CHAPTER 5

Cloud Offerings

CONTENTS

5.1 INTRODUCTION

The growth of information is at an exponential stage. It is growing day-by-day and is being generated by millions and millions of people with billions of smart devices, M2M, sensor networks, machines, and instruments. Today, there is a massive growth in unstructured data where the images, music, scans, and medical data are flowing more than the speed of imagination.

Until now, organizations could not fully or quickly synthesize and interpret all the information out there – they had to make decisions largely on the basis of instinct. But now, there are mechanisms that can capture, organize, and process the entire data scattered throughout an organization, and turn it into actual intelligence. These mechanisms enable organizations to make better business decisions.

5.2 INFORMATION STORAGE, RETRIEVAL, ARCHIVE, AND PROTECTION

There are various stages of an information management cycle. The information gets processed, managed, migrated, protected, retrieved, staged, and archived for business data. Organizations manage the data on the basis of characteristics of the data such as age, usage, compliance, policies, regulations, protection, disaster, and availability. It is done by the established practice of Information Lifecycle Management (ILM). ILM becomes even more important for cloud deployments as the process requires sharing data services between the cloud vendor and the subscribers.

ILM is not a latest data storage, retrieval, and protection solution; piece of hardware; or some software, but rather an approach to assess and manage the information across the enterprise. ILM is based on how data is used and how readily it must be available to the people who use it. It is focused on managing and storing data according to its value to business operations at any given point in time. It is concerned with the placement of data at the appropriate level of storage with an appropriate retention and retrieval policy. In response to these challenges, organizations are defining the following objectives to support and improve their information management:

1. **Cost reduction:**
   - Controlling demand for storage.
   - Reducing hardware/software costs.
   - Reducing storage personnel costs.
2. **Better system performance and personnel productivity (i.e., improved efficiency):**
   - Doing the 'right' storage activities.
   - Improvement in people, community, processes, and technologies to deliver storage services.
   - Defining and enforcing policies to manage the lifecycle of data.
3. **Increased effectiveness:** Defining and implementing an appropriate storage strategies to address current and future business requirements.

They are also coming up with new ways to generate, enhance, and sustain higher savings. These include:

1. **Activities for gaining initial savings:**
   - Reduce the amount of used storage as a result of initial clean-up.
   - Validate storage area network (SAN) requirements and reclaim used switches and switch ports.
   - Validate data replication requirements in order to reclaim used storage space and offset future growth requirements.
   - Develop and document information classification.
   - Develop and document classes of service.
   - Design and implement the tiered storage architecture.
   - Migrate existing information to lower-cost storage using a tiered storage architecture.
2. **Activities for maximizing savings:**
   - Reconfigure the current storage environment effectively to improve the available raw utilization.
   - Reclaim available storage that has been over allocated.
   - Enhance the information classification, classes of service, and tiered storage architecture.
3. **Activities for sustaining savings:**
   - Develop a storage model based on policies.
   - Implement changes to existing storage management processes, such as capacity planning, and provisioning to effectively improve capacity utilization on an ongoing basis.

While designing a target storage environment, the estimated financial impact is calculated on the basis of the following key cost components:

1. **Operating cost categories:**
   - ***Personnel:*** Storage support and contractors.
   - ***Facilities:*** Current floor space consumed by the storage. Telecommunication charges attributed to storage and tape vaulting services.
   - ***Storage hardware maintenance:*** Existing maintenance and incremental maintenance resulting from growth.
   - ***Storage software maintenance:*** Existing software maintenance and incremental requirements resulting due to innovation and development.
   - ***Outages:*** Cost avoidance associated with the reduction in unplanned outages.
2. **Investment cost categories:**
   - ***New hardware required:*** Typically includes disk, tape, and array cost but not the incremental cost of adding SAN fabric. Investment is either upfront or over a period of time if the client leases equipment.
   - ***New software required:*** New storage software required to support the target environment.
   - ***Hardware refresh:*** Investment required to refresh the existing hardware is often considered in the base case.

- *Transition services:* Incremental cost required to transform the current environment to the future environment. Not typically estimated until the scope of the third-party implementation services has been defined.

When more than 90% of the data stored on hard disks is not actively accessed by the users or applications, it is good to go for more intelligent management and migration to a less expensive storage. But the savings can go significantly beyond disk acquisition costs and annual hardware maintenance costs. Some points in information management are

1. **Data:** Discrete element, reasoning, discussion, or calculation of content created through the interactions between applications or between computing devices.
2. **Information:** Organized and structured collection of data.
3. **Information lifecycle management:** It is a combination of policies, processes, tools, and management practices to align the value of the information with the infrastructure needs, right from the time when the information is created till its disposition.
4. **Information taxonomy:** Data described in the context of business process requirements and lifecycle characteristics.
5. **Information classes**: Groups of information taxonomies with associated business value that provide the basis for storage management and service delivery.
6. **Value-driven data placement:** An event correlation framework that 'senses' the value of data changes and based on business policies 'responds' by moving data to the appropriate storage tier.
7. **Storage process:** A documented set of storage-related tasks and activities required to support a storage infrastructure.
8. **Storage service:** Capabilities provided to a customer base that is designed to meet their business requirements, wants, and needs, and enabled by a storage infrastructure.
9. **Enterprise class of service (COS):** A common set of storage services that are delivered to meet a corresponding set of storage requirements based upon key information management characteristics, and the features, functions, capabilities, processes, and governance required to deliver the required enterprise storage services.
10. **Storage tier:** A subset as a set of storage devices that are identified to store and/or maintain information for a predefined period of time based on key information management characteristics, such as performance, configuration, residency, retention, and value.
11. **Tiered storage infrastructure:** An organized collection of storage tiers reflecting the flow of all the information managed in the enterprise storage architecture.
12. **Utility-based service delivery:** The 'just-in-time' delivery of standardized storage processes, management, and infrastructure as a measurable service on a 'pay-as-you-go' basis.

5.3 CLOUD ANALYTICS

Cloud analytics is a new offering in the era of cloud computing. Cloud analytics will help in the consulting domain and will ensure better results. It provides users with a better forecasting technique to analyze and optimize the service lines, and provide a higher level of accuracy. Cloud analytics can help the users apply analytics principles and best practices to analyze different business consequences and achieve newer levels of optimization (Fig. 5.1).

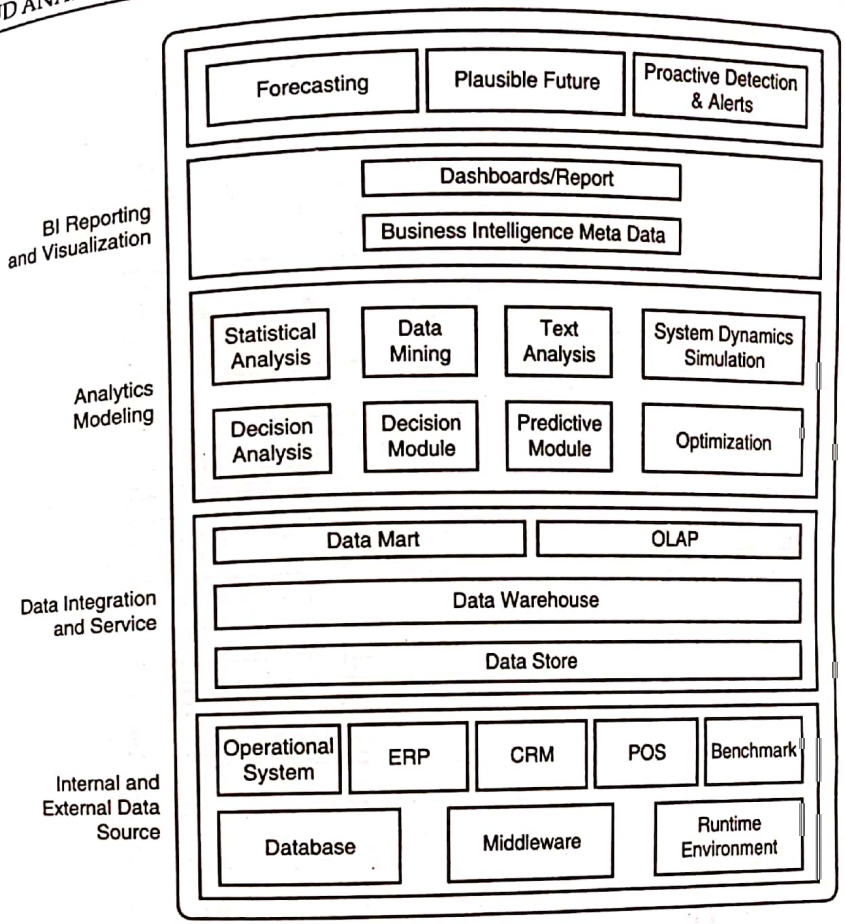

FIGURE 5.1 Cloud analytics.

It can combine complex analytics with the newer software platforms and will lead towards the predictable business situation out of every business insight.

5.3.1 *Cloud Business Analytics Competencies*

The practice of cloud analytics requires multiple service lines to strategize how client can achieve the goals in faster manner coupled with less risk. The aligned competency features business intelligence and performance management that help increase performance by providing accurate and on-time data reporting. The next is competency comprises of analytics and optimization that provide different types of modelling techniques, deep computing and simulation techniques to check for different types of 'what if' analysis to increase performance. Another competency is enterprise information management that helps in applying different architecture related to data extraction, archival, retrieval, movement, and integration. Content management system is another competency that is required for cloud analytics and includes different service architecture, technology architecture, and processes related to capturing, storing, preserving, delivering, and managing the data. It also helps to provide access in the global environment and makes it easy to share data with stakeholders across the globe.

5.3.2 *How it Works: Analytics*

Analytics works with the combination of hardware, services, and middleware. This expertise makes it best suited to help clients extract new value from their business information. Delivering business analytics and information software requires a seamless flow of all forms of data regardless of format, platform, or location. Its focus on open industry standards is the key to this effort and gives us a significant advantage (Fig. 5.2).

