Realtime publishers

The Definitive Guide™ To

Cloud Computing

sponsored by

IBM®

Dan Sullivan

Copyright Statement

Chapter 6: Identifying the Right Cloud Architecture for Your Business

Cloud computing is a general model for delivering computing and storage services. The model lends itself to a range of implementations with no single architecture constituting a "true cloud" architecture. This is hardly surprising. The defining characteristics of cloud computing (the ability to allocate and release compute and storage resources on demand, a pay-as-you-go funding mechanism, and high levels of self-service) allow cloud providers to deliver a wide range of services using a number of implementation models.

This range of variability means businesses can choose the right cloud architecture for their environments. In this chapter, we will examine several aspects of selecting a cloud architecture:

- Levels of cloud architecture
- Issues in providing compute services
- Issues in providing storage services
- Considerations for network services
- Cloud operations management
- Service layers and adapting IT operations to infrastructures
- Topics in service management

We will start with a brief review of architectural elements common to all cloud architectures.

Levels of Cloud Architecture

Cloud architectures can be thought of in terms of layers of services in which each layer depends on services provided by the next lower layer. As with other layered models of abstraction in software engineering, layers in a cloud control the potential complexity of cloud design by following a few basic principles:

- Services are provided as logical abstractions that hide implementation details. When a program needs to allocate additional storage, for example, it makes a call to a storage service requesting a particular amount of space. There is no need to delve into details about directory structures, files systems, or disk configurations.
- Services are isolated to appropriate layers in the architecture. An application programming interface (API) for storage allocation may make calls to additional services that are not available outside of the storage system. For example, when allocating new storage, an API procedure might call an isolated procedure to add the allocated disk blocks to a list of blocks that are replicated to storage devices for backup and performance reasons.
- Services are provided at a functional level appropriate to the users or services that consume the services. The higher up the stack of services we go, the broader and more business oriented the services. Although lower-level services might operate on storage blocks, upper level services might initiate business process workflows.

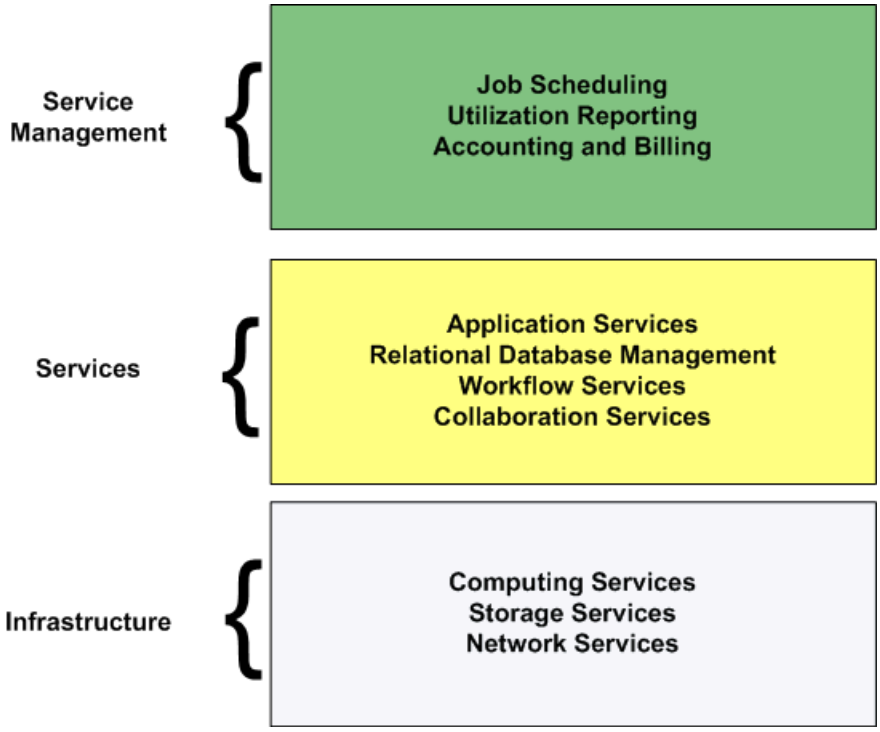

Figure 6.1: Cloud architectures can vary in detail and levels of services provided but most include some combination of infrastructure, platforms, and services management.

Broadly speaking, we can think of three coarse-grained levels of services in a cloud architecture:

- Virtualization of resources
- Services layer
- Server management processes

Each of these levels can be further subdivided.

Virtualization of Resources

One of the hallmark characteristics of a cloud is the virtualization of resources. Virtualization can be thought of as a way of abstracting computing and storage services away from implementation details and toward a more logical and less physical view of resources.

Many of us use virtual servers routinely although we might not know it. We connect to servers across the Internet that run Web sites, email servers, databases, and other business applications. Most of the time, we do not think of the implementation details about these services. Is the email server running on a single physical server? A cluster of load-balanced servers? Or perhaps the application is hosted on a virtual server that shares a physical server with several other virtual machines running an entirely different set of applications. These details are often unimportant, at least from our perspective.

The ability to hide implementation details without adversely affecting services is essential to providing cloud computing. Virtualization is especially important for efficiently using computing and storage infrastructure. (We will focus primarily on server virtualization here and address storage virtualization later in the section entitled "Providing Storage Services.")

