CHAPTER 1

FIRST DRIVE

1.1 INTRODUCTION

One of the latest drift in small and medium businesses and enterprise-sized IT is the need for a significant transformation of the IT environment. Cloud computing provides a major shift in the way companies see the IT infrastructure. This technology is primarily driven by the Internet and requires rapid provisioning, high scalability, and virtualized environments. It provides the abstraction for the business and is handled by the actual owners of the infrastructure experts. In this demanding world, the raison d'être to adopt cloud computing over standard IT deployments is flexibility, stability, rapid provisioning, reliability, scalability, and green solutions. Cloud computing can trace its intellectual roots back to grid computing, but it is often confused as the outcome of grid computing advancements and research during recent period, and that is not totally true. Grid computing paves the path for the evolution of the cloud computing concept. While these may be examples of applications of cloud computing for IT infrastructure, they are not the only ingredients of it. So, before going into the details of cloud computing, let us have a cursory glance at grid computing that gives you an immense computing grid to tap into as you need it, and scale up and down as per the requirement.

Grid computing approach starts with the breaking of the silos by inserting an additional layer on each server included in the grid. The main function of this additional layer is to create logical servers that distribute over different physical servers the computational needs (job, tasks) required by the different applications they are virtually executing. In this way, it is possible to decouple the applications from the physical systems on which they were running, and at the same time, it is possible to dynamically increase or decrease the computational power of the logical servers as per application needs.

1.1.1 Grid Computing

A grid is made up of a number of resources and layers with different levels of implementation (Figure 1.1). As said, there are different types of grid that are usually organized according to this taxonomy. Starting from the layer at the bottom – virtualization, which involves only physical resources – we may have then:

- **Information grids:** These are aimed to provide an efficient and simple access to data without worries about platforms, location, and performance.
- **Compute grids:** These exploit the processing power from a distributed collection of systems.
- **Services grids:** They provide scalability and reliability across different servers with the establishment of simulated instance of grid services.
- **A mix of them:** Each of these have specific sets of characteristics that are peculiar of the hybrid characteristics of compute and service grids.

Conceptually, we can imagine three layers, the lower being the physical one where we have the servers, storage devices, and the interconnecting network. In the second layer, we see the different operating systems, mapped one-to-one with the servers. The upper layer is the application one where we map different applications supporting the enterprise business processes.

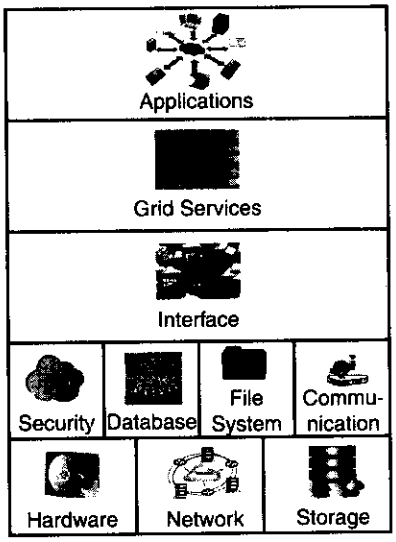

FIGURE 1.1 Simple grid architecture.

Grid computing is an evolution of distributed computing that utilizes open standards to allow you to see independent and physically scattered computing resources as though they were a unique large virtual computer.

With these concepts in mind, we can consider a 'compute grid', where the grid's goal is to exploit the processing power from a distributed collection of systems. Main functionalities of a compute grid are to manage the resources' workload, apply utilization policies and security rules, schedule and execute parallel tasks across distributed resources, and provision (reserving, adding, removing) resources according to the scheduling needs. It is a special kind of compute grid where resources – typically distributed all over the world, but could also be within an enterprise – are used by the grid only when idle, which means provisioning and scheduling policies are very 'relaxed'.

Information grid provides transparent and efficient access to data independent of their location, type, and platform, and allows end-users secure and transparent access to any information source regardless of where it exists. It supports sharing of data for processing and large-scale collaboration, and provides logical views of data without having to understand where the data is located or whether it is replicated. It manages data cache or data replication automatically to get the most efficient and secure access.

Information is usually defined as 'meaningful data' from the perspective of the end-user. An information grid provides an abstraction over disparate and distributed information sources, such as a Database Management System (DBMS), flat files (for example, comma-separated files), structured files (for example, XML documents), or a Content Management System (CMS).

An information grid also has the ability to federate or integrate data and information from heterogeneous resources into a unified virtual repository. The whole idea is to present a single view of the information.

1.1.2 *Grid – The Way to Cloud*

The concept of cloud computing can trace its roots to grid computing that provides rapid provisioning of resources. It is not mandatory that grid computing should be in the cloud; actually it depends on the type of users, whether they are consumers or administrators. Grid computing requires software that can be divided and computed or serviced on a single or multiple systems. This creates a problem of non-functioning of the overall solution if one of the components fails because of the internal dependency. With the advent of Internet, computing crossed geographical boundaries and networks and has given us the chance to exploit services and computing globally over the Internet.

Both cloud and grid services provide scalability as a functionality. This is achieved through load balancing and high availability instances of the applications running either on variety of operating systems or a single one. Both services provide on-demand services for the instances, users, storage, networks, and data transferred at a particular time, and can be de-allocated when they are not required. These computing involve the multi-tasking environments available on single or multiple instances based on single or multiple servers.

Optimization is a grid type where the primary focus is optimization of underutilized IT resources in an organization. Grids require a different way of thinking about how to deliver IT datacenter services, and resistance to changing behaviour is always the toughest hurdle to overcome in technology adoption. Lack of industry standards is a barrier to widespread adoption, as clients perceive the risk of not-protected technology investment. Security will have to be proven over time to potential customers at a number of levels for grids to be considered for adoption in shared workload environments. The cost of computational power (both CPU and storage) continues to decline, which may erode part of the financial benefits of grids. To exploit grid advantages fully, physical resources across heterogeneous systems can be virtualized building a single resource image.

The following sub-sections will help us understand the benefits of grid computing when deployed for infrastructure management and extended to cloud computing arena (Figure 1.2). This is also discussed in detail later in the chapter.

Storage/Data/Information

- Provides logical views of data without having to understand where the data is located or whether it is replicated.

System Management

- Defines, controls, configures, and removes components and/or services (could be physical) on a grid using automated or physical methods.

Metering, Billing, and SW Licensing

- Provides tools to monitor and distribute the number of licenses while using licensed software.
- Provides metering and billing techniques, such as utility-like services, so that the owners of the resources made available are accurately compensated for providing the resources.

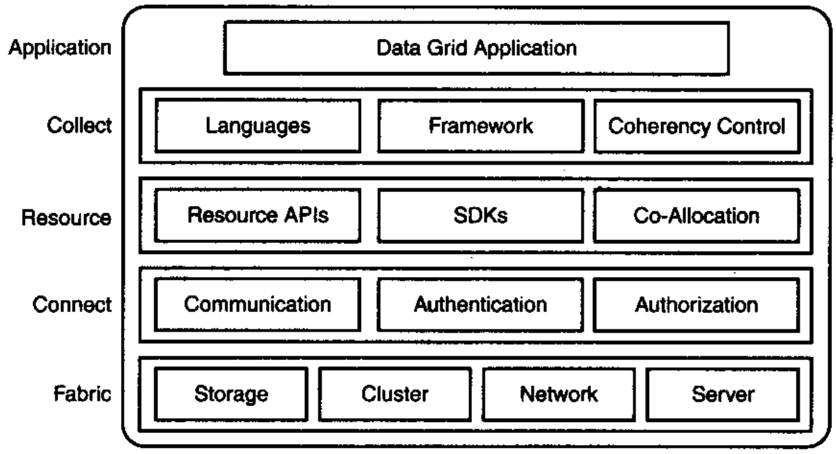

FIGURE 1.2 Standard grid architecture.

Security

- **Authentication:** The grid has to 'be aware' of the identity of the users who interact with it.
- **Authorization:** The grid has to restrict access to its resources to the users who are eligible to access it.
- **Integrity:** Data exchanged among grid nodes should not be subject to tampering.

Differing grid solutions may hit differing stages, but majority of the grid marketplace is transitioning from the 'early adoption' to the 'early majority' phase. Over the past few years, the market has evolved from specialist customers – predominantly in the academic and research sectors – using grid to accelerate internal simulations to a stage where corporate users are starting to apply grid and virtualization in a meaningful way that delivers clear business benefit (risk and portfolio analysis, seismic applications, clash analysis, etc.).

Organizations are now starting to use grid and virtualization technologies to unleash idle computing capacity to accelerate critical business processes and to optimize and improve resiliency of their IT infrastructure.

1.2 ESSENTIALS

Cloud computing is a term that describes the means of delivering any and all Information Technology – from computing power to computing infrastructure, applications, business processes and personal collaboration – to end-users as a service wherever and whenever they need it.

The *Cloud* in cloud computing is the set of hardware, software, networks, storage, services, and interfaces that combine to deliver aspects of computing as a service. Shared resources, software, and information are provided to computers and other devices on demand. It allows people to do things they want to do on a computer without the need for them to buy and build an IT infrastructure or to understand the underlying technology.

1.2.1 Emerging Through Cloud

Cloud computing is an emerging style of IT delivery in which applications, data, and IT resources are rapidly provisioned and provided as standardized offerings to users over the web in a flexible pricing model.

Cloud computing is a way of managing large numbers of highly virtualized resources such that, from a management perspective, they resemble a single large resource.

There is greater need for IT to help address business challenges and cloud computing can help you do all of these:

- **Doing more with less:** Reduce capital expenditures and operational expenses.
- **Higher quality services:** Improve quality of services and deliver new services that help the business to grow and reduce costs.
- **Reducing risk:** Ensure the right levels of security and resiliency across all business data and processes.
- **Breakthrough agility:** Increase ability to quickly deliver new services to capitalize on opportunities while containing costs and managing risk.

Cloud computing is the provision of dynamically scalable and often virtualized resources as a service over the Internet *(public cloud)* or intranet *(private cloud).*

1.3 BENEFITS

As an emerging IT delivery model, cloud computing can significantly reduce IT costs and complexities. The buzz surrounding cloud is based mostly on a new kind of user experience – particularly in the consumer Web space – for search, social networking, and retail. From the consumer perspective, cloud computing is a means of acquiring services without needing to understand the underlying technology. Many of us use cloud delivery models everyday without knowing it when we share photos online, download music, or access bank accounts using our mobile phone.

From a technology perspective, cloud computing is loosely defined as a style of computing where dynamically scalable resources (such as CPU, storage, or bandwidth) are provided as a service over the Internet. The process is typically automated and takes minutes. Cloud computing can be considered as a massively scalable, self-service delivery model that lets you access processing, storage, networking and applications as services over the Internet. Enterprises adopt cloud models to improve employee productivity, deploy new products and services faster and reduce operating costs – starting with workloads that are ripe for this environment. These typically include development and test, virtual desktop, collaboration, and analytics.

1.4 WHY CLOUD?

A cloud typically contains a significant pool of resources, which could be reallocated to different purposes within short time frames, and allows the cloud owner to benefit significantly from

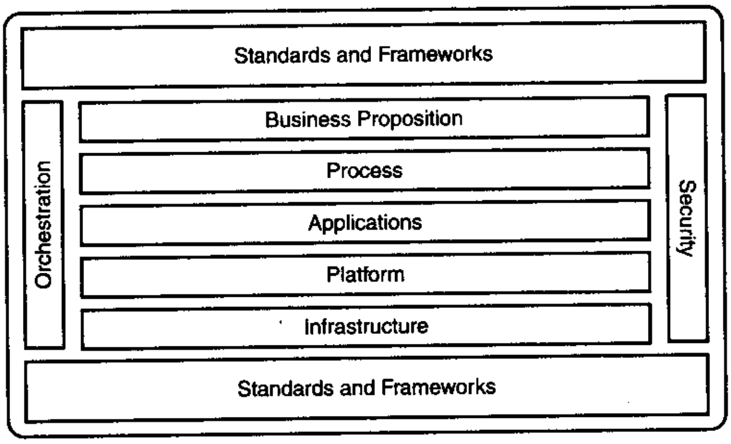

FIGURE 1.3 Basic cloud computing model.

economies of scale as well as from statistical multiplexing (Figure 1.3). The entire process of requesting and receiving resources is typically automated, and is completed in minutes.

Cloud services today are delivered in a user-friendly manner and offered on an unprecedented scale. The payment model is pay-as-you-go and pay-for-what-you-use, eliminating the need for an up-front investment or a long-term contract. This presents a less disruptive business opportunity for businesses with spiky or unpredictable IT demands, as they are able to easily provision massive amounts of resources on a moment's notice and release them back into the cloud just as quickly.

There are different reasons for adopting the cloud:

- Massive, Web-scale abstracted infrastructure.
- Dynamic allocation, scaling, movement of applications.
- Pay per use.
- No long-term commitments.
- OS, application architecture independent.
- No hardware or software to install.

This results in business- and IT-aligned benefits:

- Accelerate innovation projects that can lead to new revenue.
- Make IT an enabler of, not a barrier to, rapid innovation.
- Provide an effective and creative service delivery model.
- Deliver services in a less costly and higher quality business model, while providing service access ubiquity.
- Create a sustainable competitive differentiation.
- Rapidly deploy applications over the Internet and leverage new technologies to deliver services when, where, and how your clients want them – before your competitors do.

- Lower IT barriers to launch new business services.
- Build and integrate modular services – in record time – by leveraging 'rentable' IT services capabilities, pay only for what you use.

1.5 BUSINESS AND IT PERSPECTIVE

Businesses are now looking internally and saying to themselves that we need to deliver this same level of end-user experience with our own IT for our end-users – employees, partners, and customers. Delivering IT-enabled services via the Internet that are built for the end-user to be in control is what has come to be called 'cloud computing'.

Cloud computing is an emerging consumption and delivery model that enables the provisioning of standardized business and computing services through a shared infrastructure, where the end-user is enabled to control the interaction in order to accomplish the business task.

Computing resources such as processing power, storage, databases, and messaging are no longer confined within the four walls of the enterprise. Instead, a tightly woven fabric of abstract – or virtual – resources are tapped into whenever they are needed. Essentially, everything needed from a computing resources standpoint is provisioned by the cloud – much like the electrical power grid we all tap into.

1.6 CLOUD AND VIRTUALIZATION

Virtualization has been around for 30 years. Yet, how many have really truly virtualized at all the layers of the stack? You really cannot expect cloud to produce what a cloud is expected to produce if it is not virtualized, standardized, and automated, because people expect scalable services.

In a cloud environment, people expect self-service, being able to get started very quickly, self-provisioning, or rapid provisioning. All of those things essentially demand that you do have these very important fundamentals in place.

The only way you are going to be able to get efficiency is by virtualizing, standardizing, and automating (Figure 1.4). And that's going to drive down costs and improve service. This is really a pretty simple equation and we are seeing organizations that are doing this achieve very real measurable business results. These results include:

- **Server/storage:**
  - IT resources from servers to storage, network, and applications are pooled and virtualized to help provide an implementation-independent, efficient infrastructure, with elastic scaling – environments that can scale up and down by large factors as demand changes.
- **Automation using:**
  - Self-service portal: Point and click access to IT resources.
  - Automated provisioning: Resources are provisioned on demand, helping to reduce IT resource setup and configuration cycle times.

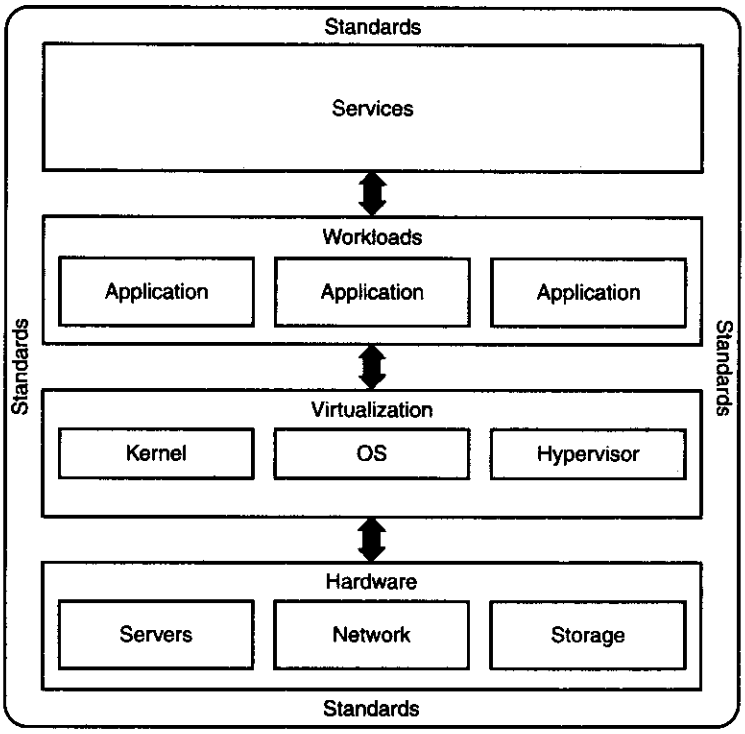

FIGURE 1.4 Datacenter clouds.

- **Standardization through:**
  - Service catalogue ordering: Uniform offerings are readily available from a services catalogue on a metered basis.
  - Flexible pricing: Utility pricing, variable payments, pay-by-consumption with metering and subscription models help make pricing of IT services more flexible.

1.7 CLOUD SERVICES REQUIREMENTS

Cloud computing is being touted as the next best thing for cutting the cost of providing first-class IT services. You can decide which workloads are right for the cloud and which may not be through an examination of your workloads – uses of IT resources for particular activities or tasks. You can also decide which workloads can go on the vendor cloud (via the Internet or a virtual private network [VPN]) and which need to remain onsite (behind the organization's firewall). This focus on outcomes and delivery models presents a new opportunity to open up competitive accounts and expand the IT optimization conversation with existing clients.

Most cloud computing vendors offer point-solution and product offerings. In contrast, one should offer comprehensive, asset-based solutions that help deploy dynamic infrastructure, which is required for a cloud delivery model. These services along with workload solutions are designed to deliver business outcomes to our clients. Any approach to cloud computing should offer the following powerful advantages:

- A proven service management system embedded with cloud services to provide visibility, control, and automation across IT and business services.
- Services targeted at certain infrastructure workloads to help accelerate standardization of services, supporting significant productivity gains and rapid client payback on their investment.

Infrastructure strategy and planning services for cloud computing should be designed to help companies plan their infrastructure workloads via appropriate cloud delivery model. Specific assistance includes cloud strategy, cloud assessment, design and development of a cloud roadmap, and return on investment (ROI) assessment by workload. Cloud leaders can help clients identify the right mix of public, private, and hybrid cloud models for infrastructure workload. Clients should be encouraged to get started with a strategy and planning consulting engagement as well as a pilot implementation of a key workload.

1.8 CLOUD AND DYNAMIC INFRASTRUCTURE

Through cloud computing, clients can access standardized IT resources to deploy new applications, services, or computing resources rapidly without re-engineering their entire infrastructure, thus making it *dynamic*.

Cloud Dynamic Infrastructure is based on an architecture that combines the following initiatives:

- **Service Management:** Provide visibility, control, and automation across all the business and IT assets to deliver higher value services.
- **Asset Management:** Maximize the value of critical business and IT assets over their lifecycle with industry tailored asset management solutions.
- **Virtualization and Consolidation:** Reduce operating costs, improve responsiveness, and utilize resources more fully.
- **Information Infrastructure:** Help businesses achieve information compliance, availability, retention, and security objectives.
- **Energy Efficiency:** Address energy, environment, and sustainability challenges and opportunities across the business and IT infrastructure.
- **Security:** Provide end-to-end industry customized governance, risk management, and compliance for businesses.
- **Resilience:** Maintain continuous business and IT operations while rapidly adapting and responding to risks and opportunities.

1.9 CLOUD COMPUTING CHARACTERISTICS

Cloud computing uses commodity-based hardware as its base. The hardware can be replaced any time without affecting the cloud. It uses a commodity-based software container system. For example, a service should be able to be moved from one cloud provider to any other cloud provider with no effect on the service.

This also requires a virtualization engine and an abstraction layer for the hardware, software, and configuration of systems. It has the feature of multi-tenant where multiple customers share the underlying infrastructure resources, without compromising the privacy and security of their data. Clouds implement the 'pay-as-you-go' pattern with no lock-in and no up-front commitment and are elastic as the service delivery infrastructure expands and contracts automatically based on the capacity needed.

1.9.1 Cloud Computing Barriers

IT organizations have identified four major barriers to large-scale adoption of cloud services. The first one is security, particularly data security. Interestingly, the security concerns in a cloud environment are no different from a traditional datacenter and network. However, since most of the information exchange between the organization and the cloud service provider is done over the web or a shared network, and because IT security is entirely handled by an external entity, the overall security risks are *perceived* as higher for cloud services. Some additional factors cited as contributing to this perception are (a) limited knowledge of the physical location of stored data, (b) a belief that multi-tenant platforms are inherently less secure than single-tenant platforms, (c) use of virtualization as the underlying technology, where virtualization is seen as relatively new technology, and (d) limited capabilities for monitoring access to applications hosted in the cloud.

The next one is *governance and regulatory compliance*. Large enterprises are still trying to sort out the appropriate data governance model for cloud services, and ensuring data privacy. Quality of service (availability, reliability, and performance) is still cited as a major concern for large organizations. Not all cloud service providers have well-defined service-level agreements (SLAs), or SLAs that meet stricter corporate standards. Recovery times may be stated as 'as soon as possible' rather than a guaranteed number of hours. Corrective measures specified in the cloud provider's SLAs are often fairly minimal and do not cover the potential consequent losses to the customer's business in the event of an outage. Inability to influence the SLA contracts is another issue. From the cloud service provider's point of view, it is impractical to tailor individual SLAs for every customer they support. The risk of poor performance is perceived higher for a complex cloud-delivered application than for a relatively simpler on-site service delivery model. Overall performance of a cloud service is dependent on the performance of components outside the direct control of both the customer and the cloud service provider, such as the network connection.

Integration and interoperability is the third concern. Identifying and migrating appropriate applications to the cloud is made complicated by the interdependencies typically associated with business applications. Integration and interoperability issues include a lack of standard interfaces or APIs for integrating legacy applications with cloud services. This is worse if services from multiple vendors are involved. It also includes software dependencies that must also reside in the cloud for performance reasons, but which may not be ready for licensing on the cloud. There are worries about how disparate applications on multiple platforms, deployed in geographically dispersed locations, can interact flawlessly and can provide the expected levels of service.