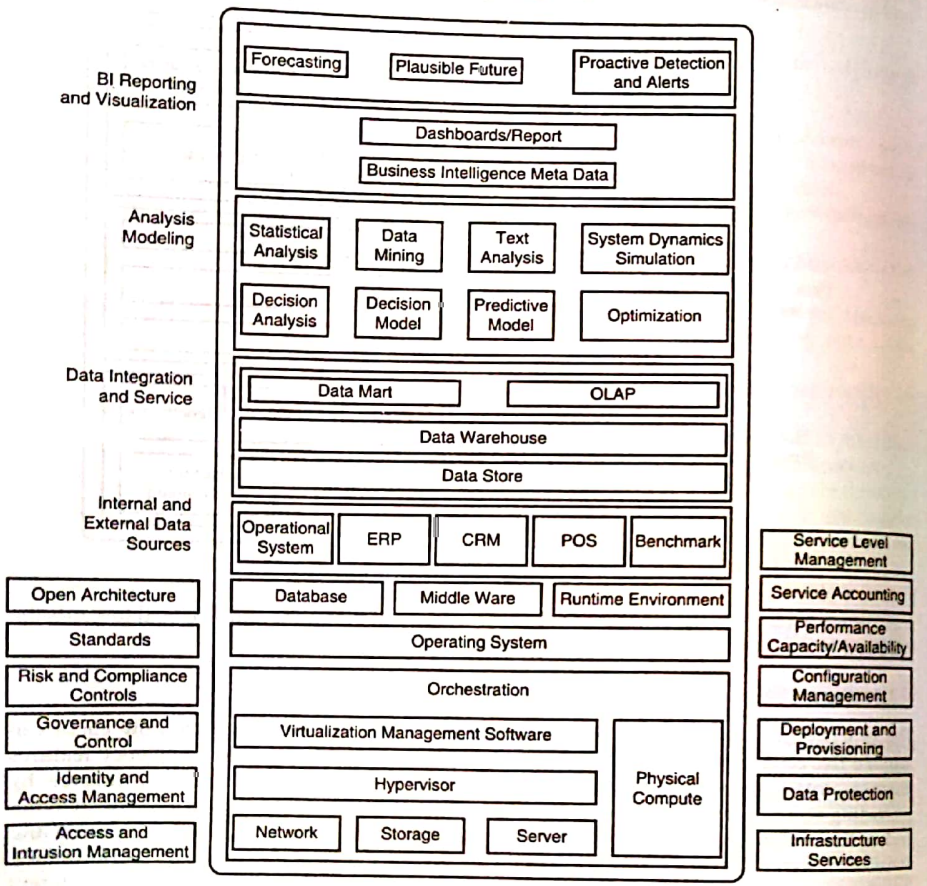

FIGURE 5.2 Cloud analytics business outcomes.

The analytics system features include the platform that provides data reporting, analytics based on text, mining activities, business intelligence dashboards, and perceptive analytics

techniques. This also takes care of the storage optimization and different high-performance data-warehouse management techniques. It also involves the umbrella activity of different installation services and a highly reliable system platform.

Analytics Business Outcomes

Analytics systems help to get the right information as and when required, identify how to get it, and point out the right sources to get it. Therefore, analytics also helps in designing the policies faster as decisions are based on the information available in the organization and decision-makers work with the exploration services available within the organization. This system also helps in gauging the business results by measuring the different metrics generated with the help of analytics. Analytics give options through which the organization can increase profitability, reduce cycle time, and reduce defects.

5.4 TESTING UNDER CLOUD

Private cloud deployments need virtualized servers that can be used for testing the resources for the cloud. This ensures more secure and scalable solutions where consumers can access the IT resources in the test environment. Therefore, testing under the cloud environment gives a very good insight by decreasing the manual intervention and reducing the processes that are involved in typical testing environment. By enabling access to resources as and when required, the cloud reduces the investment on capital, as well as enables the business to handle the ups and downs of the testing requirements. With this feature, organizations can reduce the test cycles, minimize IT costs, reduce defects, rationalize the testing environment, and hence, improve the service quality. This provides a good return on investment (ROI) on moving the typical testing environment to the cloud. This also gives the flexibility to play with the surrogate of the real system without the actual risk.

5.4.1 *Benefits*

1. Cut capital and operational costs, without affecting the mission critical production applications.
2. Offer new and innovative services to clients, present an opportunity to speed up the cycles of innovation, and improve solution quality.
3. Facilitate a testing environment based on the request, and provide a request-based service for storage, network, and OS.

5.4.2 *Value Proposition*

Business test cloud delivers an integrated, flexible, and extensible approach to test resource services and management with rapid time to value. This is an end-to-end set of services to strategize, design, and build request-driven delivery of test resources in a cost-effective and efficient manner.

These test tools allow you to orchestrate and build your services and development projects, and permits to catalog and organize all of the various software assets. These can be provisioned by the administrators, development team leads, or project team members who are permitted to do the same.

5.4.3 The Biggest Benefiters

The biggest problem for the finance heads are to reduce the operational and capital expenses of the development and testing environment. Testing under the cloud environment comes as a boon to implement this in an organization. So testing under the cloud not only helps in reducing the financial burden on the company, but also reduces the cycle time for testing and development environment without buying the infrastructure. The cloud environment provides test tools to synchronize and build the services for the project, and helps to arrange the project resources, assets and various versions of deliverables. Also, permission can be set on the basis of the roles in the organization to provide access to the assets.

So why is testing under the cloud a great on-ramp to cloud computing for organizations that want to embrace cloud computing and are trying to figure out where to start? Well, the test and development environment is rarely optimized and is usually like a low hanging fruit in terms of getting ROI and benefit out of cloud computing. These are environments that have a high degree of dynamic change. Hence, the environment is tested and developed after having gone through a lot of build-up and teardown as it moves from project to project. Typically, 30 to 50% of all servers in an IT environment are dedicated to test and development, which is a significant amount. A number of these servers are positioned at very low levels of utilization, and in fact, sit idle for a period during the project life cycle. So, a better utilization of those systems can be a huge benefit for the organization.

Many defects are introduced into test and development environments because of the fact that there is a high degree of manual configuration that typically goes on. Often, projects are backlogged because of the limited access to test environments. Thus, if you do a lot of test and development, or even if you buy your products, you still have to integrate those in a test environment before moving them into production. This intends to be an excellent entry point into cloud computing.

It is important to take a holistic look at the problem you are trying to solve. Are you building a development cloud? Are you building a test cloud? What pre-existing service management or enterprise management tools exist, so that you can identify the starburst opportunities that come off business development and test cloud services or at the cloud strategy and planning consulting workshop service? Some of the things to consider include:

1. What are the types of service management integration (SMI) that need to happen?
2. Do they want to integrate with discovery?
3. Does the SMI want to integrate with the service desk?
4. Does it want to create a help ticket whenever a provisioning step occurs?
5. Do these SMIs want to create an asset every time a new virtualized environment is created?
6. Do they want to do usage and accounting chargeback?
7. Does the SMI want a test optimization service?

These and many other possible opportunities exist in the organization, so make sure that this is the part of the discussion regarding the requirements so that it can make its way into the design and implementation piece of the engagement.

Organizations engaged in the development and testing share a common challenge in executing fluid projects within rigid infrastructures. Development projects are initiated

to meet specific 'needs' that may include introducing new applications into the market, updating or patching existing software, or developing in-house assets. There is no way to 'schedule' needs; identification and application development can be as much an art as a science with an intrinsic unpredictability that makes resource allocation very difficult within traditional infrastructures.

As needs arise erratically and projects slip within their individual timelines, there are inevitable overlaps and gaps in development resource requirements. Organizations are often faced with frustrating delays as projects wait idle due to the unavailability of the infrastructure. Alternatively, maintaining enough capacity to accommodate peak demand will leave large windows of severe under-utilization. Even the best compromise represents unwelcome cost and inefficiency.

With the ability to deploy virtual environments quickly and automatically, and redirect capacity as per the needs of the projects, cloud computing offers an ideal solution for testing and development. Cloud vendors make the transition even more appealing with solutions that allow your clients to experiment with cloud in what is already, by definition, an experimental environment–a low risk introduction to cloud and a first step toward addressing other IT challenges across the business.

The cloud testing services give the overall business transformation value that helps you to reduce the cost associated with the IT operations by prioritizing your business requirements. A strategic roadmap is required to enjoy the benefits of application virtualization for testing in the cloud environment, and requires different assessments.

The first assessment is to consolidate the server and conduct a comprehensive virtualization assessment. Getting the requirements of a cloud infrastructure is a very difficult task; hence, the first assessment provides an opportunity to study the initial requirements of the cloud environments that help to provision automatically, and gives the reason to adopt the cloud infrastructure, and appreciates the best practices of cloud. The other assessment is to determine what type of software models can be applied to the available infrastructure to win over the constraints and increase the schedules.

Any business that relies on software development for revenue probably has some benefits to gain from reducing its cycle time and improving the quality. This offering, through its automation and change, and configuration management, can reduce errors and speed up the deployment.

Businesses with a dedicated development and testing organization can be benefitted by the reduction in capital and other expenses that are the result of automation and utilization improvements. There are a growing number of organizations that no longer have space or power to expand their datacenter. By using standardization, consolidation, and virtualization, they can maximize the use of their existing assets.

Businesses that are looking to change their business model to be more line-of-business focused may, instead of doing straight allocation, want to do chargeback and have customers pay for only what they use. They may also be able to smoothen out the resources across different departments so that each department does not have to maintain an account of its own peak demand, and smoothen that demand across the peaks and valleys of different departments and different projects.

5.4.4 *Cloud Offering Key Themes*

Today, the enterprise datacenters are managed by infrastructure specialists and administrators working in the operations team. This team is responsible for the availability, archiving, security, disaster recovery, service management, and ongoing operations. The main objective of this team is risk, release, and change management, and promoting standard infrastructure requirements for the applications. Adoption of virtualization technologies has brought a significant change in enterprise datacenters over the last several years, as it has fundamentally started to change the way IT is delivered and serviced. Enterprise IT operations managers are tasked with serving two main constituencies:

1. **Application teams:** Delivery of production internal and/or external applications according to service-level requirements in a cost-effective manner. Application infrastructures are often deployed in silos and provisioned for peak demand, resulting in capacity capabilities far beyond their normal requirements.
2. **Development teams:** Development departments are usually driven by user needs for frequent delivery of new features. Development teams often want to quickly test new ideas and/or features in a realistic environment. However, there is often a significant delay between requests to IT for new environments and actual delivery (often can be several weeks). Development teams frequently request and then hesitate to return IT environments into the centralized pool due to the fear of losing access to resources needed for their development cycles.

Due to the history of delivering dedicated environments to support both groups, infrastructure resource utilization metrics are typically below targets. The IT infrastructure managers are no longer being pushed solely to reduce cost, but also to improve end-user responsiveness. The operations team wants to become more responsive to business needs and reduce the application provisioning time by some great percentage (e.g., 90%), increase the resource utilization from what are typically very low percentages (e.g., <20%) to an exponentially greater reduction (e.g., 50–60% or higher), and reduce IT operational costs by some percentage (e.g., 25% or more), all this while avoiding vendor lock-in and regaining control of their applications and infrastructure so they can use it in the most efficient manner to meet the needs of the business.

Virtualization has its benefits, and it does modestly improve resource utilization and delivery times within its hypervisor domain, but nearly all datacenters have multiple hypervisors in use as well as many other computing resources that are not virtualized. Operations teams face significant challenges with manual provisioning and management, VM sprawl, and difficulty in scaling when needed.