Logical Units of Computing Resources

Server virtualization allows us to manage compute resources in finer-grained units than just a physical server allows. One of the first advantages of this approach is that we can work with a standardized set of features, such as the number of CPU cores and amount of RAM. For example, a standard virtual server might be equivalent to a physical server with one Intel Xeon 5600 series or AMD Opteron 6000 series processor and 8GB. One could also define virtual servers in terms of performance relative to standard benchmarks, such as the Transaction Processing Performance Coucil's (http://www.tpc.org/tpcc/default.asp) online transaction processing (OLTP) benchmarks (TPC-C and TPC-E) and the ad hoc, decision support benchmark (TCP-H). How the logic units are defined is less important than the fact that we have a standard for allocating computing resources that is not tied to a particular physical implementation.

By decoupling how we allocate computing resources from the underlying hardware that provides those resources, we gain flexibility in managing how we consume compute services and manage them.

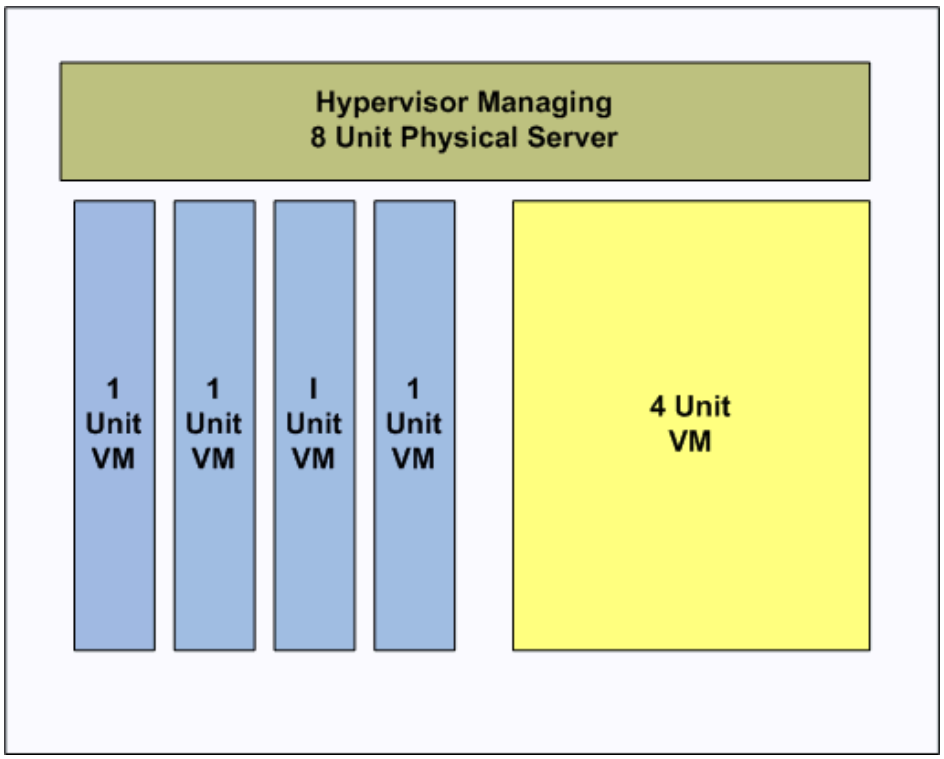

Figure 6.2: Server virtualization allows cloud service consumers to use standardized units of computing services without concern for the physical implementation details.

Hardware Independence

Another advantage of virtualization for cloud service providers is hardware independence. Cloud consumers can allocate the level of computing resources they need without having to worry about whether a particular physical server is a 2, 4, or 8 core server. Cloud providers can deliver those logical units using the most economical way possible. For example, a cloud might have several types of physical servers running in the cloud. The less energy efficient servers are only used when the more efficient servers are running at peak capacity. The first time a cloud consumer runs a job, the job might run on one of the more energy efficient servers; the next time the same job runs on the other type of server.

Standardized Service Pricing

Along with logical units of computing resources and hardware independence, virtualization allows for standardized service pricing. Although this is not a technical issue, it has a direct impact on how cloud service consumers plan and manage their use of the cloud.

Virtualization of services is an essential element of a cloud architecture. It directly enables the most efficient allocation of resources, reduces the need for cloud service consumers to understand the nuanced differences in physical servers, and provides for a straightforward pricing model that consumers can use for planning and budgeting.

Services Layer

The services layer is another common characteristic of cloud architectures. At this level, we work with not just virtualized hardware but also operating system (OS) and application services. It is certainly possible to provide a cloud that offers only infrastructure services (that is, the virtualized equivalent of bare metal machines), but for business users of cloud services, the services layer can provide additional benefits.