The next concern is whether the workloads are suitable (or not) for cloud deployment. Not every application is a suitable candidate for moving to a cloud computing environment. Whether or not a particular application is a good fit depends on a combination of the nature of the business functions it provides, the capacity characteristics it requires (some processing patterns will be more cost-effective than others in a pay-as-you-use model), and technical aspects of the application or its infrastructure requirements.

1.10 CLOUD ADOPTION

Business function that suits cloud deployment can be low-priority business applications, for example, business intelligence against very large databases, partner-facing project sites, and other low-priority services. Cloud favours traditional Web applications and interactive applications that comprise two or more data sources and services and services with low availability requirements and short life spans; for example, enterprise marketing campaigns need quick delivery of a promotion that can just as quickly be switched off. It is also helpful when high volume, low cost analytics and disaster recovery scenarios, business continuity, backup/recovery-based implementation are required. It is like a boon to one-time batch processing with limited security requirements, record retention, media distribution, and mature packaged offerings, like e-mail, collaboration infrastructure, collaborative business networks.

Based on technical characteristics, we can say that it is suitable for applications that are modular and loosely coupled; isolated workloads; single virtual appliance workloads and software development and testing; and pre-production systems. It gels well with R&D projects, prototyping to test new services, applications, and design models and applications that scale horizontally on small servers – that is, by adding more servers, rather than by increasing a server's computational capacity.

Applications that need significantly different levels of infrastructure throughout the day, such those used almost solely during the business day, should be deployed through cloud. Applications that need significantly different levels of infrastructure throughout the month, or that have seasonal demand, such as those used primarily during the end-of-the-quarter close or during a holiday shopping season, are the best examples of cloud deployments. Applications where demand is unknown in advance – for example, a Web start-up will need to support a spike in demand when it becomes popular, followed potentially by a reduction once some of the visitors turn away – can also be deployed using clouds.

It is not suitable for mission-critical and core business applications, transaction processing and applications that depend on sensitive data normally restricted to the organization, or requiring a high level of auditability and accountability as these process cannot share the high importance data, processing power, and hardware with the third party. Applications that run 24×7×365 with steady demand, applications that consume significant amounts of memory, including applications dependent on large in-memory caches, databases, or data sets are not suitable for cloud. Applications that take full advantage of multiple cores, such as those that do a significant amount of parallel processing, and thus benefit from many cores on a single server, are not recommended for cloud deployment.

It is not recommended for applications that require high-performance file system I/O needing high-bandwidth interserver communications, for example, highly distributed applications. Cloud does not work well with applications that scale vertically on single servers – that is, by increasing a server's computational capacity rather than adding more servers and applications dependent on third-party software, which does not have a virtualization or cloud aware licensing strategy.

1.11 CLOUD RUDIMENTS

Cloud delivers a software platform that will enable customer IT to build an Infrastructure-as-a-Service (IaaS) cloud. Cloud is built on the capabilities of existing virtualization management and physical server provisioning solutions to deliver application infrastructure to users that can be consumed in a self-service manner.

Cloud optimizes the usage of the physical and virtual infrastructure through intelligent resource allocation policies, and adds the ability to flex applications elastically based on demand. The high-level capabilities of any cloud include the following:

- **Resource Aggregation and Integration:** Cloud solution operates on top of existing virtualization management, physical server provisioning, and system management environments. It retrieves inventory information about machines and software templates from multiple locations, and aggregates this information into a central logical view of all resources in the infrastructure.
- **Application Services:** Rather than provide access to resources directly, cloud solutions' application 'Definitions' describes packages of machine capacity and software images that can be allocated by resource consumers. Applications can range from individual machines provisioned with an operating system image through to full multi-tier application environments that consist of collections of machines and software stacks provisioned in a specific order with network and storage dependencies handled through integration with third-party management tools. Application instances represent an agreement between the cloud provider and consumer to use capacity on a reservation or on-demand basis. Reservations allocate capacity in the resource inventory, guaranteeing that the capacity will be available to the consumer at some defined point in the future. On-demand allocations provide access to resources but do not guarantee availability. Reserved and on-demand capacity can be combined in an application, where a baseline of capacity can be elastically increased or decreased according to metrics and policies defined by the consumer.
- **Self-Service Portal:** An important principle of a cloud solution is to enable self-service access to resources with minimal IT involvement. It should support the notion of account owners signing up for contracts and then being able to delegate the use of the purchased capacity within their own groups or departments. Users can request machines

or entire multi-machine application environments and monitor and control them using a web-based self-service portal. The system will drive the workflows necessary to create the environment, and provide run-time environment management in order to support application elasticity.

- **Allocation Engine:** Dynamic Resource Management (DRM) is the automated allocation and reallocation of IT resources based on policies that express business demands and priorities. DRM is a key component of any cloud solution that maximizes the efficiency of the IaaS infrastructure. DRM policies should be applied both when initially placing applications onto machine resources and when selecting applications to migrate in order to preserve SLAs around application performance. Some of the allocation and migration strategies include advance reservation of resources, load-based placement and migration, application and resource topology constraints, energy usage optimization, etc. The use of sophisticated DRM helps to increase utilization of cloud resources, reduces overspending by effectively using existing resources, and saves costs in terms of operations, power, and cooling.
- **Reporting and Accounting:** In order to close the loop and determine how the cloud is behaving, metering information on resource allocations as well as actual usage is collected in an accounting database. The data is centrally available to create reports on inventory capacity, capacity allocated versus capacity used by contract, and usage-billing reports based on consumed resources.

The following are the *cloud features* that would help to bring in *agility* and *transparency* along with increase in the utilization of the existing resources at the datacenter of any customer.

- **Self-Service:** This feature presents an interface for separate authenticated end-users – via role-based access controls (RBAC) – to select options for deployment. It should have unique policy controls per tenant and user role, and the ability to present unique catalogues per user or group. The self-service portal is a web interface also accessible in other ways, such as through a mobile client, etc.
- **Dynamic Workload Management:** With cloud solution implemented, datacenters are enabled with automation and orchestration software that coordinates workflow requests from the service catalogue or self-service portal for provisioning virtual machines. Also each provisioned virtual machine is enabled with a life-cycle for deployment expiration which increases the efficiency of utilization of resources.
- **Resource Automation:** Using cloud solution, Admins or engineering team members of the datacenter could control the heterogeneous environment on a single pane. This feature establishes secure multi-tenancy, isolates virtual resources, and helps prevent contention in the load aware resource engine which intelligently does the workload packing or load balancing across hypervisors automatically.
- **Chargeback, Showback, and Metering:** Using this feature Admins could bring out the usage reports for cloud infrastructure service consumption and these usage reports serves as a basis for metering and billing system. Using this Admins will be able to understand if the virtual machines are attached with appropriate resources. Enabling *chargeback, showback,* and *metering* in any organization would bring in transparency to the business and environment for management to clearly see the usage and dollar value associated to it and take decision-making steps.

- **Open Architecture:** The cloud should be integrated with existing third-party products that are already installed in the datacenter. It should also be integrated to a public cloud for using additional resources and should be managed through a single cockpit. It is also possible to meter the public cloud resource usage.
- **Image Pools:** The cloud solution should have full blown service catalogue and support to most of the operating systems. It should be possible to vary the hardware configuration for the templates. It should also integrate with existing templates and images used by the development and testing teams.
- **Role-Based Access Administration:** The cloud solution should have the capability to integrate cleanly with any of the existing, LDAP, or other authentication and identity mechanisms. These features are crucial for providing secure multi-tenancy. This would also bring in security to the self-service portal.
- **Virtualization:** The cloud should extend support to virtualization layer. This implies that it should support most of the industry-proven hypervisors. This enables the Admins and engineering team of the datacenter to control them over a single pane.

1.11.1 *Cost Savings with Cloud*

Faster Time-to-Market (Missed Business Opportunity)

Deploying new application environments quickly and reliably can have a directly impact on competitiveness enabling organizations to take market share. The cloud will enable automated delivery of application environments exponentially faster than current practices.

With the cloud model, teams could be delivering fully configured, multi-component application environments to users in some minutes. This makes an immediate impact on user efficiency as well as eliminating much of the manual labor previously required of both the IT and application teams. In addition to this, ability to remove a (physical or virtual) will have similar performance, and once again allows that compute power to be available for other uses.

Public Cloud Interfaces

Cloud infrastructure with its policies should manage workload placement optimally by looking at several metrics. Cloud should also offer the capability to burst out to public cloud or internal resources when needed and cut off that link when done. The cloud should also be able to meter for the usage of the deployed instances in public cloud. Customer datacenter could use resources in public cloud for test and development environment if there are no resources available on the premise which will also help them to defer from the new hardware procurement.

Automated Scaling

The cloud solution should provide an out-of-box functionality to flex-up or flex-down an application instance or resource based on performance metrics and should also flex-up and flex-down an environment automatically or manually. The cloud solution should offer policies that can be customized to look at any metric and take action based on the threshold. These policies must be embedded in a service catalogue to monitor an application or the entire environment and flex-up or flex-down with more resources.

Business Transparency

Service Accounting helps to improve utilization of datacenter infrastructure with accurate visibility into the true costs of physical and virtualized workloads. It will enable decision makers to have full cost transparency and accountability for usage, metrics, roles and definitions. This would also help an Admin to understand whether a machine is equipped with right resources or not.

1.11.2 Benefits

Cloud brings lot of benefits for any enterprises. Let us explore these in brief here. They will be discussed in detail in the later chapter also.

- Increase agility on the IT datacenter resources and innovation.
- Enable self-service portal and thus ensure VM in less lead-times.
- SLAs are met as the VM lead-times and downtimes are significantly reduced.
- Trial and error configuration tests can be done at ease.
- Complete control over cloud usage for Admins.
- Scalability and flexibility allow the IaaS cloud to almost deliver the promise of unlimited IT services on demand.
- Pay for only what they use and are not charged when their service demands decrease.
- Significant reduction in the costs for IT datacenter.
- Private cloud enables dynamic sharing of the resources available in IT datacenter so that demands can be met cost-effectively.
- Considerable increase in the utilization of resources of IT datacenter.
- Increase in operational efficiency of the resources in the IT datacenter.
- Achieve a greener datacenter (server consolidation and virtualization enables over committed machines).
- Support for heterogeneous hardware vendors. Avoids Vendor Locking.

It will help the enterprises by

- Reducing the number of administrators required to manage a more diverse IT resource pool.
- Dramatic reduction in cycle times to provision new assets.
- Realization of an infrastructure 'pay-per-use' model.
- Reduction in planned capital spending and maintenance.
- Increased user satisfaction with IT services.
- Reduction in physical server count.
- Consolidation of enterprise application licenses.
- Flexibility to meet future demands on infrastructure goals that can be leveraged.
- Capacity on-demand (pre-provision, automate).

- Consolidated, streamline change control.
- End-to-end application provisioning.
- Allowing developers to provision development application environment autonomously.
- End-to-end performance measurement.
- Consumption-based charge back.
- Plan for active/active datacenter operations.
- Plan for increased datacenter density.
- Separate production and development networks.

1.12 SUMMARY

In this chapter, we explored cloud computing, its benefits, and its services. The chapter also gave deep insights into cloud computing models that are put into practice. The next chapter discusses the different types of cloud models and service platforms.

CHAPTER 2

2.1 INTRODUCTION

Cloud computing is an emerging style of computing where applications, data, and resources are provided to users as services over the Web. The services provided may be available globally, always on, low in cost, 'on demand', massively scalable, 'pay-as-you-grow'. Consumers of a service need to care only about what the service does for them, and not on how it is implemented. Cloud computing is a technology that allows users to access software applications, store information, develop and test new software, create virtual servers, draw on disparate IT resources, and more – all over the Internet (or other broad network).

Cloud computing is a model-driven methodology that provides configurable computing resources such as servers, networks, storage, and applications as and when required with minimum efforts over the Internet services. Cloud also indicates essential characteristics, delivery models, and deployment models.

This chapter visualizes several models for cloud computing, including private clouds (where the deployment is within the organization's firewall) and public clouds (where the application services and data are hosted by a third party outside the firewall). Consistent data availability and security is a critical success factor for any cloud deployment. Businesses need to ensure that data is adequately protected and can be restored in a timely fashion following any disruption event.

Clouds need a datacenter, but the aim of cloud computing is to eliminate the need to think about datacenters. A datacenter is a facility used to house computer systems and associated components, such as telecommunications and storage systems. It generally includes redundant or backup power supplies, redundant data communications connections, environmental controls (e.g., air conditioning, fire suppression), and security devices.

Datacenters are tied to locality, with specific components including redundant power supplies, redundant communications, environmental controls, security devices, etc.

Clouds are location-independent, providing abstracted versions of datacenter components that are not tied to a specific datacenter: virtual servers, virtual storage, virtual networking, etc. Reliability and redundancy comes from cloud providers using multiple datacenters, so clouds almost certainly span one or more datacenters, but themselves are not datacenters.

2.2 CLOUD CHARACTERISTICS

Cloud carries the basic infrastructure characteristics that are helpful to deploy cloud service in a fast and cost-effective way (Figure 2.1). The following characteristics set apart cloud from other computing techniques.

2.2.1 *On-Demand Service*

A consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service provider.

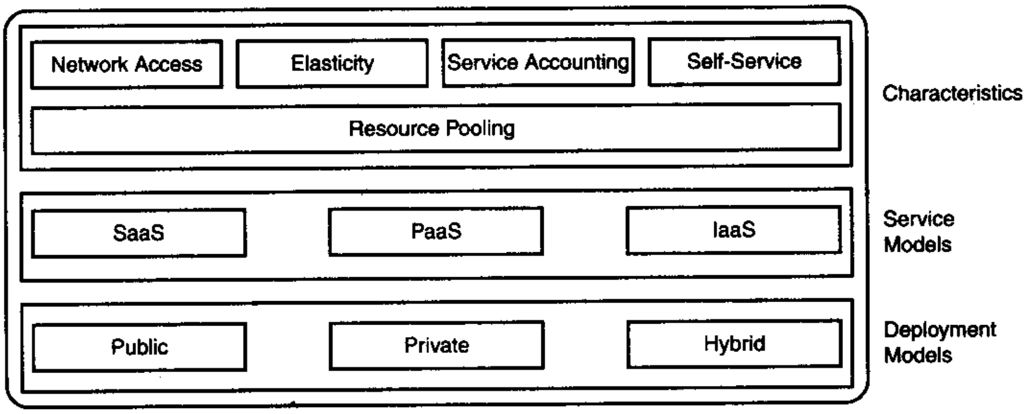

FIGURE 2.1 Cloud model.

2.2.2 *Ubiquitous Network Access*

Capabilities are available over the network and accessed through standard mechanisms that promote use by heterogeneous thin or thick client platforms (e.g., mobile phones, laptops and personal digital assistants [PDAs]).

2.2.3 *Location-Independent Resource Pooling (Multi-Tenant)*

The provider's computing resources are pooled to serve multiple customers using a multi-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to the demand. There is a sense of location-independence in that the customer generally has no control or knowledge about the location where the services are located (for example, country, state, or datacenter). Examples of resources include storage, processing, memory, network bandwidth, and virtual machines.

2.2.4 *Rapid Elasticity*

Capabilities can be rapidly and elastically provisioned, in some cases automatically, to quickly scale out and rapidly released to quickly scale in. To the consumer, the capabilities available for provisioning often appear to be unlimited and can be purchased in any quantity at any time.

2.3 MEASURED SERVICE

Cloud systems automatically control and optimize resource use by leveraging a metering capability at some level of abstraction appropriate to the type of service (e.g., storage, processing, bandwidth, and active user accounts). Resource usage can be monitored, controlled, and reported providing transparency for both the provider and the consumer of the utilized service.

The more standardization you are able to achieve within your infrastructure, the greater the economies of your operating expenses become. Similarly, the more you leverage virtualization within your IT infrastructure, the greater the economies of your capital expenditures will be. So, addressing both standardization and virtualization is the key to reduce infrastructure costs while still meeting the dynamic needs of the business.

Why are organizations taking action now to migrate to cloud computing?

- To derive the greatest flexibility and cost-reduction benefits from their cloud computing investments.
- To avoid vulnerability to costly problems and delays arising from a trial-and-error method of migrating workloads.
- To augment limited in-house resource or experience to rapidly develop an optimization roadmap to smoothly migrate workloads to a cloud computing environment.

Cloud vendors can address client's challenges by:

- Prioritizing workloads for cloud adoption based on business impact and risk.
- Maximizing business return by identifying applications that are well suited for cloud computing and have high business impact.
- Addressing problematic workloads to improve their propensity for cloud computing.
- Helping avoid costly implementation issues by identifying and addressing potential difficulties during migration.
- Mitigating the risk of costly implementation delays by identifying potential problems and addressing them before migration.
- Avoiding inadequate performance of highly complex and integrated workloads.
- Leveraging expertise to deliver an actionable roadmap to successfully migrate applications to a cloud computing environment.
- Accelerating your cloud initiatives.

2.3.1 Cost Factor

There are a number of reasons why cloud computing is popular with businesses. There is the cost aspect. By virtualizing your environment and standardizing it, you can deliver more services with fewer resources and drive up utilization. By adding automation, you can reduce labour cost giving you an additional cost benefit. This gives you a lot of flexibility because you can access cloud workloads services without thinking about the location and time of its execution. What this allows the organization to do then is to free up budget so that the money can be diverted to innovation and development of new capabilities rather than just keeping the lights on and running the IT enterprise.

The growing complexity of IT systems and soon a trillion connected things demand that sprawling processes become standardized services that are efficient, secure, and easy to access. A service management system will provide visibility, control, and automation across IT and business services to ensure consistent delivery. Self-service plus standardization will drive lower operational costs, unlock productivity, and ensure better security.

Cloud allows businesses to be smarter about how they deliver services. The first aspect of this is a self-service portal. This allows your end consumers to only see services they are allowed to have; however, it allows them to initiate the process of getting those services. Behind that service request, you could put either a very light or no-touch approval process, or you could put a more complex one in which you may need multiple levels of signature. This allows you to really fit what the business needs. In some cases where you have high security, you have high-level service-level agreements (SLAs) – you really want to be able to control how those services are distributed. In other cases, let's say you do not have an R&D development team and you want to be able to have as much flexibility as possible. This allows you to be very productive. It allows you to really make your infrastructure more dynamic, and get resources to the teams that really need them at any point in time.

Let's look at some of the major factors that are driving cloud computing economics. If you look at the infrastructure layer, first comes virtualization. By virtualizing workloads and being able to stack multiple workloads on a system to drive utilization up, you can lower your capital requirements. In a number of cases, businesses have hundreds, if not thousands, of physical servers and unless they have used virtualization and unless they are really driving that utilization, the utilization could be as low as 10 percent. So, in a lot of cases, organizations that use cloud computing are able to drive utilization up and either lower future capital requirements or even retire antiquated equipment and drive their costs down.

From a labour perspective, using a self-service portal allows your clients to help themselves. So there is less support and it makes the offering more available from a service perspective. In terms of automation, it takes tasks that are very manual and repeatable, and by automation then, it reduces your IT operations cost. In a development or test environment, you need multiple skills to get that environment to the end-user. You need operating system skills, middleware skills, database skills, and application skills. This allows you to define that environment as a repeatable, deployable resource, and it drives down your labour cost there. Of course, you need to standardize those workloads. Standardization has labour cost and quality benefits so that you can ensure consistency from environment to environment.

In many cases, you may want to use multiple models for different types of services that you want to deliver. Starting with private cloud services, the first model (which is also the most popular currently) is the private on-premise cloud. If the cloud is within the organization's datacenter, it is operated and managed by the organization itself.

The need to achieving cost optimization has also provided fertile ground for cloud computing. The cloud paradigm is an attempt to improve service delivery by applying engineering discipline and economies of scale in an Internet-inspired architecture.

Cloud computing can be an important new option in helping businesses optimize the IT expense equation while maintaining fast, high-quality service delivery.

2.3.2 *Benefits*

We can enjoy many benefits by adopting cloud:

- **Self-service capability:** Once somebody deploys the cloud services, they are capable of self-service. Now testing teams do not have to buy computing services as they can enjoy the same services over the cloud and it reduces the procurement process. Hence, they can concentrate on the testing services and efforts.

- **Resource availability:** It is the one of the most common benefit facilitated by virtualization. It also helps to track and leverage the resource pool under the same umbrella of resource units.
- **Operational efficiency:** Sometimes conventions and configurations followed by test and operation teams may differ from those followed by development teams. This can cause the application behaviour to be different from what was intended as well as delay services. The template-based approach, with its solution stacks of hardware, configurable applications, and operating systems, is more transparent and can help the teams to understand the environment better.
- **Hosted tools:** Due to these, the developers and testers need not install, configure, run, or maintain tools on their systems as they can log into the tools from any machine on the network maintaining the tools. Rather they can simply login to the tools and enjoy the services over the network.

These four benefits help the developers and testers to concentrate on their core work, retain focus, and concentrate more on their work without worrying about other jobs. This increases quality and productivity, and therefore, more developer innovation, increased test quality and coverage, etc. which are beneficial for an organization.

There are a number of major challenges developers face today in getting started and rolling out new applications and services faster. However, innovative new products and services are the lifeblood of rapidly growing companies. They represent a substantial portion of corporate sales and profits. In an environment of heightened competition, the inability to roll out new applications and services quickly means declining market share and lost revenue.

A growing application backlog leaves lines of business and end-users frustrated because they feel IT is a bottleneck and they look for ways to work around IT to roll out new products and services more quickly. Testing backlog is often very long, and a major factor in the delay of new application deployments.

A major reason testing takes so long is it takes weeks, on average, to set up application environments for test and QA as well as production. This is because of the time it takes to procure new hardware and software, and then schedule time with IT to configure and set up the systems. Configuration and setup are manual processes where errors are easily introduced. The average new application takes six to nine months to deploy, on average. This is caused by a number of factors ranging from poor governance to poor collaboration between business users and development to inflexible infrastructure and tools. Almost 30 percent of all defects are caused by wrongly configured test environments. This is a result of manual processes without any automation to replicate testing environment along with challenges organizations face in finding available resources to perform tests in order to move new applications into production. Test environments are seen as expensive and provide little real business value.

