2009 IEEE International Conference on Web Services

CCOA: Cloud Computing Open Architecture

Liang-Jie Zhang and Qun Zhou
IBM T.J. Watson Research Center, New York, USA
E-mail: {zhanglj, qzhou}@us.ibm.com

Abstract

Cloud Computing is evolving as a key computing platform for sharing resources that include infrastructures, software, applications, and business processes. Virtualization is a core technology for enabling cloud resource sharing. However, most existing Cloud Computing platforms have not formally adopted the service-oriented architecture (SOA) that would make them more flexible, extensible, and reusable. By bridging the power of SOA and virtualization in the context of Cloud Computing ecosystem, this paper presents seven architectural principles and derives ten interconnected architectural modules to form a reusable and customizable Cloud Computing Open Architecture (CCOA). Two case studies on Infrastructure and Business Cloud are used to deliver business and practical value of infrastructure and business process provisioning services over the Internet. We also present some potential value-added services of the proposed CCOA to guide strategic planning and other consulting practices of Cloud Computing.

1. Introduction

As a key service delivery platform in the field of service computing [12], Cloud Computing provides environments to enable resource sharing in terms of scalable infrastructures, middleware and application development platforms, and value-added business applications. The operation models may include pay-as-go utility models, free infrastructure services with value-added platform services, fee-based infrastructure services with value-added application services, or free services for vendors but sharing of revenues generated from consumers.

As summarized in [1], typically there are four types of resources that can be provisioned and consumed over the Internet. They can be shared among users by leveraging economy of scale. Provisioning is a way of sharing resources with requesters over the network. One of the major objectives of Cloud Computing is to leverage Internet or Intranet to provision resources to users.

The first type of resources is infrastructure resources, which include computing power, storage, and machine provisioning. For example, Amazon EC2 provides web service interface to easily request and configure capacity online [2]. Xdrive Box service provides online storage to users [3]. Microsoft SkyDrive provides free storage service, with an integrated offline and online model that keeps privacy-related files on hard drives, and enables people to access those files remotely [4]. In the area of computing power sharing, the Grid computing initiative has taken it as its major focus to use clustering and parallel computing technologies to share computing power with others, based on task scheduling when computers are idle.

The second type of resources in Cloud Computing is software resources including middleware and development resources. The middleware consist of cloud-centric operating systems, application servers, databases, and others. The development resources comprehend design platforms, development tools, testing tools, deployment tools, and open sources-based reference projects.

The third type of resources in Cloud Computing is application resources. The leading companies in the information industry are gradually moving applications and related data to the Internet. Software applications are delivered through Software As A Service (SaaS) model or mashups of value-added applications. For example, Google has used Cloud Computing platform to offer Web applications for communication and collaboration [19]. Google docs move productivity applications to the Web, gradually replacing heavy-weighted desktop applications. Therefore, developing a reusable and customizable Cloud Computing Open Architecture for enabling application development environment is a key to success of application sharing over the Internet.

The fourth type of resources in Cloud Computing is business processes. Some applications can be exposed as utilities, namely loosely coupled sub-processes or tasks inside customers' business processes. Business process sharing is the business-driven application outsourcing that supports reuse, composition, and provisioning.

This paper is organized as follows. The first part of Section 2 discusses the two key enabling technologies, Virtualization and SOA, for articulating the value of creating an open architecture for Cloud Computing. The rest of Section 2 presents a Cloud Computing Open Architecture (CCOA) and its seven principles for building extensible and flexible Cloud Computing systems and applications. Section 3 describes two case studies on Infrastructure Cloud and Business Cloud, to illustrate infrastructure service provisioning and business process as a service. Some value-added offerings based on the proposed CCOA are also depicted. Section 5 discusses

978-0-7695-3709-2/09 $25.00 © 2009 IEEE
DOI 10.1109/ICWS.2009.144

related work and future directions. Conclusions are drawn in the end of this paper.

2. Cloud Computing Open Architecture

Currently, there is no standard definition or specification for Cloud Computing. It may take some time to define the key characteristics of Cloud Computing based on practices in the field. Cloud Computing involves a set of key technologies to address resource sharing based on business requirements. Based on our practices in the areas of service provisioning and solution design, we think the following two key enabling technologies could play very important roles in this revolutionary phase: virtualization technology and Service-Oriented Architecture (SOA).

The virtualization technology handles how images of the operating systems, middleware, and applications are pro-created and allocated to the right physical machines or a slice of a server stack. The images could be moved around and put into production environment on demand. On the other hand, virtualization technology can also help reuse licenses of operating systems, middleware, or software applications, once a subscriber releases his/her service from the Cloud Computing platform.

The SOA is the evolution of a system or software architecture for addressing componentization, reusability extensibility, and flexibility. In order to construct scalable Cloud Computing platforms, we need to leverage SOA to build reusable components, standard-based interfaces, and extensible solution architectures.

