Chap-06

Advanced Topics in Cloud Computing

- Cloud computing is a rapidly moving target.

- New technological advances and application services are regularly introduced.

- There are many open challenges, especially in the context of energy-efficient management of datacenters and the marketplace for cloud computing.

- Objectives
  - Overview of various open issues in cloud computing that need long-term investigation.
  - Issues involved in energy-efficient cloud computing and presents a “green” cloud computing architecture.
  - Market models needed for realizing an open market for cloud computing systems from the perspective of federations of clouds and agreements between clouds.
  - A general overview of some of the existing standards that enable interoperation between clouds and a brief look at third-party cloud services are presented.

1. Energy efficiency in clouds

– Modern datacenters that operate under the cloud computing model are hosting a variety of applications ranging from those

  - that run for a few seconds (e.g., serving requests of Web applications such as ecommerce and social network portals)

  - to those that run for longer periods of time (e.g., simulations or large dataset processing) on shared hardware platforms.

– The need to manage multiple applications in a datacenter creates the challenge of on-demand resource provisioning and allocation in response to time-varying workloads.

- Normally, datacenter resources are statically allocated to applications based on peak load characteristics
  - in order to maintain isolation and provide performance guarantees.

- According to the McKinsey report on “Revolutionizing Data Center Energy Efficiency”, a typical datacenter consumes as much energy as 25,000 households.

- Energy costs of powering a typical data center doubles every five years.
- Because energy costs are increasing while availability dwindles,

- there is a need to shift focus from optimizing datacenter resource management for pure performance alone to optimizing for energy efficiency while maintaining high service-level performance as shown in Figure

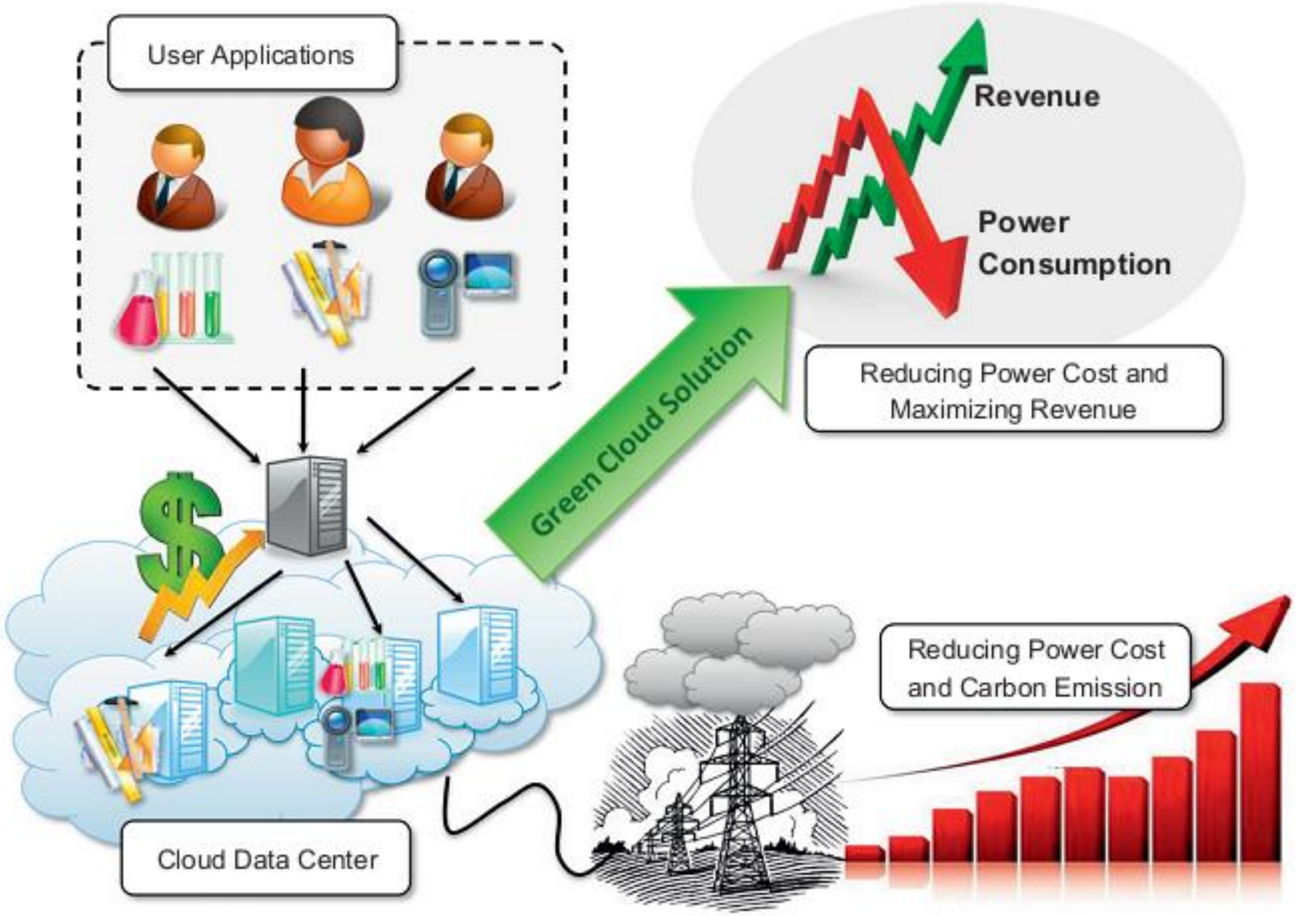


FIGURE 11.1

A "green" cloud computing scenario.

- Datacenters are not only expensive to maintain, they are also unfriendly to the environment.

- Carbon emissions due to datacenters worldwide are now more than the emissions of both Argentina and the Netherlands.

- High energy costs and huge carbon footprints are incurred due to the massive amount of electricity needed to power and cool the numerous servers hosted in these datacenters.

- Cloud service providers need to adopt measures to ensure that their profit margins are not dramatically reduced due to high energy costs.

- According to Amazon’s estimate, the energy-related costs of its datacenters amount to 42% of the total budget,
  - which includes both direct power consumption and the cooling infrastructure amortized over a 15-year period.

- As a result, companies such as Google, Microsoft, and Yahoo! are building large datacenters in barren desert land surrounding the Columbia River in the United States to exploit cheap hydroelectric power.

- There is also increasing pressure from governments worldwide to reduce carbon footprints, which have a significant impact on climate change.

- To address these concerns, leading IT vendors have recently formed a global consortium, called The Green Grid, to promote energy efficiency for datacenters and minimize their impact on the environment.

- Lowering the energy usage of datacenters is a challenging and complex issue
  - because computing applications and data are growing so quickly that larger servers and disks are needed to process them fast enough within the required time period.

- Green cloud computing is envisioned to achieve not only efficient processing and utilization of computing infrastructure but also minimize energy consumption.

- This is essential for ensuring that the future growth of cloud computing is sustainable.

- Cloud computing, with increasingly pervasive front-end client devices such as iPhones interacting with back-end datacenters, will cause an enormous escalation in energy usage.

- To address this problem, datacenter resources need to be managed in an energy-efficient manner to drive green cloud computing.

- In particular, cloud resources need to be allocated not only to satisfy QoS requirements specified by users via service-level agreements (SLAs) but also to reduce energy usage.

- This can be achieved by applying market-based utility models to accept user requests that can be fulfilled to enhance revenue along with energy-efficient utilization of cloud infrastructure.

1.1Energy-efficient and green cloud computing architecture

– A high-level architecture for supporting energy-efficient resource allocation in a green cloud computing infrastructure is shown in Figure

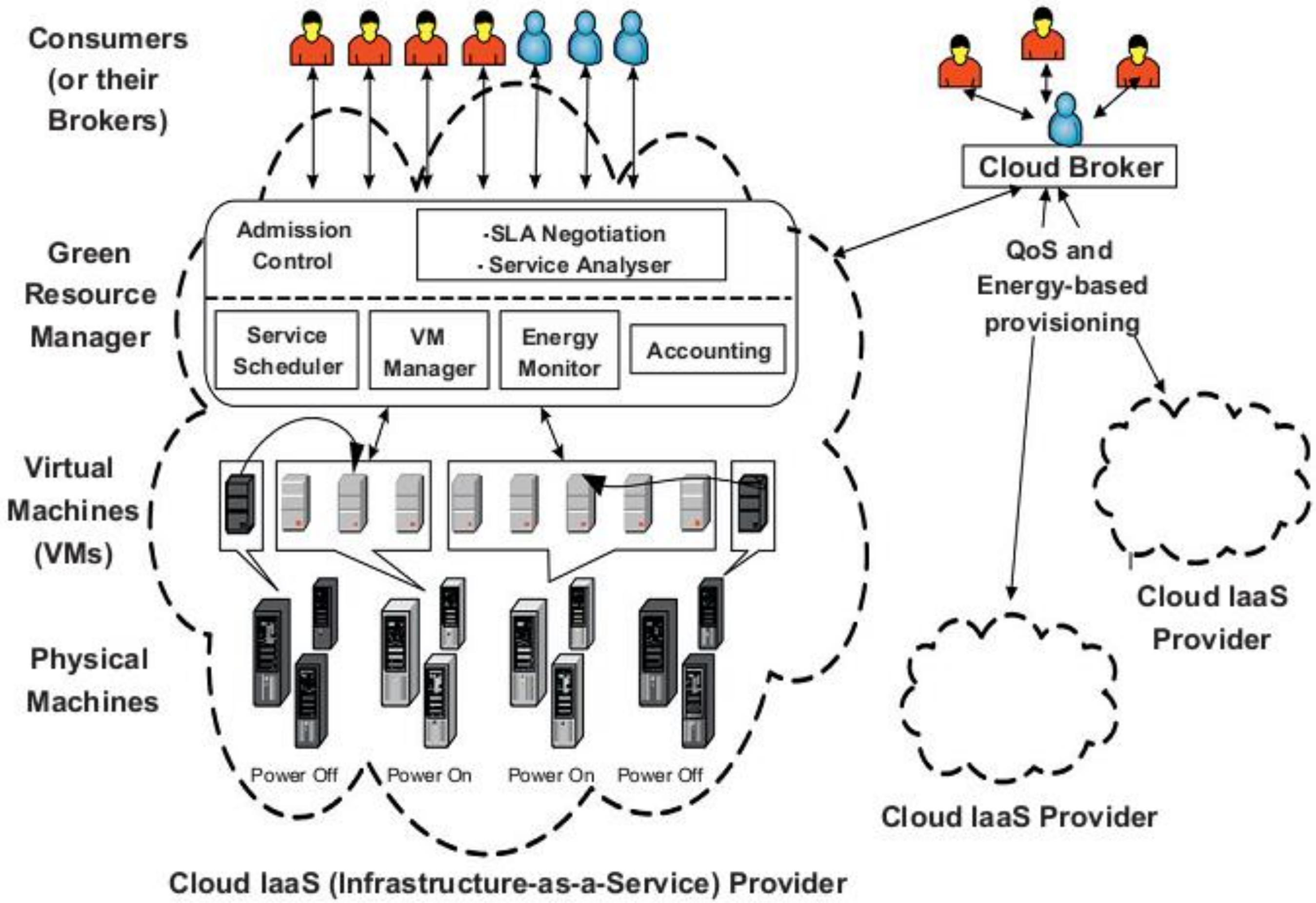


FIGURE 11.2

High-level system architectural framework for green cloud computing.