2.4 CLOUD DEPLOYMENT MODELS

Let's talk about cloud computing and the different types of cloud deployment models and different types of services that can be delivered using that model. Cloud computing is a style of computing in which business processes, application, data, and any type of IT resource can be provided as a service to users.

Cloud delivery models can be briefly classified into three types (Figure 2.2):

- **Public:** In a public cloud, a business rents the capability and they pay for what they use on-demand.
- **Private:** In private clouds, a business essentially turns its IT environment into a cloud and uses it to deliver services to their users.
- **Hybrid:** Hybrid clouds combine elements of public and private clouds.

A private cloud drives efficiency while retaining control and greater customization. Public clouds today are for processes deemed more easily standardized and a lower security risk. There some functions that already exhibit a high degree of standardization, that are more easily

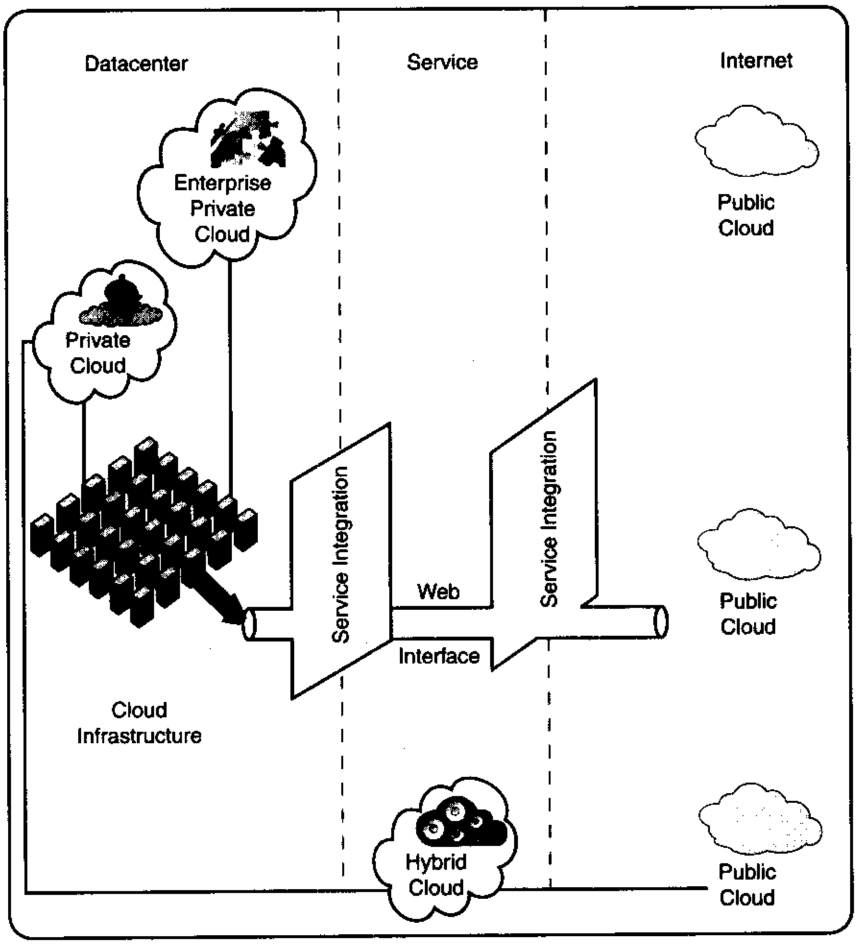

FIGURE 2.2 Private, public, and hybrid clouds.

moved to a public cloud – things such as search, e-commerce, and discreet business processes like sales force management.

There is not a one-size-fits-all model; in a number of cases, businesses may end up using all these models eventually, based on the business model for different services.

2.4.1 *Public Clouds*

Public cloud services are available to clients from a third-party service provider via the Internet. Public clouds provide an elastic, cost-effective means to deploy solutions and take care of deploying, managing, and securing the infrastructure. Companies can use it on demand, and with the pay-as-you-use option, it is much like utility consumption. Enterprises are able to offload commodity applications to third-party service providers (hosters).

The term 'public' does not mean

- That it is free, even though it can be free or fairly inexpensive to use.
- That a user's data is publically visible – public cloud vendors typically provide an access control mechanism for their users.

Public Clouds – Application Workloads

- Public facing Web pages.
- Public Wiki's, blogs, etc.
- Batch processing jobs with limited security requirements.
- Data intensive workloads.
- Software-as-a-Service applications.
- Online storage solutions.
- Online backup/restore solutions.
- Isolated workloads where latency between application components is not an issue.

Application Workloads not Ready for Public Cloud Today

Workloads that depend on sensitive data normally restricted to the organization are public today. Most companies are not ready to move their LDAP server into a public cloud because of sensitivity of employee information. Health care record – until security of cloud provider is well established – is another example. Some other examples include:

- Workloads composed of multiple, co-dependent services.
- High throughput online transaction processing.
- Workloads requiring a high level of auditability and accountability.
- Workloads based on third-party software that does not have a virtualization or cloud aware licensing strategy.
- Workloads requiring detailed chargeback or utilization measurement as required for capacity planning or departmental level billing.
- Workloads requiring flexibility and customization.

2.4.2 *Private Clouds*

Private clouds are deployments made inside the company's firewall (on-premise datacenters) and traditionally run by on-site servers. Private clouds offer some of the benefits of a public cloud computing environment, such as elastic on-demand capacity, self-service provisioning, and service-based access. They satisfy traditional requirements for greater control of the cloud infrastructure, improving security, and resiliency because user access and the networks used are restricted and designated.

Services in Private Cloud

This section highlights the services provided by private cloud and services consumed by public cloud specifically:

- Virtualization.
- Government and management.
- Multi-tenancy.
- Consistent deployment.
- Chargeback and pricing.
- Security and access control.

Consuming Services from Public Cloud

- Security and data privacy.
- Ease of access.
- Discovery of services.
- RESTful interface support.
- Lower cost.
- Speed and availability.

High 'Cost of Privacy'

Many experts believe that a private cloud implemented with internal hosting/running of the infrastructure makes it difficult to realize many key benefits of clouds, including:

- **Eliminating capital expenses and operating costs:** Ownership of the hardware or software eliminates the pay-per-use potential, as these must be upfront purchases. The full cost of operations must be shouldered, as there is no elasticity. If the private cloud hardware is sized for peak loads, there will be inefficient excess capacity. Otherwise, the owner faces complex procurement cycles.
- **Removing undifferentiated heavy lifting by offloading datacenter operations:** Utility pricing (for lower capital expenses and operating expenses) usually implies an outside vendor offering the on-demand services, and relies on the economies of multiple tenants sharing a larger pool of resources. These higher costs might be justified if the benefits of quicker and easier self-service provisioning and service-oriented access are large.

Private Clouds Provide more Control

In traditional security models, location implies ownership, which, in turn, implies control when security is location-specific. Then location, ownership, and control are aligned. Strong requirements for control and security usually drive a preference for a private cloud, where they own the cloud resources and control the location of those resources. For example, governments may not want their applications or data to reside outside certain borders. Clouds rely on virtualization, and in the public model, this loose coupling breaks the link between location and application, and this reduces the perceived ownership and control.

But control of information is not, in fact, dependent on total ownership or a fixed location. One example is public key encryption – the ownership of the key means control over the information without having to own the rest of the infrastructure. Control can be created over an untrusted infrastructure via a combination of encryption, contracts with service-level agreements, and by (contractually) imposing minimum security standards on the providers. Compliance is difficult outside traditional security models. As long as control through technology and contracts can be clearly demonstrated, it should be possible to make a public cloud computing environment as compliant and as secure as a privately owned facility. Auditors and regulators are continuously adapting to new technologies and business models.

There are two ends to the ownership spectrum – complete implementation ownership, and complete lack of ownership and control of implementation. There are many possible approaches in between, like partial control, shared ownership, etc. There are also different levels of limited access – specific departmental access, industry-only access, controlled partner access, etc.

2.4.3 Hybrid Clouds

A hybrid cloud is a combination of an interoperating public and private cloud. In this model, users typically outsource non-business-critical information and processing to the public cloud, while keeping business-critical services and data in their control. The hybrid model is used by both public and private clouds simultaneously, and is an intermediate step in the evolution process, providing businesses with an on-ramp from their current IT environment into the cloud.

It offers the best of both cloud worlds – the scale and convenience of a public cloud and the control and reliability of on-premises software and infrastructure – and lets them move fluidly between the two based on their needs. This model allows:

- Elasticity, which is the ability to scale capacity up or down in a matter of minutes, without owning the capital expense of the hardware or datacenter.
- Pay-as-you-go pricing.
- Network isolation and secure connectivity as if all the resources were in a privately owned datacenter.
- Gradually move to the public cloud configuration, replicate an entire datacenter, or move anywhere in between.

2.4.4 *Community Clouds*

A community cloud is controlled and used by a group of organizations that have shared interests, such as specific security requirements or a common mission. The members of the community share access to the data and applications in the cloud.

2.4.5 *Shared Private Cloud*

This is shared compute capacity with variable usage based pricing to business units that are based on service offerings, accounts datacenters and it requires an internal profit centre to take over or buy infrastructure made available through account consolidations.

2.4.6 *Dedicated Private Cloud*

Dedicated private cloud has IT Service Catalogue with dynamic provisioning. It depends on Standardized SO architectural assets that can be broadly deployed into new and existing accounts and is a lower cost model.

2.4.7 *Dynamic Private Cloud*

Dynamic private cloud allows client workloads to dynamically migrate to and from the compute cloud as needed. This model can be shared and dedicated. It delivers on the ultimate value of clouds. This is a very low management model with reliable SLAs and scalability.

2.4.8 *Cloud Models Impact*

Clouds will transform the IT industry. They will profoundly affect how we live and how businesses operate.

Cloud computing

- Provides massively scalable computing resources from anywhere.
- Simplifies service delivery.
- Provides rapid innovation.
- Provides dynamic platform for next generation datacenters.

Some say it is grids or utility computing or Software-as-a-Service, but it is all of those combined.

Public Clouds: Benefits

There are various ways to benefit from public clouds. Let us see some of the offering facilitated by public clouds:

- Lower barrier to entry/upfront investment.
- Offer self-service for rapid-start development.

- Deliver new pricing models for hardware, software, and service consumption.
- Increase or decrease capacity in minutes.
- Pursue new workloads and opportunities demo/sandbox, collaboration, prototypes.

Internal Private Clouds Drive Cost Savings

There are significant cost savings in implementing an internal private cloud versus a usual traditional infrastructure. With a traditional infrastructure, each server typically runs a single application and the hardware is sized to meet peak demands, which leads to very low average hardware utilization and high software costs due to the number of servers that are deployed and the lack of resource sharing. The internal private cloud uses virtualization on larger servers and leverages advanced service management capabilities to drive efficiency. Servers can be dynamically provisioned to adjust to workload changes and end-users can request the services they need through self-service portals, which drive automation.

Significant cost savings can be achieved by leveraging these capabilities to automate test and development environments. Automation drives down IT labour cost by automatically responding to changes in the environment and taking action before problems occur. Virtualization coupled with service management greatly improves server utilization and reduces software license costs since fewer machines need licenses. Automated provisioning and standardization allows systems to be provisioned in minutes by scripting the install process. In addition, end-users can now interface with IT through self-service portals to request services much like ATMs are leveraged to improve banking service. It can:

- Reduce IT labour cost by 50 percent in configuration, operations, management, and monitoring.
- Improve capital utilization by 75 percent, significantly reducing license costs.
- Lower administrative costs by 50 percent.
- Reduce end-user IT support costs by up to 40 percent.
- Reduce provisioning cycle times from weeks to minutes.
- Benefits of cloud economics with security within your firewall.
- Provide self-service for rapid-start development.
- Provide consistency of application environments.

2.4.9 Savings and Cost Metrics

Cloud computing's use of virtualization consolidates systems, which will drive reductions in hardware costs. This is often the initial appeal of funding virtualization projects.

Labour savings are even greater. Many companies still undertake the manual provisioning of IT systems, suffer long and costly delays while people wait for resources to become available, and distract highly skilled personnel from key project to focus on the mundane administration of systems. The automation of these tasks in a highly virtualized cloud environment can save significant labour costs while improving quality and productivity.

The total savings substantially off-set the small incremental increase in software costs that are usually necessary to deliver virtualization and the service management component that are elements of every cloud computing environment.

Cloud computing features two delivery models, private cloud computing and public cloud computing. Private cloud computing exists behind the firewall, while public cloud computing is accessed through the Internet. Cloud vendors believe that these three models – traditional IT, private cloud services, and public cloud services – will all co-exist as part of an overall strategy, based on application type and the business need that would dictate which model.

Hybrid clouds are services delivered to the end user that are composed of both private and public cloud computing elements.

2.4.10 *Commoditization in Cloud Computing*

When businesses started taking advantage of IT, the first organizations to computerize their business processes had significant gains over their competitors. As the IT field matured, the initial competitive benefits of computerization fell. Computerization then became a requirement just to stay on a level playing field. In essence, there is an increasing amount of IT that operates as a commodity.

For example, a paper products company needs a certain amount of unique IT to run its business and make it competitive. But it also runs a huge amount of commodity IT. The commodity technology takes time, money, people, and energy away from their business of producing quality paper products at a competitive price.

As executive management realizes it is operating a lot of commodity IT, which is not core to their competency, the debate shifts from whether cloud computing will take hold in the enterprise to a debate about how much of the organizational IT will be left internal, on-premise. IT functions should be evaluated, and a determination made as to which is a 'commodity' and which is not. Then determine where to place that function in the new IT organization.

2.5 SECURITY IN A PUBLIC CLOUD

Let us now discuss some of the security concerns that should be considered for the cloud deployments.

2.5.1 *Multi-Tenancy*

As long as the cloud provider builds its security to meet the higher-risk client, all of the lower-risk clients get better security than they would have normally. A bandage manufacturer may have a low risk of being a direct target of malfeasants, but a music label that is currently suing file sharers could have a high risk of being targeted by malfeasants. When both the bandage manufacturer and the music label use the same cloud (multi-tenancy), it is possible that attacks directed at the music label could affect the bandage manufacturer's infrastructure as well. So the cloud provider must design the security to meet the needs of the music label – and the bandage manufacture gets the benefits.

2.5.2 *Security Assessment*

Over time, organizations tend to relax their security posture. To combat a relaxation of security, the cloud provider should perform regular security assessments. The assessments should be done by someone who is experienced and able to identify issues and fix them.

The report should be provided to each client immediately after the assessment is performed so that the clients know the current state of the overall cloud's security.

2.5.3 *Shared Risk*

Sometimes, a cloud service provider may not be the cloud operator, but may be providing a value-added service on top of another cloud provider's service. For example, if a Software-as-a-Service (SaaS) provider needs infrastructure, it may make more sense to acquire that infrastructure from an Infrastructure-as-a-Service (IaaS) provider rather than building it. These cloud service provider tiers that are built by layering SaaS on top of IaaS, for example, can affect a cloud user's security. In this type of multi-tier service provider arrangement, each party shares the risk of security issues because the risk potentially affects all parties at all layers. This issue must be addressed by taking into consideration the architecture used by the cloud provider and working that information into the total risk mitigation plan.

2.5.4 *Staff Security Screening*

Most organizations employ contractors as part of their workforce. Cloud providers are no exception. As with regular employees, the contractors should go through a full background investigation comparable to the cloud user's own employees.

A cloud provider must be able to provide its policy on background checks and document that all of its employees have had a background check performed as per the policy. The contract between the user and cloud provider should bind the cloud provider to require the same level of due diligence with its contractors.

2.5.5 *Distributed Datacenters*

Disasters are a fact of life, and include hurricanes, tornadoes, landslides, earthquakes, and even fibre cuts.

In theory, a cloud-computing environment should be less prone to disasters because providers can provide an environment that is geographically distributed. But many organizations sign up for cloud computing services that are not geographically distributed, and therefore, they should require their provider to have a working and regularly tested disaster recovery plan, which includes SLAs.

Organizations that do contract for geographically diverse cloud services should test their cloud provider's ability to respond to a disaster on a regular basis.

2.5.6 *Physical Security*

Physical external threats should be analyzed carefully when choosing a cloud security provider. Do all of the cloud provider's facilities have the same levels of security? Are you being sold on the most secure facility with no guarantee that your data will actually reside there? Do the facilities have, at a minimum, a mantrap, card or biometric access, surveillance, an onsite guard, and a requirement that all guests be escorted and all non-guarded egress points be equipped with automatic alarms?

2.5.7 *Policies*

Any organization that says it has never had a security incident is either being deceptive or is unaware of the incidents it has had. It is, therefore, unrealistic to assume a cloud provider will never have an incident. Cloud providers should have incident response policies, and they should have procedures for every client that feed into their overall incident response plan.

2.5.8 *Coding*

All cloud providers still use in-house software, which may contain application bugs, so every organization should make sure that their cloud provider follows secure coding practices. Also, all codes should be written using a standard methodology that is documented and can be demonstrated to their customer.

2.5.9 *Data Leakage*

Data leakage has become one of the greatest organizational risks from a security standpoint. Virtually every government worldwide has regulations that mandate protections for certain data types.

The cloud provider should have the ability to map its policy to the security mandate users must comply with and discuss the issues. At a minimum, the data that falls under legislative mandates, or contractual obligation, should be encrypted while in flight and at rest. Further, an yearly risk assessment just on the data in question should be done to make sure the mitigations meet the need. The cloud provider also needs to have a policy that feeds into the security incident policy to deal with any data leakages that might happen.

2.6 PUBLIC VERSUS PRIVATE CLOUDS

A public cloud is a shared cloud computing infrastructure that anyone can access. It provides hardware and virtualization layers that are owned by the vendor and are shared between all customers. It is connected to the public Internet and presents an illusion of infinitely elastic resources.

Initially it does not require upfront capital investment in infrastructure. For consumption-based pricing, the user pays for resources used, allowing for capacity fluctuations over time.

Provisioning is applied through simple Web interface for self-service provisioning of infrastructure capacity. Potentially significant cost savings are possible from providers' economies of scale. Operating costs for the cloud are absorbed in the usage-based pricing. Separate provider has to be found (and paid for) to maintain the computing stack. Users have no say in SLAs or contractual terms and conditions. Sensitive data is shared beyond the corporate firewall. Distance may pose challenges with access performance and user application content for geographic locations. Support for operating system and application stacks may not address the needs of the business.

A private cloud is a cloud computing infrastructure owned by a single party. It provides hardware and virtualization layers that are owned by, or reserved for the business. It, therefore, presents an elastic but finite resource and may or may not be connected to the public Internet.

2.7 CLOUD INFRASTRUCTURE SELF-SERVICE

The cloud infrastructure has to be provisioned and paid for up-front in private clouds. Users pay for resources as used, allowing for capacity fluctuations over time. Self-service provisioning of infrastructure capacity is only possible up to a point in private clouds. Standard capacity planning and purchasing processes are required for major increases. For a large, enterprise-wide solution, some cost savings are possible from providers' economies of scale. The enterprise maintains ongoing operating costs for the cloud, and the cloud vendor may offer a fully managed service (for a price). SLAs and contractual terms and conditions are negotiable between the cloud vendors and customers to meet specific requirements. All data and secure information remains behind the corporate firewall and the option exists for close proximity to non-cloud datacenter resources or to offices if required for performance reasons for geographic locality. Private clouds can be designed for specific operating systems, applications, and use cases, unique to the business.

There is no clear 'right answer', and the choice of cloud model will depend on the application requirements. For example, a public cloud could be ideally suited for development and testing environments, where the ability to provision and decommission capacity at short notice is the primary consideration, while the requirements on SLAs are not particularly strict. Conversely, a private cloud could be more suitable for a production application where the capacity fluctuations are well understood, but security concerns are high.

Cloud computing employs a structured technique to holistically leverage IT industry best practices to uncover areas of relative strength and weakness across multiple IT domains (strategic alignment, computing system and storage, applications and data, processes, organization, finance/environment, and network) to determine readiness for a cloud computing deployment.

Infrastructure strategy and planning for cloud computing strategy gears the clients who are looking for assistance in understanding the business value that the cloud computing model can bring. It is designed to help the clients evaluate their readiness for cloud computing and possible cloud computing uses within their infrastructure. The goal is to develop a high-level vision strategy, value case, and roadmap for cloud computing.

Infrastructure strategy and planning for cloud computing employs a structured technique to holistically leverage IT industry best practices to uncover areas of relative strength and weakness across multiple IT domains to determine readiness for a cloud computing deployment. It is a business and IT executive initiatives to identify where and how cloud computing can drive business value.

2.7.1 *Infrastructure Strategy and Planning Features*

The strategy and planning has three major features:

- Assessment of the current environment to determine strengths, gaps, and readiness.
- Development of the value proposition for cloud computing in the enterprise.
- Strategy, planning, and roadmap to successfully implement the selected cloud delivery model.

Cloud-based systems have brought a new, scalable application delivery service model to the market. Cloud services promise to help reduce capital and operational costs while providing higher service levels. However, cloud services rely heavily on keeping the data and applications they are managing available at all times, and to restore operations quickly following any type of data disaster (database corruption, virus attack, hardware failure,local/regional disaster).

Cloud administrators need to ensure that a minimum of data is at risk by performing backups as frequently as possible, to meet stringent recovery point objectives. And downtime must be limited as well following an outage to meet strict recovery time objectives.

Emerging model where users can have access to applications or compute resources from anywhere with their connected devices through a simplified UI are best suitable alternatives for ease of use. Applications reside in massively scalable datacenters where compute resources can be dynamically provisioned and shared to achieve significant economies of scale. The 'pay-as-you-go' usage model enables users and companies to predict and manage expenses, reduce costs, and simplify operations better.

2.7.2 *The Path to Cloud Computing*

The path from simple virtualization to cloud computing occurs in five somewhat distinct stages.

Stage 1: Server Virtualization

Companies usually start virtualization as a consolidation attempt. The focal point tends to be on reducing capital expenses (like server, storage, and networks), reducing energy costs, and perhaps avoiding or delaying a datacenter build-out or move.