Key Themes

1. **Infrastructure to applications:** The most commonly used case for private cloud is Infrastructure-as-a-service (IaaS). This can often be as simple as managing VM images to prevent VM sprawl. The vision is to support complex multi-tier application provisioning such that the applications can be fully configured and ready-to-run. In some cases, these applications are then delivered as a Software-as-a-Service (SaaS) model. In between, enterprises create their own Platform-as-a-Service (PaaS) catalog to enforce consistency in development platforms and middleware. Since companies are using this

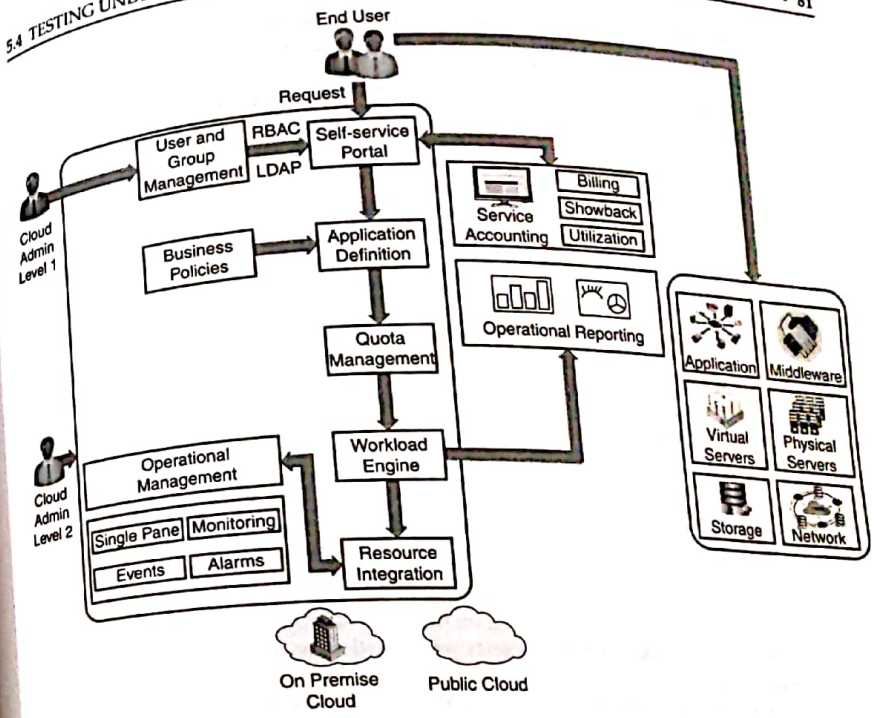

FIGURE 5.3 Cloud orchestration workflow.

for their internal use, they do not need to rewrite their internal applications to match an external PaaS custom application programming interface (API) set, that is required if a public PaaS service were to be considered.

2. **Development/testing to production:** In software development, there is a high-level lifecycle for applications. The beginning is development (dev), then testing/QA, staging, followed by production. Enterprises can also differentiate between the types of production applications between internal versus external production (customer facing) in terms of criticality. In development/testing, the orientation is toward many users with simpler requests (e.g., VMs or infrastructure) with a focus on shorter duration resource usage (reservation, quota management, and reclamation are keys to prevent sprawl). Production applications, in contrast, are more concerned about meeting runtime performance service-level agreements (SLAs) with dynamic scaling, and managing more complex, multi-tier applications. The solution should be unique in supporting the full lifecycle in one product offering in the enterprise environment (Fig. 5.3).

3. **Allocation and run-time scaling:** Allocation is the process of instantiating a service catalog item: infrastructure, platform, or application. Run-time scaling is flexing-up or flexing-down of required resource elements to meet the SLA requirements defined by the application owner according to the standard corporate standards and business policies. The solution should be unique in providing both IaaS and runtime scaling/workload management in one product offering for the enterprise environment.

Benefits

1. **Increase agility and innovation**
   - Enable self-service delivery (within minutes).
   - Deliver on SLAs.
   - Simplify process for 'what-if' experimentation.
   - Gain control over public cloud usage.
2. **Decrease costs**
   - Increase utilization.
   - Increase operational efficiency (100 servers per admin).
   - Achieve a greener datacenter.
   - Maintain vendor choice.

Offering Key Characteristics

1. **Service layer**
   - Self-service portal for different cloud users: administrators, cloud delegates, and end-users.
   - Chargeback/billing and reporting on the basis of usage and capacity.
   - Operations management self-service portal with service management.
2. **Applications**
   - Automated application provisioning and lifecycle management.
   - Dynamic scaling to meet SLAs.
3. **Allocation engine**
   - Account-based quotas, reservations, scheduling, and approvals for resource allocation.
   - Policy based migration, movement, and failover of workloads.
4. **Resource integrations**
   - Maintenance of virtualization platforms.
   - Support for popular provisioning tools.
   - Integration for popular public cloud/external services.
5. **Datacenter integrations**
   - Role-based access.
   - Adapter-based integration to accounting, asset management, change management, entitlement, service catalog systems, and ticketing systems.

5.5 INFORMATION SECURITY

Information security risks are potential damages to information assets. Successful organizations take a risk-based approach to information security. Nothing can be 100% secure – but by

knowing your current state you can take a risk-based approach. You can focus on implementing mitigating controls to address your most significant risks. The remaining minimized risk is acceptable because the likelihood of exploit and the severity of exploit versus the cost of mitigation do not have a positive cost/benefit. Risk can be quantified by the following expected (average) damage:

1. **Value of asset:** What are your valuable information assets?
2. **Vulnerabilities:** What vulnerabilities exist in your systems that can be exploited and lead to the damage of your assets?
3. **Threats:** The level of threats that aim at exploiting vulnerabilities.

Security controls are protection mechanisms to avoid information security risk and have the following features:

1. **Must be effective:** Mitigate the given risk.
2. **Should be adaptive:** Adapt to changing risks.

Three main types of security controls are as follows:

1. **Preventive:** Prevent security incidents (e.g., patching vulnerability).
2. **Detective:** Detect a security incident (e.g., monitoring).
3. **Corrective:** Repair damages (e.g., virus removal).

Successful organizations recognize risks, implement the appropriate mitigating controls, and innovate to grow their business.

5.5.1 *Expectation of Privacy*

Consumers expect that security should be built into services themselves. Over 50% of potential cloud consumers still avoid online purchases due to the fear of financial information being stolen. Expectation drives regulation, and today, vendors like automakers are expected to take a greater share of responsibility. Enterprises must shore-up their weakest supply chain partners by insuring

1. More evenly distributed security responsibilities.
2. Increased transparency from start to finish.
3. Eased burden of customer-facing unit.

5.5.2 *Security Challenges*

As the information grows day by day, datacenters and infrastructures are stretching the upper limits of these resources. They are trying to maximize the use of the power, space, and personnel; hence, CAP-EX and OP-EX are also increasing day by day. Cloud computing and virtualization give an opportunity to meet these challenges. On one hand, it gives weapons to deal with these challenges; and on the other, it also gives rise to its own problems such as veracity of the virtual environment, data integrity problems, and even security challenge.

Another area to watch is security around Web applications. An average deployed application have dozens, sometimes hundreds, of defects and a bulk of security threats today target the application layer. Companies must take a proactive action to reduce the instances for their Web applications being exploited – before a hacker even has the opportunity to

compromise their business. The most feasible and economic solution for the organizations is to catch vulnerabilities as early as possible and build a secure software from the ground up and not bolt it once it has been deployed.

We are moving toward a future where enterprises will adopt the services from cloud service providers externally. The most important point is that the workloads that will use these services will be rendering the low-risk workloads. This will also account for some of the assurances for the security, and the price of the service will be the key factor to decide whether to adopt the service or not. The workloads that possess medium-to-high risk factors will adopt private and hybrid clouds. This will also cover the workloads that contain proprietary contents and also those that need more security and depth of defense. Once these services mature and settled, the latter ones will also move towards the external cloud to enjoy the benefits of the external cloud, but without compromising the security.

5.5.3 *Security Compliance*

There is a need of policies and procedures for governance and risk factors with respect to cloud security. These services should include procedures to handle change management and incident management. These services generate reports for multi-tenant environments. Therefore, one has to bank on large log and audit files to do so. Transparency is an important factor as it is very important for public clouds and it is a black box for the service users.

It is also required to conduct third-party-based checks and audits for the agreements that are breached in the process. Also, third-party-based audits can issue non-compliance to the subjected violations. This process maintains visibility in the system.

Another method is to have strong SLAs so that the flexibility can be managed for the process based on the situations that will enable the traditional outsourcing model and service management to enjoy the benefits of the cloud.

5.5.4 *Identity-Based Protection*

The cloud environment requires extra protection levels as it works with a diverse set of groups. Therefore, it is essential to have proper authentication for getting access to resources for the environment. It also requires regulated monitoring of users, details regarding the logging to the resources, and check-up for the background verifications. One of the important aspects is to maintain the access that matches with the profile of the work and gauges the risk if something goes wrong due to the improper use of resources. There can be different classes of users, such as administrators who require the access based on the work they are doing with the cloud environment.

Maintenance of the identity is required to conduct smooth operation in the cloud deployments and authenticate the real users. This is required for both internal and external purposes for the hosted applications. The biggest hurdle is to secure the confidential data. In order to do this, one has to maintain a secure protocol over the networks and activate the firewalls to ensure the security of the confidential information and the information that is sensitive but not important for the business should be destroyed.

5.5.5 *Data Protection@Cloud*

Relevant terms that dictate the protection of data are: how it is stored, how it is accessed, what the compliances are, and what audits are required as per the SLAs. It also relates to the regulation of the breach of the data and its separation on the storage infrastructure. This even includes the archived data.

This is handled by the process of encryption and is managed by encryption keys, and the data are protected in the cloud datacenter. Another point that is not taken care of most of the times is the protection of the mobile data. It should be ensured that encryption is done for the mobile data as well. One of the biggest problems in Internet-based cloud is sending a large amount of data that is not possible with the Internet-based environments. Therefore, the data should be encrypted, and both the cloud service provider and subscriber should have the keys to encrypt it.

The movement of data between different locations of the organization depends on the cloud environment, support, SLAs, and business activities. There can be violations of the intellectual property law which should be kept in mind while working with different types of data. It must be ensured that the legal teams should review all the requirements of cloud environments and the methods to control the data that are collocated in a large geographical area.

Other important thing that can be classified is the data type and its associated risks for the protection of data. The risks levels and matrix breaches can be obtained and security mechanisms based on them can be derived. The measures can be different from domain to domain; for example, public services challenges will be different from financial services data.

5.5.6 *Application Security@Cloud Deployment*

It is incorrect to think that the protection mechanism in a cloud environment works only at the application level. In fact, it is required at the image level as well. Therefore, cloud vendors should have a clear and sound way to tackle the protection mechanism by meeting the demands of the subscriber for issuing licenses for the required period of interval, destroying them after use, and making sure that the sensitive unimportant data is also destroyed at the same time.