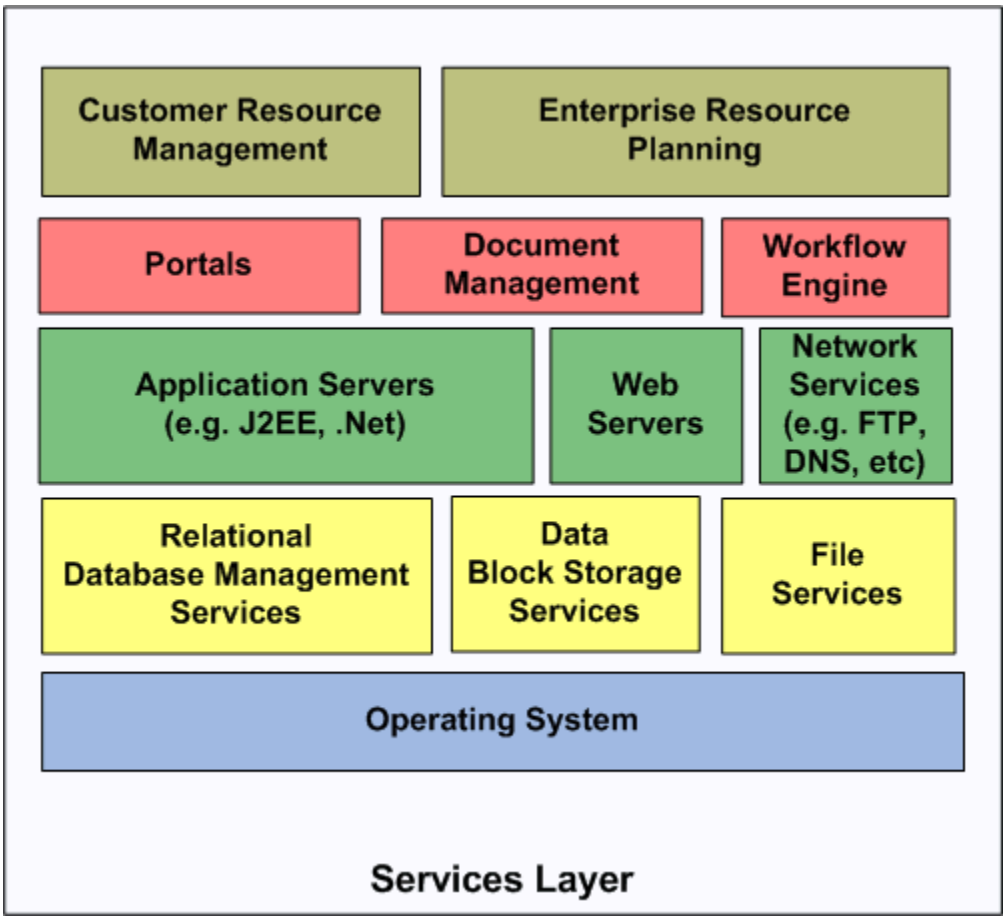

Figure 6.3: The services layer consists of a wide range of service types, some of which build on others within the same layer.

Services such as those shown in Figure 6.3 might be delivered in different ways to customers. OSs of course are included in the virtual machine images, but other services might be independent of virtual machine instances. Persistent storage services, such as block storage and relational database services, might be available as services available to all virtual machine instances running in the cloud. Higher-level services, such as application servers, portals, and workflow engines, might be embedded within virtual machine instances along with other software stack components. At the highest levels, business applications such as CRMs and ERPs may be provided as Web applications that run in the cloud. At this level, service consumers are completely divorced from implementation details and are solely concerned with business-related functionality.

Service Management Processes

A third major aspect of cloud architectures are the service management processes that support the delivery of services. These include:

- Virtual machine image management
- Image deployment
- Job scheduling
- Usage accounting
- Management reporting

The first two of these services supports a catalog of images preconfigured to particular applications, software stacks, or OSs that can be deployed by cloud service consumers.

Job scheduling applications help with routine processes that run repeatedly on a schedule as well as large, one time jobs that can be submitted to run in the cloud as services are available. Job scheduling services are especially useful when services pricing varies by point in time demand or time of day.

Usage accounting and management reporting are necessary for billing or charge-backs on the part of cloud service providers and for cloud service consumers who must plan and manage their budgets for IT services.

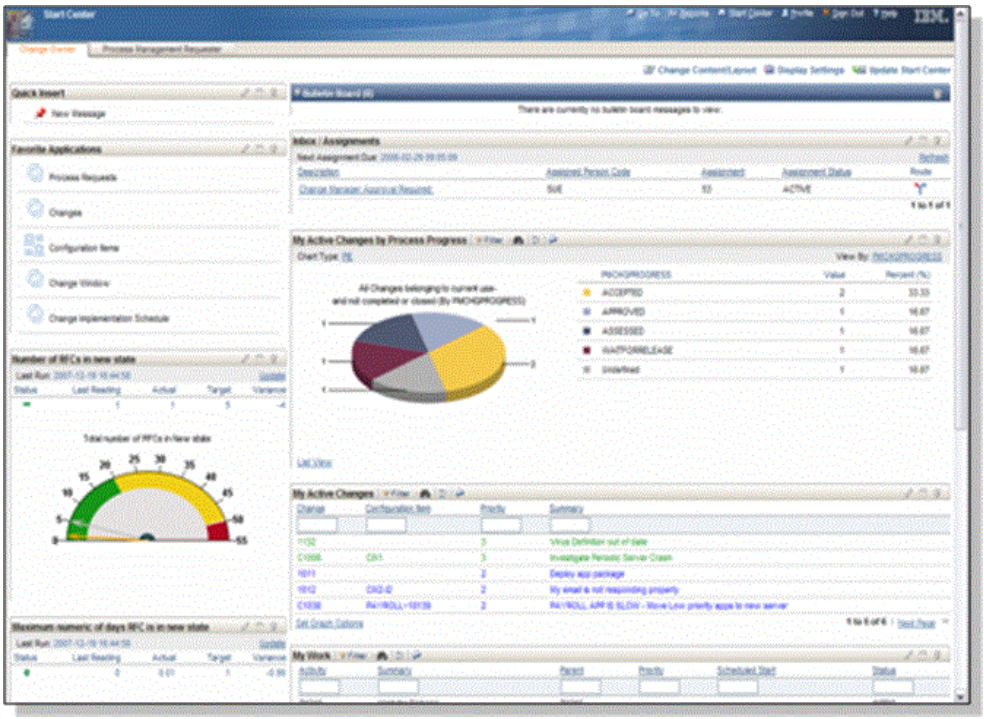

Figure 6.4: Management reporting serves the needs of both cloud service providers and consumers.

To summarize, cloud architectures can be described in terms of infrastructure, services, and services management. Variations in these layers allow for different types of cloud architectures. Multiple factors will determine the best choice of architecture for a particular set of business requirements. The remaining sections of this chapter will delve into those factors.