Creating a so-called Cloud Computing platform is easy as long as it can enable sharing of at least one of the resources. However, building a unified, scalable and reusable Cloud Computing architecture to support sharing of all types of resources still faces challenges in the areas of technology breakthrough and best industry practices.

2.1 "OSI" Model for Cloud Computing

We have identified the following three objectives to help address the above challenge of defining a good open architecture for Cloud Computing.

The first objective is to articulate a reusable way of creating scalable and configurable provisioning platform for Cloud Computing. This paper brings together the power of Service-Oriented Architecture (SOA) and virtualization to deliver business and practical value to emerging software applications, hardware, and business process provisioning services over the Internet in the context of Cloud Computing.

The second objective is to propose a set of common and shared services for building Cloud Computing platforms, to provide business services or other cloud offerings to its enterprise consumer users in a unified approach.

The third objective is to maximize the potential business value of Cloud Computing based on an extensible IT infrastructure and management system. This will lead to the value added services of business cloud, through monetization of the combined power of SOA and Cloud Computing.

As we know, OSI standards for Open System Interface [6] which has encountered some challenges to realize its value in the context of generic open systems. In this paper, we try to limit the scope of the open system to a specialized domain, and leverage service-oriented thinking to help modularize open architecture for Cloud Computing. In the following section, we present a Cloud Computing Open Architecture based on seven principles.

2.2 Seven Principles of Cloud Computing Architecture

In this Cloud Computing Open Architecture, we propose an integrated co-innovation and co-production framework to get cloud vendors, cloud partners, and cloud clients to work together based on seven principles, which are used to define ten major architectural modules and their relationships shown in Figure 1. The presented Cloud Computing Open Architecture covers cloud ecosystem enablement, cloud infrastructure and its management, service-orientation, cloud core on provisioning and subscription, compostable cloud offerings, cloud information architecture and management, and cloud quality analytics. This is a logical and modularized separation, which helps isolate concerns of details of each module during the design process. Since the connections between the identified key architectural principles for Cloud Computing are quite complex, the information exchanges are going through the Cloud Information Architecture and Cloud Ecosystem Management. In the rest of the section, we will introduce the details of each principle illustrated in Figure 1.

Principle 1: Integrated Ecosystem Management for Cloud

An architecture must support the management of the ecosystem of Cloud Computing. This ecosystem includes all involved services and solutions vendors, partners, and end users to provide or consumer shared resources in the Cloud Computing environment. Cloud vendors expose its interaction interfaces of its internal operations and product development capability to the cloud. Cloud partners provide components to cloud vendors or serve as agents to provide value-added services to the cloud clients. Cloud clients are users of the cloud services that offer business goal driven resource sharing.

From an architecture design perspective, cloud vendor dashboard provides an integrated view of interaction with vendors' frontend and backend operations. For example, in the frontend, marketing and business development activities, services delivery portals, and customer support can be enabled in the cloud vendor's dashboard. In the backend, solution design and development activities, or hosting environment are used to support the frontend's operations.

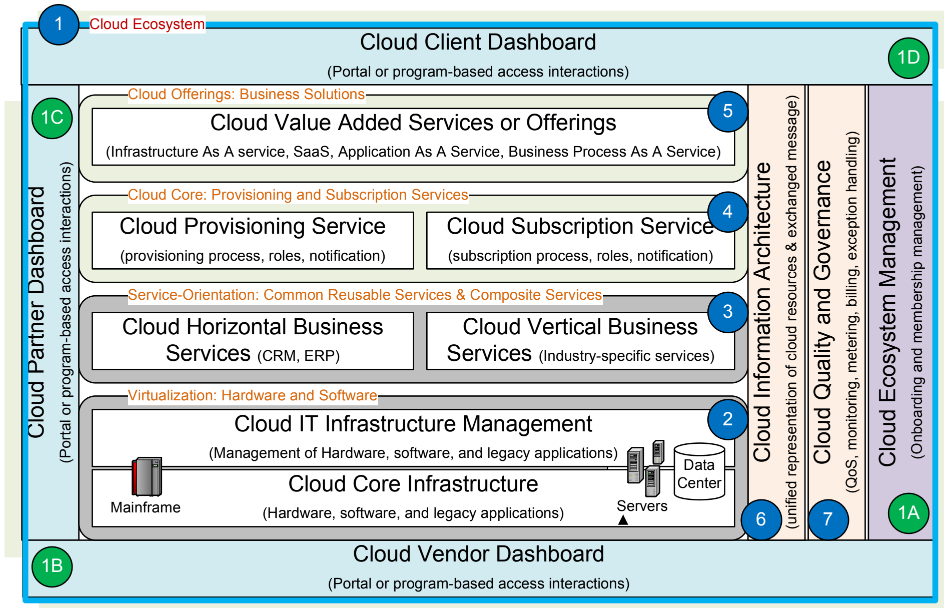

Figure 1. Cloud Computing Open Architecture Overview Diagram

Since most of the cloud vendors do not work alone anymore, they need to collaborate with their partners [7] in the value chain of Cloud Computing environment. In this regard, a partner dashboard is needed for the participating partners to interact with the cloud vendors and clients. For example, if the cloud partners serve as component suppliers for the cloud vendors, architectural building blocks for interacting with vendors and collaboration policy manager are keys to the value chain integration.