For maintaining and supporting the security, it is important for the customer to follow the standards that cloud subscribers demand. All the Web-based requirements should be coded to match the actual requirements, and the published content on the Web should adhere to the policies of the business.

In order to work with the successfully protected virtual environments, everybody in the cloud deployment should adhere to an agreed-upon basic security policy. Cloud vendors and subscribers should have the intrusion-based polices audited and should check the prevention system placed to handle this. This exercise becomes more valuable when we work in a shared environment because different subscribers on the same cloud environment should have the agreement for security and protection policies.

So far, we have talked about the security measures on the basis of software, policies, SLAs and audits; but we should also take care of the security in physical terms as well.

These security measures can include biometric systems and closed circuit television (CCTV) monitoring. These measures can restrict any unauthenticated entry into the system.

5.6 VIRTUAL DESKTOP INFRASTRUCTURE

Virtual desktop infrastructure (VDI) provides end-user virtualization solutions. This is designed to help in transforming distributed IT architectures into virtualized, open-standards-based frameworks leveraging centralized IT services. VDI combines hardware, software, and services to connect the clients' authorized users to platform-independent, centrally managed applications, and full desktop images that run as virtual machines on the servers in the datacenter.

The VDI solution introduces a new method of delivering and managing user desktop environments. In the IT industry today, several technology vendors provide components that enable building a VDI. VDI is generically used with reference to a collection of products and infrastructure components used to form a virtual desktop solution. This solution consists of the following:

1. Portal interface.
2. Thin-client.
3. PC with client components such as
   - Browsers with client messaging.
   - Security technologies.
4. IT infrastructure.
5. Varied client devices.

Project-based services provide IT consultants specializing in virtualization technologies, assistance to assess the organization's desktop and application needs, and subsequently develop a VDI solution that best meets these needs.

5.6.1 *Architecture Overview*

Desktop cloud virtualization services provide several advantages to an enterprise. One of very important advantages is the reduction of cost. By moving the core function of distributed end-user devices and applications to a centralized infrastructure, the lifecycle of the end-user device is extended, and the performance requirements are moved to a centralized infrastructure. The administration of a centralized IT infrastructure is more cost efficient than the administration of a distributed one. The implementation of virtualization with VDI solutions allows businesses to simplify their IT environment, reduce costs and complexity through consolidation of physical resources, and standardization of operating environments.

VDI creates a framework that offers many advantages to the enterprise such as follows:

1. **Cost reduction:** More efficient use of resources can increase utilization.
2. **Flexibility:** Common physical infrastructure can support a variety of end-users. New desktop images can be created dynamically without a hardware-procurement cycle. Multiple types of guest OSs can run on the virtual machines so that the physical hardware can support a wide range of end-users without any costly integration or reconfiguration of the systems between user accesses.
3. **Security:** The data remain in the datacenter with the access control.
4. **Availability:** Higher availability as VM can be quickly migrated to a different physical server in the event of a hardware failure.
5. **Efficiency:** Service delivery is more efficient when IT processes are optimized for a centralized environment.

5.6.2 *Enterprise Level*

VDI provides a set of proven integration patterns and methods for implementing a client virtualization. The VDI team works with various tools and products to help the users with an assessment of their environment to develop VDI solution requirements and a solution design. The team then develops and deploys this solution into the client environment on the basis of a common architecture that supports the four VDI solution models (shared service, virtual client, workstation blades, streaming).

VDI solution is shaped by the component selection. The base VDI architecture is designed to integrate with existing client environments and, as the name implies, provide an access method to a highly scalable virtual infrastructure to the front-end clients. The VDI server farm's back-end integrates with existing infrastructure services and legacy applications.

VDI solution is designed to reduce the dependency on distributed PCs and laptops. By placing critical applications and data in a centralized datacenter with access from a variety of client device options, VDI can significantly reduce the cost and complexity of the management, and maintenance of desktop images.

VDI end-user desktops run on virtual machines hosted in a centralized IT infrastructure in the client datacenter. The end-users access their individual desktop image or a pool of desktops through a client access device such as PC client, a thin client, or a Web-based client. The applications run on virtual machines on host servers in the datacenter from resource pools rather than on the local machine.

Additional resources can be easily and quickly added to the IT infrastructure as business requirements arise. The VDI solution provides increased security as data remain in the datacenter. Optional secure encapsulation capabilities can allow network connections to be encrypted.

VDI's virtual client solution integrates into the organization's datacenter to leverage the existing network and infrastructure services. The solution provides access services to the virtual infrastructure hosted in the VDI server farm or datacenter (Fig. 5.4).

The desktop client devices can be new or existing thin client devices, PCs with an access client, a Web browser, or various combinations depending on the client environment. The desktop client devices can vary from organization to organization, and can be provided as a part of the VDI service engagement if required.

The VDI access service or 'connection broker' provides devices and user authorization, portal integration, session management, host monitoring, application streaming, and

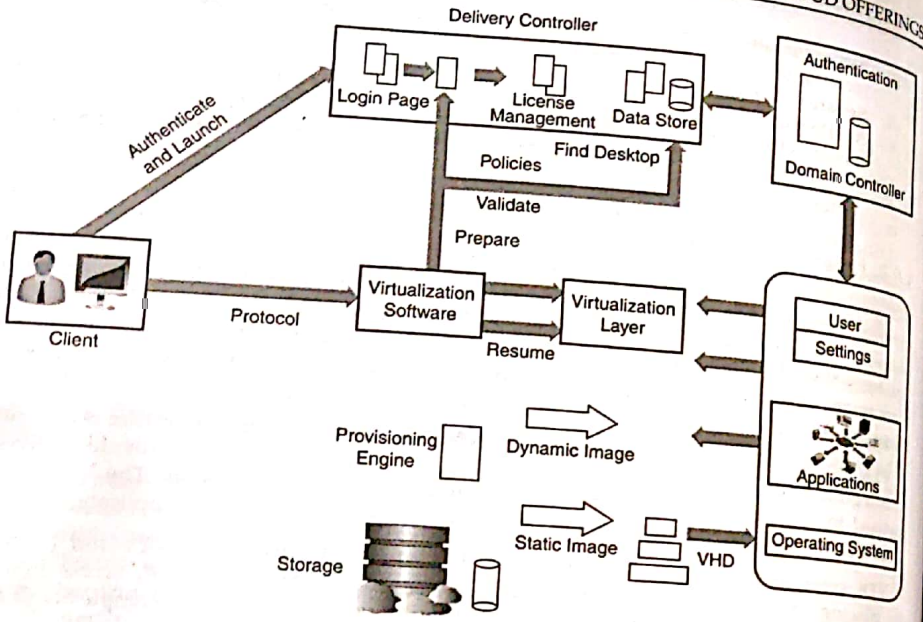

FIGURE 5.4 Virtual desktop infrastructure.

consumption-based metering. This access service also provides load balancing as needed for the front-end security servers, and for the connection servers. The security servers and connection brokers must have high availability as they are the single points of access to the virtual environment.

The VDI server farms, or datacenters, consist of a set of physical hardware platforms that facilitates the virtual environment to host shared OS, virtual machines, and dedicated desktop clients.

The infrastructure services include existing services such as activity directory, file, print, management, network, authentication services, and storage. This solution integrates with these existing services rather than introducing redundant features not required by the business. The service will not bundle in the components that are already available in the business's environment. If additional infrastructure services are required, they can be added to the project-based service for an additional charge. The cloud service project team integrates the VDI solution with the existing infrastructure services and desktop client devices as part of the project-based service engagement. Additional offerings may be combined with the cloud service product to meet the business requirements as needed.

The virtual desktop images can be configured to access the existing infrastructure services in the client datacenter as needed. The following sections briefly discuss the components of the VDI at a higher level.

5.6.3 Client Access

The users use their remote desktop client device to connect to their virtual desktop. This service is provided mostly by all virtual desktop vendors and is supported by a set of products that connect remote clients to the centralized virtual desktops. End users can use existing PC-running OSs where the user initiates a remote session to a VDI resource using a remote desktop client application or a Web browser. With some third-party products, one can have local desktop icons configured to access published applications or published desktops.

5.6.4 Desktop Virtualization Services

VDI can be viewed as a central-server-based resource pool with components connecting end users to applications, networking, and storage resources. VDI uses vendors to centrally host and deliver a cost-efficient desktop environment from the datacenter, and uses the management server to provide the virtual desktop management.

Virtualized clients and desktop management proactively manage diverse desktop environments and VDI server-based client technology. The end user retains the features and flexibility of the traditional desktop.

Management server authenticates users, determines the pool they belong to, and using predetermined policies provisions a desktop for those end users, with complete users' specifications and privileges, and finally deploys it. These pools can be persistent or non-persistent.

Individual desktop assignment is a static, one-on-one relationship between a user and a specific virtual desktop. This configuration is good for power users where the desktop is specifically configured for a particular user. This configuration can include specific applications, data access, and resource allocations. Individual desktops give users a high degree of customization.

A virtualized IT environment helps in providing security-rich, anywhere-anytime availability and access to applications, information, and resources. VDI is a unique end-user virtualization solution that helps businesses transform their distributed IT architectures.

Desktop cloud project-based service can substantially reduce the total cost of ownership by reducing the effort required for desktop PC deployment and management, software distribution, desk-side support, and help-desk required to support and maintain desktops. Businesses can avoid significant upfront investment and continuing cost for developing and maintaining the necessary skills, knowledge, and experience in systems management and desktop virtualization technologies.

5.6.5 Desktop Management

VDI brings together the desirable features of traditional terminal server while retaining the important features of distributed computing.

5.6.6 Pool Management for Virtual Desktop Infrastructure

The role of a management server is to maintain the authentication of the user, check the pool for the user, he belongs to, the policies he is entitled for, and ultimately provision the desktop

for the specific user. These pools can be persistent or non-persistent. The connection server allocates entitled users to a virtual desktop from the non-persistent pool as requested. This allocation is not retained when the user logs off and the virtual desktop is placed back into the non-persistent pool for re-allocation to other entitled users. When the user connects to the non-persistent pool on subsequent occasions, the management connection server connects the user to any virtual desktop in the non-persistent pool. This is typically a many-to-many relationship.

When a group of users is entitled to the persistent pool, every user in the group is entitled to any of the virtual desktops in the pool. The management connection server will allocate users to a virtual desktop as requested. This allocation is retained for subsequent connections.

Maximizing the use of non-persistent pools for all users who do not have a desktop customization requirement is recommended. Typically, the task users could do with a non-persistent desktop.

VDI also offers managed services for businesses that wish to derive benefits of virtual infrastructure access but lack the necessary skills and expertise required for the ongoing management of the virtual infrastructure. Businesses can avoid significant up-front investment and continuing cost for developing and maintaining the necessary skills, knowledge, and experience in systems management and desktop virtualization technologies.

5.7 STORAGE CLOUD

For any type of cloud deployment – private, public, or hybrid cloud – the environments are built using key foundation building blocks such as servers, storage, applications, and infrastructure. Storage and compute resources scale together, and the failure to manage them efficiently results in the failure of cloud services.