Providing Compute Services

There are compelling reasons to adopt a cloud architecture that include an internal or private cloud. Businesses maintain total control over computing resources with a private cloud. This can significantly reduce compliance issues with cloud computing. Private and confidential data is not moved outside the company, data destruction policies and procedures are defined by the business, and systems are not shared with outsiders, including potential competitors. With the advantages come additional functional responsibilities.

Businesses that choose to provide private clouds or hybrid private/public clouds must be in a position to provide the physical infrastructure and basic management services needed in a cloud. (Businesses can provide higher-level services, such as enterprise applications, as cloud applications while using a public or other third-party physical infrastructure.) Those that will deliver computing services directly though a private cloud should consider:

- Hardware selection
- Implementing virtualization
- Failover and redundancy
- Management reporting

A business' ability to address each of these issues can strongly influence their success in delivering computing services in a cloud.

Hardware Selection

Hardware selection for clouds depends upon two competing interests: controlling costs by redeploying existing hardware versus acquiring a standardized server platform that is configured specifically for cloud computing. Using existing hardware can lower initial capital expenditures but might lead to higher costs over the long term. Older machines that require more maintenance, need parts that are difficult to procure, or consume more electricity can have a larger total cost of ownership than new servers. One option is to use existing hardware initially and replace it over time as the cost of new servers becomes competitive with the cost of continuing to operate the older devices.

An advantage of new hardware is that the cloud can be configured with standard servers optimized for cloud computing: large numbers of CPU cores, significant amounts of memory, high speed I/O for connections to disk arrays, and so on. Standardization also helps reduce maintenance costs.

Implementing Virtualization

Many organizations use virtualized servers outside of clouds; however, virtualization in the cloud requires more management services than typical in IT environments. Conventionally, managed virtual servers are installed by systems administrators and run for extended periods carrying out a fixed set of functions. Additional controls are available in some environments that support virtual machine migration from one physical server to another. This is especially useful in situations in which a single server is running at or near capacity and one or more of the virtual machines needs to be moved to a less utilized physical server. Even this, though, does not meet the level of virtualization management needed in a cloud.

Clouds should support end user management of computing resources. A knowledgeable user should be able, for example, to select a virtual machine image from the catalog and instantiate a specified number of virtual servers.

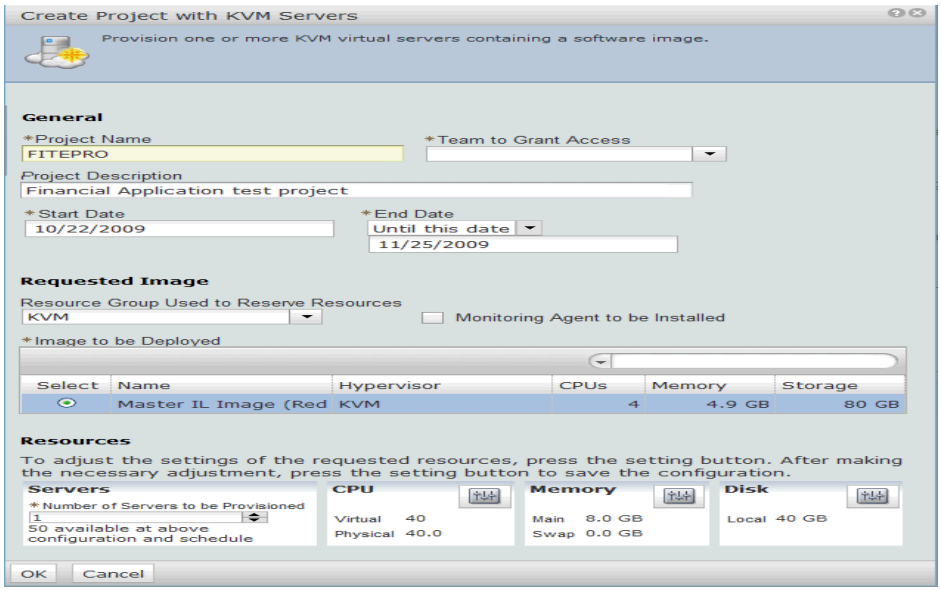

Figure 6.5: Providing computing services in a cloud requires significant support software, such as services for selecting and instantiating virtual machine instances.

Failover and Redundancy

An advantage of cloud architectures is that we move away from tightly coupling applications and services to particular physical or virtual servers. Applications are run on virtual servers that meet a set of configuration requirements defined by the cloud service user. Applications that are well suited for the cloud do not need specialized hardware or a particular server. This reduces the challenge of providing failover services.

Clouds are inherently redundant. If a physical server fails for any reason, it can be removed from the pool of available resources. Virtual machine images are deployed to other physical servers until the failure is corrected. This type of failover and redundancy is at the server level, not the application level.

If a physical server were to fail while an application were running on it, recovery would depend on the application. For example, if the application provided stateless Web services, it could be restarted on another virtual machine instance on another physical server and start responding to service requests again. In cases where the application writes state information to persistent storage and checks for prior execution information each time the application starts, the application could also recover fairly robustly on another virtual machine.

Management Reporting

Different types of management reports are required when providing computing services in a cloud. In a traditional "one server-one application" approach, the business owner of a process is responsible for identifying the servers needed to support a business process and covering the cost of the servers, either virtual or physical. In this model, there is fairly little to report outside of utilization rates. The business process owner is paying for sole use of servers, so there is not much incentive to monitor server use as long as it does not adversely affect performance.