Clients or end users of Cloud Computing can be grouped into two classes: enterprise and consumer users. The cloud client dashboard provides a focal point for all kinds of users to interact with Cloud Computing services or offerings. This focal point provides a unified framework for users to consume cloud services via multiple channels such as Web portal, program-based business to business collaboration channel, or phone-based customer representative channel. There are opportunities to explore a converging software and services architecture for enterprise and consumer users based on various pricing strategies, security enablement, and other features of software and services. Since enterprise users or consumer users are co-existing role players in the service ecosystem, an enterprise user may have multiple consumer users. In the end, they are just consumers of Cloud Computing resources at different levels.

Putting all those dashboards together, the Cloud Computing ecosystem management layer (1A) provides an integrated on-boarding process and common utilities to support the seamless collaboration and message exchanges among cloud vendors, partners, and clients. For example, the onboard progress covers the registration of business entities and users. The business entities include cloud vendors, cloud partners, and enterprise cloud clients. The user entities are end users within a certain business entity (e.g. an employee of a company, or a member of a registered community like a social network), or consumer users in the open Internet space.

Principle 2: Virtualization for Cloud Infrastructure

There are two basic approaches for enabling virtualization in the Cloud Computing environment. The first approach is hardware virtualization that is to manage hardware equipments in plug-and-play mode. Hardware equipments can be added or removed without affecting the normal operations of other equipments in the system. Of course, performance or storage spaces may be dynamically changed due to those add and remove actions.

The second approach is software virtualization, i.e., to use software image management or software code virtualization technology to enable software sharing. Specifically, software images can be created based on the degree of reusability of a set of software systems including operating system, middleware, and applications. The other software virtualization technology is dynamic code assembly and execution. In this case, there are no software images. Code elements will be dynamically copied from repositories and pasted in right places based on business logic. For instance, in an Internet application, some JavaScript code elements can be dynamically retrieved and inserted into a right Ajax package to create new functions or features for a Web client. By the same token, server side programs can be dynamically assembled and executed based on the composition of reusable code elements and just-in-time complier technologies. With the development of multi-core technology and parallel programming, this dynamic code assembly and execution technology will have great advantages moving forward. Especially, it will get rid of the requirements of huge storage spaces for software images for middleware and development tools. But in today's environment, both software virtualization technologies can co-exist and support each other based on usage scenarios.

In short, the Cloud IT Infrastructure Management module covers software image management, hardware virtualization, and legacy application packaging.

The target resources being managed by the Cloud IT Infrastructure Management module is the Cloud Core Infrastructure, which comprises all supporting hardware, software, and legacy applications for operating a Cloud Computing environment. For example, hardware may include mainframe, distributed servers or clusters, and data storages. Software may include database packages and related middleware. Legacy applications may involve home-grown applications or ISV applications that are part of the infrastructure.

It is noted that this virtualization principle in the Cloud Computing Open Architecture is an extension of the operational system layer in the SOA Solution Stack (a.k.a. SOA Reference Architecture) [8] in the context of Cloud Computing enablement. For example, a rack manager can be leveraged to handle wire hardware systems. The interface of dynamically assembling IT resources enables infrastructure resource provisioning, and can be used to host or deploy business applications that leverages distributed cloud infrastructure resources.

Principle 3: Service-Orientation for Common Reusable Services

As introduced before, in addition to the virtualization characteristic, service-orientation is another driving force to enable Cloud Computing to further realize the business value from asset reusability, composite applications, and mashup services. There are two major types of common reusable services: Cloud Horizontal and Vertical Business Services.

The Cloud Horizontal Business Services consist of various platform services that hide the complexities of middleware, database, and tools. In addition to offering middleware or development tools as services in the Cloud Computing environment, some common utilities such as on-boarding, provisioning, monitoring, billing tools, or cross-industry services like Customer Relationship Management (CRM) and Enterprise Resource Planning (ERP).

The Cloud Vertical Business Services include all domain specific or industry-specific utility services. Examples are shipping and payment services.

Both common reusable services in CCOA can be reused to enable Cloud core's provisioning and subscription services, as well as to build cloud offerings such as Infrastructure As A Service, SaaS, Application As A service, Business Process As A Service, which will be further described in Principle 5.