- It consists of four main components:
  - Consumers/brokers: Cloud consumers or their brokers submit service requests from anywhere in the world to the cloud.
    - It is important to note that there can be a difference between cloud consumers and users of deployed services.
    - For instance, a consumer can be a company deploying a Web application, which presents varying workloads according to the number of “users” accessing it.
  - Green Resource Allocator: Acts as the interface between the cloud infrastructure and consumers.
    - It requires the interaction of the following components to support energy-efficient resource management:

- Green Negotiator: Negotiates with the consumers/brokers to finalize the SLAs with specified prices and penalties (for violations of SLAs) between the cloud provider and the consumer, depending on the consumer's QoS requirements and energy-saving schemes.
  - In Web applications, for instance, the QoS metric can be 95% of requests being served in less than 3 seconds.

- Service Analyzer: Interprets and analyzes the service requirements of a submitted request before deciding whether to accept or reject it.
  - Hence, it needs the latest load and energy information from VM Manager and Energy Monitor, respectively.

- Consumer Profiler: Gathers specific characteristics of consumers so that important consumers can be granted special privileges and prioritized over other consumers.

- Pricing: Decides how service requests are charged to manage the supply and demand of computing resources and facilitate prioritizing service allocations effectively.

- Energy Monitor: Observes and determines which physical machines to power on or off.

- Service Scheduler: Assigns requests to VMs and determines resource entitlements for allocated VMs. It also decides when VMs are to be added or removed to meet demand.

- VM Manager: Keeps track of the availability of VMs and their resource entitlements. It is also in charge of migrating VMs across physical machines.

- Accounting: Maintains the actual usage of resources by requests to compute usage costs. Historical usage information can also be used to improve service allocation decisions.

- VMs: Multiple VMs can be dynamically started and stopped on a single physical machine to meet accepted requests,
- hence providing maximum flexibility to configure various partitions of resources on the same physical machine to different specific requirements of service requests.

- Multiple VMs can also run concurrently applications based on different operating system environments on a single physical machine.

- In addition, by dynamically migrating VMs across physical machines, workloads can be consolidated
- and unused resources can be put on a low-power state, turned off, or configured to operate at low performance levels to save energy.

- e.g., using Dynamic Voltage and Frequency Scaling, or DVFS

- Physical machines: The underlying physical computing servers provide hardware infrastructure for creating virtualized resources to meet service demands.

1.1.1 Energy-aware dynamic resource allocation

- Recent developments in virtualization have resulted in its use across datacenters.
- Virtualization enables dynamic migration of VMs across physical nodes according to QoS requirements.
- Unused VMs can be logically resized and consolidated on a minimal number of physical nodes, while idle nodes can be turned off (or hibernated).
- Through consolidation of VMs, large numbers of users can share a single physical server, which increases utilization and in turn reduces the total number of servers required.
- Moreover, VM consolidation can be applied dynamically by capturing the workload variability and adapting the VM placement at runtime using migration.

- Currently, resource allocation in a cloud datacenter aims at providing high performance while meeting SLAs, with limited or no consideration for energy consumption during VM allocations.
- However, to explore both performance and energy efficiency, two crucial issues must be addressed.
- First, turning off resources in a dynamic environment puts QoS at risk; aggressive consolidation may cause some VMs to obtain insufficient resources to serve a spike in load.
- Second, agreed SLAs bring challenges to application performance management in virtualized environments.
- These issues require effective consolidation policies that can minimize energy use without compromising user QoS requirements.

- The current approaches to dynamic VM consolidation are weak in terms of providing performance guarantees.

- One of the ways to prove performance bounds is to divide the problem of energy-efficient dynamic VM consolidation into a few subproblems that can be analyzed individually.

- It is important to analytically model the problem and derive optimal and near-optimal approximation algorithms that provide provable efficiency.

- To achieve this goal, clouds need novel analytical models and QoS-based resource allocation algorithms that optimize VM placements with the objective of minimizing energy consumption under performance constraints.

1.1.2 InterClouds and integrated allocation of resources

- Cloud providers have been deploying datacenters in multiple locations throughout the globe.

- For example, Amazon EC2 Cloud services are available via Amazon datacenters located in the United States, Europe, and Singapore.

- This disbursement is leading to the emergence of a notion, called the InterCloud, supporting scalable delivery of application services by harnessing multiple datacenters from one or more providers.

- In addition to enhancing performance and reliability, these InterClouds provide a powerful means of reducing energy-related costs.

– One reason is that the local demand for electricity varies with time of day and weather.

– This causes time-varying differences in the price of electricity at each location.

– Moreover, each site has a different source of energy (such as coal, hydroelectric, or wind), with different environmental costs.

– This gives scope to adjust the load sent to each location, and the number of servers powered on at each location, to improve efficiency.

- In such environments, algorithms that make routing decisions by considering the location of the user, the energy-efficiency of the hardware at each site, the energy mix, and the number of servers currently on at each location are needed.

- A particularly promising approach is to use this routing to make work "follow the renewables."

- A major problem with renewable energy is that most sources are intermittent and uncontrollable.

- Dynamically routing requests to locations with available renewable energy can greatly reduce the nonrenewable energy used and facilitate the widespread use of clean energy.

- Sending loads to remote datacenters incurs both delay costs and energy costs due to the increased amounts of data that are transferred over the Internet.

- Improvements in energy-efficient transport technology should lead to significant reductions in the power consumption of the cloud software services

11.2 Market-based management of clouds

- Cloud computing is still in its infancy, and its prominent use is twofold:
  - (1) complete replacement of in-house IT infrastructure and services with the same capabilities rented by service providers;
  - (2) elastic scaling of existing computing systems in order to address peak workloads.
- The efforts in research and industry have been mostly oriented to design and implement systems that actually enable business vendors and enterprises to achieve these goals.
- The real potential of cloud computing resides in the fact that it actually facilitates the establishment of a market for trading IT utilities.
- This opportunity until now has been mildly explored and falls in the domain of what it is called market-oriented cloud computing

2.1 Market-oriented cloud computing

- Cloud computing already embodies the concept of providing IT assets as utilities.
- Then, what makes cloud computing different from market-oriented cloud computing?
- First, it is important to understand what we intend by the term market.
- The Oxford English Dictionary(OED) defines a market as a “place where a trade is conducted”.

- More precisely, market refers to a meeting or a gathering together of people for the purchase and sale of goods.

- A broader characterization defines the term market as the action of buying and selling, a commercial transaction, a purchase, or a bargain.

- Therefore, essentially the word market is the act of trading mostly performed in an environment either physical or virtual that is specifically dedicated to such activity.

– If we consider the way IT assets and services are consumed as utilities, it is evident that there is a trade-off between the service provider and the consumer;

– this enables the use of the service by the user under a given SLA.

– Therefore, cloud computing already expresses the concept of trade, even though the interaction between consumer and provider is not as sophisticated as happens in real markets:

– Users generally select one cloud computing vendor from among a group of competing providers and leverage its services as long as they need them.

- Moreover, at present, most service providers have inflexible pricing, generally limited to flat rates or tariffs based on usage thresholds.

- In addition, many providers have proprietary interfaces to their services, thus restricting the ability of consumers to quickly move—and with minimal conversion costs—from one vendor to another.

- This rigidity, known as vendor lock-in, undermines the potential of cloud computing to be an open market where services are freely traded.

- Therefore, to remove such restrictions, it is required that vendors expose services through standard interfaces

- This enables full commoditization and thus would pave the way for the creation of a market infrastructure for trading services.

- What differentiates market-oriented cloud computing (MOCC)from cloud computing is the presence of a virtual marketplace
- where IT services are traded and brokered dynamically.

- This is something that still has to be achieved and that will significantly evolve the way cloud computing services are eventually delivered to the consumer.

- More precisely, what is missing is the availability of a market where desired services are published
- and then automatically bid on by matching the requirements of customers and providers.

- At present, some cloud computing vendors are already moving in this direction;
- this phenomenon is happening in the IaaS domain—which is the market sector that is more consolidated and mature for cloud computing

- We can clearly characterize the relationship between cloud computing and MOCC as follows:
  - Market Oriented Computing has the same characteristics as Cloud Computing;
  - therefore it is a dynamically provisioned unified computing resource allowing you to manage software and data storage as on aggregate capacity resulting in “real-time” infrastructure across public and private infrastructures.
  - Market Oriented Cloud Computing goes one step further by allowing spread into multiple public and hybrid environments dynamically composed by trading service.

2.2A reference model for MOCC

- Market-oriented cloud computing originated from the coordination of several components:

- Service consumers, service providers, and other entities that make trading between these two groups possible.

- Market orientation not only influences the organization on the global scale of the cloud computing market.

- It also shapes the internal architecture of cloud computing providers that need to support a more flexible allocation of their resources,
- which is driven by additional parameters such as those defining the quality of service