Cloud service consumers can use and should expect detailed usage reporting. With a pay-as-you-go pricing model, there is an incentive to allocate the fewest number of virtual servers and run them for the shortest time possible while still meeting business requirements. Cloud service consumers can use reports detailing:

- Number of virtual servers allocated to a job and the time the servers ran
- Peak and average utilization rates of servers
- Amount of data stored persistently
- Amount of data transferred across the network
- Charges for compute, storage, and network services

Detailed utilization information will help business process owners optimize their applications. For example, if analytic servers are running at 40% utilization because they are dependent on data preprocessing operations that are not processing data fast enough, additional servers could be instantiated for preprocessing. Presumably the cost of running the additional preprocessing servers would be offset by reducing the length of time the servers have to run. The analytic servers would run at higher utilization and for shorter periods of time reducing the overall cost of the process.

Providing computing services in a private or hybrid cloud requires a combination of hardware, virtualization management and deployment systems, a server configuration that supports failover and redundancy, as well as robust management reporting.

Providing Storage Services

If a business moves forward with providing private cloud computing services, it will have to provide storage services as well. This would require additional support services:

- Storage virtualization
- Backup or other redundant storage
- Disaster recovery

Storage Virtualization

Storage virtualization, like server virtualization, abstracts the services provided by hardware. Consumers of these services can allocate resources without concern for implementation details. For example, details like the logical unit number (LUN) mappings to storage volumes and storage devices are managed by storage virtualization software. When persistent storage is needed, the cloud services consumer simply makes a call to a programming interface specifying the amount of storage required.

> **Local vs. Cloud Storage**
>
> Virtual machine instances can provide local storage for temporary storage during the life of the virtual machine instance. The data in this storage is lost when the virtual machine is shut down. The persistent cloud storage described here is provided by devices that are independent of virtual machines. Multiple virtual machines can access the same storage blocks and the data continues to exist regardless of how virtual machines are started and stopped.

The advantages of virtualized storage are similar to those of virtualized servers:

- More efficient use of storage—rather than dedicating large units of storage to a single use for extended periods of time, storage is allocated in smaller increments and for only as long as it is needed
- Lower capital expenditures for individual projects and business units that do not have to acquire storage hardware
- Lower operating costs associated with the pay-as-you-go model typical in cloud computing storage
- More efficient delivery of backup and recovery services

This last benefit is especially important.

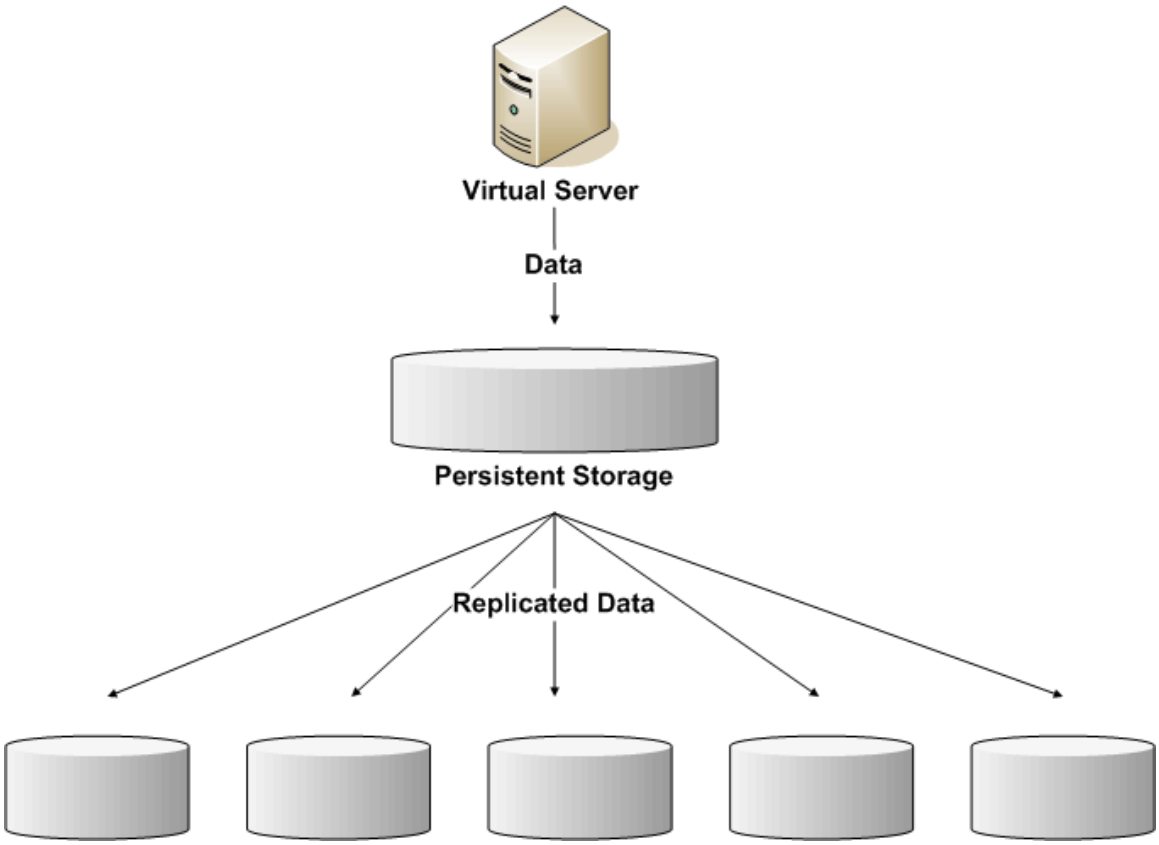

Figure 6.6: Cloud storage systems can use data redundancy to improve data management performance and reliability of data services.