Storage management in the cloud can help organizations to address their challenges related to data and storage management in their clouds, for example availability of data at all times, storage resource utilization, application performance, longer restore times, higher storage costs, low productivity of storage personnel, and increased risk of data loss and downtime.

5.7.1 *Value Proposition*

Data and storage management within a cloud environment is a critical necessity to provide a reliable, on-demand service experience while at the same time reducing the costs and minimizing the risks. Streamlining the data to target applications plays an important role, which means the data has to be made available at all times and the storage should be provisioned rapidly to the applications built into the cloud for delivering efficient services. Often, storage administrators spend over 50% of their time on manual repetitive tasks. They find it difficult to meet stringent rules that are essential for restoring operations quickly after any disruption (database corruption, virus attack, disaster, and hardware failure) in a cloud. Failure to ensure data availability at all times can lead to a significant failure of a cloud service.

Also, the proper placement of data on different tiers of storage within the cloud, if done efficiently, helps to minimize the overall costs of hardware, software, and administration.

Cloud vendors offer technologies – storage, hardware, and software – as well as key storage services to support subscribers in their journey to leveraging cloud computing. They can assist in planning, designing, building, deploying, and even managing and maintaining storage solutions, whether on their premise or someone else's.

Cloud technology is helping organizations to build a smarter business infrastructure with immense flexibility and scalability – one that could result in improving service levels while reducing capital and operational costs. Today, many organizations in various industries including media and banking, which deal with large amounts of data, are increasingly adopting the cloud technology to address their needs of delivering faster services, protecting data in real time, seamless communication between employees, partners, and suppliers, for business continuity, and, of course, to become more energy efficient – to be a greener organization.

5.7.2 *Challenges*

Cloud services majorly focus on keeping the data and applications they are managing available at all times, and even more essential is to develop the ability to quickly restore the operations following any type of data disaster. Storage management is an important function to ensure that the data are available, capacity is provisioned rapidly, and storage resources are utilized effectively. Cloud administrators often find it difficult to meet all the following challenges they face concerning storage and data management:

1. Data availability and application performance.
2. High capital and operating costs, less ROI.
3. Utilization of storage resources.
4. Lack of automation – low productivity of storage personnel with specialists doing mundane tasks.

For customers, the drivers for adopting cloud technologies have been cited as follows:

1. Paying for only what they use.
2. Reduced costs.
3. Monthly payments instead of all up front
4. Having a standardized system
5. Always having the latest software version, since nothing is installed locally

5.7.3 *Business Drivers*

1. Need for standardization and automation of storage services.
2. Need to meet service levels consistently – provisioning on-demand computing capacity and storage capacity.
3. Need for simplified management of storage infrastructure – quick provisioning and redeployment of resources, built-in data reduction capabilities.
4. Data security and compliance issues.
5. Need to lower costs – lack of upfront capital, lower utilization of hardware resources
6. Recovery point objectives (RPOs), recovery time objectives (RTOs).

5.7.4 *Benefits*

1. Improved service levels by ensuring data availability and application performance and by quick provisioning through automation
2. Reduced capital and operational expenses by leveraging standardization and automation
3. Optimized utilization of storage resources and built-in data reduction capabilities to manage more storage with less hardware
4. Reduced hardware, software, and administration costs with policy-based data storage management.
5. Managed risk and streamline compliance through real-time data protection.

5.7.5 *Product/Solutions Overview*

Storage management software and services solutions for the cloud help in ensuring that business and IT are fully aligned and supported by integrated service management. They help in delivering a workload-optimized approach and offer a choice of implementation options for superior service delivery with agility and speed.

Cloud vendors offer second-generation storage management technology for cloud environments, delivering faster ROI. Cloud storage services include worldwide capability and capacity to provide integrated cloud service offerings to meet your storage management needs.

Cloud vendors reduce the complexity of managing cloud environments by offering a complete portfolio of automated solutions for managing data and storage infrastructure, enabling better efficiency for business resiliency, reducing costs, and improving security, while increasing visibility, control, and automation of the cloud storage infrastructure. There is a need of the broadest, most scalable, and reliable set of storage solutions available to keep the cloud services functional. These storage solutions should have the complete portfolio for protecting, managing, and virtualizing the environment.

5.7.6 *Product/Solution Description*

Cloud vendors should offer a complete portfolio of software solutions and services for the storage management in the cloud, designed to help in streamlining the storage resources in order to support cloud services, protection, and management of data, being able to virtualize the entire storage infrastructure, and offer it as a single resource to the cloud services bundle.

Cloud-based deployment and service models use new and scalable delivery models to offer cost-effective solutions. Data availability and application performance are critical factors to migrate the applications in the cloud. Additionally, a cloud environment cannot afford longer downtime as after any disruption, it is critical to restore operations quickly. Whether it is an organization managing its own private compute/storage cloud environments or a managed services provider offering cloud-based services (private or public), the key concern is how well the existing resources are optimized in their infrastructure, to improve end-user experience – whether they are able to provide flexibility, speed, reliability, and efficiency. Data and storage management play a critical role in improving on-demand service experience and reducing costs and risks in the cloud.

Often, the absence of sophisticated storage management systems in a cloud results in the lack of visibility into the storage utilization and provisioning, costs, and associated risks. Cloud administrators face difficulty in understanding how much capacity is available, where it is, which applications are accessing it, how secure they are, and whether they are able to meet stringent RTOs. Gaps that may exist between the unpredictable demand for data availability and the ability of the business to support the same in an efficient way, results in unmet service levels, additional downtime, new hardware and operational costs, and lower customer satisfaction.

5.8 SUMMARY

Today, as we are becoming interconnected, instrumented, and intelligent – the world is becoming smaller – more data is being generated within the operations of all organizations, and they are struggling to manage the complexity in their storage environments. The costs of backup and recovery, archiving, expiration, and storage resource management are exploding. It is not just about increasing the capacity but managing the data efficiently, reducing the data, ensuring adequate protection of the data, and quick restoration of the data for better business performance. We need better implementation of VDI, test cloud environments, and analytics to handle cloud offerings.

CHAPTER 6

Cloud Management

CONTENTS

6.1 INTRODUCTION

Companies and their IT vendors are increasingly focusing on virtualization-based cloud services and consolidation solutions, and the potential benefits they provide businesses. But this growing popularity sometimes obscures the fact that managing cloud virtualized infrastructures can present organizations significant challenges in areas related to implementation and service management. In particular, it is often difficult for businesses to determine how and for what purposes the employees and groups are utilizing virtualized IT assets.

What is required is a cloud solution that aims to help overcome these problems by providing insight into the relationships between virtualized and physical IT assets – who is utilizing shared resources, and what and how much they are using. Such information is vital in a number of ways. For organizations such as IT outsourcers, a well-suited solution will serve as an accurate measurement tool underlying billing processes and service-level agreement (SLA) compliance.

Innovative cloud virtualization technologies from cloud vendors extend that service accounting concept far beyond simple partitioning on a single server to a systems virtualization platform. This platform includes server and storage technologies and common tools to deliver workload and platform management across your IT environment.

Over the past decade, the number of department servers and department storage has proliferated, creating a challenge for IT management. The top driver for consolidation is the reduced cost, followed by improved system performance, ease of management, high availability (HA), security, and disaster recovery (DR). In addition to consolidation, many enterprises are interested in managing the growth of their IT resources to maximize return on investment (ROI). To do this effectively, they require usage information of all resources on the network (storage, server, network, and application) to build a complete picture. This information can then be used by the IT staff for optimizing their use of existing resources, improving their service level (through better performance and availability), and proactively managing their capacity planning activities.

On the other hand, users are facing some uncertainties to implement such solutions. Lack of skills to realize such virtualization concepts, or the inability to qualify the value, are the main inhibitors for effective implementation. The biggest problem for these services is to know the delivery mechanism – how it is used and how it is charged on the basis of usage. Automated processes for cost reallocation and analysis of security and misuse would result in a high level of cost savings.

IT services are viewed as critical to a business. Increase in the number of users, demand for new technologies, and complexities of client–server systems frequently cause IT service costs to grow faster than others. It is, therefore, common that enterprises are mostly not able to give business justification for service improvement investments which are then viewed as an expensive exercise.

The CPU, data store, and bandwidth usage based costing can be done by IT accounting. But it is not recommended to use it directly to charge the end-users as it loses its benefits. It is good for all to reduce the cost of the overall service rather than working on the IT accounting of the computer resources.

Current leading practice uses IT accounting to aid investment and renewal decisions, and to identify inefficiencies or poor value. A fixed amount is charged for a set capacity determined by the level of service detailed in the SLAs.

To provide cloud business with a clear understanding of the value they receive from IT, the cost model must be:

1. **Equitable:** The chargeback approach must allocate costs proportional to each unit's true consumption of IT services.
2. **Controllable:** Business units should have a degree of control over, and input into, IT spending decisions.
3. **Repeatable and predictable:** Charges for a given service should be consistent over a six- to twelve-month period enabling a business unit to forecast its IT costs over the period.
4. **Simple:** The chargeback algorithm should be easy to understand, implement, and administer to minimize confusion and overhead expenses.
5. **Comprehensive:** All IT costs must be associated with a service. There should be no 'tax' or overhead bucket to account for infrastructure.

A cost model works best when customers understand the pricing structures and their limitations, have some control or influence over the consumption of IT and thus, their cost for IT, and believe that the value is reasonable and equitable.

6.1.1 *Service-Based Model*

Recently, there has been a strong push for IT to invoice business units for services described in business terms instead of IT terms. This service-based approach has been driven in part by cost transparency and cost reduction requirements.

The success of a service-based model depends largely on business managers and IT managers working together to define the Service Portfolio, which includes the services the IT organization provides and the cost of these services to the business units. Making IT services understandable to business managers gives them a clear window into the infrastructure and application reinvestment. A business director may be persuaded to fund or support infrastructure changes that will drop or increase the consumption or price of services in order to meet business needs in a better fashion.

6.2 RESILIENCY

Resiliency is the capacity to rapidly adapt and respond to risks as well as opportunities. This maintains continuous business operations that support growth and operate in potentially adverse conditions. The reach and range step of the assessment process examines business-driven, data-driven, and event-driven risks. The objective here is to explore the risk situation for the company, processes, people, or whatever that affects the business of that organization. This can be divided further down after thorough investigation because risk in one section will not be same as risk in another section.

So we will be looking across different parts of the company. We like to focus on one specific area first – maybe a specific business process. By doing so, we usually arrive at the 80/20 rule which says that about 80% of issues are going to be common across all business processes, all business entities, and all buildings.

When you use the resilience framework to look at different parts of the company, you are trying to understand whether you have a risk that you can accept, or whether you have a risk that you want to avoid and mitigate. In other words, you may choose to do nothing about a risk, or you may improve your infrastructure to help ensure that you can handle events if they occur.