Principle 4: Extensible Provisioning and Subscription for Cloud

Extensible service provisioning is the unique feature of a Cloud Computing system. Without extensibility, the provisioning part of the Cloud Computing architecture can only support a certain type of resource sharing. This implies that the service provisioning architecture for free-use users and paying users are the same. Both types of users can be service providers or consumers from time to time. From service consumers' perspective, they are interested in how to easily access services based on their own business logics and goals. From service providers' perspective, three levels of service provisioning described in Section 1 will be the target offerings.

The Cloud Core shown in Figure 1 includes a set of provisioning and subscription services. This addresses a key question of how to handle service providers' provisioning process and service consumers' subscription process in the Cloud Computing architecture. We have categorized Cloud Core into Cloud Provisioning and Subscription Service. The key architectural elements of Cloud Provisioning Service include provisioning process, role definitions, and notification framework. Cloud Subscription Service involves subscription process, role definitions, and notification framework. The role defining framework and attributes as well as notification framework can be shared for both provisioning and subscription services.

Principle 5: Configurable Enablement for Cloud Offerings

Cloud offerings are the final products or services that are provisioned by the Cloud Computing platform. Since all cloud offerings should address certain business goals, cloud offerings are also known as cloud business solutions. In alignment with the categorization of resource sharing described in Section 1, CCOA defines its offering aspects at the following four levels: infrastructure as a service, software as a service (SaaS), application as a service, and business process as a service.

Leveraging the SOA reference architecture's extensibility and configurability, the proposed CCOA pursues configurable enablement of Cloud Computing platforms and services. The modularized cloud ecosystem management, virtualization, service-orientation, and cloud core have formed a solid foundation to ensure a computing platform to be configurable, compostable, and manageable.

In the category of infrastructure as a service, CCOA leverages the cloud core to support provisioning and subscription of the virtualized IT infrastructure resources by exploiting some common reusable services and cloud ecosystem management capabilities. Some example cloud offerings are storage cloud and infrastructure cloud.

Most cloud offerings are delivered or accessed through Web browsers. For instance, we can use web browser to upload photos, audio files, video files, and other documents to a storage cloud. In all other cloud offerings, Web interfaces have been proven effective channels as well to enable cloud clients to interact with cloud partners and vendors in the lifecycle of service delivery.

In the area of software as a service, lots of success stories demonstrate common reusable services such as CRM as a service (e.g. Saleforce.com), payment as a service (e.g. eBay's PayPal), and shipping as a service by leveraging the cloud core to manage the provisioning and subscription. It is emphasized that lots of value added services can be built on top of the common reusable services in CCOA before they are provisioned to cloud clients as cloud offerings. Therefore, composite applications or service composition technologies in SOA could play an important role in building value added cloud services in the context of CCOA.

In the application as a service area, lots of standalone applications or network-based applications can be provisioned as cloud services for sharing in the Cloud Computing value chain. For example, Web-based development tools could be good cloud offerings for the application development community, to design and develop applications for cloud or standalone applications by leveraging the Cloud Computing's development resources, such as enterprise modeling software, business process modeling software, architecture modeling software, application development software, and related asset repositories. Specifically, in the service-oriented solution design space, offering SOMA-ME [13] service as a design in the Cloud Computing environment is a natural way to empower mass community members of application design and development. There is no absolute boundary between software as a service and application as a service. In CCOA, an application is the integration of a set of software packages based on certain business logics and goals.

In the Cloud Computing environment, business process as a service is a new model for sharing best practices and business processes among cloud clients and partners in the value chain. An example business process as a service offering in the Cloud Computing environment is software testing, which is a very important business process in the lifecycle of software development. Since software systems or applications need to be deployed and run on different operating systems, middleware environments, or different versions or configurations of them, it is very hard for individual developers or companies to set up a testing "factory" to address those testing concerns in the area of application services. Upgrading or maintaining testing environment involves the migration of hardware, software, and testing knowledge repositories. They are the two major expenses in addition to the testing engineers who manipulate the environment to conduct the testing.

Principle 6: Unified Information Representation and Exchange Framework

Information representation and message exchange of Cloud Computing resources are very important to enable the collaborative and effective features of Cloud Computing. In CCOA, Cloud Computing resources include all business entities (e.g. cloud clients, partners, and vendors) and the supporting resources such as virtualization related modules, service-orientation related modules, cloud core, and cloud offerings. For example, in cloud ecosystem management, country, site, and organization are associated with the business entities in the Cloud Computing. Role players are used to define their functions in the dynamics of the Cloud Computing ecosystem. Project, task, documents, transactions, business process, reference links, annotation, and events are potential resources for supporting collaboration among various business entities in the Cloud Computing environment.

In CCOA, the cloud information architecture module enables representation of those cloud entities (business entities and supporting resources) in a unified Cloud Computing entity description framework. For example, technologies such as Resource Description Framework (RDF), Web Services Resource Framework (WSRF), and XML are candidates for implementing this unified framework. The messages exchanged between cloud entities form message exchange patterns. The message format and message exchange patterns can be reused to support various business scenarios. The message routing and exchange protocols as well as message transformation capability form a foundation for cloud information architecture.