One of the advantages of virtualized storage is the ability to provide large amounts of storage through a single logical device—the storage cloud. Behind the scenes, of course, we have multiple disk arrays possibly located in different facilities. This setup creates an opportunity to replicate data across multiple storage arrays to improve reliability and performance.

Reliability is preserved because multiple copies of data are available. If a storage device should fail, there is no need to restore from backup tape; the data is immediately available from another device. The particular device that returns the data is irrelevant to the user. Replication can be done asynchronously so that I/O operations return as soon as data is written to the primary storage device. A background replication process can add new or changed blocks to a queue of blocks that will be copied to devices.

Users can also benefit from improved performance with multiple copies. Data warehousing and business intelligence applications often query large amounts of data. Users contending for access to a single copy might experience bottlenecks and associated drops in performance. In the cloud, different queries can be served by different copies of the database, relieving contention for the same resource.

This type of replication also supports disaster recovery. In the event of a catastrophic failure in one data center, users could be re-routed to another data center that maintains replicated copies of the lost data.

Figure 6.7: Storage virtualization supports data replication across data centers, which improves reliability and performance.

This type of replication does not eliminate the need for backup, however.

Backups and Cloud Storage

Data replication as just described is a valuable asset in cases of disaster recovery, but it cannot meet all recovery requirements. The ideal replication solution maintains multiple copies of data in near real time, so any errors generated in the source system will be replicated to other storage devices as well. Without a separate backup copy of data, there would be no way to restore the database back to a point in time before the error was introduced.

Backup services are generally specified in terms of recovery point objectives (RPOs) and recovery time objectives (RTOs). An RPO defines points of time in history that can be restored; examples include previous day at midnight, previous end of week, or in the case of highly volatile databases, a previous time in the same day. RTOs define the maximum period of time between request of a restore operation and the time the restore operation completes.

Traditional backups are easily accommodated in the cloud. Source data is backed up from the cloud and written to cloud storage. The process could be as simple as copying and compressing data files or block storage from one storage area to another. If backup software supports direct reads and writes to cloud storage, backup processes can take advantage of incremental and differential backups reducing the total amount of space needed to store backup files.

Management Reporting for Storage Virtualization

A reporting framework, similar to one needed for server virtualization, is required for storage virtualization as well. Businesses that deploy shared disk arrays will probably have a storage reporting system in place that provides much of the needed functionality:

- Reporting on storage used by project, department, or other billable unit
- Cost of storage by type, such as primary storage versus archival storage
- Trending reports on growth in storage use

Infrastructure managers should have additional detailed reports on such things as replication performance.

Storage virtualization and server virtualization share many of the same benefits and management requirements. Together with networks services, they constitute the core infrastructure for cloud services.

Network Services for Cloud Computing

Networking can be the most resource constrained part of cloud infrastructure. Public cloud providers are necessarily dependent on public Internet providers for connectivity between their data centers and their customers. Private cloud providers might also depend on public Internet providers, especially for access from remote offices or smaller corporate facilities. Dedicated network connections can be employed between sites, but cost is a limiting factor. The key issues we must consider when evaluating different cloud architecture options are:

- Capacity
- Redundancy
- Management reporting

Capacity

Network capacity limits the amount of data that can move between data centers and between cloud service consumers and the cloud. This directly affects a number of services within the cloud.

Intra-Cloud Replication

From an infrastructure management perspective, network capacity and speed directly affect replication. As noted earlier, replication is an essential element of creating and maintaining a reliable, high-performance cloud. Heavy demands for loading data into the cloud not only create demand to get data into the cloud but also lead to additional network I/O due to replication. Cloud administrators might determine, for instance, that given the mean time between failures (MTBF) on disk drives, cloud-stored data should be replicated four times to reduce the probability of data loss to whatever threshold they have defined. This means that all data loaded into the cloud plus data generated or updated by cloud-based operations will need to be copied over the network four times.

Loading Data into the Cloud

Cloud computing is an ideal approach to analyzing large amounts of data. In fact, the phrase "Big Data" has become a moniker for use cases where traditional data management methods break down. The need to deal with multi-terabyte and even petabytes of data used to be a problem limited to specialized niches, such as national intelligence and astrophysics; today, the problem spans industries such as financial services, retail, pharmaceuticals, government, and life sciences.

Businesses with large data sets can leverage large numbers of servers to process and analyze "Big Data" in parallel using platforms such as Apache Hadoop (http://hadoop.apache.org/). It is not always practical to move large amounts of data over networks to load it into the cloud. In such cases, it is best to bypass the network and employ a cloud version of "sneaker net" (that is, ship hard drives to data centers).

> **Hadoop and Related Tools**
>
> Hadoop is an open source implementation of the map reduce model made famous by Google. In addition to supporting massively parallel processing over clusters of computers, it includes a scalable database (HBase), a data warehouse infrastructure (Hive), a high-level data flow language (Pig), and a coordination service for distributed applications (ZooKeeper).

Network capacity can be a limiting factor in cloud architectures if a large amount of data (relative to network capacity) has to be moved into the cloud. In some use cases, this is only a problem during the transition to cloud computing when initial data is loaded; after that, data is generated in the cloud using cloud-based servers. In other cases, data may be generated outside the cloud by sensors and other instrumentation; in such cases, we would need to design network capacity to meet large-scale data transfers over the long term.

Redundancy in the Network

Both computing and storage services in the cloud use redundancy to mitigate the risk of failures. When servers fail, they are removed from the pool of available resources. When storage devices fail, data is retrieved from another device with a redundant copy of the data. Network services require similar redundancy to avoid a single point of failure.