2.2.1 A global view of market-oriented cloud computing

- A reference scenario that realizes MOCC at a global scale is given in Figure

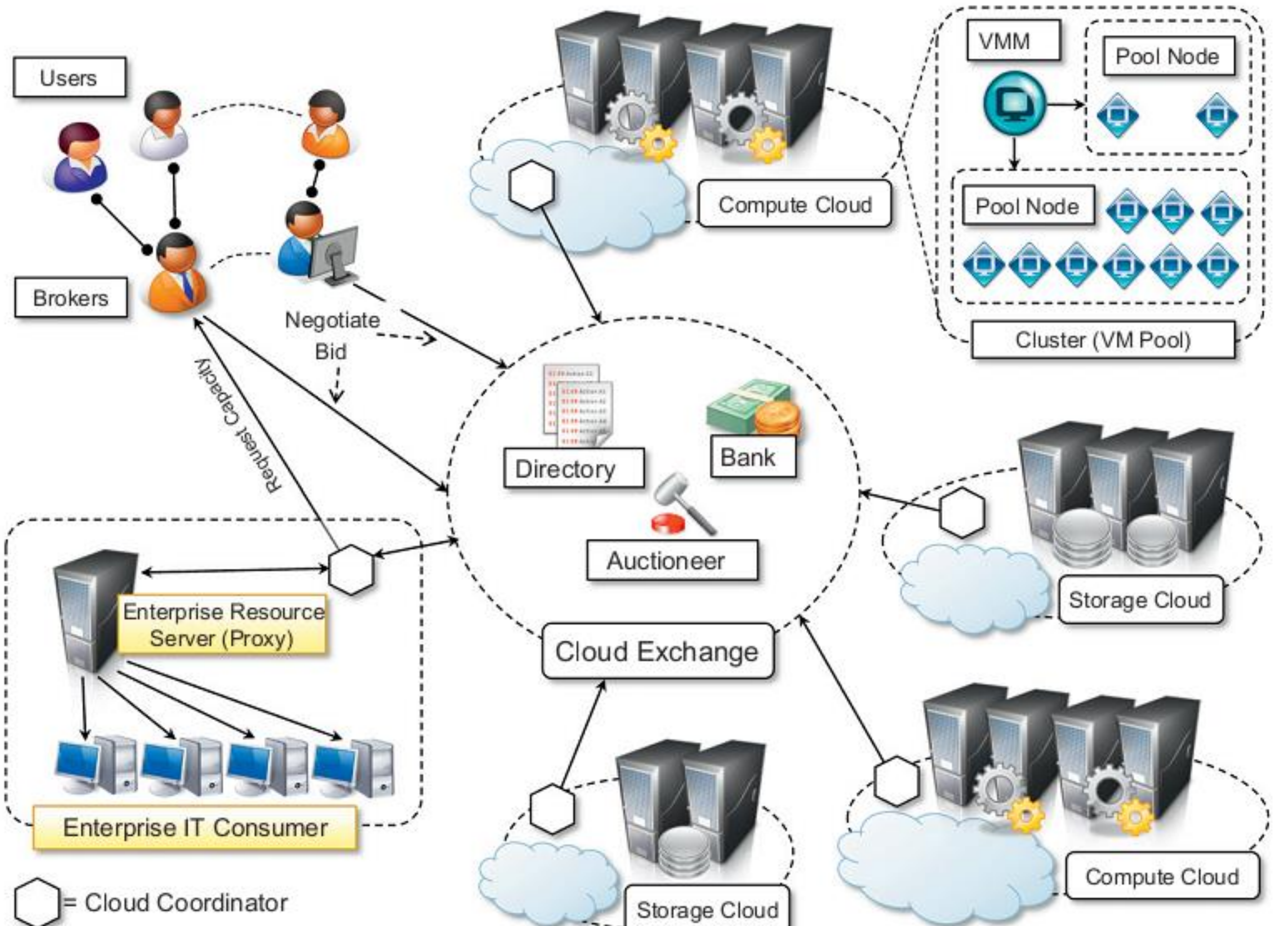


FIGURE 11.3

Market-oriented cloud computing scenario.

- Several components and entities contribute to the definition of a global market-oriented architecture.

- The fundamental component is the virtual marketplace—represented by the Cloud Exchange (CEx)—which acts as a market maker, bringing service producers and consumers together.

- The principal players in the virtual marketplace are the cloud coordinators and the cloud brokers.

- The cloud coordinators represent the cloud vendors and publish the services that vendors offer.

- The cloud brokers operate on behalf of the consumers and identify the subset of services that match customers' requirements in terms of service profiles and quality of service.

- Brokers perform the same function as they would in the real world:

- They mediate between coordinators and consumers by acquiring services from the first and subleasing them to the latter.

- Brokers can accept requests from many users.

- At the same time, users can leverage different brokers.

- A similar relationship can be considered between coordinators and cloud computing services vendors.

– Coordinators take responsibility for publishing and advertising services on behalf of vendors and can gain benefits from reselling services to brokers.

– Every single participant has its own utility function, that they all want to optimize rewards.

– Negotiations and trades are carried out in a secure and dependable environment and are mostly driven by SLAs, which each party has to fulfill.

– There might be different models for negotiation among entities, even though the auction model seems to be the more appropriate in the current scenario.

- The same consideration can be made for the pricing models: Prices can be fixed, but it is expected that they will most likely change according to market conditions.

- Several components contribute to the realization of the Cloud Exchange and implement its features.

- Directory.
  - The market directory contains a listing of all the published services that are available in the cloud marketplace.
  - The directory not only contains a simple mapping between service names and the corresponding vendor (or cloud coordinators) offering them.
  - It also provides additional metadata that can help the brokers or the end users in filtering from among the services of interest those that can really meet the expected quality of service.
  - Moreover, several indexing methods can be provided to optimize the discovery of services according to various criteria.
  - This component is modified in its content by service providers and queried by service Consumers.

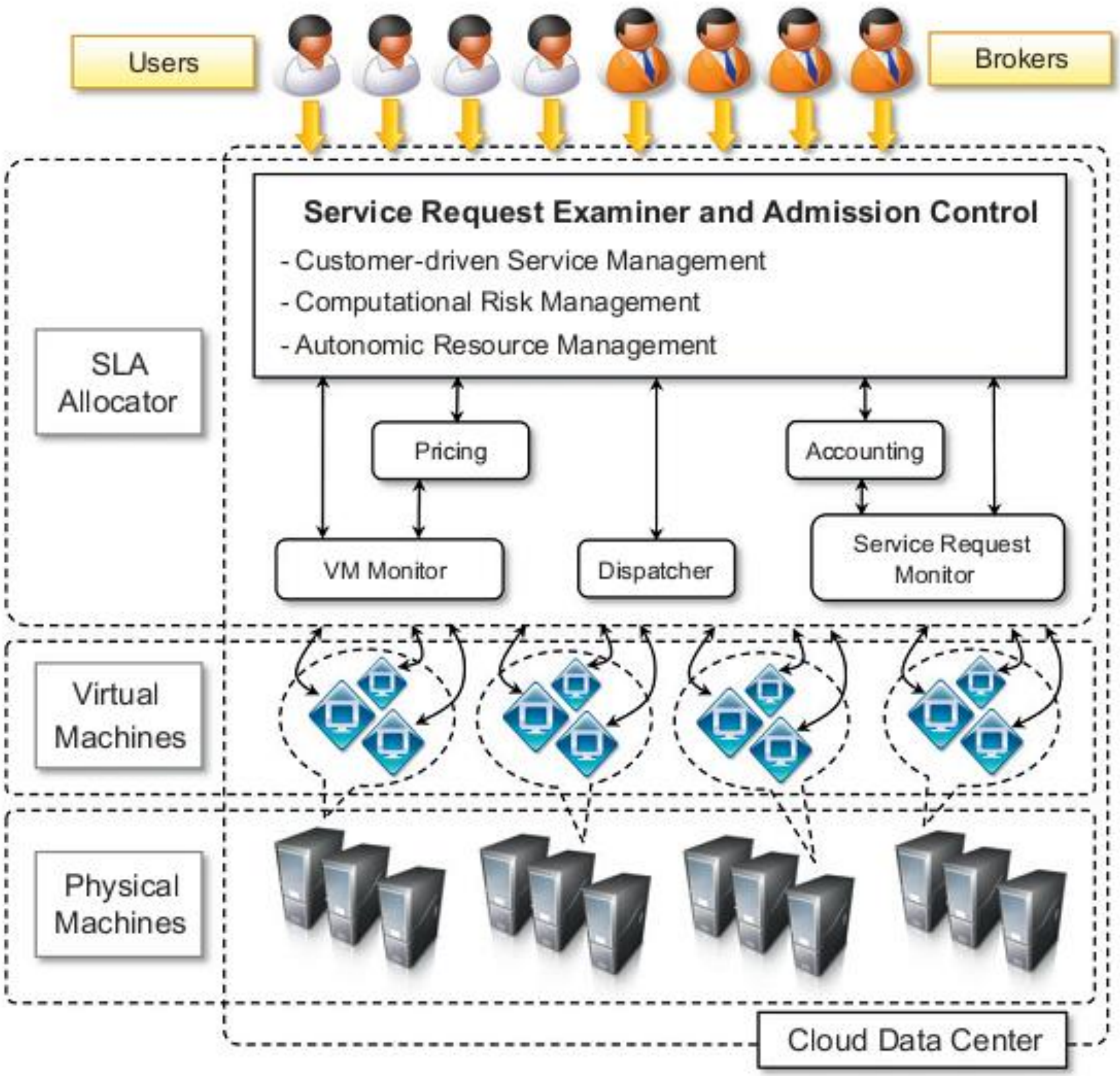
- Auctioneer.
  - The auctioneer is in charge of keeping track of the running auctions in the marketplace and of verifying that the auctions for services are properly conducted and that malicious market players are prevented from performing illegal activities.
- Bank.
  - The bank is the component that takes care of the financial aspect of all the operations happening in the virtual marketplace.

  - It also ensures that all the financial transactions are carried out in a secure and dependable environment.

  - Consumers and providers may register with the bank and have one or multiple accounts that can be used to perform the transactions in the virtual marketplace.

- 2.2.2 Market-oriented architecture for datacenters
  - Datacenters are the building blocks of the computing infrastructure
  - that backs the services offered by a cloud computing vendor, no matter its specific category (IaaS, PaaS, or SaaS).
  - Following figure provides an overall view of the components that can support a cloud computing provider in making available its services on a market-oriented basis

– The model applies to PaaS and IaaS providers that explicitly leverage virtualization technologies to serve customers' needs.



- There are four major components of the architecture:
  - Users and brokers: They originate the workload that is managed in the cloud datacenter.
  - Users either require virtual machine instances to which to deploy their systems (IaaS scenario) or deploy applications in the virtual environment made available to them by the provider (PaaS scenario).
  - These service requests are issued by service brokers that act on behalf of users and look for the best deal for them.

  - SLA resource allocator: The allocator represents the interface between the datacenter and the cloud service provider and the external world.
  - Its main responsibility is ensuring that service requests are satisfied according to the SLA agreed to with the user.

- Several components coordinate allocator activities in order to realize this goal:
  - Service Request Examiner and Admission Control Module : This module operates in the front-end and filters user and broker requests in order to accept those that are feasible given the current status of the system and the workload that is already processing.
  - Pricing Module: This module is responsible for charging users according to the SLA they signed
  - Accounting Module: This module maintains the actual information on usage of resources and stores the billing information for each user
  - Dispatcher: This component is responsible for the low-level operations that are required to realize admitted service requests
  - Resource Monitor: This component monitors the status of the computing resources, either physical or virtual
  - Service Request Monitor: This component keeps track of the execution progress of service requests.

- Virtual machines (VMs).: Virtual machines constitute the basic building blocks of a cloud computing infrastructure, especially for IaaS providers.
- VMs represent the unit of deployment for addressing users' requests.
- Infrastructure management software is in charge of keeping operational the computing infrastructure backing the provider's commercial service offering.
- Physical machines.: At the lowest level of the reference architecture resides the physical infrastructure that can comprise one or more datacenters.
- This is the layer that provides the resources to meet service demands.