Just like blood in human bodies, the cloud information architecture uses its information "blood" to form "blood stream" to get all various modules to communicate with each other in an effective way in CCOA.

Principle 7: Cloud Quality and Governance

The last and most important module in CCOA is the Cloud Quality and Governance shown in Figure 1. This module is responsible for the identification and definition of quality indicators for Cloud Computing environment and a set of normative guidance to govern the design, deployment, operation, and management of the cloud offerings.

From quality indicators' perspective, Quality of Services (QoS) parameters can be directly used to define cloud entities' reliability, response time, security, and integrity. The integrity can be checked through traceability enablement and compliance validation. Security is a very important aspect of the cloud quality. Only authorized business entities or users can access the right Cloud Computing resources. Access control, privacy, and protection of the cloud entities form the aspects of the trust and security foundation in CCOA.

From the governance perspective, lots of best practices from SOA governance can be borrowed to enable Cloud Computing environment and services offerings. For example, in order to launch a cloud initiative, a center of excellence or competence can be formed to better communicate and coordinate between business leaders and technical teams. In the project execution phase, monitoring, metering, billing, and exception handling are to be enabled and coordinated across line of businesses which produce, operate, and manage various modules of the cloud offerings.

3. Case Studies of CCOA

There are multiple scenarios of using the proposed CCOA, and no particular constraints with the design of any specific portions of the CCOA. For example, we can use CCOA to do high-level strategic planning for an enterprise to execute Cloud transformation initiative. We can also use CCOA to build an infrastructure stack to reduce operating cost. By the same token, we can use CCOA to guide the development or deployment practices of a Cloud Computing application. We can also create new value-added Cloud Computing applications based on the existing assets in a systematic way by following the principles of CCOA.

For an inter-connected Cloud Computing scenario, CCOA can be used as the architectural foundation to guide the design, development, deployment, and management of collaborative service delivery in the cloud value chain. From the methodologies' perspective, we can use the bottom-up approach to identify capabilities and provisioning of infrastructure. We can also use top-down approach to create cloud offerings and leverage or create cloud infrastructure to support the offerings.

In this section, we present two example cloud offerings based on CCOA. The first case is about Infrastructure Cloud, which is an example infrastructure as a service. The second case studies Business Cloud, which is an example of business process as a service in Cloud Computing environment.

3.1 Infrastructure Cloud

In this case study, we build our own private cloud to enable computing resource sharing. The usage scenario is to offer service requesters a pay-to-use model to provide servers with a selected set of preinstalled software packages to shorten time to production based on request submissions.

We use the 7 principles to instantiate CCOA for an Infrastructure Cloud to supply Server Provisioning Offering to cloud clients. Within the cloud ecosystem (Module 1), the Cloud Ecosystem Management module (1A) provides Web services user account management. Cloud Vendor Dashboard (1C) is responsible for handling the sign in process and provides a Web portal for cloud vendors to access the cloud ecosystem. The Cloud Partner Dashboard module (1C) in Figure 2 is not used in this case. The reason is that all virtualization capability of hardware and software is provided purely by cloud vendors in house. No cloud partners are involved in this Infrastructure Cloud. Cloud Client Dashboard supports sign-in process and a Web-based portal for cloud clients to submit order request and access related information on the provisioned servers. WebSphere Portal Server in Module 2 has been used to deploy UI related portlets.

In the virtualization module (Module 2), the Cloud Core Infrastructure includes a large number of servers, data centers, and supporting software packages such as WebSphere Application Server (WAS), DB2 database, and LDAP. In this case study, we use WebSpere Process Server to support BPEL processes and WebSphere Portal Server to support portlets. LDAP is used to manage users and access control information. Cloud Infrastructure Management is enabled by Tivoli Enterprise Portal Server and Tivoli Provisioning Manager, to handle the hardware virtualization and provisioning in this Infrastructure Cloud.

In the Service-Orientation module (Module 3), we have identified and defined a set of common reusable services based on service-oriented thinking. Cloud Horizontal Business Services contain cloud order management service, resource management service, Email notification service, and workflow management service. Here they are all implemented as Web services. The cloud vertical business services of module 3 are not used in this case study, since this infrastructure provisioning offering cloud does not leverage or provide domain-specific or industry-specific business services.

In the Cloud Core module (Module 4), Cloud Provisioning Services are supported by a set of business processes such as Service Provisioning (ServiceP), Service De-provisioning (ServiceDeP), and Contract Change including Contract Extension and Contract Termination. Cloud Provisioning Services are mainly used by Cloud Vendors to selectively share cloud resources. Cloud Subscription Service is enabled by Offering Creation and Order Creation business processes.