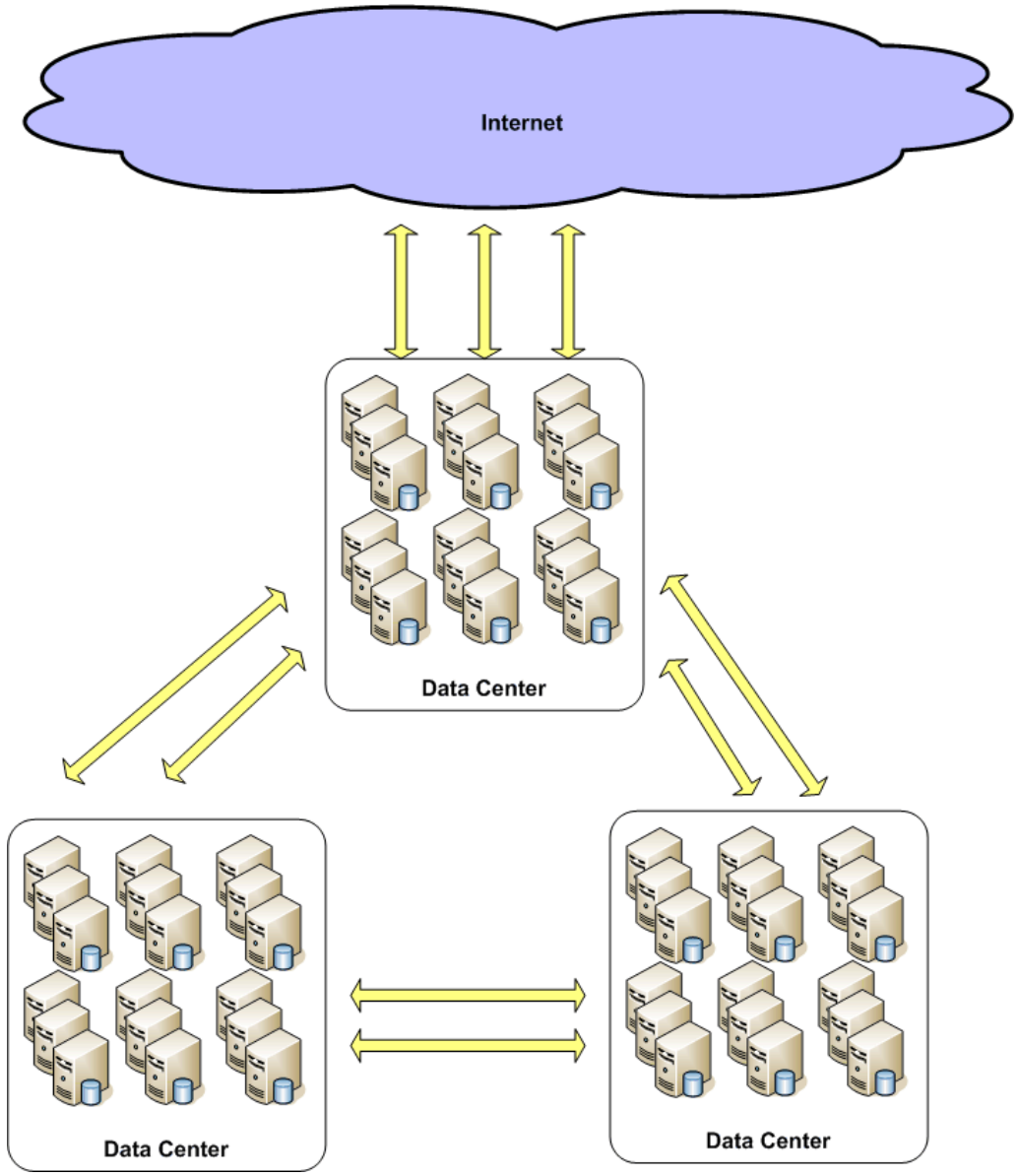

Figure 6.8: Redundant network connections are necessary between data centers as well as to the Internet.

Management Reporting

Cloud service users will be interested in network use reporting as a means to control costs and to monitor trends in network usage. We should not underestimate the cost of network services. For example, when dealing with "Big Data," the cost of network I/O can exceed the cost of computing and storage resources. Management reports can be especially useful if they provide a detailed breakdown of network use by time period and by job. Aggregate reporting over extended periods of time are also needed to determine baseline usage rates, cyclical patterns of variation in network utilization, and long-term growth trends.

Network services, computing services, and storage services are the foundation of cloud computing. Each of these components are provided in redundant manners supporting reliability and increased performance. Management reporting is required in all three areas. In addition to the requirements mentioned, there are further demands for operations support.

Cloud Operations

Maintaining an efficient cloud operation requires management support mechanisms in addition to those previously described; in particular, image management and workload management. These are tasks associated more with overall cloud management than with individual uses of cloud services.

Image Management

A cloud can only instantiate the virtual machine images available in the cloud's catalog. The catalog constitutes the baseline set of services provided in the cloud. Users can install additional services, of course, but once a virtual machine is shut down, those changes are lost. The next time that system is required, the additional software must be installed again. For many situations, the cloud catalog constitutes the set of applications and platforms that can run in the cloud.

Machine images can include a fairly wide range of software in addition to the base OS:

- Application servers
- Software libraries
- Analytic software
- Business-specific applications

The base OS as well as the optional software will need to be maintained over time. Each image in the catalog will have to be routinely patched, scanned for vulnerabilities, and rebuilt as new versions of core components become available.