Stage 2: Distributed Virtualization

Once companies start down the virtualization way, and start to achieve capital expense improvements (like server, storage, and networks), the next focus tends to be on elasticity, operational improvements, rapidity, and organizing downtime more efficiently.

Stage 3: Private Cloud

Once processes are designed for alacrity and standards are in place to enable broad automation, the company is ready to look at introducing self-service capabilities based on the virtualization architecture.

Stage 4: Hybrid Cloud

Private clouds will not be the only answer for any enterprise. The self-service portals and interface introduced by private clouds should enable IT enterprises to leverage public cloud services when they make logic without affecting end users.

Stage 5: Public Cloud

Virtualization is not the must thing or are not the stepping stones before companies use public cloud services. Actually, some companies will attempt with cloud in the public cloud arena first, and use their lessons to establish private clouds for their enterprises.

2.8 SUMMARY

We have discussed several models for cloud computing, including private clouds (where the deployment is within the organization's firewall) and public clouds (where the application services and data are hosted by a third party outside the firewall). Consistent data availability and security is a critical success factor for any cloud deployment. Businesses need to ensure that data is adequately protected and can be restored in a timely fashion following any disruption.

CHAPTER 3

CLOUD AS A SERVICE

3.1 INTRODUCTION

In today's economy, many businesses are faced with the challenge of 'taking cost out' of their infrastructure while continuing to deliver new, innovative business services – basically they need to 'do more with less'. These days, many businesses face the necessities of a fast change to their IT infrastructure to manage requests during peak times. Organizations are dealing with IT resource optimization and lowering cost, and are looking for a way to manage these resources to meet such requirement. Also, they are trying to add rental-style capability to IT resource usage.

Cloud computing defines a new way to manage IT resources enabling self-service provisioning of IT resources, metering-style accounting based on use/time, automation of IT management in a standard process environment.

Cloud computing is a user experience and a business model. It is an emerging style of computing in which applications, data, and IT resources are provided as services to users over the network. Cloud computing is also an infrastructure management methodology. It is a way of managing large numbers of highly virtualized resources such that, from a management perspective, they resemble a single large resource, which can then be used to deliver services.

This chapter visualizes the different cloud models with respect to services. It also takes into account what service is all about and the different type of infrastructure services that can be offered as cloud as a service.

Common attributes of a cloud infrastructure are defined as:

- **Flexible pricing,** which means utility pricing, variable payments, pay-by-consumption; subscription models make pricing of IT services more flexible.
- **Elastic scaling,** which means resources scale up and down by large factors as the demand changes.
- **Rapid provisioning,** which means IT and network capacity and capabilities are – ideally automatically – rapidly provisioned using Internet standards without transferring ownership of resources.
- **Standardized offerings,** which mean uniform offerings readily available from a services catalogue on a metered basis.

There are two primary levers to achieve cost optimization – operating expense (Op-ex) and capital expense (Cap-ex) – and for many businesses, it is not just a question of lowering costs. It is also important to strike the right balance between Op-ex and Cap-Ex.

When you look at the different cloud types, the common terminology that comes up is 'as a service', with infrastructure as a service being the most basic type of service. It includes compute power, storage, and file systems as a service. At the next level, you have platform as a service, where a compute platform or middleware is provided.

At the next level, this is software as a service, and this is where you take software capability that would typically be in a package – like customer relationship management or e-commerce – and you deliver that as a service. At the highest level is business process as a service. This is where a business can take a function that it considers to be a commodity, and not a differentiator, and just completely outsource it and buy it as a service. So cloud computing has really taken a hold because of the fact that you take virtualization, standardization, and automation, and drive this automated delivery of services at a reduced cost.

Cloud computing introduces the concept of 'IT-as-a-Service'. To support this service, the cloud infrastructure must deliver:

- **Abstraction:** Alleviate IT consumers from the operations of applications, allowing end-users to focus instead on the execution of high-value activities.
- **Virtualization:** Access to business services on-demand independent of location and resource constraints.
- **Dynamic allocation:** Dynamically provisions, configures, reconfigures, and de-provisions IT capability as and when needed, transparently and seamlessly.
- **Data management:** Fast, secure, reliable data access, and mobility, with integrated data protection and recovery management.

Because all data reside on the same shared storage systems, effective and efficient data and storage management become critical in a cloud deployment.

3.2 GAMUT OF CLOUD SOLUTIONS

Even within the cloud computing space, there is a spectrum of offering types. There are five commonly used categories.

- **Platform-as-a-Service (PaaS):** This is the provisioning of hardware and OS, frameworks and database, for which developers write custom applications. There will be restrictions on the type of software they can write, offset by built-in application scalability.
- **Software-as-a-Service (SaaS):** This is the provisioning of hardware, OS, and special-purpose software made available through the Internet.
- **Infrastructure-as-a-Service (IaaS):** This is the provisioning of hardware or virtual computers where the organization has control over the OS, thereby allowing the execution of arbitrary software.
- **Storage-as-a-Service (SaaS):** This is the provisioning of database-like services, billed on a utility computing basis, for example, per gigabyte per month.
- **Desktop-as-a-Service (DaaS):** This is the provisioning of the desktop environment either within a browser or as a terminal server.

The distinction between the five categories of cloud offering is not necessarily clear-cut. In particular, the transition from Infrastructure-as-a-Service to Platform-as-a-Service is a very gradual one.

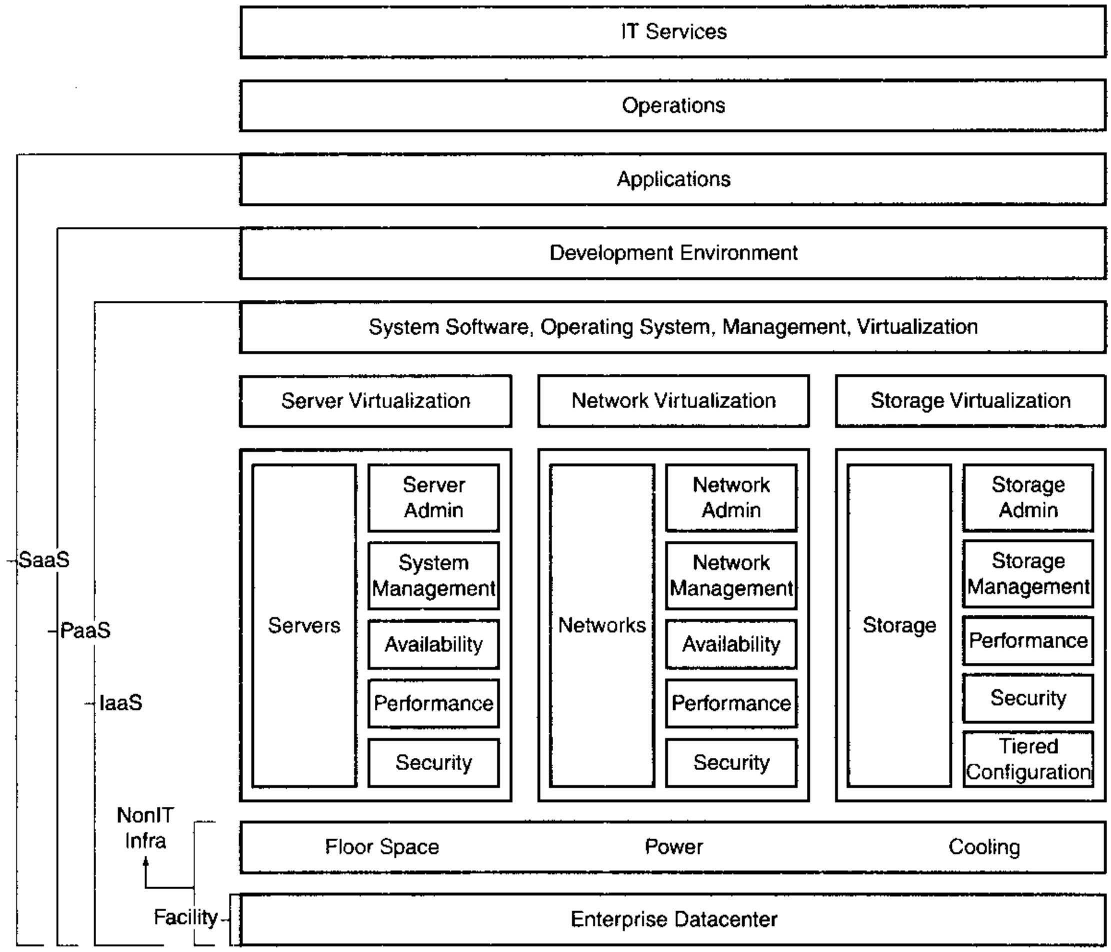

FIGURE 3.1 Cloud@datacenter.

3.2.1 Platform-as-a-Service

Instead of just offering applications over the Web in the form of Software-as-a-Service (SaaS), PaaS public cloud players are actually offering an entire Platform-as-a-Service (PaaS). They provide the foundation to build highly scalable and robust Web-based applications in the same way that the traditional operating systems like Windows and Linux have done in the past for software developers. What is very different about this model is that no longer is the platform itself 'sold' to the customer who is then responsible for running and maintaining it. In this model, it is this very operational capability of the platform hosting that is of primary value here (and that is how such platforms are typically billed). This has far-reaching implications to both the business models of PaaS vendors as well as their customers. One can use private clouds to speed application deployments for fast deployment in minutes. It will help tracking usage for the chargeback and will give the option of cost effective and secure appliance.

In order to optimize deployments, many organizations are looking to extend SOA to cloud services.

Cloud capabilities can improve the productivity of your development and test teams to roll out new applications and SOA services faster and reduce application backlog. It provides a catalogue of virtual images, and patterns all ready for immediate use. Patterns define a cluster of servers working together.

PaaS saves costs by reducing upfront software licensing and infrastructure costs, and by reducing ongoing operational costs for development, testing, and hosting environments.

PaaS significantly improves development productivity by removing the challenges of integration with services such as database, middleware, web frameworks, security, and virtualization. Software development and delivery times are shortened since software development and testing are performed on a single PaaS platform. There is no need to maintain separate development and test environments.

PaaS fosters collaboration among developers and also simplifies software project management. This is especially beneficial to enterprises that have outsourced their software development.

There is a challenge for tight binding of the applications with the platform which makes portability across vendors extremely difficult. PaaS in general is still maturing, and the full benefits of componentization and collaboration between services is still to be demonstrated. PaaS offerings lack the functionality needed for converting legacy applications into full fledged cloud services.

SaaS, PaaS, and IaaS suit different target audiences. SaaS is intended to simplify the provision of specific business services. PaaS provides a software development environment that enables rapid deployment of new applications. IaaS provides a managed environment into which existing applications and services can be migrated to reduce operational costs.

3.2.2 *Software-as-a-Service*

SaaS saves costs by removing the effort of development, maintenance, and delivery of software; eliminating up-front software licensing and infrastructure costs; and reducing ongoing operational costs for support, maintenance, and administration.

The time to build and deploy a new service is much shorter than for traditional software development. By transferring the management and software support to a vendor, internal IT staff can focus more on higher-value activities.

Applications that require extensive customization are not good candidates for SaaS. Typically, this includes most complex core business applications that will not be the best suit for SaaS.

There are also issues involved in moving to SaaS. Moving applications to the Internet cloud might require upgrades to the local network infrastructure to handle an increase in network bandwidth usage. Normally only one version of the software platform will be provided. Therefore, businesses are obliged to upgrade to the latest software versions on the vendor's schedule. This could introduce compatibility problems between different vendor offerings.

3.2.3 *Infrastructure-as-a-Service*

IaaS saves costs by eliminating the need to over-provision computing resources to be able to handle peaks in demand. Resources dynamically scale up and down as required, reducing capital expenditure on infrastructure and ongoing operational costs for support, maintenance, and administration. Organizations can massively increase their datacenter resources without significantly increasing the number of people needed to support it (Figure 3.2).

The time required to provision new infrastructure resources is reduced from typically months to just minutes – the time required to add the requirements to an online shopping cart, submit it, and have it approved. IaaS platforms are generally open platforms, supporting a wide range of operating systems and frameworks. This minimizes the risk of vendor lock-in.

Infrastructure resources are leased on a pay-as-you-go basis, according to the hours of usage. Applications that need to run 24×7 may not be cost-effective. To benefit from the dynamic scaling capabilities, applications have to be designed to scale and execute on the vendor's infrastructure. There can be integration challenges with third-party software packages. This should improve over time, however, as and when independent software vendors (ISVs) adopt cloud licensing models and offer standardized APIs to their products.

SaaS is considered to be considerably more mature as a cloud offering than PaaS or IaaS. Even then, it is mainly Small & Medium Businesses that have adopted cloud services. Adoption by the larger enterprises is still extremely low. A perception of cloud services as a high-risk technology option has led the large organizations to restrict the use of cloud services

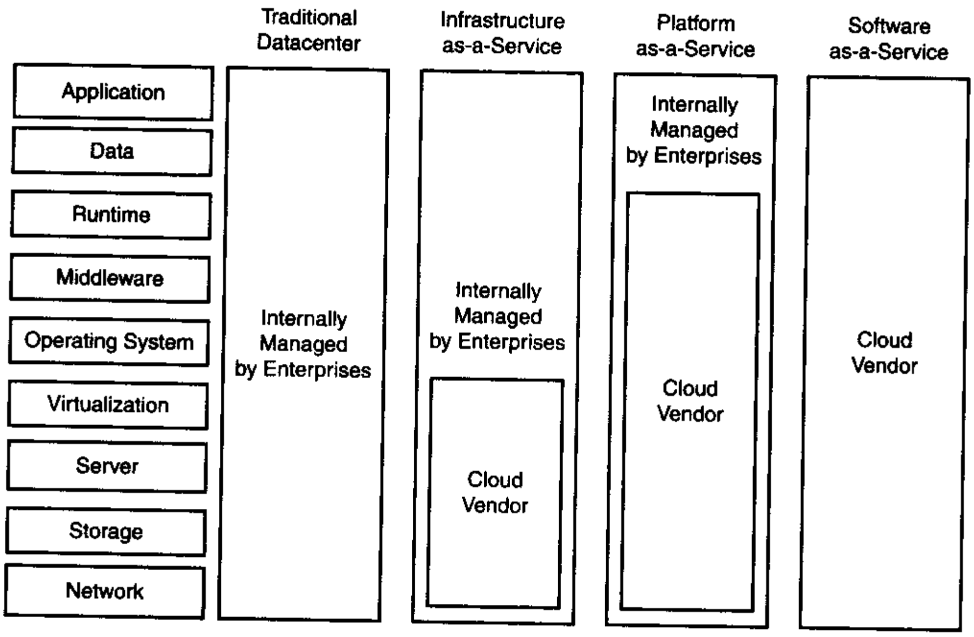

FIGURE 3.2 Cloud taxonomy.

to a limited number of projects. PaaS is a more sophisticated service platform, and is still an emerging product. It will need to stabilize and mature before developers can use it for extensive building of new SaaS applications. For IaaS, the entry of large vendors such as Amazon and other cloud vendors is driving up the maturity of the offering rather quickly, with Amazon leading efforts to define the IaaS market.

There are now hundreds of vendors offering some flavour of cloud computing. Other vendors have attached themselves to the 'Cloud bandwagon' by providing ancillary services to available offerings.

3.3 PRINCIPAL TECHNOLOGIES

The key to being able to provide the dynamic cloud infrastructure is the virtualization layer that sits between the cloud instances and the physical hardware it runs on. The platform virtualization software – the hypervisor – allows multiple operating system instances to run as guests on the same server.

The main drivers for cloud computing are *cost, agility*, and *time to market*. By building cloud infrastructures using any cloud orchestrator and provisioning engine one can realize cost savings and improve time to market. This sits on top of the virtualization layer working on network, server, and storage. It is a layer of software that (a) interacts with multiple servers, (b) enables IT departments to pool resources together across servers, and (c) defines standardized tiers of services called *virtual compute centres*. This helps break down infrastructure silos and drives sharing of infrastructure.

Using cloud orchestrator and provisioning engine, IT departments can define organizations and on-board users that can share the underlying cloud infrastructure in a secure multi-tenant fashion. IT can then create standardized collections of VMs and set policies on how users can use these VMs. Users can login into orchestrator and self-provision respective workloads which IT has setup already. This effectively removes IT involvement each time users require infrastructure-enabling agility and faster time to market for applications.

These orchestrators also transport with the API which allow cloud administrator and users to interrelate with the cloud infrastructure in a systematic way. In addition, cloud orchestrator and provisioning engine allows writing workflows to automate creation of cloud infrastructure.

The increased pooling and sharing of resources, self-provisioning, and increased automation deliver greater cost savings in IT infrastructure, agility, and faster time to market for applications. One can deliver cloud benefits for today's applications and for applications developed in the future as cloud orchestrator and provisioning engine builds the top of the hypervisor layer.

Virtualization is the foundation for cloud. It consists of physical hardware with hypervisor layered on top of it. Cloud orchestrator and provisioning engine consists of one or more cells that communicate with a single database and offer a web portal. Using the web portal, cloud administrators create cloud infrastructure resources and users' self-provision cloud infrastructure resources in a secure multi-tenant fashion, thus enabling Infrastructure-as-a-Service (IaaS).

Chargeback and metering is a key piece of the on-premise cloud solution. This server talks to own database, server database, and cloud orchestrator and provisioning engine databases and allows to associate costs with the cloud and generate usage and billing reports. This should also integrate with workflow systems, LDAPs, approval process, etc. to provide the lifecycle management of the cloud environment.

3.4 CLOUD STRATEGY

This section provides a high-level guidance to define the cloud strategy and the artefacts that capture the architecture of a cloud-enabled application. These architectural artefacts are meant for the implementation planning phase of cloud, enabling an application. Only the high-level architecture of the system that is to be cloud-enabled is captured in these artefacts.

The implementation planning phase of cloud enables an application lying between the business strategy definition for the adoption of cloud and the design, development, and implementation phases of the application that is being implemented to be offered on the cloud platform. It takes care of linking the business strategy that is defined for a business to adopt a cloud-based strategy and the IT requirements for the applications on the cloud that are needed to support this strategy. So, this critical piece of the implementation planning translates the business intent to a set of IT requirements for the cloud-based application, deriving the high-level structures of the cloud-based application and defining a roadmap for the implementation of the application.

So, the primary input for the cloud implementation planning phase comes from the cloud strategy for the business that is driving the cloud-based implementation of one or more applications on the cloud.

The key steps in cloud implementation planning are as follows:

- Understand cloud strategy.
- Define cloud application requirements.
- Assess cloud readiness.
- Define high-level cloud architecture.
- Identifying change management requirements.
- Develop roadmap and implementation plan.

Of these phases, the artefacts defined here relate to the phase of 'Defining the high level cloud architecture'. In this phase, the high-level structure of the cloud-based application is defined from the inputs of the previous step, the 'Define cloud application requirements'.

The first step in the architecture development is the usage of existing asset analysis to understand which of the components, including components of the existing application, can be used for building the application and how they can be leveraged in the cloud environment. The next step is to derive the high-level structure of the solution from the IT requirements for cloud in the form of the following artefacts. The non-functional characteristics that the application must support and deliver are captured in the artefact 'Non-Functional Requirements'.

Infrastructure strategy and planning for cloud computing helps you develop a cloud strategy, plan, and roadmap:

- Business and IT executive workshop to identify where and how cloud computing can drive business value.
- Develop the value proposition for cloud computing in the enterprise.
- Identify priority of workloads to migrate to cloud.
- Assess the current environment to determine strengths, gaps, and readiness.
- Strategy, plan and roadmap to successfully implement the selected cloud.
- Analyse cloud computing opportunity.
- Analyse IT environment and capability gap.
- Assess cloud readiness.
- Develop high-level cloud roadmap and value proposition.

This helps to deploy the cloud deployment with following benefits:

- **Reduced risk and faster deployment:** It leverages cloud vendor assets, skills, and experience to reduce risk. It accelerates development and implementation by identifying the gaps, activities, and risks and defines mitigation strategies within an implementation roadmap.
- **Improve service:** It identifies the optimal delivery model mix and prioritizes the workloads to migrate to cloud to achieve your business and IT objectives.
- **Lower cost:** It identifies opportunities to reduce capital and operating expense across the infrastructure.

3.5 CLOUD DESIGN AND IMPLEMENTATION USING SOA

Service-Oriented Architecture (SOA) is a very useful architectural style for implementing applications in the cloud. Adoption of SOA would provide the best way to leverage and consume the application services provided by the cloud. A cloud-based application consists of many granular coarse-grained services offered on the cloud. These cloud offerings may in turn integrate and leverage services and systems from different environments. The coarse-grained services that the cloud offerings leverage can come from traditional IT environment in the form of standard services already provided and are internal to the cloud.

Standardization across these different environments is not possible, and hence, they mostly consist of heterogeneous environments. So a cloud service offering should be based on open standards that can be consumed and leveraged by this environment. Language- and platform-independent services can be provided using standards-based, platform-agnostic SOA architecture, with support for appropriate industry and technology standards.

The cloud-based application services consist of a coherent set of business processes that are aligned to the business boundaries, provide a coherent, integrated set of operations that adheres to the business intent and provides value to the users. They also comprise the consumable interfaces whether they are user interfaces on different devices, coarse-grained Web service interfaces, feeds, or widgets.

Map the cloud application services to the business processes that they consist of. These business processes could be at different levels. Drill down the processes to the levels that make business sense.

These processes, along with business strategy, the modular business alignment model, and the process scenarios to be supported form the input to Service-Oriented Modeling and Architecture (SOMA) methodology. SOMA is applied with a meet-in-the-middle approach. The business processes and business strategy along with business goals stated are used to arrive at the service portfolio using SOMA. While deriving the services, a bottoms-up approach is also taken, taking into consideration the existing assets, which consist of the existing application components and services available on the public clouds and internal to the cloud as well as industry cloud component maps.

3.5.1 *Architecture Overview*

The purpose of the architecture overview artefact in a cloud-based implementation is to communicate to the sponsor and external stakeholders a conceptual understanding of the architectural goals of the cloud implementation. It offers a layered conceptual model of the application services to be cloud-enabled and provides a high-level vision of the cloud architecture and its scope to developers. It is easy to explore and evaluate alternative architectural options for the cloud implementation. This stage enables early recognition and validation of the implications of the cloud-based architectural approach and facilitates effective communication between different communities of stakeholders and developers.

