keep the workload at an efficient level to maintain a certain cost of services
- **Quality of service guarantees**—To address availability and performance of applications and systems and the appropriateness of the applications for accomplishing business tasks

The strategies for resource management will differ depending on the cloud deployment and delivery models used by the enterprise. However, the most common resource management issue is fluctuation of workload. Sometimes a spike can be predicted, as in seasonal activity, in which case the appropriate resources can be provisioned in advance. In other cases, when the spike is unplanned and unforeseen, auto-scaling may be employed, provided there is an available pool of resources that can be released or allocated and a monitoring system is in place to allow a control loop to decide in real time to reallocate resources.[17]

Whatever load-balancing techniques are selected, they must consider the factors/targets summarized in **figure 4.1**.

Figure 4.1—Load-Balancing Factors/Targets

Factor/Metric	Definition	Target
Overhead	• Determines the amount of overhead involved while implementing load-balancing algorithms • Composed of overhead due to movement of tasks, interprocess and interprocessor communication	• Should be minimized so that load-balancing techniques can work efficiently
Throughput	• Used to calculate the number of tasks whose execution has been completed	• Should be high to improve the performance of the system
Performance	• Used to check the efficiency of the system	• Must aim for improvement at a reasonable cost; for example, reducing response time while keeping some acceptable delays
Resource utilization	• Used to check the utilization of resources in a system	• Should be optimized for efficient load balancing
Scalability	• The ability of an algorithm to perform load balancing for a system with any finite number of nodes	• Must aim for improvement
Response time	• The amount of time taken to respond by a particular load-balancing algorithm in a distributed system	• Should be minimal, to improve efficiency
Fault tolerance	• The ability of an algorithm to perform uniform load balancing despite arbitrary node or link failure	• All systems expected to be highly fault tolerant

Source: Mahendran, D.; M. Gopi; S. Priyadharshini; R. Karthick; "A Study on Cloud Management Techniques," *International Journal on Innovative Research in Computer and Communication Engineering*, volume 1, issue 1, 2013, p. 60-61

Cloud computing initiatives tend to impact a wider variety of people inside (and sometimes even outside) the enterprise, compared to other services being outsourced. Most enterprises outsource services that are not core to them, thus impacting a smaller amount of people. However, with cloud computing, enterprises most frequently target their core services and take them to the cloud to enjoy the benefits of cloud computing for their core businesses. Obviously, this tends to impact a significant number of people throughout the entire enterprise, thus requiring adequate preparation and management of all resources to keep ensuring operability.

4.6 Storage and Data Availability

Availability—of storage and data—is a significant driver for enterprises to move certain activities to the cloud. They want a place to process and store their data without having to invest in the infrastructure to do so in-house. Therefore, in selecting a CSP and managing the relationship with that CSP, agreement on availability expectations is critical.

Availability—the amount of time the provider guarantees that the enterprise's data and services are available to the enterprise—is generally documented as a percentage of time per year. For example, a 99.9 percent guarantee from the CSP tells the enterprise that its data/storage will be available to the enterprise from the data center on nearly a full-time, 24/7 basis, or conversely that the enterprise will be unable to access the data for no more than about 10 minutes per week, or less than nine hours per year.

A 100 percent guarantee is not given, as the CSP must plan on downtime for maintenance. This should be acceptable to the enterprise, if the downtimes planned by the CSP do not coincide with the enterprise's need for its data. (Availability does not include guarantees related to connectivity issues that are out of the control of the provider, such as Internet outages or power losses.)

Each enterprise must determine what level of availability is required for its operation. Enterprises that depend on real-time transactions (such as an e-commerce site) will need a higher degree of availability than enterprises that are simply storing backups. Enterprises processing real-time transactions can attach a very real financial impact to periods of unavailability.

As devastating as the direct financial impact may be, however, other losses that may arise due to unavailability, such as a loss of reputation and customer trust, can be even more harmful, long-lasting and less conducive to amelioration.

The enterprise's CSP should publish and guarantee availability and that guarantee should specify the compensation that will be provided to the enterprise if the provider falls short of the guaranteed availability metrics. The enterprise must be sure to read the SLA closely, to see how the CSP defines "availability." Does the CSP consider it not a downtime if even one Internet client can access even one service? Or does the CSP consider availability to exist only when multiple Internet service providers and countries can access all services? The difference between these two interpretations can be staggering and must be evaluated in terms of the enterprise's needs in achieving business objectives.

4.7 Business Continuity and Disaster Recovery

Most enterprises have detailed plans to deal with disasters or disruptions within "the walls" of the enterprise itself. They should have equally comprehensive plans for their cloud computing service. Indeed, many enterprises consider cloud computing services as backup and a plan for disaster recovery. And, while it can serve that purpose, it requires considerable communication between the client enterprise and the CSP to ensure clear, agreed understanding about the assignment of responsibilities and the expectations for performance.

In preparing BCPs and DRPs, the enterprise should plan for the worst-case scenario: complete disruption of the CSP. Maintaining availability of data should be a core element in the plans and expectations should be clearly spelled out, with recovery point objectives (RPOs) and recovery time objectives (RTOs) defining the data that must be restored and when that data must be accessible for the enterprise to resume operations. (The CSP's RPOs, RTOs, backup testing frequency and procedures must be consistent with the enterprise's security policy.)

Other factors the enterprise should consider in preparing BCPs/DRPs include:[18]

- Maintaining backup copies of enterprise data so failure of a storage component or the deterioration of stored data does not result in permanent loss of the information. This should be supplemented with automatic failover, which will automatically replace an affected component with a backup.
- Creating multiple copies of enterprise data and multiple access routes to the data to avoid inaccessibility if any one network component, storage device or server fails.
- Avoiding the temptation to try to increase data availability through in-house, ad hoc "solutions" rather than simply using tools specifically designed for the purpose.