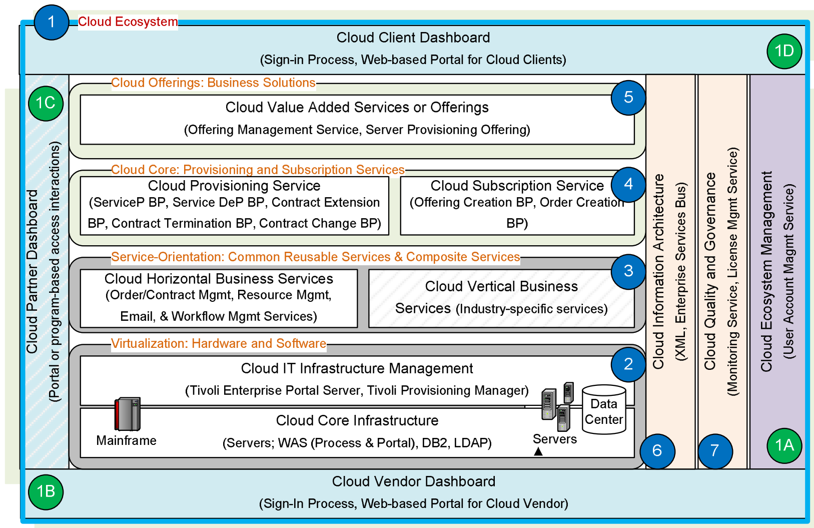

Figure 2. Instantiated Architecture for Infrastructure Cloud

In the Cloud Offering module (Module 5), the Infrastructure Cloud only provides Server Provisioning Offering to cloud clients. More value-added services can be enabled and offered in this Infrastructure Cloud environment without changing the basic modules of its architecture. For example, storage cloud offering can be created for the cloud vendor to share its data centers with cloud clients. All the creation and provisioning businesses of cloud value-added services are managed by Offering Management Service, which is realized in a Web service in this case study.

In the Cloud Information Architecture module (Module 6), we define data structures for subscription orders, contracts, SLAs, project information, and business scenarios. Those data structures and data are stored in DB2 database in the Virtualization module (Module 2). We also define data structures for users and groups. In the Virtualization module, LDAP is used to capture the user related information and access control. Enterprise Services Bus (ESB) technology is used to implement message routing and transformation in the cloud information architecture module.

In the Cloud Quality and Governance module (Module 7), service-level agreements, contracts, and resource statistics are defined based on the data structures defined in the Cloud Information Architecture module. For example, the start and end time of provisioning or de-provisioning services are captured in this module to ensure that services are delivered according to contracts. In the case of exceptions, we can also use this module to track where they come from and to provide solutions. From the enablement perspective, we use Tivoli Enterprise Portal Server to monitor the service operations in the Virtualization module (Module 2). Meanwhile, Tivoli Provisioning Manager can schedule and monitor all provisioning tasks. In the governance aspect, organizations like the Center of Excellence on Cloud Computing and Cloud Operations Team are created to support its overall strategic planning, architecture design and review, alignment, and coordination. The overall productivity, reusability, cost-effectiveness are also identified and reflected in the design of the Cloud solution architecture based on CCOA's seven principles. For example, some policies on when to use Web services and where to allocate machines are also part of the governance process.

As an example, we use the Cloud Vendor Dashboard to monitor the usages and resource statistics which are granted by the Cloud Quality and Governance module. Figure 3 illustrates the statistic data chart of the Cloud Computing resources. This usage diagram shows how many resources are available for cloud clients to subscribe. The resources are categorized based on server types and operating systems (e.g. Windows and Linux). In this case, there are Xen-VM, xSeries, Xen Cluster, VMware, DynamicP6, and DynamicP. We have two operating systems (Windows and Linux) for provisioning. As shown in Figure 3, there are one subscribed Xen-VM machine with Linux and one Xen-VM for subscription. There are 17 xSeries for subscription. There are 13 subscribed Xen Cluster machines with Linux and 9 machines for subscription. There are one subscribed

VMware machine with Windows and one VMware machine for subscription. There are 1 subscribed DynamicP6 with Linux and 5 machines for subscription. There are 4 DynamicP servers for cloud clients to subscribe.