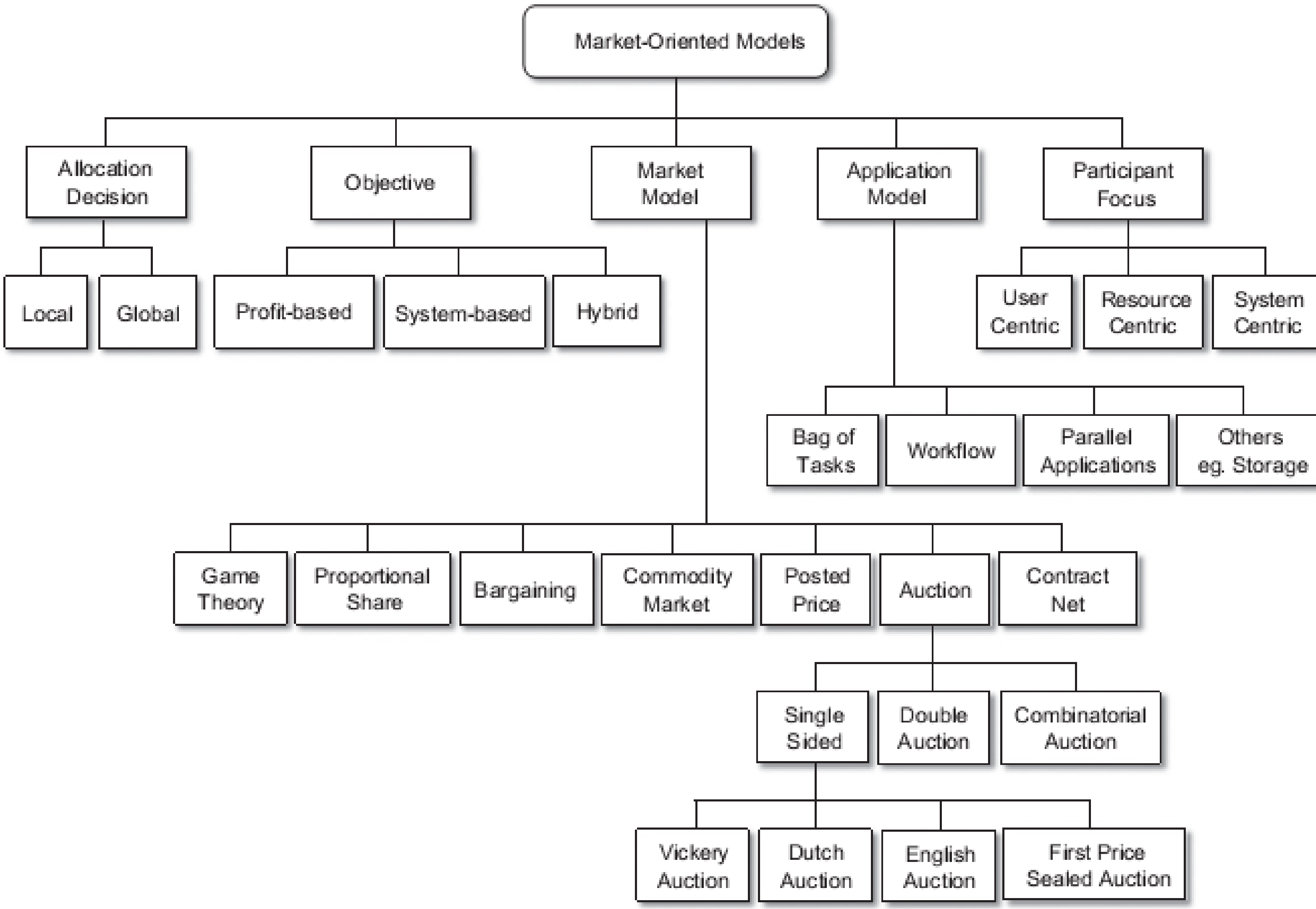
- 11.2.3Technologies and initiatives supporting MOCC
  - Existing cloud computing solutions have very limited support for market-oriented strategies to deliver services to customers.
  - Most current solutions mainly focused on enabling cloud computing concern the delivery of infrastructure, distributed runtime environments, and services.
  - Since cloud computing has been recently adopted, the consolidation of the technology constitutes the first step toward the full realization of its promise.

- Until now, a good deal of interest has been directed toward IaaS solutions, which represent a well-consolidated sector in the cloud computing market, with several different players and competitive offers.

- New PaaS solutions are gaining momentum, but it is harder for them to penetrate the market dominated by giants such as Google, Microsoft, andForce.com.

- 11.2.3.1 Framework for trading computing utilities
  - Computing grids aggregate a heterogeneous set of resources that are geographically distributed and might belong to different organizations.
  - Such resources are often leased for long-term use by means of agreements among these organizations.
  - Within this context, market-oriented schedulers, which are aware of the price of a given computing resource and schedule user’s applications according to their budgets, have been investigated and implemented.
  - The research in this area is of relevance to MOCC, since cloud computing leverages preexisting distributed computing technologies, including grid computing.

- Complete taxonomy and analysis of such schedulers, which is reported in Figure



- A major classification categorizes these schedulers according to allocation decision, objective, market model, application model, and participant focus.
- Along this dimension, it is possible to classify the schedulers into the following categories:
- Game theory.
- Proportional share.
- Commodity market.
- Posted price
- Contract-Net.
- Bargaining.
- Auction.

- The most popular and interesting market models for trading computing utilities are the commodity market, posted price, and auction models.

- Commodity market and posted price models, or variations/combinations of them, are driving the majority of cloud computing services offerings today.

- Auction-based models can instead potentially constitute the reference market-models for MOCC, since they are able to seamlessly support dynamic negotiations.

- 11.2.3.2 Industrial implementations
  - Flexible pricing models: amazon spot instances
  - Amazon Web Services (AWS), one of the biggest players in the IaaS market,
  - recently introduced the concept of spot instances, which allows EC2 customers to bid on unused Amazon EC2 capacity and run those instances for as long as their bid exceeds the current spot price.
  - The spot price varies periodically according to the supply of and demand for EC2 instances and is kept constant within a single hour block.
  - Spot instances can be terminated at any time, and they are usually priced at a lower price with respect to the traditional (on-demand and reserved) instances,
  - since they rely on exceeding capacity available in the EC2 infrastructure.

- Virtual market place: SpotCloud
- SpotCloud is an online portal that implements a virtual marketplace, where sellers and buyers can register and trade cloud computing services.

- The platform is a market place operating in the IaaS sector.

- Buyers are looking for compute capacity that can meet the requirements of their applications,

- while sellers can make available their infrastructure to serve buyers’ needs and earn revenue.

- SpotCloudprovides a comprehensive set of features that are expected for a virtual marketplace. Some of them include:
  - Detailed logging of all the buyers’ transactions
  - Full metering, billing for any capacity
  - Full control over pricing and availability of capacity in the market
  - Management of quotas and utilization levels for providers
  - Federation management (many providers, many customers, but one platform)
  - Hybrid cloud support (internal and external resource management)
  - Full market administration and reporting
  - Applications and pre-build appliances directories

- Market directories: AppSpot, the cloud market
  - SpotCloud is implemented as an application hosted on AppSpot.
  - This is a huge portal serving applications built on top of the GoogleAppEngine infrastructure: appspot.com is a namespace under which all the scalable Web applications developed with the Google AppEngine technology are made available to the community of Internet users.
  - A solution that is more oriented toward listing available cloud building blocks is The Cloud Market, which features a comprehensive listing of Amazon EC2 images.

11.3Federated clouds/InterCloud

- Cloud computing strongly implies the presence of financial agreements between parties, since services are available on demand on a pay-per-use basis.
- Nonetheless, the concepts characterizing cloud federation and the InterCloud are applicable, with some limitations, to building aggregations of clouds that belong to different administrative domains.

- 11.3.1Characterization and definition
  - The terms cloud federation and Inter Cloud, often used interchangeably
  - The term federation implies the creation of an organization that supersedes the decisional and administrative power of the single entities and that acts as a whole.
  - Within a cloud computing context, the word federation does not have such a strong connotation but implies that there are agreements between the various cloud providers, allowing them to leverage each other's services in a privileged manner.

- A definition of the term cloud federation was given by Reuven Cohen, founder and CTO of Enomaly Inc.

" Cloud federation manages consistency and access controls when two or more independent geo-graphically distinct Clouds share either authentication, files, computing resources, command and control or access to storage resources"

- InterCloud is a term that is often used interchangeably to express the concept of Cloud federation.

- It was introduced by Cisco for expressing a composition of clouds that are interconnected by means of open standards to provide a universal environment that leverages cloud computing services.

- By mimicking the Internet term, often referred as the “network of networks,” InterCloud represents a “Cloud of Clouds”

- and therefore expresses the same concept of federating together clouds that belong to different administrative organizations.

- The primary difference between the InterCloud and federation is that the InterCloud is based on future standards and open interfaces,
- while federation uses a vendor version of the control plane.

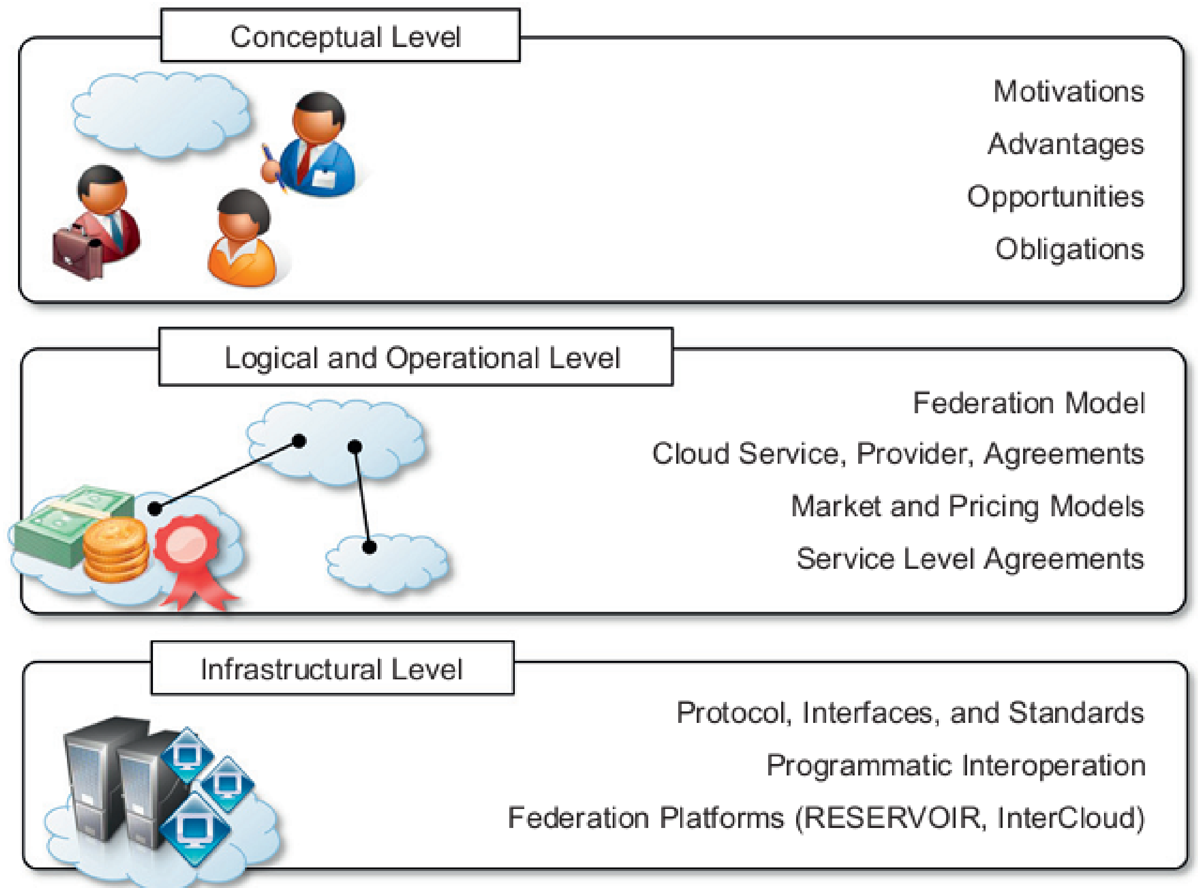
- With the InterCloud vision, all Clouds will have a common understanding of how applications should be deployed.