Workload Management

Workload management functions can vary from basic job scheduling to job optimization. Job scheduling software is useful for queuing large jobs or for repeated jobs in the cloud. The information managed in the job scheduler is useful for tracking future use of cloud services. If metadata about previously run jobs—such as number of servers used, duration of jobs, amount of network I/O, and so on—is collected, it can provide data for estimating future demands on various cloud resources.

Clouds, like any other IT resource, can be optimized. All things being equal, users might prefer to run large jobs overnight and shorter jobs during the workday. This may lead to peak demands that are significantly higher than low demand periods. For example, users may run most data loading jobs at night, leading to periods where demand exceeds capacity while network capacity is underutilized during the day. This type of skewed demand schedule may be smoothed by adjusting price of services. If network resources are in high demand at night, the price is higher than in the day. If demand for computing servers is low in the early hours of the business day, the hourly price for servers is reduced.

Figure 6.9: Demand for cloud resources can be smoothed by varying the price of services to shift demand away from peak periods to low-demand periods.

Software for cloud operations such as image management and workload management are necessary to ensure clouds operate in an efficient manner. Up to this point in the discussion, we have focused on lower-level services, such as virtual servers, storage, and network I/O, and management of those low-level services. Information technology services also provide high-level functions.

Services Layer: Adapting IT Operations to Cloud Infrastructure

The cloud is an ideal platform to run many, although certainly not all, business services. Applications written to take advantage of mainframe capabilities and tuned to run on mainframe OSs are probably best run on that platform. Many business applications are already running on distributed platforms, taking advantage of clusters of servers, shared storage devices, and high-speed network interconnections. These applications are ideal candidates for deploying to a cloud, but there are still additional factors that should be considered when moving systems to the cloud:

- Designing for recoverability
- Managing workload
- Performing maintenance and upgrades
- Maintaining security

These are all considerations in service delivery, but cloud architectures influence how we address them.

Designing for Recoverability

Recoverability is an issue at the application level as well as at the infrastructure level. An application that depends on a large pool of servers to analyze data should address questions such as:

- What happens if a single server fails?
- Will the job have to be restarted from the beginning?
- Is there a way to detect what data was being analyzed when the server failed?
- Is there a way to roll back to a prior state without starting from the beginning?

There are many design choices for addressing these types of questions. For example, each server can receive a subset of data from a distribution node. The distribution node maintains a queue of data sets to distribute to analysis servers. When the distribution node receives a message that a data set has been analyzed, it is removed from the queue. In this way, if a server fails while analyzing data, the data will simply be sent to another server for processing. To avoid a single point of failure, this solution would also require a failover mechanism to start another distribution node should the primary one fail. Alternatively, multiple distribution nodes could run simultaneously and use persistent cloud storage to maintain the queue of data sets that could be read by any of the distribution nodes. This is just one example of a resilient application design for distributed computing; there are many others.

Managing Workload

Providing services through the cloud will require us to think of jobs and workloads in ways that we do not necessarily need to when we have full control of dedicated servers. In particular, we will want to maximize server utilization when we run our jobs while ensuring jobs finish in whatever time window required. If, for example, our cloud charges a minimum of 1 hour of server time for each instance, and we have several small workloads, we should run those in tandem on a single virtual server rather than run them on different servers each incurring the minimum charge.

Performing Maintenance and Upgrades

Maintenance and upgrades of applications will have to be coordinated with the cloud service provider. When departments or projects manage their own servers, they can determine their own upgrade schedule (within broader company policies, anyway). In the cloud, applications are delivered through virtual machine images maintained in the centrally managed image catalog. Similarly, patching and other maintenance decisions will have to be coordinated with the cloud provider.

Maintaining Security

Fundamental security considerations continue to persist in the cloud. Of particular importance is the need to manage identities and entitlements in the cloud. If private information is stored in the cloud, appropriate application-level controls will have to be in place to prevent unauthorized access. Direct access to the private data via the persistent storage API will also have to be blocked through authentication mechanisms and access control lists (ACLs) or other authorization control.

In addition to access controls, we must consider application-level security issues such as vulnerability scanning. Ideally, security concern is addressed by the cloud service provider, but customizations might be the responsibility of the application owner.

Service Management Layer

A final piece of the software and infrastructure architecture that makes up a cloud is the service management layer. Throughout this chapter, we have considered core computing, storage, and network services from both the service provider and the service consumers' perspective. We have seen the overlap in concerns between both parties for issues such as image management, workload management, and optimization of resources. This overlap and shared need for support service continues as we consider the service management layer.

Service management includes additional services necessary for managing the business of providing and using a cloud. These include:

- Provisioning, which are services that allow non-IT professionals to deploy cloud services as needed
- Performance management, which provides additional management reporting and monitoring services that allow cloud providers to understand detailed operations in the cloud as well as plan for longer-term management issues
- Usage accounting, which is necessary for tracking who uses which services and for how long; this is essential for proper cost allocations or billing for cloud services
- License management services are important for compliance; running a cloud does not necessarily entitle one to run as many instances of a commercial off-the-shelf product as one would like—cloud service consumers cannot not be expected to monitor the number of copies of licensed software running in the cloud or to know licensing details, thus license management systems are needed to ensure compliance

Support services such as these, and others related to service monitoring and availability, provide the higher-level management services necessary, especially when running a private cloud.

Summary

Cloud services can be provided with a number of architectures, and a wide range of factors need to be considered when choosing to deploy a cloud. Issues related to providing computing services, storage services, and network services all come into consideration at the most fundamental levels. Reliability, performance, and management reporting are recurring themes when considering those three core services. In addition, cloud operations management, adapting IT operations to cloud architectures and topics, and service management must be examined as businesses choose the right cloud architecture for their situations.