You may also decide that the risk is one that you would prefer to transfer to somebody else, such as business continuity and resiliency services. A lot of organizations feel more comfortable transferring risks associated with business continuity to cloud vendors rather than handling risks themselves, as recovery centers are designed to be robust and ensure resilience in the face of a disruption. Additionally, transferring the risk can be accomplished through managed security or resiliency services. This allows you to concentrate on strategic initiatives and leaves the day-to-day management and monitoring of your availability and security configurations to staff locations. So, what can we recommend to create a framework of resiliency? The resiliency blueprint includes different layers – facilities, technology, applications and data processes (both IT and business), organization, and finally, strategy and vision.

The framework enables us to examine the business, understand what areas of vulnerability you might have come across – business-driven, data-driven, and event-driven risks – and quickly pinpoint areas of concern and help you understand what actions you can take to reduce the risks associated with those areas.

6.2.1 *Resiliency Capabilities*

The strategy combines multiple parts to mitigate risks and improve business resilience in the following manner:

1. From a facilities perspective, you may want to implement power protection.
2. From a security perspective, that is, to protect your applications and data, you may want to implement a biometrics solution. You might want to implement mirroring, remote backup, identity management, e-mail filtering, or e-mail archiving.
3. From a process perspective, you may implement identification and documentation of your most critical business processes. You may split the functions of processes. You may also want to implement specific requirements confirming to government regulations and standards.
4. From an organizational perspective, you may want to take an approach that addresses the geographic diversity and backup of workstation data. You may want to implement a virtual workplace environment.
5. From a strategy and vision perspective, you would want to look at the kind of crisis-management process you should have in place. You may also want to examine how you can clearly articulate your security policies to everybody and how you implement change management.

Resilience tiers can be defined as a common set of infrastructure services that are delivered to meet a corresponding set of business availability expectations. The criteria describing

resilience tiers were developed by the lines of business and include characteristics/attributes for business impact (e.g., revenue), risks (e.g., legal), application availability (e.g., 24 x 7), and agility (e.g., multiple physical instances).

6.3 PROVISIONING

The provisioning process is a service that uses a group of compliant processes called 'Solution Realization'. Environment provisioning roles separate the preparation tasks and assurance tasks from provisioning tasks. Provisioning design decouples provisioning build and integration activities from requirements, design, procurement, and hardware setup. The process formalizes Quality Assurance (QA) testing in preparation for turning over the provisioned product to the customer. Provisioning is a broad-based service that begins with a Request for Service (RfS) to build a fully provisioned environment for the purpose of hosting an application and database. Provisioning can also be invoked when a major modification must be made to the existing environment. Provisioned environments include development, test, QA, production, and DR. Provisioning defines and communicates the information that is required to begin provisioning. The output from provisioning is an environment configured and tested with an appropriate hardware platform, storage, network, operating system, middleware, other system software, backup capability, monitoring capability, and with the application installed per requirement. In this way

1. Provisioned products are servers built with all the software and infrastructure required to support a business application.
2. Standard solutions are defined so that standard workflows can be derived.
3. Design is completed with due diligence before the RfS is accepted, including documentation of all specifications.
4. Server hardware is assembled, cabled, and connected to the network and storage area network (SAN) before work orders are released to the provisioners.

6.3.1 *Characteristics*

There is an owner who provides technical oversight for the lifecycle of each project lifecycle defined from the initial request for comments (RfCs) through to the delivery to the customer. Specifications are reviewed for completeness and accuracy before the release of work orders to provisioners. Missing and incorrect information is resolved before provisioning begins.

Provisioner roles for each part of the stack perform build, installation, configuration, and interim verification activities (no change). A status of 'Hold' with a reason code indicates when work orders and the RfS itself are stopped, awaiting a response from an external process. The provisioned product is tested, assured for quality, and signed off by the technical owner before being turned over to the customer.

6.3.2 *Approach*

The environment provisioning process takes an assembly line approach to building a server and integrating its components. To prevent interruption of provisioning tasks due to unforeseen

or redundant work, the process defines that upstream activities be completed and signed off before starting the downstream activities. Following are the activities discussed:

1. Planning precedes execution.
2. Validating build specifications precedes building.
3. Packaged software installation procedures, being tried and tested, precede the installation of the package on a server.
4. Having servers racked, stacked, cabled, and connected to storage and network precedes issuing work orders for provisioning the OS and base software image.

Measuring achievement is easier without having to account for stops and starts caused by handoffs. It becomes possible to automate building the stack and integrating more components with provisioning tools.

6.3.3 *Benefits*

This section discusses the benefits of provisioning.

1. **Ability to measure progress of all the work related to one RfC:** It supports the ability to deliver to service levels.
2. **Continuous improvement activities based on process measurements:** It enables eliminating delays and learning to continuously provision servers rapidly to shorten the time to deliver.
3. **Isolation of the build, install, configure, and customize tasks from requirements, design, and hardware setup activities:** It provides focus for leveraging provisioning automation tools.
4. **Role players performing a finite set of repeatable activities:** It enables the collection of intellectual capital necessary for beginning to automate their activities, and for planning full automation.
5. **An assembly line approach to provisioning:** It facilitates automation of piece parts of the process in an incremental approach to self-service.

Long-Term Goals

1. Achieve operational efficiencies by using a common set of processes and procedures to deliver provisioning services to the enterprise.
2. Achieve target environmental defect rate.
3. Establish and achieve service-level objectives for delivery of provisioned environments.
4. Reduce time to set up development and test environments.
5. Reduce hardware/software spending through optimization of all environments and reuse of assets.
6. Enforce enterprise provisioning standards.

Short-Term Goals

1. Reduce the defect rate for the setup of the development and test environments.
2. Improve and provide consistency in the provisioning of environments for all platforms.
3. Transfer skills and knowledge of new standard processes and procedures to the provisioning teams.
4. Gain stakeholder agreement before deployment of a provisioned product that all requirements have been met.
5. Reduce rework.
6. Improve quality of work experience for process participants.

6.4 ASSET MANAGEMENT

Asset management and change management interact frequently. Several of the activities required to provision an environment rely on RfCs in order to get approval to change known configurations of infrastructure and software components. There are different factors that help to develop the asset management strategy:

1. **Software packaging:** Asset management relies on software packaging. The output from software packaging will be used on a daily basis during the installation and configuration of the various software packages requested by the customers. Asset management will only engage software packaging directly when there is an exception. It will pass information so that new or modified packages can be built to enable provisioning.
2. **Incident management (IM):** It is used to track any interruptions or issues to the asset management service. These are most likely to be encountered during the OS or application installation, or during the verification of other provisioned components. IM will also be used as an entry point to Problem Management, which will not be engaged by the Asset Management directly. IM's 'business as usual' escalation of recurring incidents as potential problems will contribute to resolve problems related to asset management.
3. **Pool management:** Pool management works with asset management to make sure that the products requested are available on the requested date and for the specified duration. Pool management serves as the intermediary process between asset management and the infrastructure-on-demand (IOD) process and activities.
4. **Release management (RelM):** It controls the scheduling and testing of additions and updates to environments.
5. **Configuration management:** It helps with its own repository for assets and inventory items in the absence of a process.
6. **Systems management (SysM):** It is both a process and a service. In order to interface with asset management, it provides all of the information on what attributes of OS, middleware, and business application components need to be monitored. A mature SysM process determines triggers, thresholds, event generation, severity, event correlation automated response, and the tools that will be used.
7. **Operational readiness (OPR) management:** Asset management interacts with OPR much as other projects and services do. To prepare for release into an environment, it is necessary that the documentation describing and supporting the provisioned product aligns with the enterprise standards.
8. **Backup management:** Enterprise performance management (EPM) links to backup management after the new server is added to the backup script, along with any customizations to the backup job.

6.5 CLOUD GOVERNANCE

One of the major components of any governance model is the proper definition of roles and responsibilities within an appropriate organizational structure. The domain owners within the organization own and are accountable for the business functionality within their proper business domain. These domain owners report to the head, but they also have direct reporting responsibilities within their business domain. These technical roles, along with the domain owners, strive to achieve a confluence between business and IT. One of the major aspects of cloud governance is to ensure that the lifecycle of services maximizes the value of service-oriented architecture (SOA) to the business. In order for governance to be effective, all aspects of the service lifecycle need to be properly handled.

The process transcends all phases of the service lifecycle: model, assemble, deploy and manage. Each task is numbered based on the phase it falls under. The cloud governance scenario should be broken down into realizations (see Fig. 6.1). They can be:

1. Regulation of new service creation.
2. Getting more reuse of services.
3. Enforcing standards and best practices.
4. Service management and version control.

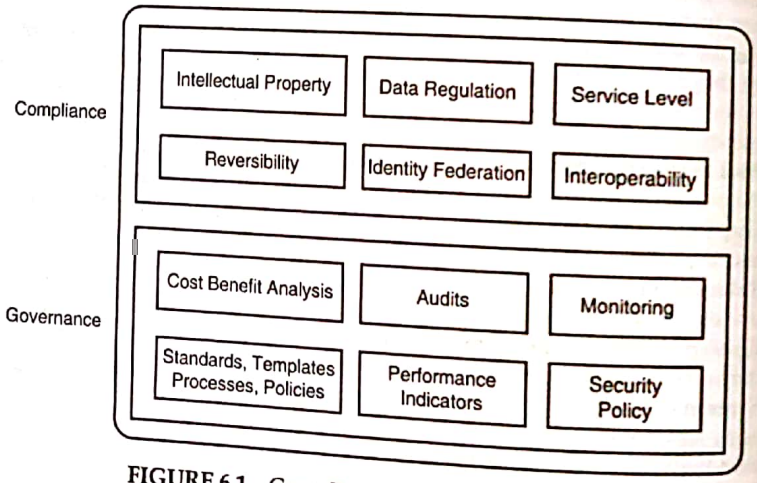

FIGURE 6.1 Compliance and governance.

6.6 HIGH AVAILABILITY AND DISASTER RECOVERY

High availability (HA) and disaster recovery (DR) are some of the important factors for cloud deployments. As the cloud is based on service models, different SLAs govern the service-based models to avail the service. HA and DR go together and define the factors for different SLAs between vendor and subscriber to ensure service availability and trust, and help develop

credibility for the cloud vendor. Availability is not just a technology issue – it is a business issue as well. It is sometimes easy for executive management to take infrastructure availability for granted – by the business executives, not just the IT executives. The business must be able to make IT investment decisions based on business value. Achieving very high levels of availability usually requires substantial investment (and not just in technology). IT must manage the infrastructure to deliver the required and funded level of availability. But 'the business' must determine the level of availability required to support its business objectives and make the appropriate investments to support that level of availability.