- Eventually workloads submitted to a Cloud will include enough of a definition (resources, security, service level, geo-location, etc.) that the Cloud is able to process the request and deploy the application.

- This will create the true utility model, where all the requirements are met by the definition and the application can execute "as is" in any Cloud with the resources to support it.

11.3.2Cloud federation stack

- Creating a cloud federation involves research and development at different levels: conceptual, logical and operational, and infrastructural.
- Figure provides a comprehensive view of the challenges faced in designing and implementing an organizational structure.



11.3.2.1 Conceptual level

- The conceptual level addresses the challenges in presenting a cloud federation as a favorable solution with respect to the use of services leased by single cloud providers.

- In this level it is important to clearly identify the advantages for either service providers or service consumers in joining a federation
- and to delineate the new opportunities that a federated environment creates with respect to the single-provider solution.

- Elements of concern at this level are:
  - Motivations for cloud providers to join a federation
  - Motivations for service consumers to leverage a federation
  - Advantages for providers in leasing their services to other providers
  - Obligations of providers once they have joined the federation
  - Trust agreements between providers
  - Transparency versus consumers

- It is possible to identify functional and nonfunctional requirements that cloud service providers have behind these motivations.

- The functional requirements include:
  - Supplying low-latency access to customers, regardless of their location
  - Handling bursts in demand.
  - Scaling existing applications and services beyond the capabilities of the owned infrastructure
  - Make revenue from unused capacity.
- nonfunctional requirements are the following:
  - Meeting compulsory regulations about the location of data.
  - Containing transient spikes in operational costs.
  - Disaster recovery.

- 11.3.2.2 Logical and operational level

  – The infrastructural level addresses the technical challenges involved in enabling heterogeneous cloud computing systems to interoperate seamlessly.

  – It deals with the technology barriers that keep separate cloud computing systems belonging to different administrative domains.

  – By having standardized protocols and interfaces, these barriers can be overcome.

  – In other words, this level for the federation is what the TCP/IP stack is for the Internet: a model and a reference implementation of the technologies enabling the interoperation of systems.

- At this level it is important to address the following issues:
  - What kind of standards should be used?
  - How should design interfaces and protocols be designed for interoperation?
  - Which are the technologies to use for interoperation?
  - How can we realize a software system, design platform components, and services enabling interoperability?
  - Interoperation and composition among different cloud computing vendors is possible only by means of open standards and interfaces.
  - Moreover, interfaces and protocols change considerably at each layer of the Cloud Computing Reference Model

- 11.3.3Aspects of interest
  - Besides motivation and technical enablers, other elements should be considered.
  - In particular, standards for interoperability, security, and legal issues have to been taken into consideration while defining a platform for interoperability among cloud vendors.

- 11.3.3.1 Standards
  - Standards play a fundamental role in building a federation.
  - Their main role is to organize a platform for interoperation that goes beyond ad hoc aggregations and private settlements between providers.
  - Standardized interfaces and protocols facilitate the realization of an open organization where providers can easily join.
  - The advantages are primarily technical; standards facilitate the development of software and services that interconnect systems.
  - Furthermore, they help in defining clear paths for new providers to join, thus contributing to the realization of open systems.
  - Interoperation between vendors has always been an element of concern for enterprise

- Open cloud manifesto
  - The Open Cloud Manifesto constitutes the first step toward the realization of a cloud interoperability platform.
  - The manifesto was drafted in 2009 as a result of the coordinated activity of different cloud vendors and currently lists more than 400 cloud computing services providers that support the vision it embodies.
  - More than proposing standards, the manifesto is a declaration of intent, endorsed by commercial players in the field of cloud computing, to realize an interoperable and open cloud computing platform.

- It introduces the goals of an open cloud platform, which can be summarized as follows:
  - Choice
  - Flexibility
  - Speed and Agility
  - Skills.
- The manifesto ends with a set of recommendations for cloud computing vendors to pursue an open collaboration and an appropriate use of standards.

- Distributed management task force
  - The Distributed Management Task Force (DMTF)is an organization involving more than 4,000 active members, 44 countries, and nearly 200 organizations.
  - It is the industry organization leading the development, adoption, and promotion of interoperable management standards and initiatives.
  - With specific reference to cloud computing, the DMTF introduced the Open Virtualization Format (OVF) and supported several initiatives for interoperable cloud technologies,
  - such as the Open Cloud Standards Incubator, the Cloud Management Working Group (CMWG), and the Cloud Audit Data Federation Working Group (CADFWG).

- The main purpose of the specification is to provide a platform-neutral format for distributing packaged software systems.
- Therefore, the key features of the format are the following:
  - Optimized for distribution.
  - Optimized for a simple, automated user experience.
  - Supports both single VM and multi-VM configurations.
  - Portable vendor- and platform-independent VM packaging.
  - Extensible.
  - Localizable
  - Open standard.

- From a technical point of view, the OVF defines a transport mechanism for virtual machine templates.

- One single OVF package can contain one or more virtual machine images, and once deployed to the host system, it adds a self-containing, self-consistent software solution for achieving a particular goal.

- Portability is a fundamental property of the OVF.
- To fully address the potential issues arising in deploying a software appliance over a wide range of virtual hardware, different levels of portability are identified:
  - Level 1:The packaged appliance runs on a particular product and/or CPU architecture and/or virtual hardware selection.
  - Level 2: The packaged appliance runs on a specific family of virtual hardware.
  - Level 3: The packaged appliance runs on multiple families of virtual hardware.

- Open cloud computing interface

  - The Open Cloud Computing Interface (OCCI) is an open organization that comprises a set of specifications driven by the community and delivered through the Open Grid Forum.

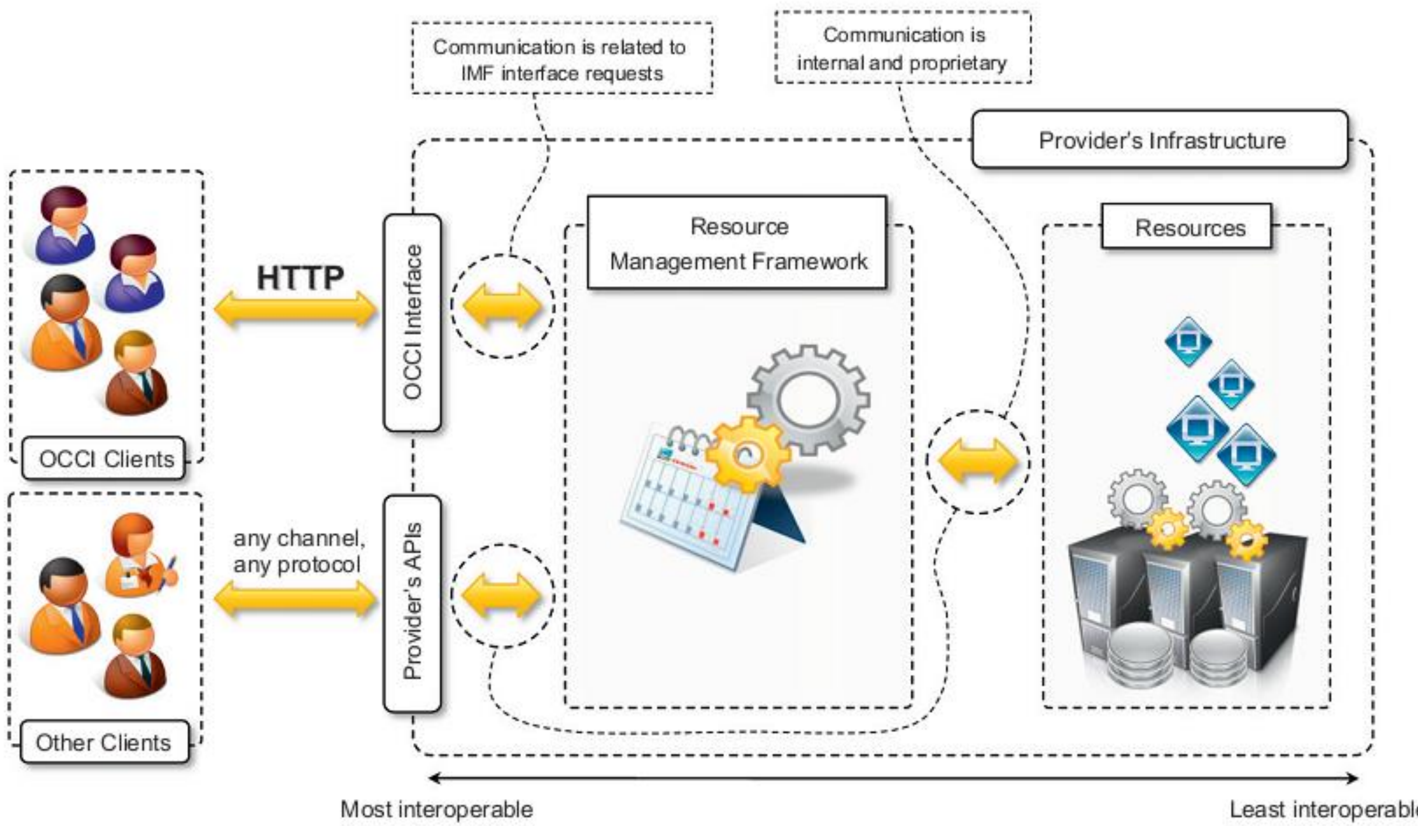
  - These specifications define protocols and APIs for several kinds of management tasks.

  - Initially conceived to create a remote management API for IaaS-type services, the OCCI has evolved into a wider set of APIs focusing on integration, portability, and interoperability.

  - The current set of specifications covers all three market segments of cloud computing: IaaS, PaaS, and SaaS.

- OCCI has currently delivered
  - three documents that define:
    - The formal definition of the OCCI core model
    - The definition of the OCCI infrastructure extensions for the IaaS domain
    - The OCCI rendering model, which defines how to interact with the OCCI core model through REST over HTTP

- Besides these, the OCCI is also working on other specifications that cover the following topics:
  - billing, monitoring, advanced reservation, negotiation, and agreement features.
- Figure provides a reference scenario for the Open Cloud Computing Interface, which defines a single access point to the resources managed by a given cloud resource provider.



- Cloud data management interface
  - The Cloud Data Management Interface (CDMI)is a specification for a functional interface that applications will use to create, retrieve, update, and delete data elements from the cloud.
  - This interface also provides facilities for discovering the properties of a given cloud storage offering.
  - CDMI has been proposed by the Cloud Storage Technical Working Group of the Storage Network Industry Association (SNIA),
  - an association promoting standards in the management of IT information with a particular focus on data storage.
  - SNIA has also produced a reference implementation of the CDMI, thus facilitating the process of quickly producing a working standard by means of feed-back from the community.

- The specification introduces and defines the concept of cloud storage as a “delivery of virtualized storage on demand,” also known as Data storage-as-a-Service (DaaS).

- The main concept of DaaS is to abstract data storage behind a set of interfaces and to make it available on demand.

- This definition encloses a considerably wide range of storage architectures.

- Figure provides the overall context in which cloud storage interfaces will operate.