This artefact provides an overview of the main conceptual elements of the cloud implementation and relationships within and outside the cloud infrastructure, which may include other cloud and non-cloud environments, candidate offerings, cloud components, nodes, connections, data stores, users, external systems, and technical components to support requirements. As such, it represents the governing ideas and candidate building blocks of the cloud implementation.

This artefact can also provide key stakeholders with the first high-level view of the cloud architecture landscape of their transformed cloud-based implementation.

Typically, the artefact is produced over multiple iterations and as the project moves through the solution definition and design phases, the conceptual models get clearer and this document is kept up-to-date and governed to always have the conceptual elements current.

An architecture summary depiction represents the governing ideas and candidate building blocks of a cloud-based offering and enterprise architecture. It also provides an overview of the main conceptual elements and relationships in the architecture. The main purpose of such a depiction is to communicate a simple, brief, clear, and understandable overview of the target IT system.

At the enterprise-level, an architecture summary depiction is often produced as part of an overall IT strategy to move to a cloud-based offering model. In this instance, it is used to describe the vision of the business and IT capabilities required by an organization that needs the offering to be hosted on the cloud. It provides an overview of the main offerings and the relationships with other offerings, external systems, components, nodes, connections, data stores, users, external systems, and a definition of the key characteristics and requirements.

At a system- and component-level, the architecture summary depiction is developed very early in the project (possibly pre-proposal), and influences the initial cloud hosting vision with the component model and operational model. It is intended that design commitments be based on this conceptual overview as the (more detailed) component model and operational model are developed and validated. Subsequently, the component model and the operational model are the primary models used for implementation activities, while their compliance with the architecture summary depiction is maintained continually. Changes to the architecture summary depiction are made using a governance process.

For most infrastructure cloud engagements, the project scope is something less than the client's enterprise architecture. Therefore, the depiction will represent the governing ideas and candidate building blocks of the offerings that are to be hosted on the cloud and represents the focus of the engagement. There can be, then, multiple views of the cloud-based IT environment represented by architecture summary depiction: the current and future view, views by waves of transformation or views by types of clouds, for each solution alternative.

3.6 CONCEPTUAL CLOUD MODEL

So far, we have discussed the different service paradigms for cloud deployment. Now let us discuss the conceptual cloud model based on cloud services.

The conceptual cloud model describes the structure of the cloud-based services as a system in terms of its software components with their responsibilities, interfaces, relationships, and the way they collaborate to deliver the required functionality. The cloud component model for implementation planning is specified at the conceptual level.

The highest level of the conceptual model is the set of cloud-based offerings that make up the cloud-based business solution. These highest-level conceptual components are derived based on the business intent and business functionality. These conceptual components, which form the offerings, align tightly to the business intent that initiated the creation of the cloud-based offerings.

The conceptual offerings that provide the cloud-based implementation of the solution can be further broken down and depicted in a layered composition. The next level of conceptual component model broken down below the offering consists of the following elements:

- High-level service components that form the services provided by the offerings.
- The resources that support the cloud services.
- The technical components that provide the technical underpinnings of the cloud service and support the non-functional needs of the cloud application.
- External and internal services that are leveraged by the cloud application services.

The high-level conceptual component model is the main artefact that provides an abstract view of the design of a cloud application to business stakeholders. This abstract view describes how the business needs will be met by the cloud application components without delving into their technical details. Components identified can be decomposed into further layered conceptual component structures to convey further details of the respective components (Figure 3.3).

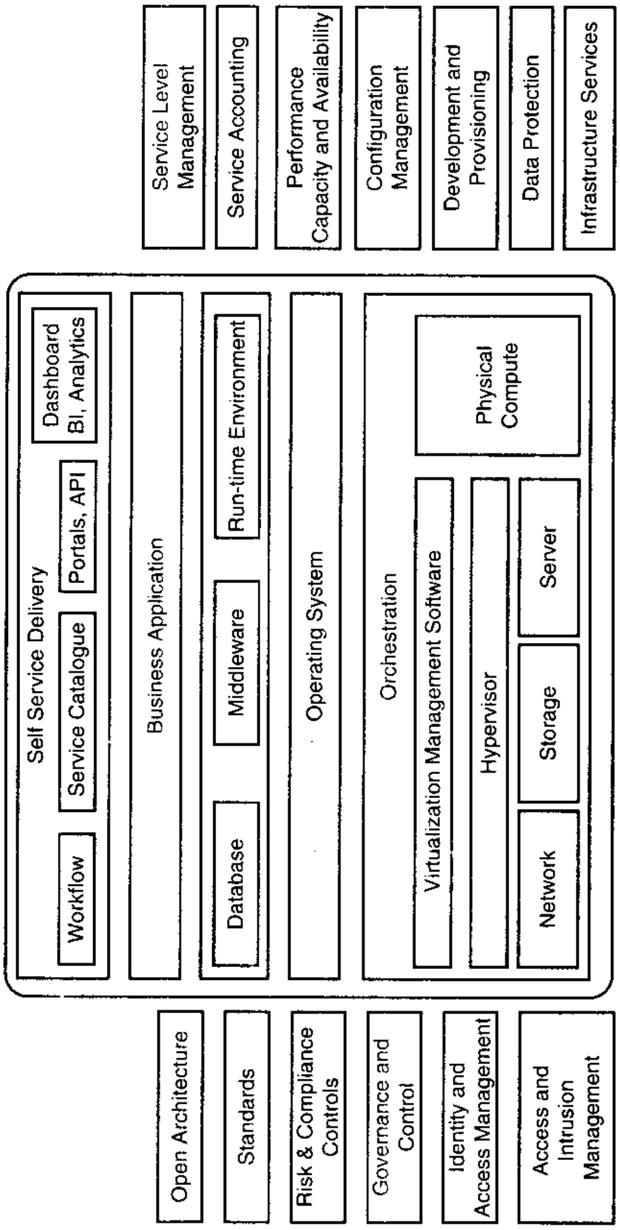

FIGURE 3.3 Cloud model.

The cloud conceptual component model should contain the following elements:

- The conceptual structure of the cloud application.
- The dynamic interactions and dependencies between various conceptual components.
- The components that comprise the cloud services provided by the application, each of which may be made up of sub-components.

At this level, the conceptual components are not converted to physical components but retained at the conceptual level only.

3.6.1 *Cloud Application Security and Privacy Principles*

The cloud application security principles contain the high-level guidance to cover the security and privacy characteristics identified in the Information Asset Profile. This is a summary document of the information asset required for the architecture. Cloud application security principles deliverable contains a number of security principles that occur in a typical application. Typically, quite a number of the characteristics identified in the information asset profile effort will relate to existing enterprise security elements in the client.

In addition to the elements from the existing enterprise security program, consider the following areas that tend to change as part of moving to the cloud environment.

3.6.2 *Governance*

How will decisions like different SLAs be made between cloud vendors and cloud customers for the cloud environment given the new stakeholders? These decisions are typically made by either another infrastructure group within the client IT organization or an external vendor who provides cloud infrastructure management for either a managed private cloud or public cloud environment.

Authentication and Access Control

As an application moves to the cloud, the methods used to authenticate and authorize users require examination. Existing methods may require additional support to work in the cloud environment, such as providing connectivity to an existing server, and may require significant extension to accommodate the requirements of an external vendor providing infrastructure support.

Data Protection

With the move to a cloud environment, the existing methods of data protection – protecting the data from disclosure or modification both on the network and on a storage medium – are likely to require change with the move to virtualized storage and with the introduction of additional infrastructure administrators, often from a vendor organization.

Logging and Alerting

Logging is the ability to tie actions to an individual. Alerting is the method of recognizing activities that may indicate a malicious act and bringing those to the attention of the security staff. With the move to a cloud infrastructure, with different network and additional administrators, often from an external vendor, tying actions to specific individuals requires additional attention.

The focus in this work product is to provide the broad guidance as to the security and privacy elements that must be present in and around the cloud application.

3.7 CLOUD SERVICE DEFINED

This section highlights the aspects of service, its scope, and cloud-based different platform integration and deployment services.

3.7.1 Service Definitions

Let us begin with a cursory glance on what 'Service' is.

- **Service:** A specific IT deliverable that provides customer value. It is measurable in customer terms and provides the basis for doing business with the customer. It is delivered through a series of processes and/or activities.
- **Services Portfolio:** The collection of services provided by IT that in their aggregation represents all the 'value add' activities performed by IT.
- **Service Component:** A logically grouped set of activities that represent part of a service that touches the customer. Service components are grouped together to create deliverables that form the basis for doing business with customers.
- **Service Owner:** The individual accountable for ensuring the customer receives the identified value of the service. They take the customer's end-to-end view of IT activities by working with process owners to ensure all the required delivery components fit together smoothly.
- **Process:** A collection of related activities that take inputs, transforms them, and produces outputs that support an enterprise goal.
- **Enablers:** The decomposed components of a service (process, organization, technology) that are combined to create *Service Deliverables*. The collection of activities (and their supporting roles and technology) become the workflow needed for *Service Delivery*.
- **Service Level Agreements:** A grouping of Services or Service Components that have had specific delivery commitments and roles identified with the customer. SLAs can be grouped together in different ways to represent various products. Examples would include things like e-mail and service delivery.
- **Service Level Management:** Service Level Management governs the planning, coordinating, drafting, agreeing, monitoring, and reporting on Service Level Agreements (SLAs), and the on-going review of service achievements to ensure that the required and cost-justifiable service quality is agreed to, maintained, or where necessary improved. SLAs provide the basis for managing the relationship between the provider and the customer. Service Level Management is essential in any organization so that the level of IT Service needed to support the business can be determined and monitoring can be

initiated to identify whether the required service levels are being achieved – and if not, why not.

- **Service Level Management Objectives:** SLA objective is to maintain and improve IT Service quality, through a constant cycle of agreeing, monitoring, and reporting upon IT service achievements and instigation of actions to eradicate poor service – in line with business or cost justification. It develops a better relationship between IT and its customers.

SLAs should be established for all IT services being provided. Underpinning Contracts (UCs) and Operational Level Agreements (OLAs) should also be in place with those suppliers (external and internal) upon whom the delivery of service is dependent.

SLAs are likely to be a service... if the 'what' is separated from the 'how', and we can change the underlying assets – processes, technologies, data – as well as suppliers, but still provide the promised business outcomes and value. It is a service if its description and design highlight separate roles for customers, users, providers, and suppliers, and it is offered with a pre-defined value proposition stating specific price points and performance metrics tied to business (not just IT operational) outcomes. It is something that makes sense as a customer/user-selectable item, with corresponding service requests available in a menu or service catalogue.

The facts here are startling – inefficiency is prolific – clearly, progress is needed.

3.7.2 Services Scope Overview

Platform Integration and Deployment Services

These provide a set of project services for the planning, design, procurement, assembly and integration, site installation, and project management of the deployment mainstream and special-purpose end-user devices, and also includes a number of asset lifecycle services.

It integrates and customizes multiple devices – including PCs, wireless and mobile devices, kiosks, ATMs, point-of-sale (POS) devices, and printers – to end-user specifications. These services utilize a factory approach for the off-customer-site build and integration services, using build and integration centres around the globe.

Software Platform Management Services

It provides a set of project and annuity services to manage end-user software platforms, including image development and management, application software packaging and distribution, and services to manage the availability of the end-user platform proactively. This includes services to design and migrate end-user platforms, for example, Microsoft XP to Vista, or Linux. It utilizes a factory approach for the off-customer-site platform management services, using management and configuration centres.

3.7.3 Platform Integration and Deployment Component Services

This section discusses some of the very important platform integration and deployment component services.

Order Management

This service handles the procurement of hardware and/or software on behalf of the customer (regardless of asset ownership), and the fulfilment (delivery) of that hardware and software to the central build centre prior to Platform Build and Pre-Load.

Warehousing and Stock Management

This gives provision of central warehousing facilities to store hardware and other agreed components before and after Build and PDP and before shipment to site. It provides services for receiving and warehousing, and ensures that inventory is sufficiently maintained and protected while in storage.

Platform Build and Test

This service provides services to build, integrate, customize, prepare, and test the hardware and software platform before shipment to the customer site or end-user location.

Base Backup

This service provides a base backup to be taken during platform pre-build (assumes the software platform supports this feature).

Data and Personality Migration

It migrates data and personality settings (e.g., desktop wallpaper, Internet Favourites, Desktop layout) from original to replacement platform.

Asset Tagging and Custom Labelling

This service provides for custom asset tagging of hardware components during Build and pre-delivery preparation at the central build facility.

Asset Inventory Update and Report

This service includes adding any new asset to the customers or managed asset database.

Logistics and Delivery

This service includes the packaging and shipment to site of the user hardware platform following Platform Build and Test services.

Installation

This service provides the deployment of the platform into the customer's operational environment. This could be either at the end-users' desk, or agreed deployment centre location for machines that are to be collected or deployed by the customer.

Extended Project Management

This is an extension to the base project management services that cloud uses to manage internal functions and activities. This service extends project management scope to cover overall management of the platform deployment program, including customer and customer third-party resources and activities.

Platform Removal and Return

This includes decommission of platform from customer location, and return to cloud vendors for refurbishment or disposal (but not including for refurbishment or disposal activities).

Asset Refurbishment

This service checks for suitability of the hardware platform, refurbishment, and upgrade as appropriate, and integration into the deployment process.

Asset and Data Disposal

This service provides removal of sensitive data from the hardware platform to varying levels of security (e.g., to department of defence standards), as well as safe environmental disposal to international standards and/or asset value recovery.

Emergency Replacement

This service provides for the emergency replacement of like for like hardware platforms (pre-built to the customers standards) to agreed service levels.

Software Platform Management Services

This section explains the software platform management services in detail.

Software Platform Design Consulting

These services help the client understand the business and technology needs for a new software platform, and provide the design specification.

Software Platform Creation and Customisation

This service is to create and test a software platform to support the needs of the business.

Software Platform Support and Maintenance

It is for ongoing support and maintenance of the client's software platform, including platform updates and management, and ongoing problem support.

Application Scripting

This is the creation of application software unattended installation scripts to accommodate the customer environment. This service can include new software applications and updates to existing packages.

Application Discovery

This service is to discover what applications are in use across the organization.

Application Portfolio Management

This service is to help the clients manage their application portfolio.

Software Delivery

It services the schedule and transport application packages to target end points. This service includes the capability to PUSH software out to user PCs from a central distribution and/or the ability for the users to PULL Software to their PCs using a Web interface.

Antivirus Management

These are the services to manage the delivery of antivirus signature files. It provides real-time protection from malicious virus attacks that can cause potentially disastrous system.

Patch Management

It services to ensure end-user devices have the latest patch releases installed. It provides real-time deployment of critical Operating System patches to help protect against flaws and vulnerabilities.

Health Check Services

These services ensure that end-user devices are in a good state of health of the system, like infrastructure and PC checks. It performs remote monitoring of supported workstations for critical hardware alerts and initiates scripted responses as alerts are received. Examples are identification of low memory, detection of spyware, identify low disk space.

Compliance Services

These services are to ensure that end-user devices are compliant with client standards. They perform remote monitoring of supported workstations to perform activities such as detecting peer-to-peer software detection of games, etc.

Until there is an agreement to industry standard definition of what an IT service is, you must agree with the client on a definition of 'service'. So, we will consider different type of the services visibility with respect to customer.

External Services

These services are visible and seen by the customer, and include business services (business intelligence, logistics, receiving orders, marketing services, invoicing, accounting, etc.) and user services (desktop support, maintenance, education, etc.).

Internal Services

These services are invisible or less visible to the customer, but essential to the delivery of IT services. They include infrastructure services (hosting services, storage, availability, data retention, or recovery) and network services (network, remote access, mobile or wireless services, etc.). It takes care of application services (integration, testing, design, maintenance, optimization, etc.).

User-Initiated Service Request

It includes service request handling in incident management, for example, the service request progresses through its lifecycle exactly as an incident. Most companies separate user service requests and incidents. For example, the service requests follow a different process and use a different tool to track it throughout its lifecycle. Significant IT workload is responding to user-initiated requests for some work to be done.

Users request services for which their businesses have already contracted with a service provider, and to which they are already entitled. Some people have referred to the list of services from which a user can order services as a service catalogue. Perhaps it is a 'User Services Catalogue'. These service requests and this type of listing of services are appropriately offered through the single point of contact for users of IT services, the service desk.

Customer-Initiated Service Request – A Service Catalogue-Based Request

It includes the concept of a service catalogue as a list of services that the customer can order. The customer is the one who pays for services. When customers order a service, their users are entitled to receive services under that agreement. Each individual request made by a user is a user initiated service request. The user is entitled to services that their business has ordered through a service catalogue request. This type of request is from the customer to some account rep or 'business relationship manager' who responds to this request and initiates the service provisioning for that customer.

3.8 SUMMARY

This chapter visualizes the different cloud models with respect to services. It also takes into account what service is all about and the different types of infrastructure services that can be offered as cloud as a service.

CHAPTER 4

CLOUD SOLUTIONS

4.1 INTRODUCTION

Cloud environment presents an opportunity to enhance the user experience by providing a broader communication path for reaching out to the user or for providing a series of business services to the user via the application features.

Deploying the application to the cloud is somewhat different since the deployment process will not be done locally within the enterprise and the existence of the provisioned image and a series of deployment steps is needed to deploy the application and validate the deployment.

Development and testing environments are readily available within the cloud environment. The advantages of these environments, especially from a costing perspective, are numerous as there is no need to purchase and deploy servers within the normal enterprise environment. If a POC was being developed and project was cancelled, no software, hardware, or even development tools would have to purchased, only to be thrown away later as the cloud supports development and testing of applications.

4.1.1 Cloud Application Planning

The design and development of cloud applications requires many unique considerations:

- Business functions.
- Application architecture.
- Security for cloud computing.
- Cloud delivery model.
- User experience.
- Development, testing, and run-time environments.

Application architecture is selected through some sort of criteria evaluation.

The key thing to talk about from a security aspect is the enhancements to the existing security model where data protection and isolation of the data from other areas of the cloud environment. Encryption is one possibility to further enhance the security model whereas the enterprise would not necessary invoke that option. Other aspect of security would be to further authenticate and authorize users of the application and the services the users have been entitled to use.

4.1.2 Cloud Business and Operational Support Services (BSS and OSS)

Business Support Services (BSS) are the components that cloud operators use to run their business operations. The term BSS applies to service providers in all sectors such as utility providers. Typical types of activities that count as part of BSS are taking customer orders, managing customer data, managing order data, billing, rating, and offering services.

Operational Support Services (OSS) are computer systems used by cloud service providers. The term OSS most frequently describes 'network systems' dealing with the network itself, supporting processes such as maintaining network inventory, provisioning services, configuring network components, and managing faults.

BSS and OSS components need to be externalized so that the supporting services of the application being transformed to the cloud environment can capitalize on the various functions the OSS and BSS provides. For example, provisioning can be adapted to support the applications provisioning requirements instead of creating self-provisioning from scratch. The ability to tap into the monitoring, metering events and keep track of the activity within the cloud environment can assist the application to continue to provide the service levels and quality of service. Each of the OSS and BSS offered by the cloud environment will continue to support the application and the consumers of the application in maintaining the key characteristics of cloud computing.

The cloud application architecture brings together the business services, security, infrastructure, and integration required for an optimal solution. Cloud services represent any type of IT capability that is provided by the cloud service provider to cloud service consumers. Typical categories of cloud services are infrastructure, platform, software, or business process services. In contrast to traditional IT services, cloud services have attributes associated with cloud computing, such as a pay-per-use model, self-service usage, flexible scaling, and shared of underlying IT resources.

The cloud vendor is responsible for delivering instances of cloud services of any category to cloud service consumers, the ongoing management of all cloud service instances from a provider perspective, and allowing cloud service consumers to manage their cloud service instances in a self-service fashion. The technical aspects of a cloud service are captured in a service template, which is also the artefact that describes how the OSS capabilities of the cloud vendor are exploited within the context of the respective cloud service.

For most cloud services, specific software are required for implementing cloud service specifics: For IaaS, these are typically hypervisors installed on the managed infrastructure; for PaaS, this would a multi-tenancy enabled middleware platform; for SaaS, a multi-tenancy enabled end-user application; and for BPaaS, multi-tenancy enabled business process engine. Depending on the nature of the respective cloud service, the notion of a cloud service instance represents different entities.

Cloud services can be built on top of each other, for example, a software service could consume a platform or infrastructure service as its basis, and a platform service could consume an infrastructure service as its foundation. However, this is not required, that is, a software service could also directly be built on top of 'traditional' infrastructure, clearly inheriting all constraints associated with such an infrastructure. In general, basic cloud architectural postulates to share as much as possible across cloud services with respect to management platform and underlying infrastructure. However, it does not require to only having one single, fully homogeneous infrastructure – of course, this would be the ideal goal, but given different infrastructure requirements, this is not possible. For example, if a particular cloud service has very specific infrastructure needs, it is clearly allowed to run this cloud service on a dedicated infrastructure (e.g., the Google search engine or HPC cloud services would always run on a purpose-built physical infrastructure for performance and efficiency reasons; they wouldn't run virtualized compute cloud service).

In the context of building cloud services on top of each other, it is important to distinguish the sharing of a common OSS/BSS structure across multiple cloud services and the usage of the actual cloud service capability by another one.

In any case, each cloud service offered by a cloud service provider is 'known' to the BSS and OSS of the cloud vendors. Consequently, a cloud service provider offers cloud services as a result of very conscious business decisions, since taking a cloud service to market must be supported by a corresponding solid business model and investments for the development and operations of the cloud service.

4.2 CLOUD ECOSYSTEM

It is very important to understand the relationship between a cloud service and the artefacts that can be developed based on and within the boundaries of an ecosystem-focused IaaS or PaaS cloud service. Bringing any cloud service to market requires corresponding pre-investment, along with respective metering and charging models in support of the corresponding business model. Therefore, making the characteristics flexible to artefact developers is not possible as it would be very hard to make the corresponding costs flexible and by that very hard to predict.