Of course, the negotiation and agreement phases of initiating a relationship with a CSP should include confirming that the CSP already has disaster recovery and backup procedures in place. Various provisions can be included in the contract to ensure both parties agree about processes and responsibilities in the event of a disaster or disruption, including the following activities:

- Document the technical mechanisms and procedures that prevent data loss, such as contractor/enterprise responsibilities and routines for backup, failover or redundancy.
- Provide for continuity of accessibility, usability and preservation of all agency data, regardless of any migration of data to other formats during the contract. Terms should provide for appropriate testing to ensure data integrity prior to any migration.
- Specify provisions and procedures for backup, restoration of services and disaster recovery.

- Upon transfer of data, ensure that technological parity with other service providers is guaranteed.
- Guarantee the preservation of data and provide for routine monitoring of data to identify formats that are at risk of obsolescence.
- Include provisions relating to the migration of data to new formats, when appropriate, and the provision of proper documentation about migration activities to the enterprise.
- Include provisions for the safe return/transfer of data if the CSP is the subject of a takeover.
- Specify what will happen to the data when the contract terminates; for example, transfer to a new provider, return to the enterprise or permanently delete.
- Specify remedies for service-provider mistakes or breaches.
- Identify penalty provisions imposed by the service provider, such as suspension of enterprise access to the data because of nonpayment.
- Precisely specify the terms for the disposal of specific data during the term, at the request of the enterprise and at the end of the term, including a warranty in relation to technological parity/obsolescence.
- Limit suspension and termination rights that are available to the CSP.
- Allow subscription levels to be scalable up and down according to demand.
- Specify reporting and audit rights of the enterprise and vendor.

While there may be overlap between a DRP and a BCP, the BCP focuses on the responsibilities, processes and activities that ensure sustained execution of the enterprise's business activities within its environment, its infrastructure operations and its IT controls. Most BCPs cover responsibilities, accountabilities and authorities; the frequency and processes involved in testing the BCP; and the overall process itself: identifying and declaring a disaster/interruption, carrying out appropriate internal and external communication, restoring activities, analyzing the disaster/interruption, and declaring closure.

The CSP's role in business continuity is to train its employees to imagine all reasonable disaster and business interruption eventualities and devise alternate methods to restore the cloud services necessary to maintaining customers' business and infrastructure needs, thereby minimizing business impact. In other words, when problems occur at the CSP, the customer should be harmed as little as possible.

4.8 Chapter 4 Knowledge Check

REVIEW QUESTIONS

1. CSPs usually allow enterprises full access to assets for testing.
   A. True
   B. False

2. The **BEST** way for an enterprise to ensure that it has some access for testing and validation of cloud assets is through:
   A. OLAs
   B. SLAs
   C. Safe-haven agreements
   D. SOC2 reports

3. A cloud deployment for a company that deals with health information of US citizens will need to comply with:
   A. CCPA
   B. GDPR
   C. HIPAA
   D. PCI-DSS

4. All of the following reflect misconfigurations that can lead to attacks except:
   A. Lack of validation
   B. Improper storage access
   C. Lack of audit clause
   D. Disabled logging

5. Cloud servers and services are _____ to the customer by the CSP.
   A. Sold
   B. Rented
   C. Described in detail
   D. Escrowed

6. A target of the performance metric is:
   A. System fault tolerance
   B. Reducing response time
   C. Load balancing
   D. Migration time minimization

7. One of the roles that the CSP plays in business continuity is to:
   A. Ensure customers have well-conceived BCPs and are prepared
   B. Provide services free of charge during disasters
   C. Write BCP clauses into SLAs
   D. Ensure their employees have prepared for eventualities

Answers on page 64

Chapter 4 ANSWER KEY

Review Questions

1. **A. True. Refer to page 56.**
   B. False

2. A. OLAs
   B. SLAs. Refer to page 56.
   C. Safe-haven agreements
   D. SOC2 reports

3. A. CCPA
   B. GDPR
   C. HIPAA. Refer to page 56.
   D. PCI-DSS

4. A. Lack of validation
   B. Improper storage access
   C. Lack of audit clause. Refer to pages 57-58.
   D. Disabled logging

5. A. Sold
   B. Rented. Refer to page 58.
   C. Described in detail
   D. Escrowed

6. A. System fault tolerance
   B. Reducing response time. Refer to page 59.
   C. Load balancing
   D. Migration time minimization

7. A. Ensure customers have well-conceived BCPs and are prepared
   B. Provide services free of charge during disasters
   C. Write BCP clauses into SLAs
   D. Ensure their employees have prepared for eventualities. Refer to page 61.

[14] Mitchell, T.; "Why You Need to Validate (and Re-Validate) a Cloud-Based System," Montrium, 8 January 2020, blog.montrium.com/experts/why-you-need-to-validate-and-re-validate-a-cloud-based-system

[15] Smith, P.; "5 Common Cloud Configuration Mistakes," DARKReading, 17 September 2019, www.darkreading.com/cloud/5-common-cloud-configuration-mistakes/a/d-id/1335768

[16] ScienceDirect, "Cloud Resource Management and Scheduling," www.sciencedirect.com/topics/computer-science/cloud-resource-management

[16] ScienceDirect, "Cloud Resource Management and Scheduling," www.sciencedirect.com/topics/computer-science/cloud-resource-management

[17] *Ibid.*

[18] Precisely, "Data Availability 101: What Data Availability Means and How to Achieve It," 16 June 2020, blog.syncsort.com/2018/06/data-availability/data-availability-101/