- A cloud data management ser-vice provides a CDMI made available to clients through REST ful interfaces.

- Such an interface provides access to information, data, and storage services that can be leveraged to access storage clouds.

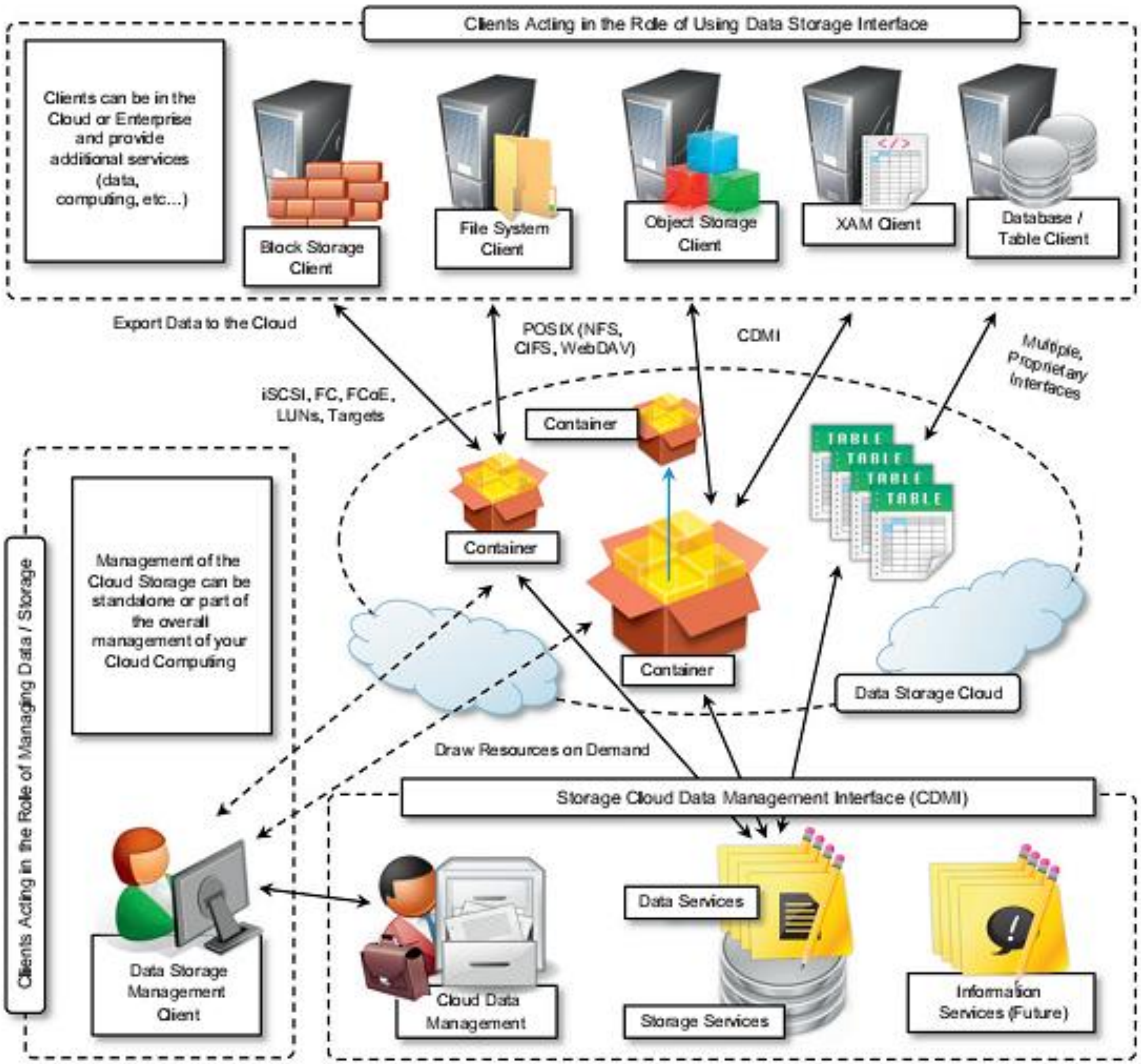


FIGURE 11.9

Cloud storage reference model.

- The object model contains the following components:
  - Data objects: These are the fundamental storage component in CDMI and are analogous to file in a file system.
  - Container objects: Container objects are the fundamental abstraction used to group stored data.
  - Domain objects: Domain objects are quite similar to container objects and they are used to represent administrative ownership stored within a CDMI storage system.
  - Queue objects: Queues are a special class of containers that are used to provide first-in, first-out (FIFO) access when storing and retrieving data.
  - Capability objects: Capability objects are a special class of container object that allow a CDMI client to discover what subset of the CDMI standard is implemented by a CDMI provider.

- Cloud security alliance
  - The Cloud Security Alliance (CSA)is nonprofit organization with the mission of promoting the use of best practices for providing security assurance in cloud computing and education on the use of cloud computing to help secure all other forms of computing.
  - Rather than acting as a standardizing body, CSA offers a context in which to discuss security practices and provide guidance for developing reliable and secure cloud computing systems.
  - The most relevant initiative of the CSA has been the Cloud Controls Matrix (CCM).
  - The matrix is specifically designed to provide fundamental security principles for guiding cloud vendors and for assisting prospective cloud service consumers in assessing the overall risks implied in leveraging a cloud service provider.

• Customer Responsibilities for Security

Table 11.2 Customer Responsibilities for Security Management in the Cloud [156]

Activities	IaaS	PaaS	SaaS
Availability management	Manage VM availability with fault-tolerant architecture	Manage this activity for applications deployed in the PaaS platform (the provider is responsible for runtime engine and services)	Provider responsibility
Patch and configuration management	Manage VM image hardening Harden your VMs, applications, and database using your established security hardening process Manage activities for your VMs, database, and applications using your established security management process	Manage this activity for applications deployed in the PaaS platform Test your applications for OWASP Top 10 vulnerabilities[20]	Provider responsibility
Vulnerability management Access control management	Manage OS, applications, and database vulnerabilities leveraging your established vulnerability management process Manage network and user access control to VMs, secure privileged access to management consoles, install host Intrusion Detection System (IDS), and manage host firewall policies	Manage this activity for applications deployed in the PaaS platform (the provider is responsible for their runtime engine and service) Manage developer access to provisioning Restrict access using authentication methods (user- and network-based controls) Federate identity and enable SSO if SAML[21] is supported	Provider responsibility Manage user provisioning Restrict user access using authentication methods (user- and network-based controls) Federate identity and enable SSO if SAML is supported

- Cloud federation security

- two main approaches can be considered:

  - Centralized federation model: This is the approach taken by several identity federation standards.
  - It distinguishes two operational roles in an SSO transaction: the identity provider and the service provider.

  - Claim-based model: This approach addresses the problem of user authentication from a different perspective and requires users to provide claims answering who they are and what they can do in order to access content or complete a transaction.

- Privacy, security, and intellectual property related issues

- This category includes all the legal issues arising from the management of data, applications, and services by the cloud service provider on behalf of the customer as part of its service offering.

- In particular, data breach, protection, and accessibility are objects of current legislation:

- Business and commerce-related issues
  - Legal issues can also arise as a result of specific contractual terms and arrangements between the cloud provider and the user.
  - Contractual terms need to be able to deal with issues such as:
    - What is defined as a service that is agreed between the two parties?
    - What is the liability of each of the parties?
    - What risk are users running while relying on a given provider?
    - What are measures that are put in place to guarantee users access to their data?

- Jurisdictional and procedural issues
  - Jurisdictional and procedural legal issues arise from two major aspects of cloud computing:
  - the geo-location of data and applications served by cloud providers and the laws
  - and regulations that are applied in case of litigation.
  - Jurisdictional issues are mostly related to location of data and the specific laws that apply in that location.
  - Cloud service providers locate their datacenters in order to reduce their operational costs.
  - The placement of datacenters is influenced by the desire to optimally serve customers on a global scale.
  - For this reason it is quite common to distribute the infrastructure of a single cloud provider over the globe.

11.3.4Technologies for cloud federations

- 11.3.4.1 Reservoir

  – Resources and Services Virtualization Without Barriers, or RESERVOIR, is a European research project focused on developing an architecture that supports providers of cloud infrastructures to dynamically partner with each other to extend their capabilities while preserving their administrative autonomy.

  – RESERVOIR defines a software stack enabling interoperation at the IaaS layer and providing support for SLA-based execution of applications on top of the infrastructure overlay that results from the federation of infrastructure providers.

- RESERVOIR is based on the concept of dynamic federation:

- Each infrastructure provider is an autonomous business with its own business goals and that might decide to partner with other businesses when needed.

- The federation is obtained by means of the RESERVOIR middleware that needs to be deployed at each site.

- The IT management at a specific site is fully autonomous and dictated to by policies that depend on the site's business goals.

- When needed, internal IT resources are leased to other providers within the context of a negotiated SLA.

- The role of RESERVOIR is to orchestrate this process and to minimize the barriers obstructing interoperation among different administrative domains

- Figure provides a general overview of a RESERVOIR cloud.

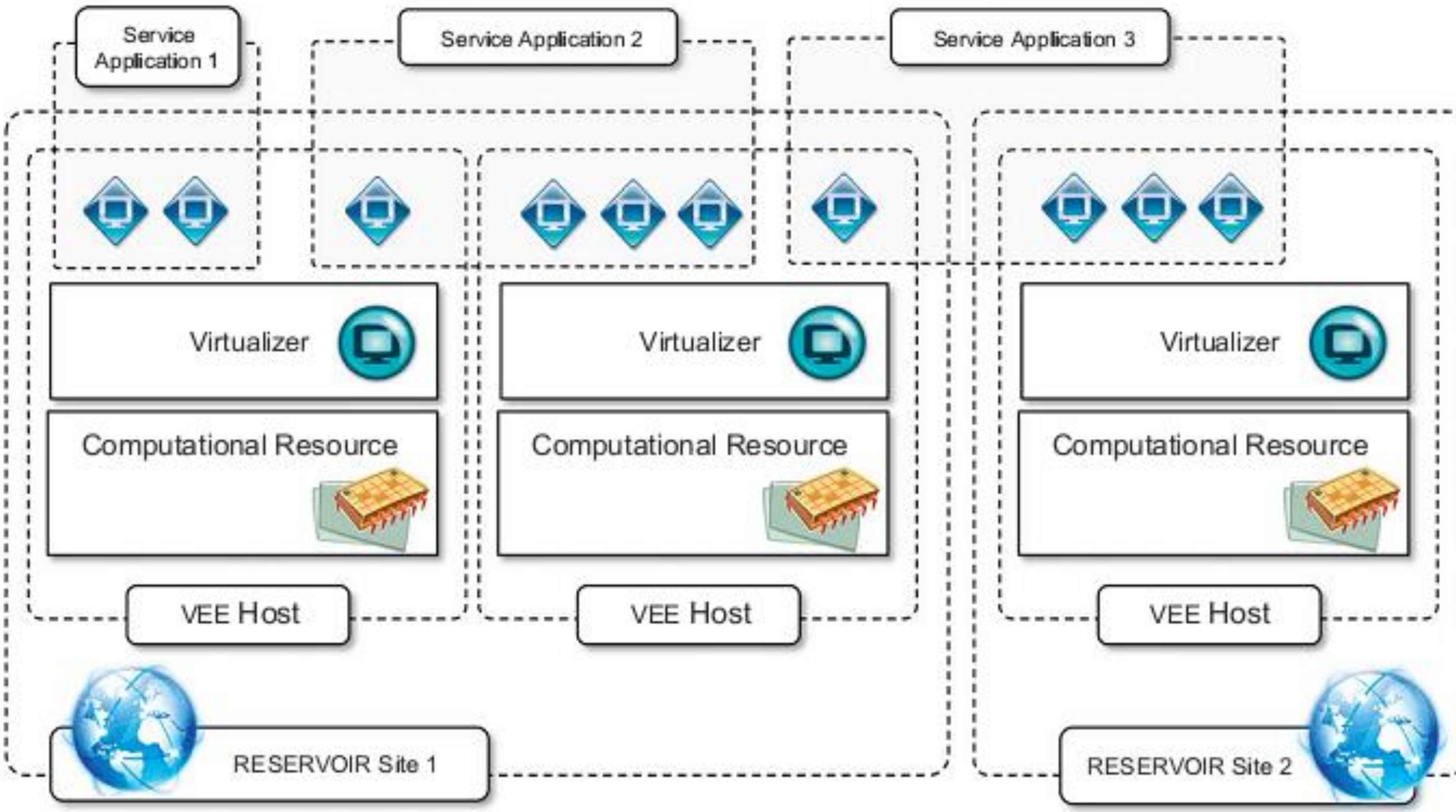


FIGURE 11.10

RESERVOIR cloud deployment.