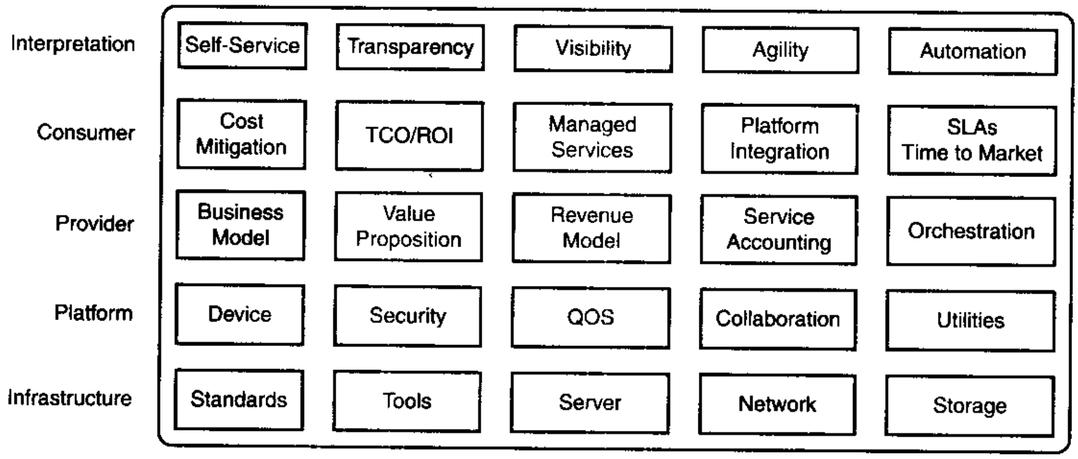

FIGURE 4.1 Cloud ecosystem.

This illustrates that defining and delivering a cloud service requires nailing down all corresponding functional and non-functional requirements. The artefacts developed on top of an ecosystem-focused cloud service have then only very minimal room to change how these functional and non-functional requirements are addressed (Figure 4.1). Note that this is not to be viewed as something negative, but rather as something very positive from an ecosystem perspective – it is a core value proposition of ecosystem-focused cloud services to provide pretty strict guidelines with respect to how they can be exploited as this is the main factor driving a reduction in cost of artefact development. The easier it is to develop artefacts for such a cloud service, the more likely the cloud service is successful.

As a summary, it is important to note that there is a difference between developing cloud services as a very conscious technical and business decision and developing artefacts on top of ecosystem focused cloud services prescribing the boundaries for how these artefacts can run.

Note that sometimes the concept of a 'Cloud Service' is also referred to as a 'Cloud Service Product'. The cloud's capability to have multiple environments to deploy the application is a major advantage because it can be a mix and match condition that best fits the business function and the application. This means that the organization is not tied to one solution for cloud computing but rather multiple solutions. So, depending on the business needs, the application, and what the application has to offer, cloud requirements will help to select the proper cloud environments.

Cloud-based environments come handy especially when used to develop, test, and run your application for the following reasons:

- Available in your own private cloud environment or on the public cloud.
- Rapid access to a configurable development and test environment to speed time to market.
- Self-service Web portal for enterprise account management and provisioning in minutes.
- Pay-as-you-go pricing, with the choice of preferred pricing through reserved capacity packages.
- Security-rich environment designed to protect your systems and data.
- Access to a rich catalogue of software images for improved flexibility and rapid provisioning.
- Rapid provisioning and faster time to value.

4.3 CLOUD BUSINESS PROCESS MANAGEMENT

Business Process Management (BPM) governs an organization's cross-functional, customer-focussed, end-to-end *core business processes*. It achieves strategic business objectives by directing the deployment of resources from across the organization into efficient processes that create customer value. Its focus is on driving overall bottom line success by *integrating verticals and optimizing core work* (for example, order-to-cash, integrated product development, integrated supply chain). This is what differentiates BPM from traditional functional management disciplines.

In addition, intrinsic to BPM is the principle of 'continuous improvement', perpetually increasing value-generation and sustaining market competitiveness (or dominance) of the organization. It clearly defines and aligns operations, organizations, and information technology. The cloud environment can help in the following ways:

- **Integration of core business:**
  - Holistic.
  - Crosses organizational functions and boundaries (height and breadth).
  - Includes business and technology.
- **Value-focused efficiency:**
  - Customer-centric perspective.
  - Bottom-line success.
  - Speed at which ROI is delivered.
  - Performance measurement.
- **Continuous:**
  - This is based on longer period of intervals pertaining to cloud business.
  - Continual improvement.

- **Cultural:**
  - Cultural considerations of the organization and geographical area kept in mind at the time of due diligence of the requirement.

4.3.1 Identifying BPM Opportunities

This section discusses the opportunities required for successful cloud business process management and the characteristics of cloud deployment offerings. The answers to the following questions can help you identify cloud opportunities better:

- Are the strategic value proposition and capabilities defined for your organization?
- How does your overall strategy drive the design and execution of your business processes? Is there a traceability of execution to goals?
- How do you manage your core business processes?
- What are your current process initiatives?
- What are your current process governance facilities?
- Are your existing organizational structures aligned to enable efficient process operations?
- How do your customers measure and assess the performance of your processes?
- How does your process performance compare to your competitors?
- How effectively does your current technology (information, systems, tools, machines, etc.) enable the enterprise's core business processes?
- What risks and challenges does your current technology present for current and future process capabilities?
- What products does your organization have? What type of products?
- What are the notable pieces of your IT portfolio?
- Has your organization adopted SOA?
- How are processes currently modelled in your organization? What's included in the model?
- What design/development tools are currently used in the organization?
- What testing tools are currently used by the organization? What are the strengths/weaknesses of the tools?
- Please describe the different business processes that are automated in the organization?

Cloud application development offerings provide:

- Cloud application reference architecture.
- Unmatched experience developing high-performing, secure applications across a wide range of technologies of the cloud vendor.
- Unmatched application security expertise.
- Leadership in cloud related technologies -- multi-tenancy, virtualization, pervasive computing.
- Significant expertise with cloud business models.
- Ability to integrate a portfolio of related cloud services.

4.3.2 Cloud Technical Strategy

This section gives the technical strategy on how the cloud customers can enable the cloud deployment.

Cloud services enable our cloud users to build middleware clouds in their datacenter and utilize public clouds, where it make sense by providing the following cloud-enabled middleware services:

- Infrastructure Services.
- Platform Services.
- Application Services.

Cloud strategy enables our organizations to do the following:

- Build middleware clouds in their datacenter.
- Utilize public clouds where it makes sense.

It does so by providing support in the following areas:

- **Cloud-enabled middleware services:**
  - Infrastructure Services.
  - Platform Services.
  - Application Services.
  - Serving the on premise and public clouds.

So, what does it mean to develop an application service for the cloud? For one, it means product features development similar to on-premise software. This means:

- **Enabling the software for cloud essentially implies:**
  - Support for collaborative multi-tenancy.
  - Self-service registration.
  - Managing customers and their entitlements.
  - Single sign on.
  - Additional security concerns.
- **Integration with the datacenter:**
  - Firewalls, reverse proxy configurations.
  - Fully qualified domain names, certificates.
  - Management of services, patch procedures.
  - Isolation, recovery, backup issues.

4.3.3 *Cloud Use Cases*

Infrastructure as a Service (IaaS) or Test/Development

Problem: Development teams require unpredictable amounts of infrastructure to get their job done. In majority of the cases, getting all of these resources in place before they are required in the development cycle can be quite a challenge. Purchasing the hardware consumes project budget and procurement of it is often quite slow. Static development and testing resources require manual re-provisioning in order to re-purpose resources for use, or new resources need to be purchased to meet demand. In cases where project timelines are short, project managers often choose not to set up much of an environment because it depletes their budget or jeopardizes the project's delivery schedule. Actual usage of the system(s) can be quite short in terms of absolute time. Different types of projects need different kinds of development components (like SQL server, SharePoint, BizTalk, etc.) depending on the architecture of the solution. Besides the testing environment, the team usually needs a system that looks quite similar to the production environment in order to perform simulation, stress and load tests, for example, or to delivery end-user training, etc.

Solution: Companies can create standardized service catalogue items for common infrastructure requirements and enable development and test teams to access infrastructure in a self-service model (IaaS). Overall control over the process is maintained with business policies around quotas, reservations, reclamation, and standardized offerings.

Standardized Development Platforms/Middleware (PaaSEnable)

Problem: Developers are often not concerned about the impact of their code on Operations. They deliver their code without involving Operations into architectural decisions or code reviews. Enterprise architects recognize the large costs associated with non-standard development platforms. To simplify the ongoing maintenance and streamline development operations, many companies are creating corporate standards around development stacks that include middleware/applications. However, most companies are hesitant to an external PaaS offering due to the constraints of having to rewrite their internal applications to fit the external PaaS API sets.

Solution: Companies can create standardized development platform definitions for use by development teams to standardize and streamline their efforts. This improves corporate IT productivity by helping them build a 'private cloud' that provides a common foundation for building custom applications which run securely behind their own firewall.

Application Cloud

Problem: Companies want to move beyond self-service for infrastructure and provide application owners the ability to define, instantiate, and manage complex multi-tier applications. This includes configuring the application for production usage and monitoring performance within the application for SLA optimization.

Solution: End users can access complete application definitions and manage them according to their quotas and preferences defined by cloud administrators. Applications in production can be monitored across multiple factors and automatically scaled up and down according to business policies.

Software-as-a-Service (SaaS) to End Customers

Problem: Many companies or ISVs want to deliver their applications to end users as a service. However, creating a multi-tenant SaaS offering requires substantial development to support security, performance, and scalability needs. Due to these high costs, companies cannot offer new services based upon existing applications.

Solution: Companies can provision a unique application instance per customer with private cloud automation capabilities. Environments are provisioned according to business policies and unique SLAs can be delivered per customer according to the business arrangements.

The cloud engine should provide the automation to provision on-demand complex application and configuration environments required along with the dynamic application scaling. It should also deliver unique business policy selections including custom placement and high availability across multiple datacenters.

4.4 CLOUD SERVICE MANAGEMENT

A service management system provides the visibility, control, and automation needed for efficient cloud delivery in both public and private implementations:

- **Simplify user interaction with IT:**
  - User-friendly self-service interface accelerates time to value.
  - Service catalogue enables standards which drive consistent service delivery.
- **Enable policies to lower cost with provisioning:**
  - Automated provisioning and de-provisioning speeds service delivery.
  - Provisioning policies allow release and reuse of assets.
- **Increase system administrator productivity:**
  - Move from management silos to a service management system.

The emergence of cloud deployments is prompting enterprises to either assemble in-house teams to manage specialized cloud service providers or look to third-party cloud brokers chiefly due to the following reasons:

- Every service-oriented approach needs a mechanism to enable discovery and end-point resolution.
- Registry/repository technology provides this where service delivery is inside the firewall.
- Cloud services delivered across firewalls need something similar – a third party that serves as a 'service broker'.

Leveraging service brokers will probably become a critical success factor in cloud computing as cloud services multiply and expand faster than the ability of cloud consumers to manage or govern them. The growth of service brokerage businesses will increase the ability of cloud consumers to use services in a trustworthy manner. Cloud service providers are expected to begin to partner with cloud brokerages to ensure that they can deliver the services they promote. These cloud intermediaries will help companies choose the right platform, deploy apps across multiple clouds and perhaps even provide cloud arbitrage services that allow end-users to shift between platforms to capture the best pricing.

There can be three categories of opportunities for cloud brokers:

- **Cloud service intermediation:** Building services atop an existing cloud platform, such as additional security or management capabilities.
- **Cloud aggregation:** Deploying customer services over multiple cloud platforms.
- **Cloud service arbitrage:** Supplying flexibility and 'opportunistic choices' – and fostering competition between clouds.

It will be similar to cloud services under one umbrella except that the services being aggregated won't be fixed. This flexibility will be important while doing chores such as providing multiple e-mail services through one service provider or providing a credit-scoring service that checks multiple scoring agencies and then selects the best score.

The ability to federate an application across multiple clouds will become important – if one service goes down, another can be started – and the service broker will just simplify it. To help federate the clouds, a 'storefront' (Apps.gov) site can be created with services that are pre-screened to meet government procurement guidelines. The new site can be expected to cut red tape and make it easier for government agencies to quickly deploy the latest technology.

4.4.1 Key Cloud Solution Characteristics

The essential cloud orchestrator and engine key characteristic capabilities are:

- **Scalability:** Cloud orchestrator should maintain an index of the resources that are acquired from the hypervisor, giving the master a low overhead and enabling it to scale across tens of thousands of machines across multiple geographies.
- **High Availability:** Cloud orchestrator should play for the master node to support 'Active–Passive' as well as 'Active–Active' scenarios for availability and Disaster Recovery (DR). Cloud orchestrator should also monitor individual physical server for availability and in case of a physical resource server failure, should restart the VM on another running server to meet the requirements.
- **Application Lifecycle:** Cloud orchestrator should offer complete application lifecycle support from the creation of infrastructure to installation, configuration, and launching an application to deletion or expiration. This allows applications to be instantiated, removed, or flexed very quickly to respond to real-time demand for those applications.
- **Multi-tenancy/Role-based Administration:**
  - Cloud orchestrator should support multi-tenant capability with specific user permissions. Cloud orchestrator should have multiple personas which are like Cloud Admin, Account Owner, and User. Application definitions are only 'published' to specific users. The application owner or administrator logs in with his/her credentials and can view (User) the application VMs that have been allocated to him/her and do admin operations (Account Owner) on those VMs if needed.
  - Role-based administration allows fine-grained control of what each person can or cannot do in terms of cloud orchestrator features.
- **Policies:** Cloud orchestrator should provide rich set of policies that can be enabled. These policies can be modified or new ones can be created to take effect at the global level on applications, VMs, hosts, etc. These policies can also be embedded in the service definitions to take effect automatically depending on metric threshold. For example, a policy can be created to allow an application to flex up to 10 VMs during high load or demand and to be reduced to only 2 running VMs during low load or demand. This frees up resources that can be used by other applications that are experiencing high load.
- **Alarms:** Cloud orchestrator should provide pre-defined alarms that can be set at the global level for applications, VMs, hosts, etc. These alarms can be used to notify individual users or application owners regarding the application thresholds being reached.

For example, an alert can be sent if the response time of an application is below an SLA but there are no more resources for the application to flex-up.

- **Application Awareness and Policy-based Allocation:** Cloud orchestrator should be aware of application requirements and optimizes the placement of the application accordingly, e.g. placing the VMs running the application close to each other to reduce latency. Cloud orchestrator should support the major application servers.
- **Resource Awareness and Policy-based Allocation:** Cloud orchestrator should optimize the usage of the cloud infrastructure through intelligent resource allocation policies and allows load balancing of the VMs
- **Elasticity Based on Performance (Flex-up/Flex-down):** Cloud orchestrator should provide out-of-box functionality to flex-up or flex-down an application instance or resource based on performance metrics.
- **Reporting and Accounting:** Cloud orchestrator should provide metering and billing reports on resource allocation and actual usage. Additionally, this data can be used to create reports on inventory capacity and consumption. This allows the different business owners to create reports on how much or how little the application is used, and administrators can then adjust the resources allocated to each application accordingly.
- **Self-Service Portal:** Cloud orchestrator should enable a self-service portal for application owners. Application owners can request machines or entire multi-machine application environment, monitor, and control them through this portal. It should drive the workflows necessary to create the environment, and provide run-time environment management in order to support application elasticity. For example, the owner of the auditing application may request more resources for his application during a busy period.

4.5 ON-PREMISE CLOUD ORCHESTRATION AND PROVISIONING ENGINE

On-Premise Cloud Orchestration and Provisioning Engine can be a bundled offering that includes hardware, software, and the services one needs to get started with cloud computing. It should include all the elements in a services ecosystem. It must have a self-service portal, it should include the automation, and it should track and control all of resources.

The other objective is to completely integrate and include a service and then, on top of that, users can add additional services to do integration or other types of cloud work. It can be a pre-packaged private cloud offering that can bring together the hardware, software, and services needed to establish a private cloud to accelerate selling efforts and effectiveness.

Cloud Orchestration and Provisioning Engine should be designed from client cloud implementation experiences and integrated with the service management software system with servers, storage, and services to enable aprivate cloud in IT environment. This will help to remove the guess work of establishing a private cloud by pre-installing and configuring the necessary software on the hardware and leveraging services for customization to your environment. All that is required is to install your applications and start leveraging the benefits of cloud computing, such as virtualization, flexibility, scalability, and a self-service portal for provisioning new services.

Cloud Orchestration and Provisioning Engine should provide an alternative to traditional IT infrastructure for IT executives seeking to enhance delivery of services and to transform the datacenter into a cost-effective Dynamic Infrastructure. Cloud Orchestration and Provisioning Engine ecosystem should be 'Built for performance' and should be based on architectures and configurations required by specific workloads. It should enable the datacenter to accelerate the creation of services for a variety of workloads with a high degree of flexibility, reliability, and resource optimization.

4.5.1 *Benefits/Value Proposition*

Powers faster time to innovation, lowers cost per unit of innovation

- **Innovation:** It should dramatically improve business value and IT's effect on time-to-market by enabling the business workloads to rapidly and accurately be deployed when and where they are needed.
- **Decrease Operational Expenses:** It must gain productivity increases in IT labour costs through automation. Maximize capital usage and reduce added capital expense.
- **Reduce Complexity and Risk:** With automation and standardization, the human error-factor should be minimized.

4.5.2 *Cloud Orchestration and Provisioning Requirement Analysis*

In order to understand the Cloud Orchestration and Provisioning Engine requirements, we need to understand the test and development requirements of the cloud deployment. Therefore, we should be sure about the automation of the testing and development cycles to reduce the deployment time. In order to do so, we can initiate the process of discussion and try to talk to cloud customer about the opportunity for the deployment of the Cloud Orchestration and Provisioning Engine. After this, we can set the boundaries of the environment. About 30–50 percent of any given IT environment is devoted to test/development purpose. It can take developers days, weeks, or even months to procure and configure appropriate hardware, networking, management software, storage, with which they can test. With the Orchestration and Provisioning Engine, a developer can log into a self-service portal, select resources required and timeframe, select an image to provision from the service catalogue, and be ready to go in hours as opposed to months. Customer datacenters support hundreds or thousands of distinct composite applications representing business workloads. They leverage multiple types of servers, storage, networks, middleware, and operating systems. About 70 percent of the IT datacenter expense is spent assembling and re-assembling existing infrastructure, leaving fewer resources for innovation.

Cloud solutions are based on its service management solutions. The first solution in a series of Cloud Orchestration and Provisioning Engine solutions that we should consider should be workload-specific deployments. It will be a great entry point for users who want to get into cloud computing as you can roll a Cloud Orchestration and Provisioning Engine into their environment without effecting anything else in the environment, and use it as your initial pilot project to start running cloud services.

Cloud Orchestration and Provisioning Engine in any IT environment represents an alternative entry point into cloud computing. So, in some cases, the users want to turn their existing environment into a cloud. In that case, we go into the datacenter, install a cloud management platform, and assign the existing resources that are there to the cloud. One scenario for this is that the organization already has plenty of equipment that they are just not using efficiently, and they see a lot of benefit from turning their existing investment into a cloud.

Cloud Orchestration and Provisioning Engine can be a great door opener and a great way to jump-start your cloud services. It can be bundled with the hardware, software, and services that you need to quickly getup and running; you can use this as a seed-and-grow model.

Entry points:

- Turn existing environment into a cloud:
  - Install cloud management platform and assign existing resources to the cloud.
  - Scenario – you already have enough equipment or Cloud Orchestration and Provisioning Engine offering is not the right platform.
- Jump start your cloud
  - Hardware + Software + Services required for a quick start up.
  - Can use a 'seed and grow' model – start with a Cloud Orchestration and Provisioning Engine and then addmore.

So what solution should you use? Let's look at some scenarios. So when can you opt for Cloud Orchestration and Provisioning Engine? This should be when you wish to get started with cloud computing and see the advantage of having a cloud management platform bundled along with the resources. This is the most rapid way to get a cloud platform up and running. In a scenario where you want to transition your existing resources into a cloud, you use this service. Also, a user who uses Cloud Orchestration and Provisioning Engine will benefit when cloud vendor can help them plan, design, and implement services, whether they want to implement infrastructure as a service, platform as a service, or turn it into a production cloud.

So, in some cases as you can see, it's appropriate to opt for Cloud Orchestration and Provisioning Engine and the business development and test services, because even if you just start with provisioning engine, you may eventually want to build out your service catalogue, integrate it into your directory, integrate it into cloud vendor or third-party service management products, and extend the capability of the platform.

4.5.3 *Cloud Infrastructure Security*

The security aspect of the cloud infrastructure goes side by side with Service Oriented Architecture (SOA) security. We can introduce it as a layered approach. At the top we can see the service layer with the run-time secure virtualized environment. This is available as the distributed service environment. These services include administrative and security aspects across different clouds as well as within a single cloud. This should gel with Web services stack and it is important that we should bind the internal resources with the other cloud services to offer better hybrid cloud services for successful cloud deployments. We will discuss more about this in Chapter 9.

One of the key aspects of SOA is the ability to easily integrate different services from different providers. Cloud computing is pushing this model one step further than most enterprise SOA environments, since a cloud sometimes supports a very large number of tenants, services and standards. This support is provided in a highly dynamic and agile fashion, and under very complex trust relationships. In particular, a cloud SOA sometimes supports a large and open user population, and it cannot assume a pre-established relationship between cloud provider and subscriber.

The Secure Virtualized Runtime layer on the bottom is a virtualized system that runs the processes that provide access to data on the data stores. This run-time differs from classic run-time systems in that it operates on virtual machine images rather than on individual applications. It provides security services such as antivirus, introspection, and externalized security services around virtual images. While the foundations of Secure Virtualized Runtime predate SOA security and are built on decades of experience with mainframe architectures, the development of Secure Virtualized Runtime is still very much in flux. Cloud vendors continuously invest in research and development of stronger isolation at all levels of the network, server, hypervisor, process, and storage infrastructuie to support massive multi-tenancy.

Cloud Orchestration and Provisioning Engine Integrated service management is offered with network, servers, storage, services, and financing as an integrated offering for client test platforms.

- **Improved time to value:** Quickly deliver a cloud using a preloaded and integrated system.
- **Improved innovation:** Dramatically improve business value and IT's effect on time-to-market by delivering services faster.
- **Decrease capital expenses:** Maximize capital usage and reduce added capital expense.
- **Reduce complexity and risk:** With automation and standardization the human error factor is minimized.
- **'Fit for purpose':** Based on architectures required by specific workloads.
- **Self-contained:** Service management, software, hardware, storage, networking.
- **Modular:** Automatically expandable and scalable.
- **Virtualized:** End-to-end across server, network, and storage.
- **Self-service:** Ease of consumption.
- **'Lights-out':** Zero touch automated operations.