HA and DR have often been treated as separate disciplines, but are converging. HA traditionally focused on avoiding/recovering from non-catastrophic disruptions – server failures, software failures, power failures, network disruptions, denial of service attacks, and viruses/worms – often relatively short in duration (minutes or hours). It may involve moving workload (dynamically) to another location, but typically does NOT involve moving people. DR traditionally focused on planning for and recovering business operations following catastrophic disruptions:

1. Site/facility destruction, hurricanes, tornadoes, floods, and fire.
2. Often long duration (days to weeks).
3. Often involves shifting work (and people) to alternate facilities for some period of time.

Similar disciplines are required for both HA and DR, but with different emphasis. This is the ability where the service or any component is required to execute its function at a state or desired time. It is based on the availability ratio such as the proportion of the service actually available with respect to the agreed service hours. There are various proportions to talk about, which are:

1. **Mean time between failures:** It is the average time for the failures occurring successively in a system.
2. **Mean time to recover:** It is the average time taken to recover from a system disaster.
3. **High availability (HA):** It is the functionality of the system that provides the agreed service levels to end-users during scheduled periods.
4. **Continuous operations:** This is the feature that gives continuous access to the end-user at any time, $24 \times 7 \times 365$.
5. **Continuous availability:** This is the characteristic to deliver the agreed service level at any time, $24 \times 7 \times 365$.
6. **Availability management:** This is the process of managing the resources such as people and technology to ensure the agreed service levels to meet the metrics and need of the organization.

Recovery capability is the process of planning for and implementing expanded operations to address less time-sensitive business operations immediately following an interruption or disaster. Recovery time objective (RTO) is the period of time within which tools and equipment, applications, or compute must be recovered after an outage (most of the time it is agreed time). It is the foundation for recovery strategies related to any development. It gives good points to think about how to implement the technology to meet the consequences of the disaster situation. It is that point in time where the system and data will be recovered after the disaster or outage. Recovery point objective (RPO) is the basis for strategizing the backup

and determines how much data is to be recreated after the functions are recovered. DR is the process of creating, verifying, and maintaining an IT continuity plan that is to be executed to restore service in the event of a disaster. The objective of the DR plan is to provide for the resumption of all critical IT services within a stated period of time following the declaration of a disaster and perform the following activities:

1. Protect and maintain currency of vital records.
2. Select a site or vendor that is capable of supporting the requirements of the critical application workload.
3. Provide a provision for the restoration of all IT services when possible.

A DR plan includes procedures that will ensure the optimum availability of the critical business functions and the protection of vital records necessary to restore all services to normal. The DR plan is dependent upon and uses many of the same recovery procedures as those defined and developed by the recovery management process. The execution of the DR plan will use many of the same policies, procedures, and staff as defined in the crisis management process. The DR event is primarily a 'crisis' of greater magnitude and scope than the situations that are routinely managed on a day-to-day basis.

The true business need for HA of IT systems, including rapid recovery for disaster situations, must be determined and justified. The cost of down-time must be understood by business units to establish true business need for HA and rapid DR. An availability strategy is required to guide the organization in implementing HA and support rapid recovery from a disaster. An availability strategy should:

1. Support the IT strategy with the business policies and requirements.
2. Justify investment in HA and DR initiatives.
3. Ingrain HA in the IT culture.
4. Define a robust IT architecture and invest in building HA into the design of the infrastructure.

When DR plans fail, the failures primarily result from lack of HA planning, preparation, and maintenance prior to the occurrence of the disaster. Lack of an IT architecture employing hot back-up components and hot back-up sites inhibits the achievement of continuous availability across component failures or site failures resulting from disaster. Recovery severely delays when many back-up components have to be rebuilt from scratch. Change management processes fail to ensure backup components, and recovery documents are updated simultaneously with primary component upgrades.

The technology must fully exploit HA design techniques such as redundancy with hot back-up capabilities to support rapid recovery.

Application and data interdependencies are important considerations in determining business function priorities. Network connectivity must consider more than connectivity between the datacenter recovery site and the user site – it should consider connectivity to business users, system to system, customers, and outside agencies, and consider failure of multiple sites and setup for connectivity from one back-up site to another back-up site. Where critical business unit users must support the recovery effort, for example to prepare for end of day processing, immediate access to workstations is required. If the business processes are dependent on printing, printer recovery must be treated with appropriate priority.

The events of previous disasters confirm that effective and rapid recovery from any disaster is dependent on mature processes supporting HA. Service-level requirements and business requirements must be understood and objectives negotiated. An infrastructure supporting HA is essential for a rapid DR. The system and application designs must be built to support HA and rapid DR. Complete configuration information is necessary to reconstruct all system platforms following a disaster. Adequate testing must validate the capability of the plans and the ability to perform the procedures, whether for day-to-day HA or for disaster.

To prevent gaps in DR plans, recovery procedures, technology platforms, and DR vendors, contracts must be updated concurrently with changes. Fast and effective recovery from a component outage or from a disaster requires well thought out, pre-developed, tested, documented, and practiced recovery. Defects and shortcomings must be resolved quickly to ensure that the plan will work.

6.7 CHARGING MODELS, USAGE REPORTING, BILLING, AND METERING

Today, enterprise business units' budgets fund 60–70% of central ITs services. The other 30–40% is funded by other means, so it is clear that in general, organizations do not use a single charging mechanism, but a combination of mechanisms for different purposes to achieve an overall solution. Existing processes were institutionalized in many large organizations decades ago and the responsibility has been passed down from employee to employee over generations. The pitfalls of chargeback are well-documented; they include user architectural rebellion, IT investment vacillation, bureaucratic excess, and malicious obedience to IT standards. These pitfalls fall into an IT-centric view of providing service; these arguments and others like them seem shallow when presented with the business imperatives that are often at the root of maintaining a chargeback system.

6.7.1 *Challenges*

Many organizations do not implement sophisticated internal chargeback mechanisms due to their complexity. You should able to determine all the metrics, and in order to be able to break them out by the user, you have to keep track of what organizations the users are in, which is not a simple task. This creates a large volume of data for the items that you can directly tie to a user (e.g., CPU, memory, and disk that are associated with a particular transaction). While this is reasonably easy to do in a dedicated workload environment, enterprise environments add another layer of complexity.

Further, you have to make additions in the overhead (OS, program products, network, support, and processes such as space management). When you introduce the allocating of details, for example, the cost for SAN ports in a switch, you may have more of a political discussion than a technical one, as you may not be able to tie back specific items to transactions.

6.7.2 *Benefits*

The benefits from implementing a more effective system can be enormous. The following are the advantages for managers looking for the benefits of implementing a more comprehensive

chargeback system. When forced to confront the issues of chargeback implementation or a chargeback system changes, managers should align their practices with the benefits of a chargeback system.

Charging for services will not solve all the problems of the IT department, nor will it be the source of all service problems when dealing with business managers. IT managers must leverage a chargeback system to harvest opportunities for improving and streamlining service delivery.

6.7.3 *Cloud Chargeback Models*

In consolidated environments, IT accounting service employs a cost recovery mechanism called chargeback. Chargeback is the system that is devised to put across the fee for the services provided in a cloud-based model. This allows the services in various bundles of value proposition to recover the cost for different offerings at different service levels.

In order to have the actual benefits of a chargeback model, organizations need to understand their own cost structures, and break down the components of the services and resources. The inner depth of devising an effective financial model requires functionalities such as the utility-based model to charge the various resources based on the billing costs (Fig. 6.2).

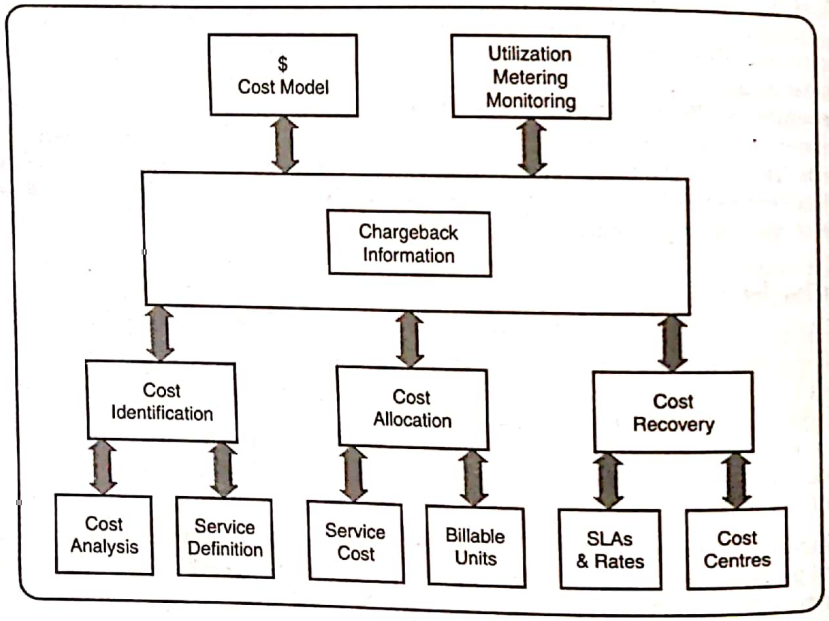

FIGURE 6.2 Chargeback model.

It is important to note that the chargeback models will not come with a silver lining and solve all the problems that are related to any organization's cloud commodity costing model. It comprises many proven, tested models in the industry, but each model is specifically based on its situation and costing mechanism. It is advised to evaluate the various models and use them based on the culture, technology, financial or social boundaries. In the following subsections we will discuss some models of chargeback that can be adopted as stand-alone or as hybrid costing models.

Subscription Model

It is the simplest model. It is actually derived by dividing the total operational cost of the organization that runs IT and the total applications hosted in the environment. The cost mechanism is very simple and recovery is easier, because of which it is more convenient to adopt. Hence it is finding its way in many organizations. This type of model requires a constant addition in the costing model by adding the upgraded cost to the model with the subscriber. This is chargeback model, but at the same time it has a drawback that it provides subsidy and unbalanced allocation of resources. In other words, some applications that are not performing well are subsidized with the high performing application and infrastructure.

Pay-Per-Use Model

This chargeback model is good in an environment where the organization has multiple lines of businesses or tenants of different size and scale. The costing model is based on the actual application consumption based on agreed SLAs. The same application can be charged differently at different service levels. If the architecture of one application is not good and consumes more resources and time, it will be charged at the upper side. It can be a complicated model as it requires managing and monitoring the service levels and resource consumption. It is good from the recovery perspective, but it is difficult to get common levels with different teams to finalize the cost and financial models.

Premium Pricing Model

In this model service and availability is guaranteed for any critical service deployment. It is devised on the line of business priorities and preferential services rendered to specific tenants. This comes at a premium pricing and supports specific policies, resource reservations, and service levels. It is also dependent on the deployment model such as a dedicated or shared model. Therefore, based on isolation and separation, a shared service-costing model can be devised. It is always preferred by those units that are involved in mission-critical applications with a high revenue generation environment. It will never exist alone in any organization. It will come in combination with other models discussed above.