- Figure describes the internal architecture of a RESERVOIR site that enables the federated and SLA-based execution of applications.

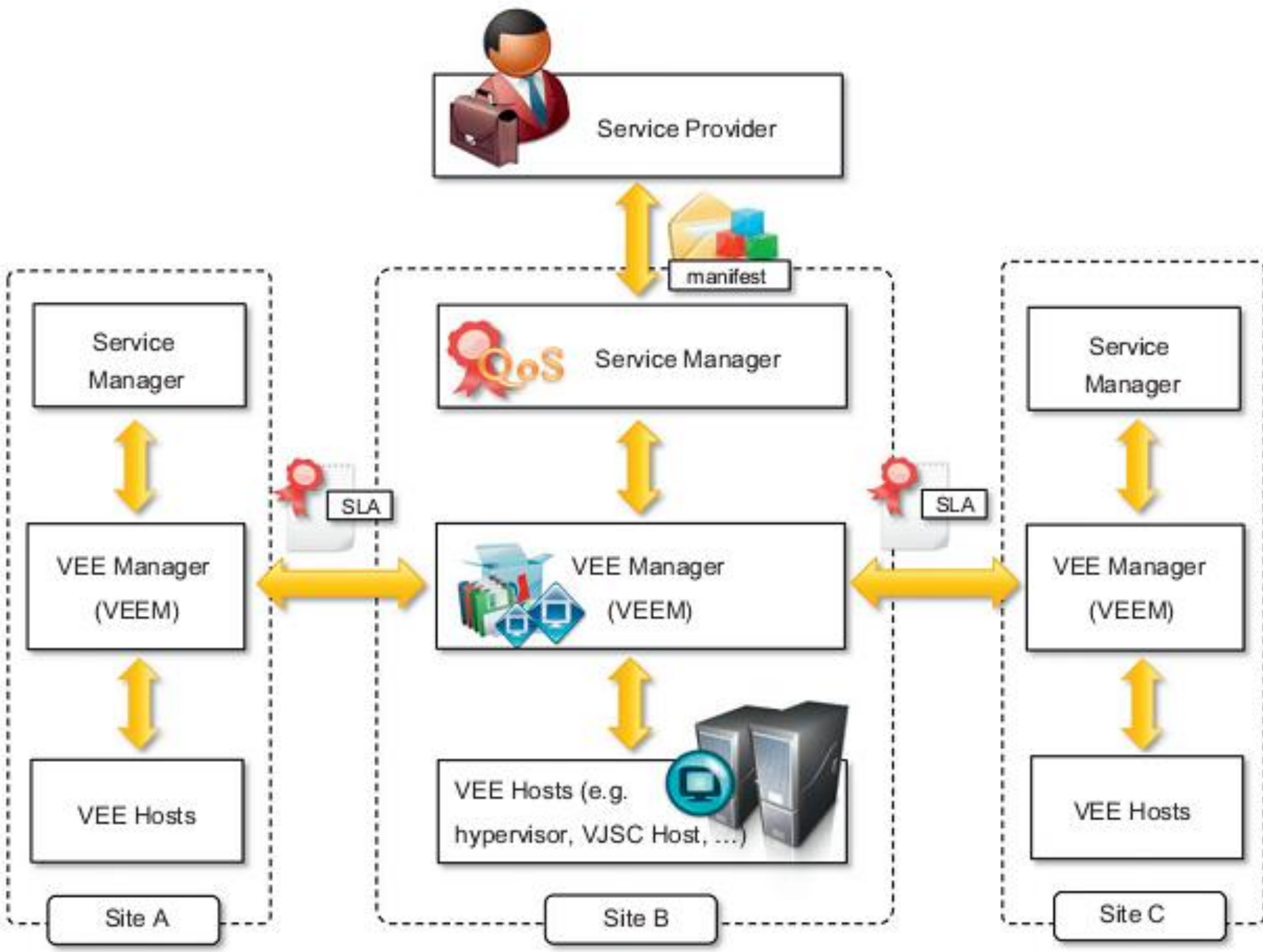


FIGURE 11.11

RESERVOIR architecture.

The RESERVOIR stack consists of three major components:

- Service Manager: The Service Manager is the highest level of abstraction and constitutes the front-end used by service providers to submit service manifests, negotiate pricing, and monitor applications.

- Virtual Execution Environment (VEE) Manager: This component is the core of the RESERVOIR middleware and is responsible for the optimal placement of VEEs into VEE hosts according to the constraints expressed by the Service Manager.

- VEE Host (VEEH): This is the lowest level of abstraction and interacts with the VEE Manager to put into practice the IT management decisions regarding heterogeneous sets of virtualization platforms.

11.3.4.2 InterCloud

- InterCloud is a service-oriented architectural framework for cloud federation that supports utility-driven interconnection of clouds.

- It is composed of a set of decoupled elements that interact via a market-oriented system to enable trading of cloud assets such as computing power, storage, and execution of applications.

- As depicted in Figure, the InterCloud model comprises two main elements:
  - CloudExchange
  - and CloudCoordinator:

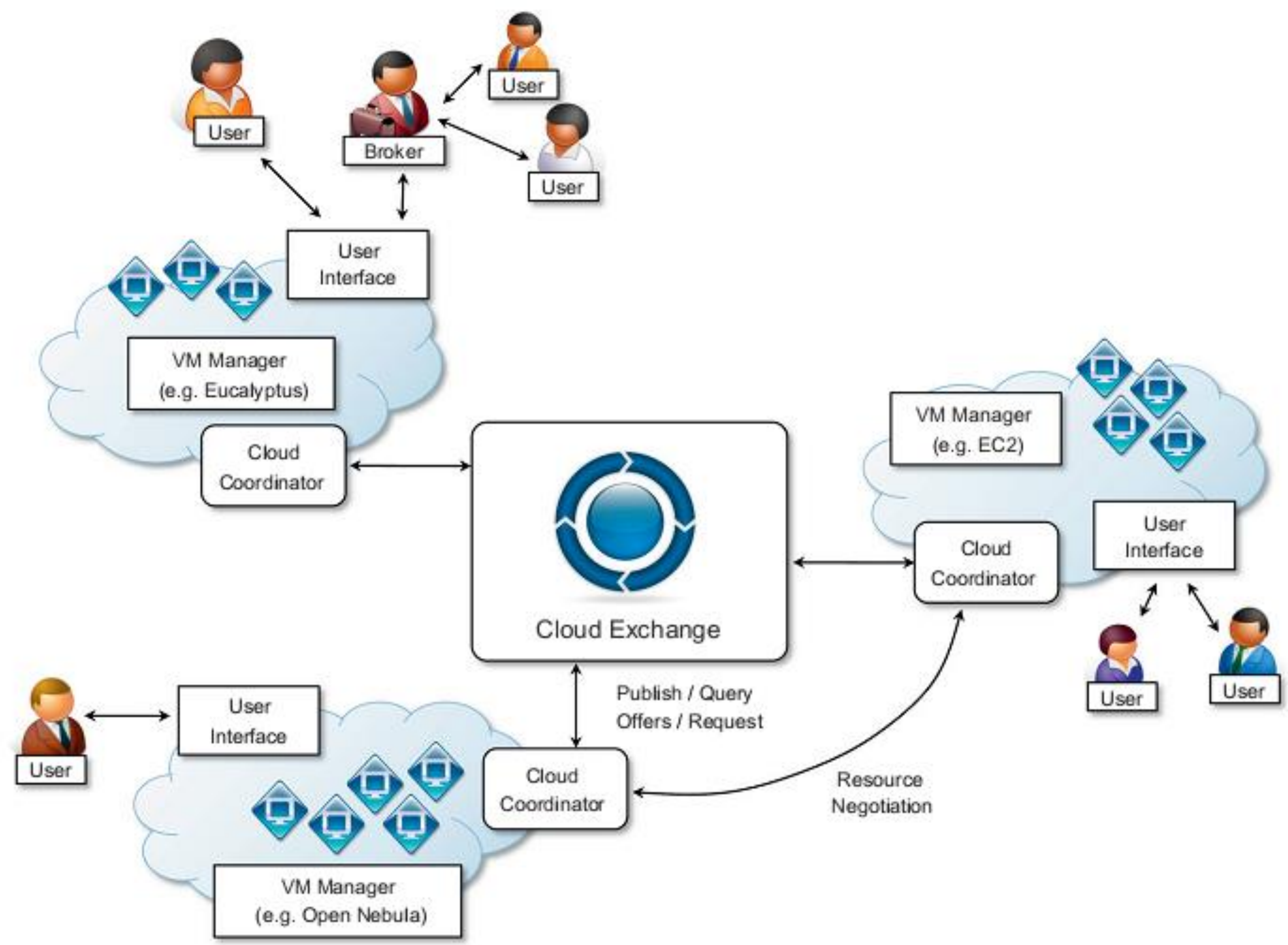


FIGURE 11.12

InterCloud architecture.

- CloudExchange: This is the market-making component of the architecture.

- It offers services that allow providers to find each other in order to directly trade cloud assets, as well as allowing parties to register and run auctions.

- In the former case, CloudExchange acts as a directory service for the federation. In the latter case, it runs the auction.

- For offering such services to the federation, CloudExchange implements a Web service-based interface that allows datacenters to join and leave the federation;
  - to publish resources they want to sell;
  - to register their resource requirements so that parties interested in selling providers are able to locate potential buyers for their resources;
  - to query resource offers that match specific requirements;
  - to query requirements that match available resources from a party;
  - to withdraw offers and requests from the coordinator;
  - to offer resources in auctions;
  - to register bids;
  - and to consult the status of a running auction.