Cloud Orchestration and Provisioning Engine is offered as a services engagement which can build a solution to a client's needs, including creation of custom virtual images for dispensing. Summary points are given below

- **Drastically reduce set-up and configuration time:**
  - New environments in minutes!
- **Reduce risk by codifying infrastructure:**
  - Freeze-dry best practices for repeated, consistent deployments.
- **Security throughout the entire lifecycle.**
- **Simplify maintenance and management:**
  - Flexibly manage and update the components of your patterns.
  - Ensure consistency in versions across development, test, production.
- **Spend less time administering, more time developing new solutions.**

4.6 COMPUTING ON DEMAND (CoD)

On-demand computing is the need of the hour. It is very essential even in a supercomputing environment. On-demand computing can be implemented using various virtualization techniques. Cloud gives you an option to leverage the computing infrastructure without actually buying the hardware. This helps you to transfer the workload if your resources are not able to support it, and at other times, lets others utilize your resources that are lying idle. In this way this makes it possible to use the resources in most efficient ways. It may be possible that there can be many spikes that can come for the utilization but cloud experts can make it smooth.

The uniquely rich set of features that on-demand computing can offer enables service seeker to deploy a true utility. The platform allows users to:

- **Align cost with utilization** so that users can scale costs down as well as up. This allows a workload to start with minimal upfront costs and scale as the demand grows without paying a penalty to increase capacity. Additionally users can benefit from not incurring the disruption to move to a larger machine.
- **Increase end-users availability significantly.** As workload can be moved dynamically, it is possible to move workload from one server to another without interruption so remedial work can be carried out if server down time is required.
- **Balance workload dynamically across multiple servers** without taking applications offline. Using the workload mobility features customer can align their costs by ensuring that workload is deployed in such a way as to optimise systems resource.
- **React to short-term resource requirements** almost instantly. If a workload has to be deployed at short notice, a virtual machine (VM) can be created on the server and resources allocated instantaneously using the dynamic capacity model.
- **Reduce the physical foot print** in the datacenter. Consolidation of workload on to a smaller number of servers will improve space, power, and cooling metrics.
- **Confidently increase system utilization** to over 75 percent without fear of degrading performance for end-users.
- **Develop a simple charging model that reflects usage** for end users as the service delivery culture continues to mature.
- **Double the workload delivered** in the power and cooling envelope.

4.6.1 Pre-Provisioning

For the on-demand computing requirement pre-provisioning is the viable option as it helps organization meet the requirement of the dynamic datacenter requirements. Organizations would like to reduce the time they take to commission servers when a new workload is to be deployed.

This approach is ideal when:

- The sizing and capacity planning is fully understood.
- The workload is fairly constant, ensuring good utilization levels are achieved.
- There are business reasons that require the physical separation of workload.
- Workload can be scaled horizontally.

4.6.2 *On-Demand CPU/Memory/VM Resources*

In the dynamic environment it is important to track the requirement of CPU, Memory, and VMs. It is based on the common pool concept where resources are allocated and de-allocated as the requirement is over. This approach is ideal when:

- Workloads are trending upwards so investment can be aligned with utilization.
- Peeks in workload are longer term.
- Workload scales vertically.
- It is more economically advantageous to 'buy out' dynamic capacity.

4.6.3 *Dynamic Capacity*

Utility CoD is used to automatically provide additional processor capacity on a temporary basis within the shared processor pool. Usage is measured in processor-minute increments, and is reported via a Web interface or collection of report by cloud vendor engineer. Billing is based on the reported usage.

This approach is ideal when:

- The workload is very variable and multiple workloads can be hosted on a single machine so that the utilization can be levelled out.
- The workload has short periods where system utilization increases massively, but for the majority of the time it is not resource-intensive.
- Workloads can share a physical platform.
- The workload is designed to scale vertically only.
- Users want to dynamically balance workloads across servers.
- Users want to continue to run very small workloads without incurring the overheads associated with running a physical server to support it.

Benefits

- Partition mobility, significantly reduced power/cooling footprint, donation of unused processor cycles of VMs with dedicated processors to uncapped partitions and at the same time guaranteed performance of these VMs.
- Very short deployment time (time-to-market optimized).
- Lowest possible cost for deployment of small workloads.
- Less management effort, for example, when using VMs.
- Most granular charging scheme, pay for the CPU and memory cycles actually used.
- Complete decommissioning of partitions; resources are available for other purposes.
- Flexible workload management, workloads can compensate for each other, thus reducing overall utilization.
- Ideal for environments with identical systems management, utilities for development and testing.

Limitations

- Short peaks, must not exceed certain limits, and needs to be monitored (via Web interface) to ensure best value is obtained.
- Utility CoD provides processor resources only to the uncapped partitions.

However, one of the most important advantages of the Dynamic Capacity model is to be flexible, that is, to allow the switching of CPU capacity on and off as needed, which can reduce costs significantly. This not reflected in this calculation as it requires input from application owners.

4.6.4 *Cloud Platform Characteristics Based on CoD*

This section discusses cloud platform characteristics on the basis of low-end, on-demand, and dynamic-capacity–based servers.

Low-End Servers

- Physical segregation of servers.
- High administration cost due to management of more physical servers.
- Limited and complex scalability, process – maximum 8 processors per server – slower turn-around time for server deployment.
- Longer lead time for server deployment from ordering of servers to setting up of infrastructure.
- Not ideal for short product lifecycle application due to fixed cost expenditure for hardware.
- Wastage of hardware resources for applications that reacts to volatile market.
- Unable to share resources between applications.
- Wastage of un-used processing cycles if the application does not fully utilized the resources.
- No hardware/application interdependency forcing down time on application owners.
- No capacity on demand capability. Downtime is required for adding new hardware.
- Low price per CPU cycle purchased but higher cost per CPU cycle actually used.

On-Demand Platform

- Physical or logical segregation of servers or partitions implementation.
- Lower administration cost due to less physical servers' management.
- Can cater for quick turn-around time for new application deployment or increase capacity due to business requirements.
- Enhanced time to market for new product launch with immediate availability of CPU/ memory capacity.
- Not ideal for short product life cycle application due to fixed cost expenditure for hardware.
- Wastage of hardware resources for applications which reacts to volatile market.
- Able to share I/O, CPU, and memory resources between applications.
- Able to take advantage of un-used CPU/memory if dynamic VM reallocation or share pool methodology is implemented.
- To provide an environment in which there are no hardware/application interdependencies forcing down time on application owners care full capacity planning and management is required.
- Capacity on demand capability – no downtime is required if COD CPU/memory is sufficient.
- Higher price per CPU cycle, lower cost per CPU cycle actually used.

Dynamic Capacity Platform

- Choose virtual machine or workload virtual machine implementation for application consolidation.
- Lower administration cost due to less physical and logical servers management.
- Can cater for quick turn-around time for new application deployment.
- Enhanced time to market for new product launch with immediate availability of infrastructure and setup.
- Ability to scale up and down which will be ideal for application with short product life cycle.
- Able to cater for applications which react to volatile market, i.e., scaling up and down capacity.
- Able to share I/O, CPU, and memory resources between applications.
- Able to take advantage of un-used processing cycle of other applications.
- No hardware/application interdependency forcing down time on application owners as workload can be dynamically moved to facilitate maintenance, etc.
- Capacity on-demand capability. No downtime is required as the machine is fully configured.
- Higher price per CPU cycle, lower cost per CPU cycle actually used. Average price due to the ability to optimize utilization and rapidly deploy workload.

4.7 CLOUDSOURCING

Today we are living in the era of optimizing the hardware resources and moving towards the large enterprise day by day so cloud computing is becoming the ingredient part of infrastructure deployments. Now you may not need only cloud computing, you may need the entire consulting, implement and management solutions.

The new wave that is igniting the cloud deployments in the service industry as a new trend is Cloudsourcing – outsourcing the end to end solution using cloud methodology using public cloud, infrastructure and platforms. This will be more planned approach as it will comprise the whole service cycle of outsourcing business with cloud principles with the help of strategized connected cloud platforms that will match the overall enterprise requirements.

This includes the whole cloud implementation, IT business consulting, integration, and configuration of the business. This will give the option through which we can enjoy the benefits of service industry with the benefits of the clouds that gives the innovative approach of paying the resources over subscription.

Real deployment of the Cloudsourcing will requires the business model with the impact of cloud customer and cloud vendor requirements.

With respect to cloud customers, it is important to note there is no control on the infrastructure layout of the cloud deployments. Even there is no control over the place from where the data services are offered from the cloud vendor. It is also to known that cloud customer don't have think about the operational staff for the deployments.

Thus, Cloudsourcing will be playing the vital role in the next generation of cloud implementation. With the availability of new open source tools it is like icing on the cake, integrated with partner cloud solutions, platforms and infrastructure. Also the new charging models like the services on both a project and subscription basis will give new wave to deploy and adopt the cloud sourcing models.

This will help to customize application on cloud infrastructure. This will be primarily being offered as a public cloud and all these offering will be available as managed services. These services will be prototype based that is developed internally on the product and working applications. Therefore, it will give a good chance to use the intellectual property for developing different business vertical solution easily.

4.8 SUMMARY

In this chapter, we pointed out the main features of Cloud Orchestration and Provisioning Engine, BPM clouds, cloud sourcing, and requirements of service management. Next chapter will discuss about different types of cloud offerings.

CHAPTER 5

CLOUD OFFERINGS

5.1 INTRODUCTION

Information is pouring in faster than we can make sense of it. It is being authored by billions of people and flowing from a trillion intelligent devices, sensors, and instrumented objects. With 80 percent of new data growth existing as unstructured content from music files, to 3D images, to medical records, to e-mail keystrokes, and more, the challenge is trying to pull it all together and make it useful.

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

5.2 INFORMATION STORAGE, RETRIEVAL, ARCHIVE, AND PROTECTION

Organizations process, manage, move, protect, and archive various business data according to unique characteristics such as age, usage patterns, compliance and archiving policies, security and disaster protection rules, and value. Information Lifecycle Management (ILM) is a growing set of recommended practices and technologies to manage data more efficiently and effectively. Lifecycle management becomes even more important for cloud deployments as we would be sharing the data services between the cloud vendor and subscribers.

ILM is not the latest data storage, retrieval, and protection solution; piece of hardware; or some software, but rather an approach to assess and manage information across the enterprise. ILM is based on how data is used and how readily available it must be 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 placement of data on the appropriate level of storage with the appropriate retention and retrieval policy. In response to these challenges, organizations are defining specific objectives to support and improve their information management:

- **Cost reduction:**
  - Controlling demand for storage.
  - Reducing hardware/software costs.
  - Reducing storage personnel costs.
- **Better system performance and personnel productivity (i.e., improve efficiency):**
  - Doing the storage activities 'right'.
  - Improving the current people, processes, and technologies being utilized to deliver storage services to the business.
  - Defining and enforcing policies to manage the lifecycle of data.
- **Increased effectiveness:**
  - Doing the 'right' storage activities.
  - Defining and implementing the appropriate storage strategy to address current and future business requirements.

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

- **Activities for gaining initial savings:**
  - Reduce the amount of used storage as a result of initial clean-up.
  - Validate 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.
- **Activities for maximizing savings:**
  - Reconfigure the current storage environment effectively improving the available to raw utilization.
  - Reclaim available storage that has been over allocated.
  - Enhance the information classification, classes of service, and tiered storage architecture.
- **Activities for sustaining savings:**
  - Develop a storage architecture governance model.
  - Implement changes to existing storage management processes like capacity planning and provisioning that to effectively improve capacity utilization on an ongoing basis.

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

- **Operating cost categories:**
  - **Personnel:** Storage support and contractors.
  - **Facilities:** Current floor space consumed by 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 maintenance and incremental maintenance resulting from growth.
  - **Outages:** Cost avoidance associated with the reduction in unplanned outages.
- **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 migrate 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 percent of the data stored cn hard disks is not actively accessed by users or applications, it is obviously ripe for more intelligent management and migration to less expensive storage. But the savings can go significantly beyond disk acquisition costs and annual hardware maintenance costs.

Some points in Information Management:

- **Data:** Discrete element, reasoning, discussion, or calculation of content created through the interactions between applications or interactions between computing devices.
- **Information:** Organized and structured collection of data.
- **Information Lifecycle Management:** The policies, processes, practices, and tools used to align the business value of information with the most appropriate and cost-effective IT infrastructure from the time information is conceived through its final disposition.
- **Information Taxonomy:** Data described in the context of business process requirements and lifecycle characteristics.
- **Information Classes:** Groups of information taxonomies with associated business value that provide the basis for storage management and service delivery.
- **Value-driven Data Placement:** An event correlation framework that 'senses' when the value of data changes and based on business policies 'responds' by moving data to the appropriate storage tier.
- **Storage Process:** A documented set of storage-related tasks and activities required to support a storage infrastructure.
- **Storage Service:** Capabilities provided to a customer base designed to meet their business requirements, wants, and needs that is enabled by a storage infrastructure.
- **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.
- **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, value, etc.
- **Tiered Storage Infrastructure:** An organized collection of Storage Tiers reflecting the flow of all information managed in the enterprise storage architecture.
- **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 the new offering in the new era of cloud computing. This 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 them apply analytics principles and best practices to analyze the different business consequences and achieve newer levels of optimization (Figure 5.1). This can combine complex analytics with the newer software platforms and will lead towards the predictable business situation out of every business insight.

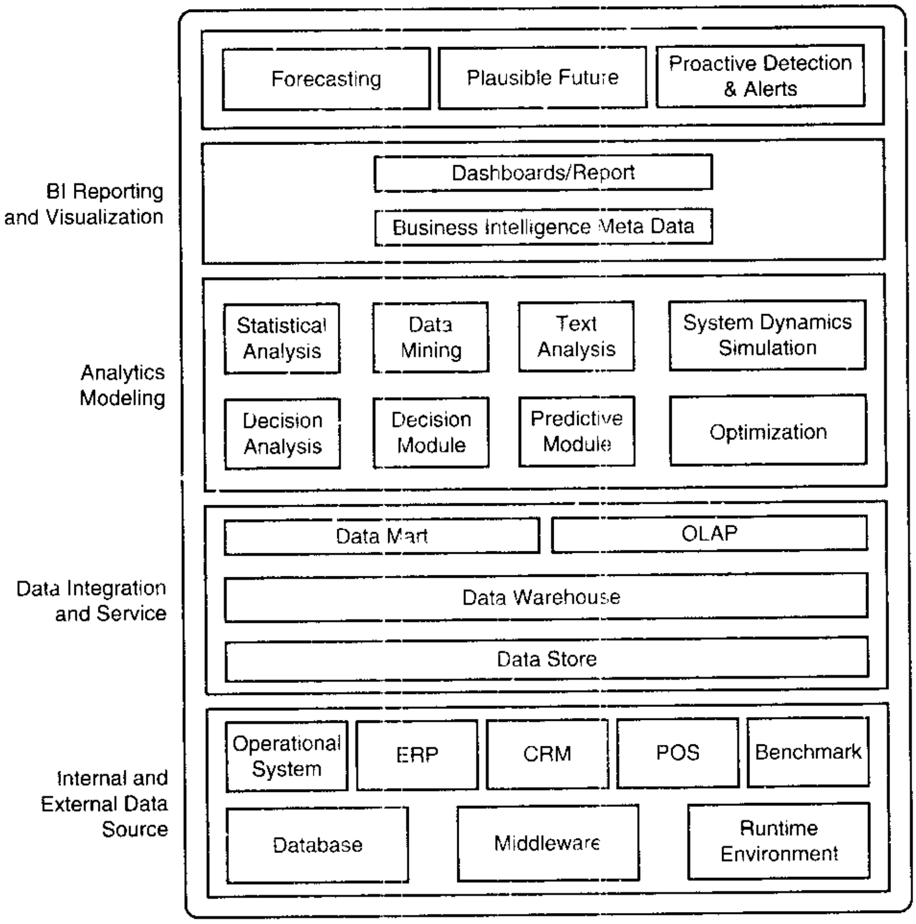

FIGURE 5.1 Cloud analytics.

5.3.1 Cloud Business Analytics Competencies

The cloud business analytics service line is supported by different types of competency areas. One of them is cloud business analytics strategy that helps clients achieve their business objectives faster, with less risk, and at a lower cost by improving how information is recognized and acted upon across the enterprise or within a business function. The next competency is business intelligence and performance management that helps increase performance by providing accurate and on-time data reporting. The next is analytics and optimization that provides different type of modelling techniques, deep computing and simulation techniques to check for different type of 'what if' analysis to increase performance. The other competency is enterprise information management that lets you apply different architecture related to data extraction, archival, retrieval, movement, and integration. Another competency that is required for the cloud analytics is the content management system that includes the different service architecture, technology architecture, and process 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 key to this effort, and gives us a significant advantage (Figure 5.2).

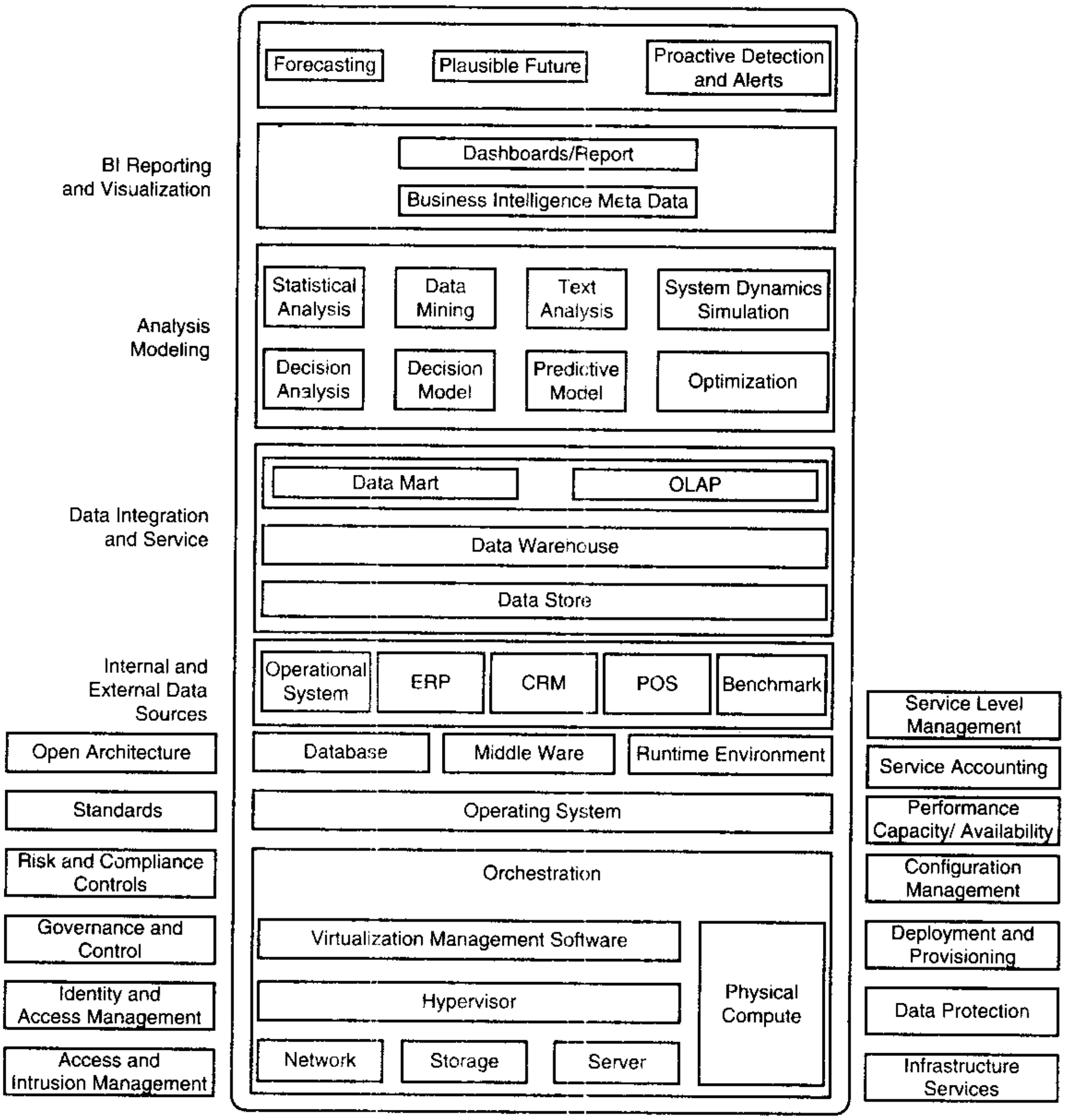

FIGURE 5.2 Cloud analytics business outcomes.

The 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. This also has 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 based on the information available in the organization as decision-makers work with the exploration services available within the organization. This also helps in gauging the business results by measuring the different metrics generated with the help of analytics. This gives the option through which the organization can increase the 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 in typical testing environment. By enabling access to resources as and when required, it reduces the investment on capital as well as enables the business to handle the ups and downs of the testing requirements. With this, organizations can reduce the test cycles, minimize the IT cost, 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 cloud. This also gives you the flexibility to play with the surrogate of the real system without the actual risk.

5.4.1 *Benefits*

- Cut capital and operational costs and not affect mission critical production applications.
- Offer new and innovative services to clients, and present an opportunity to speed cycles of innovation and improve solution quality.
- Facilitate a test environment based on request and provide 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, efficient manner.

These test tools allow you to orchestrate and build your services and development projects and allow you to catalogue and organize all of the various software assets that you have. These can be provisioned by administrators, development team leads, or project team members that you give permission.

5.4.3 The Biggest Benefitters

The biggest problems for the finance heads are to reduce the operational and capital expense of the development and testing environment. Testing under cloud environment comes as a boon to implement this. This not only helps financially to reduce the burden on the company as well as reduce the cycle time for testing and development environment without buying the infrastructure. This environment provide test tools to synchronize and build the services for the project and helps arrange the project resources, assets and versions of deliverables. Also, permission can be set based on the roles to provide the access to the assets.