'Hybrid' Model

This chargeback model is an exercise to carry the best practices of all the models and create the best costing and chargeback model for any organization. This can combine two or more chargeback models, for example, a combination of a flat fee for registering for the applications services for the first time and then charge based on the resource usage. This way we can bring transparency in the system by showcasing the flat fees as the fixed price for hosting the service, and then actual charges based on the resource utilization and usage scenarios as a variable cost.

Similarly, if we combine the subscription and as pay-per-use model, it works like the utility services provided to any home by many agencies, such as water and electricity. In total, we can say that there is no one size that fits all situations. One model can be useful for one organization and the same will not work in other organizations. It is now customary to launch service

catalogs for services offered, and any line of businesses (LOBs) can use any of the service pay back based on the agreed financial chargeback model. This gives freedom of choice to the LOBs to choose their services with their own flexibility. Cost will change from one LOB to the other if the volume, quality of service, policies, allocation, and service levels change. This freedom will help the LOBs to choose their model and add value to their services for end-users.

Aligning Chargeback with Business Perspective

In order to align the IT services with business objectives, chargeback tools become vital options. They bring together the IT systems and business users to realize the value of the infrastructure laid down to be shared by the community. The final goal viewed by the chargeback model is to provide the services at competitive prices with resiliency and robust value-added services. It is achieved by the optimized and efficient use of the computer resources, and enabling virtualized environment in the shared environment.

The chargeback models require good knowledge with respect to the organization. They require both technical and financial expertise to convert the service request into effective costing models. The other important point is the agreement between the IT and business community for values and services given by the chargeback models.

Therefore, it is very important to simplify the chargeback that can be understood by the business community, as well as it should be in agreement with the IT infrastructure managers. So, education of the models is very important to adopt and diffuse the characteristics to the indented users in the first step. Next, it is very important to divide the IT costs based on the organization's LOBs. This will help to understand the LOBs requirement and operating costs against the designated budgets to ensure transparency across the organization. The transparency characteristic motivates LOBs to work for the common objective of chargeback models across the organization. This can be the iterative process and will be reviewed by various participants from LOBs and finally agreeing to the common acceptable financial terms. It is common to use hybrid chargeback models to come to the agreement stage. Once all the models are reviewed and analyzed, the most suitable model will be opted and the most fair cost recovery model will be visualized. At the same time, organizational financial factors also play a vital role in the exercise with appropriate reporting practices. This will involve both operational and financial teams to work together and deploy mutually accepted chargeback models.

Governance is very important, especially in the shared infrastructure, as it is used by various business units, and requires policies and control to maintain the boundary conditions of any business units. The isolation of the infrastructure is also maintained for a dedicated set of requirements by the specific business units or application requirements. Under these situations, where shared, isolated, and dedicated environment is required, governance becomes supreme. It also helps to do the change management and RM to provide more satisfaction in the shared infrastructure

It is visualized that there exists flexibility in the virtualized environment with features like effective productivity and manageability. Governance helps to establish the following features:

1. Application enhancement.
2. Chargeback models.
3. Administration.

Also, it includes various features, such as responsibility and self-service controls. The best way to accomplish the governance is to categorize the policies and controls based on the services and chargeback models between the IT systems and the business community. This will help to adopt the model and any changes in the system can absorbed easily without any contention by the business community. Once established, other complex models can be introduced, such as chargeback, showbacks, transparency reports, and even penalties for operational and financial controls. This also follows the cyclic process and requires time for maturity under the supervision of the executive management.

6.7.4 Basic Requirements

The area of business unit contribution to IT funding can cause significant friction between business units and IT managers. For this reason the business units need a documented understanding of what they are getting (i.e., value) for their money. The chargeback metrics used in determining the individual business units' share of funding contribution should be directly tied to the SLA between central IT and the business unit and should reflect the following elements:

1. **Fairness:** The chargeback approach must be seen as allocating costs in proportions that reflect each unit's true consumption of information and communication services. One group should not be subsidizing IT usage at the expense of another.
2. **Control:** Business units should have a degree of control over, or input into, IT spending decisions.
3. **Repeatability and predictability:** Charges should be repeatable (there should be consistency in the application of data collection and charging methods so that charges are consistent) and predictable (a business unit should be able to create a reasonable forecast of its expected charges over a six- to twelve-month period).
4. **Simplicity:** The chargeback algorithm needs to be easy to understand, and simple and inexpensive to implement.

Chargeback works best when customers understand the pricing structures and their limitations, have some control or influence over costs, and believe that the policy is reasonably fair, given the limitations of a particular system and the underlying business rationale behind chargeback.

Chargeback Schemes

Possible chargeback approaches are listed below. It may well be that the best solution is to use a combination of these for different aspects of the IT infrastructure.

Allocation Based

In this model, IT service costs are buried in corporate overhead as a budget line item, usually determined one year at a time. The model has nothing to do with usage; instead, it charges business communities based on their position within the enterprise (e.g., the share of employees, unit shipment volume, or total revenue). This model is attractive as it is the simplest one and costs least to implement. However, the following are some weaknesses:

1. The difficulty of rebalancing the scale when the business measures change.
2. The lack of incentive for end-users to control their resource usage.
3. The frustration of business managers unable to control or influence their budget share – although, at least, it will be predictable.

Flat Fee

This model adds elements of negotiation and capacity planning. The IT organization determines what percentage of the IT service workload a business area represents, calculates a preliminary package rate for that area, then negotiates a rate with business managers. For example, if finance represents 8% of the IT workload, it might pay a proportional fee. Because the flat fee is tied to usage, it gives business managers a chance to understand what they are paying for. Flat fee is appropriate for environments in which third-party application packages are used heavily. Variations such as access fees and subscription fees can be relevant to certain components of the system, such as the network and specific end-user services.

Resource or Usage Based (Direct Cost Recovery)

Resource-based costing and its most common form of implementation, usage-based costing, focus on developing a standard unit cost for each major resource type or category that best represents the use of that resource. For example, the measure for CPU usage could be CPU seconds consumed by an application; for storage usage, it could be number of gigabytes of storage occupied by an application or business; for the network, it could be number of bytes transferred. The basic idea is that the costing unit represents some measure of the resource consumed that can be traced back to the user of that resource.

This method requires that all elements of the IT infrastructure and associated software specific to the application be identified and are directly charged to the end-user on a per-user basis. The cost per unit (whatever unit is chosen) needs to cover all IT-related costs – operations, support, buildings, and networks. There may be parts of the 'enabling infrastructure' that are chosen to be recovered through other methods, such as allocation, flat fee, or a per-user charge.

This cost-recovery model is still widely used as a traditional mainframe approach. However, bundling mainframe computing services into resource-based charges can create bloated CPU fees, which prompt users to purchase their own systems. This approach is not always effective in the complex PC-based and distributed computing environments where the mechanics and time involved in tracking usage may cost more than what the IT organization recovers. Moreover, the language in a resource-based chargeback scheme is so techno-centric that the bill mystifies business managers.

Product or Service Based

In the commercial environment, there has been a strong push for IT to invoice the lines of business or business units in business terms instead of IT terms. This means that instead of charging a business unit for CPU seconds consumed (or in the case of networks, the number of bytes transferred), this model defines IT costs in measurable events, transactions, and functions that are relevant to the business and outside the IT organization, for example, invoices produced, cheques written, e-mail messages sent, reports delivered, number of claims processed, number

of policies written, or some other metric that represents the work performed. This method could be called a business product-based approach, whereas, the resource-based method is an IT product-based approach (e.g., bills are expressed in IT terms like CPU seconds).

In any case, the product-based approach requires all the data collection instrumentation and methods used in a usage-based approach to be in place, and then expanded to include the mapping of that usage data to the product and service categories in addition to department categories.

The success of the model depends on business managers and IT organizations together defining specific services that the IT organization agrees to provide and for which the business agrees to pay. Making IT services understandable to managers gives them a clear window into infrastructure and application reinvestment.

Activity Based

Activity-based costing (ABC) is the most difficult of all the methods to develop and implement. There are almost no large IT organizations that have a truly activity-based approach to IT costing. The IT organizations that claim to have ABC usually do not – they have product-based costing. Quite a few organizations started ABC efforts within IT, but stopped them before completion and settled them according to product-based costing.

ABC assigns costs to each activity that goes into delivering a product or service. ABC is a cost methodology which

1. Derives the costs of an organization's outputs (products and services).
2. Identifies the activities and tasks (processes) used in the production and delivery of the outputs.
3. Identifies the resources consumed in the performance of these processes and instruments these activities so that the cost per task can be rolled up into a charge per major activity by the department.

ABC takes resources (i.e., expenses from the general ledger accounting system), moves them to activities (i.e., moves those costs to activities), and then moves the costs from activities to cost objects (i.e., product and services).

Activity-based management (ABM) is the method or process of using the data produced by ABC. So ABC produces cost information, ABM takes that information and uses it to find ways to improve those costs and the overall operations of the organization.

While IT product-based costing produces a charge by product (e.g., claim or policy), IT ABC produces a charge by activity (e.g., printing a claim or handling a policy or printing a report). As can be seen by this example, the level of detail required and reported is significantly higher than product-based costing.

ABC is the 'premiere' approach of cost accounting options. Its strength is that costs can be managed very well since each activity has a cost driver that can be measured. So, a charge area such as 'handling a claim' may be broken down into 10 IT service activities. These activities can be ranked by their total contribution to the overall cost of 'handling a claim'. This allows cost managers to focus their time on the largest contributors to cost by activity.

Even though this information is extremely valuable to decision-makers, the cost of getting this information can be prohibitive for all but the most disciplined of institutions. It requires a major investment of time, people, and resources to build and maintain an ABC system, with an extended implementation period.

External Pricing Model/Market Based

This model is geared towards turning a profit. The model presumes that an IT organization operates in a fairly open market inside and outside the enterprise, and requires some 'market testing' for the cost of services. Pricing is determined by what is available in the outside market. Although a small percentage of midsize enterprises use this model, it has clear advantages; it may well become a hallmark of IT organizations recognized as value-generating service providers.

Determining the correct pricing can be expensive if a survey is required to support the scale of charges. However, basic comparison with contract, staff, and consulting rates and high-level assessment of IT spends against benchmarks is sufficient to support the cost model.

Very often, profit center cost models will lead to over recovery. This can be corrected with a simple adjustment within the accounting process. However, care must be taken not to under-recover costs. Year-end upwards corrections will often cause friction with the business units. Most organizations look to over-recover by a small percentage for cost center accounting.

6.8 SUMMARY

Today IT delivers technology to the business units and assesses charges based on the number of devices provided. The business units do not have the ability to identify the elements of this cost and cannot manage their consumption of the technology.

This chapter addresses this problem by bundling the technology IT offers into services for the business units to purchase as needed. The cost model strategy detailed in this chapter provides recommendations about how to design and implement an equitable, accurate, and auditable method of charging for services that provide value to customers.