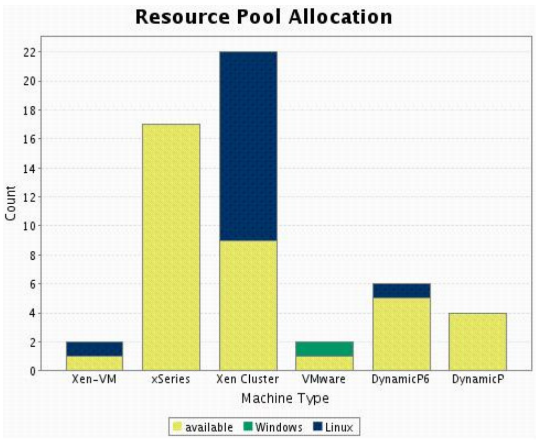

Figure 3. Cloud Resource Statistics

Figure 4 displays a screen of order creation in the Cloud Client Dashboard. Cloud clients can create orders after reading and accepting the license agreement which is maintained in the Cloud Quality and Governance module. The cloud user can pick available resources and their configurations such as operating system (OS), server type, number of CPUs, size of memory (GB), CPU sped (MHz), storage size (GB), and quantity shown in Figure 4.

The corresponding number of available resources is also shown in the dashboard based on the selection of the configurations. There are two categories of server offerings: standard and clustered. For clustered offerings, the cloud users can choose number of nodes in the order request. Once the configuration selection process is done, the order can be added to the shopping cart. The subsequent screens in the dashboard enable users to pick middleware and view the price information.

The dashboard also allows users to select team member for managing or using those resources. Once an order is submitted, the project managers or administrators of the Cloud Vendor Dashboard can review, approve, reject or delete it. After the order is approved, the user can view its status. All business logics and screen flows are implemented as BPEL processes and portlets. The data structure of an order in Figure 4 is defined in the Cloud Information Architecture module (Module 6) shown in Figure 2.

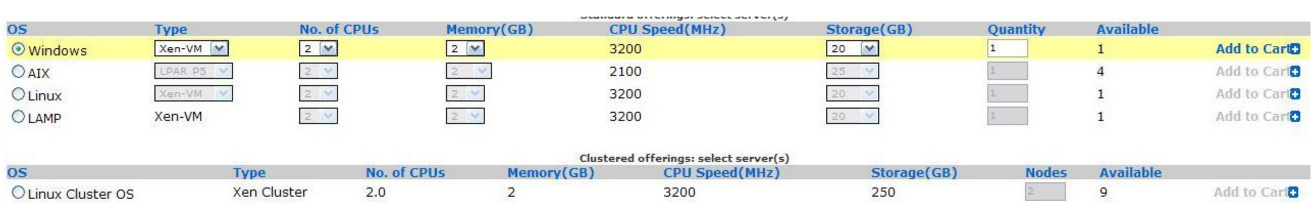

Figure 4. An Example Screen of the Cloud Client Dashboard

3.2 Business Cloud

Business Cloud covers all scenarios of business process as a service in the Cloud Computing environment. CCOA can be used to support the cloud offerings of business cloud. In this section, we employ a public cloud to demonstrate the usage of CCOA to enable business cloud offerings. In order to make a business process smarter, human intelligence can be introduced into it as a sub-process, human task, or decision making within a task.

As illustrated in Figure 5, a company called Cloud Publishing Business (CPB) would like to exploit the Cloud Computing platform to help extend its copyright protection business process that includes three major sub-processes. The first sub-process is Automatic Checking that triggers a third-party's plagiarism detection tool to generate an alert with a summary report when a new manuscript is submitted. A lot of scenarios need human intelligence to analyze the summary report by validating or correcting the findings and regenerate a report. For example, a published paper in a conference proceedings may be enhanced and submitted to a journal for consideration as long as this paper has at least 30% new contents. Or this paper may be enhanced based on his or her own blog entries in the social networking space. In those cases, human beings can perform a more effective job than computers in terms of validation, correction or enhancement of a manuscript. This is the second sub-process named Intelligent Validation and Analysis.

In this business cloud case study, we assume CPB has been convinced to execute this Intelligent Validation and Analysis sub-process in a scalable way. Then the challenge becomes how to quickly access qualified human beings to conduct a quality-guaranteed validation and analysis service. As an example, CPB leverages workforce marketplace such as Amazon Mechanical Turk (AMT) [11] to implement this sub-process. We use CCOA to separate the concerns of business execution and supporting infrastructure. As shown in Figure 5, the workforce marketplace is enabled by Amazon Mechanical Turk which implements key modules of CCOA to support cloud-based publishing business. What CPB needs to do is to leverage Mechanical Turk APIs in its Cloud Client Dashboard to build its own CPB Portal which can co-exist with Amazon's client portal. From the business model perspective, CPB can create Copyright Detection Service to be used internally or as a value-added cloud offering for other publishers or authors to subscribe. The CPB Portal provides interfaces for CPB's customers to submit requirements and job descriptions. CPB launches its overall Copyright Detection business process and generates the initial summary report after the first sub-process completes. Then CPB Portal creates corresponding Human Intelligence Tasks (TITs) based on Amazon's Mechanical Turk APIs or Web services.