So why is a test 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 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. So, it is a lot of build-up and teardown as you move from project to project that you test and develop. Typically, 30 to 50 percent of all servers in an IT environment are dedicated to test and development, which is a significant amount. A number of these servers are sitting there at very low levels of utilization, and in fact, sit idle for a period during the project life cycle. So, being able to better utilize 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 limited access to test environments. So, 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 they move 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 think about include:

- What are the types of service management integrations that need to happen?
- Do they want to integrate with discovery?
- Do they want to integrate with the service desk?
- Do they want to create a help ticket whenever a provisioning step occurs?
- Do they want to create an asset every time a new virtualized environment is created?
- Do they want to do usage and accounting chargeback?
- Do they want a test optimization service?

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

Organizations engaged in development and test share a common challenge in executing fluid projects within rigid infrastructures. Development projects are initiated to meet specific 'needs', whether bringing new applications to 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 art as 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 for infrastructure availability. Alternatively, maintaining enough capacity to accommodate peak demand will leave large windows of severe underutilization. Even the best compromise represents unwelcome cost and inefficiency.

With the ability to deploy virtual environments quickly and automatically and redirect capacity as needed, cloud computing offers an ideal solution for testing and development. Cloud vendors makes 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 gives the overall business transformation value that helps you reduce the cost associated with IT operations by helping you prioritize your business requirements. A strategic roadmap is required to enjoy the benefits of application virtualization for testing in cloud environment and requires different assessments.

The first assessment is to consolidate the server and conduct the comprehensive virtualization assessment. Getting the requirements of 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 give the reason to adopt cloud infrastructure and appreciate 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 their cycle time and improving their quality. This offering, through its automation and change and configuration management, can reduce errors and speed up deployment.

Businesses with a dedicated development and test organization can benefit from capital and expense reduction afforded by automation and utilization improvements. There is a growing number of organizations that no longer have datacenter space or power to expand their datacenter. By using standardization, consolidation, and virtualization, they can maximize 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 smooth out resources across different departments so that each department does not have to do account for its own peak demand, and smooth that demand across the peaks and valleys of different departments and different projects.

5.4.4 Cloud Offering Key Themes

Nowadays enterprise's datacenter is managed by Operations organizations (likely composed of Infrastructure professionals or System Administrators) teams. Operations departments focus on availability, stability of IT services, and IT cost efficiency. Operations department's goal is to minimize risk for delivering on non-functional requirements by avoiding unnecessary change and promoting standard infrastructure requirements for applications. Adoption of virtualization technologies has brought a significant change in the enterprise datacenter 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:

- **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.
- **Development teams:** Development departments are usually driven by user needs for frequent delivery of new features. Development teams often want to test new ideas and/or features quickly 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 back into the centralized pool due to 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. No longer are IT infrastructure managers 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 application provisioning time by some great percentage (example 90 percent) and increase resource utilization from what are typically very low percentages <20 percent to an exponentially greater reduction(50–60 percent or higher as an example), reducing IT operational costs by some percentage (example 25 percent or more); all while escaping 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 scaling when needed.

Key Themes

- **Infrastructure to Applications**
  - The most common use 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 are then delivered as a Software-as-a-Service (SaaS) model. In between, enterprises create their own Platform-as-a-Service (PaaS) catalogue to enforce consistency in development platforms and middleware. Since companies are using this for their own internal use, they do not need to rewrite their internal applications to match an external PaaS's custom API set, as is required if they were to consider a public PaaS service.
- **Dev/Test to Production**
  - In software development, there is a high level lifecycle for applications. The beginning is development (dev), then test/QA, staging, followed by production.

Enterprises can also differentiate between the types of productions applications between production internal versus external (customer facing) in terms of criticality. In dev/test, the orientation is towards many users with simpler requests (e.g., VMs or infrastructure) with a focus on shorter duration resource usage (reservation, quota management, and reclamation are key to prevent sprawl). Production applications, in contrast, are more concerned about meeting runtime performance 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 (Figure 5.3).

- **Allocation and Runtime Scaling**
  - Allocation is the process of instantiating the service catalogue item – infrastructure, platforms, or applications. Runtime scaling is the flexing up or down of required resource elements to meet SLA requirements defined by the application owner according to 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 enterprise environment.

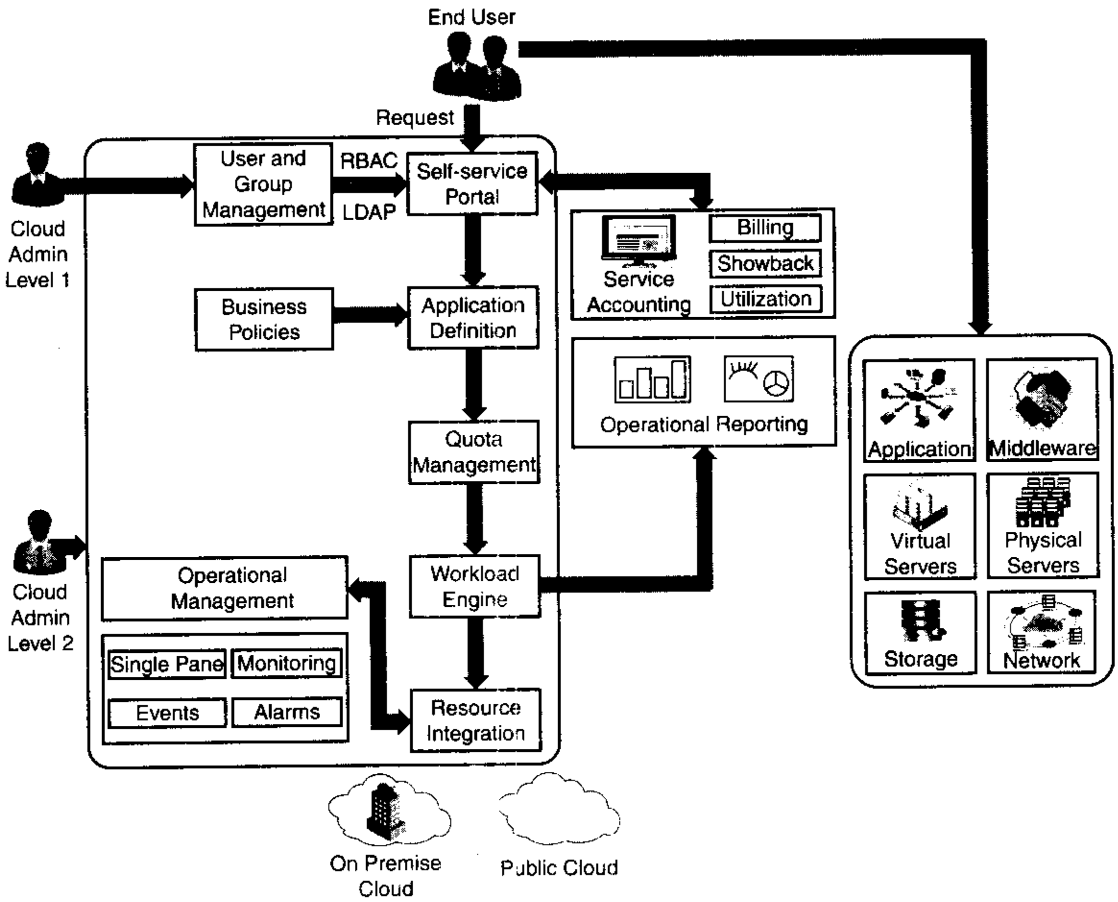

FIGURE 5.3 Cloud orchestration workflow.

Benefits

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

Offering Key Characteristics

- Service layer
  - Self-service portal for different cloud users: Administrators, Cloud Delegates, and End-Users.
  - Chargeback/billing and reporting based on usage and capacity.
- Operations management portal with workload and resource monitoring, alerting, and troubleshooting.
- Applications
  - Automated application provisioning and lifecycle management.
  - Dynamic scaling to meet SLAs.
- Allocation engine
  - Account-based quotas, reservations, scheduling, and approvals for resource allocation.
  - Policy-driven automation of placement, migration, failover.
- Resource integrations
  - Support for popular virtualization platforms.
  - Support for popular provisioning tools.
  - Integration for popular public cloud/external services.
- Datacenter integrations
  - Role-based authentication and authorization.
  - Adapter-based integration to accounting, asset management, change management, entitlement, service catalogue 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 percent 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 expected (average) damage:

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

Security controls are safeguards or countermeasures to avoid or minimize information security risks:

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

Three main types of controls:

- **Preventive:** Prevent security incidents (e.g., patching a vulnerability).
- **Detective:** Detect a security incident (e.g., monitoring).
- **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

Consumer expectation is that security should be built into services themselves. Over 50 percent of potential cloud consumers still avoid online purchases due to 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:

- More evenly distributed security responsibilities.
- Increased transparency from start to finish.
- 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. It is evident that CAP-EX and OP-EX are increasing day by day. Cloud computing and virtualization give them an opportunity to meet with these challenges. On one hand, while it gives us weapons to deal with these challenges, it also gives rise to its own problems like veracity of the virtual environment, data integrity problems, and even security challenge.

Another area to watch is security around Web applications. Average application deployed will have dozens, sometimes hundreds, of defects and a bulk of security threats today target the application layer. Companies must take proactive action to reduce the opportunities for their Web applications to be exploited – long before a hacker even has the opportunity to compromise their business. The most feasible and economic solution is for organizations to catch vulnerabilities as early as possible – and build secure software from the ground up – not bolt it once it has been deployed.

We are moving towards 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 security, and price of the service will be the deciding factor to whether adopt or not. The workloads that possess medium-to-high risk factors will adopt the private and hybrid cloud. This will also cover the workloads that contain proprietary contents and also those that need more security and depth of defence. Once these services mature and settle, 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 procedure to handle change management, incident management, etc. It generates reports for multi-tenant environments. So, we have 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 the non-compliance to the subjected violations. This maintains the visibility in the system.

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

5.5.4 *Identity-Based Protection*

Cloud environment requires extra protection levels as it works with 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 like 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 problem is to make confidential data secure. In order to do this, we have to maintain a secure protocol over the networks, and firewalls should be active to ensure the security of the confidential information; information that is sensitive but not important for the business should be destroyed.

5.5.5 *Data Protection@Cloud*

The 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 regulation of the breaching of the data and its separation on the storage infrastructure. This even includes data that is archived.

This is handled by encryption and managed by encryption keys, and data is 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. One of the biggest problems in Internet-based cloud is sending the large amount of data and that is not possible with the Internet-based environments. Therefore, the data should be encrypted and both the cloud service provider and subscriber have the keys to encrypt it.

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

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

5.5.6 *Application Security@Cloud Deployment*

If you think that the protection mechanism in cloud environment is only there at application level, it is wrong. Actually, it is required for the image level. Therefore, cloud vendors should have a clear and sound way to tackle this by meeting the demands of the subscriber for issuing the licenses for the required period of interval, destroying then after use, and making sure that the sensitive unimportant data is also destroyed at the same time.

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

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

So far, we have talked about the security measures based on software, policies, SLAs, audits, etc., but we should take care of the security in physical terms as well. These security measures can include biometric and security through closed circuit television (CCTV) monitoring. This can restrict unauthenticated entry.

5.6 VIRTUAL DESKTOP INFRASTRUCTURE

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

The notion behind the virtual desktop infrastructure is to run desktop operating systems and applications inside virtual machines that reside on servers in the datacenter. This is referred to as virtual desktops. Users access virtual desktops and applications from a desktop PC client or thin client using a remote display protocol. The Desktop users get almost all local desktop features as if the applications were loaded on their local systems, the difference being that the applications are centrally hosted and managed.

This solution:

- Consists of optional portal interface, thin-client, and PC with client components or Web browsers with client messaging and security technologies, delivered through a single, consistent framework.
- Delivers a common, standards-based, resilient IT infrastructure that is security-rich and scalable and provides authorized users with single-point, consistent access from a wide choice of 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 virtual desktop infrastructure solution that best meets these needs.

5.6.1 Architecture Overview

Desktop cloud virtualization services provide several advantages to the enterprise. One very important advantage 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 devices 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 virtual desktop infrastructure solutions allows businesses to simplify their IT environment, reducing cost and complexity through consolidation of physical resources and standardization of operating environments.

Virtual desktop infrastructure creates a framework that offers many advantages to the enterprise such as:

- **Cost reduction:** More efficient use of resources can increase utilization.
- **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 be run on the virtual machines so that the physical hardware can support a wide range of end users without costly integration or reconfiguring the systems between user accesses.
- **Security:** The data remains in the datacenter with access control.
- **Availability:** Higher availability as VM can be quickly migrated to a different physical server in the event of a hardware failure.
- **Efficiency:** Service delivery is more efficient when IT processes are optimized for a centralized environment.

5.6.2 *Enterprise Level*

Virtual desktop infrastructure provides a set of proven integration patterns and methods for implementing a client virtualization. The virtual desktop infrastructure team works with various tools and products to help users with an assessment of their environment to develop virtual desktop infrastructure solution requirements and a solution design. The team then develops and deploys this solution into the client environment based on a common architecture that supports the four virtual desktop infrastructure solution models.

Virtual desktop infrastructure solution (shared service, virtual client, workstation blades, streaming) is shaped by the component selection. The base virtual desktop infrastructure 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 virtual desktop infrastructure server farm's back-end integrates with existing infrastructure services and legacy applications.

Virtual desktop infrastructure solution introduces a new method of delivering and managing user desktop environments. Virtual desktop infrastructure is a service offering to create a virtual desktop infrastructure. In the IT industry today, several technology vendors provide components that will enable you to build a Virtual Desktop Infrastructure (VDI). VDI is generically used to reference the collection of products and infrastructure components used to form a virtual desktop solution.

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

The virtual desktop infrastructure 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. 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 virtual desktop infrastructure solution provides increased security as no data leaves the datacenter. Optional secure encapsulation capabilities can allow network connections to be encrypted.

Virtual desktop infrastructure virtual client solution integrates into the organization's datacenter to leverage existing network and infrastructure services. The solution provides access services to the virtual infrastructure hosted in the virtual desktop infrastructure server farm or datacenter (Figure 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 virtual desktop infrastructure service engagement if required.

The virtual desktop infrastructure access service or 'connection broker' provides device and user authorization, portal integration, session management, host monitoring, application

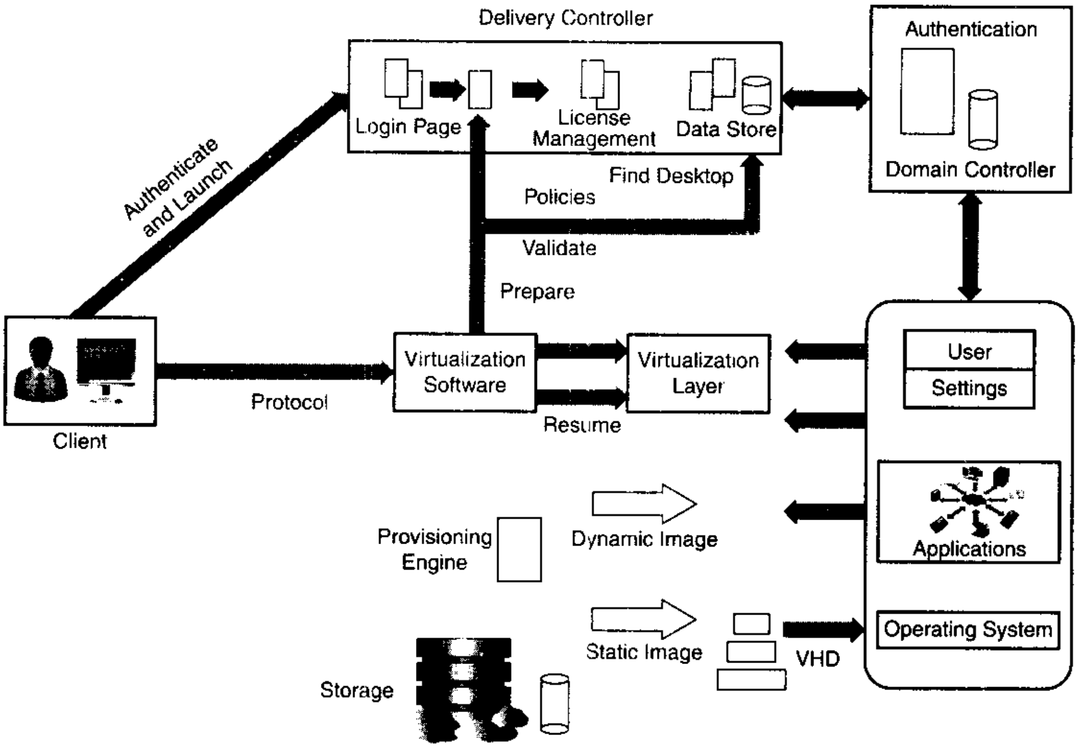

FIGURE 5.4 Virtual desktop infrastructure.

streaming, and consumption-based metering. This 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 single points of access to the virtual environment.

The virtual desktop infrastructure 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.

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 components that are 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 virtual desktop infrastructure 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 business requirements as needed.

The virtual desktops images can be configured to have access to the existing infrastructure services in the client datacenter as needed The following sections briefly discuss the components of the virtual desktop infrastructure at a high 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 supported by a set of products that connects remote clients to the centralized virtual desktops. This process is generically known as connection brokering. The logic controlling which virtual desktop a client should connect is handled by VDI product solution. This makes the process of connecting toVDI Environment simple for the end-user and tightly controlled for the IT administrator.

End-users can use existing PC running operating systems 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, you might have local desktop icons configured to access published applications or published desktops.

5.6.4 *Desktop Virtualization Services*

The core of the virtual desktop infrastructure can be viewed as a central-server-based resource pool with components connecting end-users to applications, networking, and storage resources.

Virtual desktop infrastructure uses vendors to centrally host and deliver a cost-efficient desktop environment from the datacenter, and uses Management server to provide the virtual desktop management.

5.6.5 *Desktop Management*

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

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

5.6.6 *Pool Management for Virtual Desktop Infrastructure*

Management server authenticates the user, determines the pool they belong to, and using predetermined policies, provisions a desktop for that end-user, complete with that user's specifications and privileges, and finally deploys it. These pools can be persistent or non-persistent. Non-persistent pools contain multiple hosted virtual desktops, which are initially identical and cloned from the same template. 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 the desktop 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. The persistent pool contains multiple hosted virtual desktops, which are initially identical because they are cloned from the same template. 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.

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.

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.

A virtualized IT environment helps provide security-rich, 'anytime, anywhere' access to applications, information, and resources. Virtual desktop infrastructure is a unique end-user virtualization solution that helps businesses transform their distributed IT architectures into virtualized, open-standards-based frameworks leveraging centralized IT services. It combines hardware, software, and services to connect authorized users to platform independent, centrally managed applications and full client images running in virtual machines.

Desktop cloud project-based service can substantially reduce 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.

Virtual desktop infrastructure 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, whether a 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 failure of the cloud services.

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

5.7.1 *Value Proposition*

Storage cloud reduces 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.

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

Also, 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 offers 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 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, banking, etc. that deal with large amounts of data increasingly are adopting cloud technology to address their needs around delivering faster services, protecting data in real time, seamless communication between employees, partners, and suppliers, for business continuity and, of course, the pressure to become more energy efficient – to be a greener organization.

5.7.2 *Challenges*

Cloud services rely heavily on keeping the data and applications they are managing available at all times, and the ability to restore operations quickly following any type of data disaster (database corruption, virus attack, hardware failure, local/regional disaster) is essential. Storage management is an important function to ensure that data is available, capacity is provisioned rapidly and storage resources are utilized effectively – cloud administrators often find it difficult to meet all the challenges they face concerning storage and data management:

- Data availability and application performance.
- High capital and operating costs, less return on investment.
- Utilization of storage resources.
- 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:

- Paying for only what they use.
- Cutting costs.
- Monthly payments instead of all up front.
- Having a standardized system.
- Always having the latest software version, since nothing is installed locally.

5.7.3 *Business Drivers*

- Need for standardization and automation of storage services.
- Need to meet service levels consistently – provisioning on-demand computing capacity and storage capacity.
- Need for simplified management of their storage infrastructure – quick provisioning and redeployment of resources, built-in data reduction capabilities.
- Data security and compliance issues.
- Need to lower costs – lack of upfront capital, lower utilization of hardware resources.
- Recovery Point Objectives (RPOs), Recovery Time Objectives (RTOs).

5.7.4 *Benefits*

- Improve service levels by ensuring data availability and application performance and by quick provisioning through automation.
- Reduce capital and operational expenses by leveraging standardization, automation, virtualization.
- Optimize utilization of storage resources and built-in data reduction capabilities to manage more storage with less hardware.
- Reduce hardware, software, and administration costs with policy-based data storage management.
- Manage risk and streamline compliance through real-time data protection.

5.7.5 *Product/Solutions Overview*

Storage management software and services solutions for cloud help ensure that business and IT are fully aligned and supported by integrated service management. They help deliver 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 the need of the broadest, most scalable, and reliable set of storage solutions available to keep cloud services. They 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 storage management in cloud, designed to help streamlining of storage resources 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.

Cloud-based systems have brought a new, scalable application delivery service model to the market. They can help clients save money and increase flexibility. However, for any type of cloud deployment, streamlining data to target applications in the cloud plays an important role – cloud services rely heavily on data availability and application performance. Additionally, a cloud environment cannot afford longer downtime – 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 providers offering cloud-based services (private or public), a key concern is how well the existing resources are optimized in their infrastructure, to improve end-user experience – be able to provide flexibility, speed, reliability, and efficiency. Data and storage management are critical to improve on-demand service experience, reduce costs, and reduce risks in the cloud.

Often, the absence of sophisticated storage management systems in a cloud results in lack of visibility into the storage utilization and provisioning, costs, and associated risks. Cloud administrators have 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 business to support the same in an efficient way, is resulting in unmet service levels, additional downtime, new hardware and operational costs, and lower customer satisfaction.

5.8 SUMMARY

Today, we are becoming more interconnected, instrumented, and intelligent – the world is becoming smaller. As a result, more data is being generated within the operations of all organizations, and they are struggling with managing 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 capacity but managing data efficiently, reducing data, ensuring adequate protection of data, and quick restore for better business performance. We need better implementation of virtual desktop infrastructure, test cloud environments, and analytics to handle cloud offerings.