CloudCoordinator.: This component manages domain-specific issues related to the federation.

- This component is present on each party that wants join the federation.
- CloudCoordinator hasfront-end components (i.e., elements that interact with the federation) as well as back-end components (i.e., components that interact with the associated datacenter).
- Front-end components interact with the CloudExchange and with other coordinators.
- The former allows datacenters to announce their offers and requirements, whereas the latter allows the Coordinator to learn about the current state of the datacenter to decide whether actions from the federation are required or not.

- Therefore, wherever the Coordinator detects that extra resources are required by the datacenter, it triggers the process of discovery of potential providers (by interacting with the cloud federation).

- Once potential providers are discovered and the preferred one is selected, the Coordinator contacts the remote Coordinator and negotiates.

- Similarly, when the Coordinator detects that local resources are underutilized, they can publish an offer for resources in the CloudExchange or they can look for matches among requirements registered in the Exchange service.

11.4Third-party cloud services

- One of the key elements of cloud computing is the possibility of composing services that belong to different vendors or integrating them into existing software systems.

- The service-oriented model, which is the basis of cloud computing, facilitates such an approach and provides the opportunity for developing a new class of services that can be called third-party cloud services.

- These are the result of adding value to preexisting cloud computing services, thus providing customers with a dif-ferent and more sophisticated service.

11.4.1MetaCDN

- MetaCDN[158] provides users with a Content Delivery Network (CDN) service by leveraging and harnessing together heterogeneous storage clouds.

- It implements a software overlay that coordinates the service offerings of different cloud storage vendors and uses them as distributed elastic storage on which the user content is stored.

- MetaCDN provides users with the high-level services of a CDN for content distribution and interacts with the low-level interfaces of storage clouds to optimally place the user content in accordance with the expected geography of its demand.

- By leveraging the cloud as a storage back-end it makes a complex—and generally expensive—content delivery service available to small enterprises.

- The architecture of MetaCDN is shown inFigure

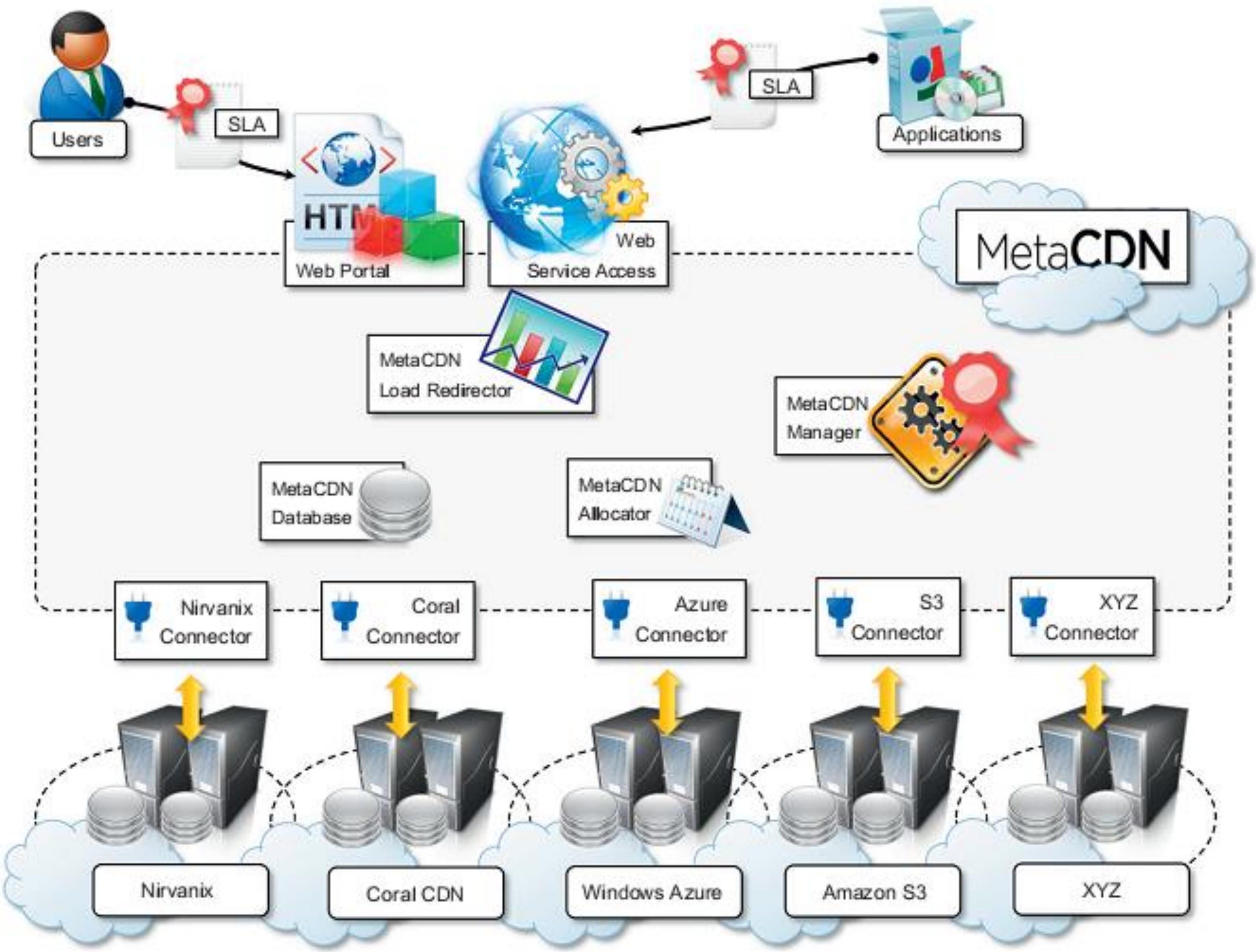


FIGURE 11.13

MetaCDN architecture.

- Four different deployment options can be selected:
- Coverage and performance-optimized deployment.: In this case MetaCDN will deploy as many replicas as possible to all available locations.
- Direct deployment.:In this case MetaCDN allows the selection of the deployment regions for the content and will match the selected regions with the supported providers serving those areas.
- Cost-optimized deployment.:In this case MetaCDN deploys as many replicas in the locations identified by the deployment request. The available storage transfer allowance and budget will be used to deploy the replicas and keep them active for as long as possible.
- QoS optimized deployment.:In this case MetaCDN selects the providers that can better match the QoS requirements attached to the deployment, such as average response time and throughput from a particular location.

- 11.4.2SpotCloud
  - SpotCloud has already been introduced as an example of a virtual marketplace.
  - By acting as an intermediary for trading compute and storage between consumers and service providers, it provides the two parties with added value.
  - For service consumers, it acts as a market directory where they can browse and compare different IaaS service offerings and select the most appropriate solution for them.
  - For service providers it constitutes an opportunity for advertising their offerings.
  - In addition, it allows users with available computing capacity to easily turn themselves into service providers by deploying the runtime environment required by SpotCloud on their infrastructure as shown in fig.

• SpotCloud Market Architecture

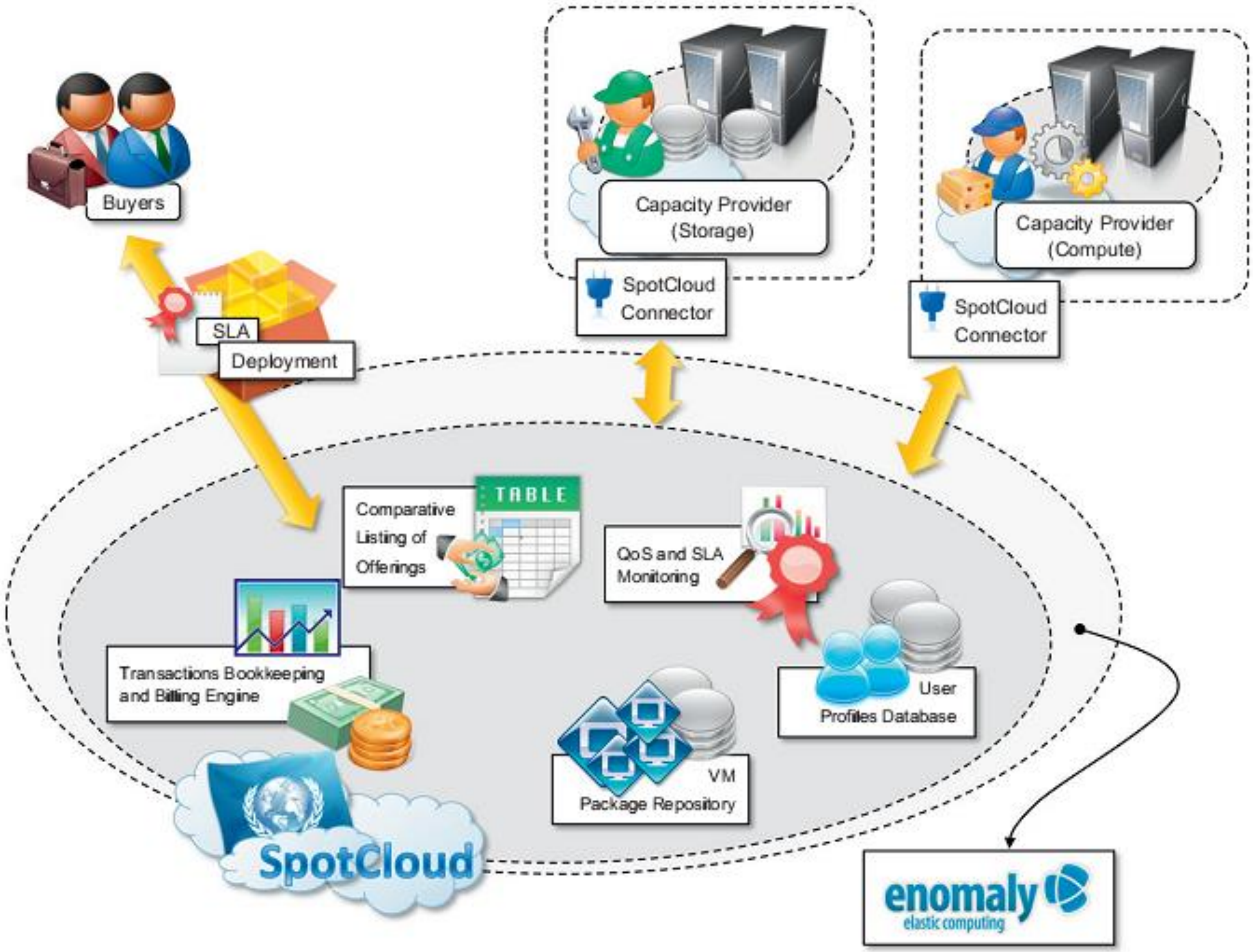


FIGURE 11.14

SpotCloud market architecture.

- SpotCloud is not only an enabler for IaaS providers and resellers, but its intermediary role also includes a complete bookkeeping of the transactions associated with the use of resources.

- Users deposit credit on their SpotCloud account and capacity sellers are paid following the usual pay-per-use model.

- SpotCloud retains a percentage of the amount billed to the user.

- Moreover, by leveraging a uniform runtime environment and virtual machine management layer, it provides users with a vendor lock-in-free solution, which might be strategic for specific applications.

End