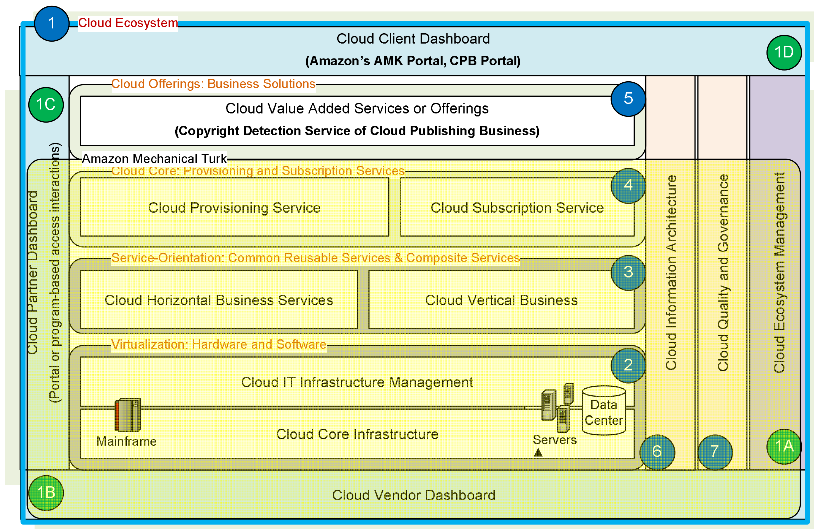

Figure 5. Architecture for Cloud Publishing Business

Theoretically, millions of HITs and qualification requirements can be automatically created in CPB Portal and delivered to Amazon's Mechanical Turk workforce marketplace. Only the qualified work will be approved and paid through AMT. The results of the second sub-process (Intelligent Validation and Analysis) can be programmatically returned to CPB Portal.

In the third sub-process Decision Making and Notification, CPB's case administration team uses a decision making template to summarize the investigation results with a recommendation, and attached analysis reports from the previous two sub-processes and deliver them to the decision making team. Once a decision is made, the case administration team will send out notifications to the business cloud consumer who requested this Copyright Protection Service in CPB's business cloud.

It may be a good exercise for the technical professionals to map AMT's core capabilities to the corresponding modules in CCOA. The connection between AMT and Amazon Elastic Compute Cloud (Amazon EC2) [2] can be enabled in CPB's business cloud offering through Amazon web services. However, it is not covered in this paper due to space constraints.

4. Related Work and Discussions

In the area of application or business process provisioning, we have published several papers on the profiling framework [10][16], provisioning and subscription processes, batch requests handling and monitoring [9][17]. In the application development for virtualized Grid computing environment, we have proposed Grid Solution Sphere concept to unleash the business value of Grid Computing [14][15][17]. The practices accumulated from those exercises have been used to articulate the provisioning and application composition aspect of the CCOA.

In the area of service-orientation, SOA Solution Stack (S3) defines an SOA reference architecture in the context of solution design and delivery [8]. The proposed CCOA is also a natural evolution of these practices and S3.

Once a Cloud Computing system is put into operations, there are two major costs associated, i.e., maintenance service fee and energy consumption cost. There is a trend that building new Internet data centers sharply reduces power consumption and the use of floor space. Ensuring Green technologies to be part of the design phase is very important for effective operations for a Cloud Computing platform in the future. IBM Research's 2008 Global Technology Outlook has identified *Internet Scale Data Centers* as an answer to address infrastructure business around Cloud Computing [18].

The purpose of the proposed approach is to bring the power of SOA and virtualization together to help realize and explore business value of Cloud Computing. We would like to encourage the research community to work together to refine CCOA and focus on the theoretical analysis of CCOA to build solid foundations for Cloud Computing. As for future research topics, we will work on the detailed architectural building blocks and interaction patterns for all ten modules defined in CCOA, as well as standard service interfaces (or APIs) between solution-oriented architectural building blocks and their hosting Cloud Computing infrastructure. Through these interfaces, the hosting infrastructure and value-added cloud services or offerings cooperatively determine proper resource provisioning and management actions. For example, a Cloud Computing platform provides basic services (e.g., Hadoop style [5] and beyond) to help build ultra scalable and resilient data processing capabilities in business solutions. Moreover, one can quickly build new or composite services that leverage existing service-oriented assets and applications.

5. Conclusions

In this paper, we have proposed the Cloud Computing Open Architecture (CCOA) based on seven architectural principles and ten architectural modules, by integrating the power of service-oriented architecture (SOA) and virtualization technology of hardware and software. We have also presented an Infrastructure Cloud as a case study to illustrate how to use CCOA to enable infrastructure-level resource sharing as a cloud offering. We have also studied a business cloud to illustrate how to use CCOA to separate the concern of business design from infrastructure enablement through Cloud Publishing Business. The case studies and analysis have shown that the proposed CCOA is an extensible and configurable architecture for providing normative guidance and enabling infrastructure, software, application, and business process sharing in a unified manner.

References

[1] Liang-Jie Zhang, Carl K Chang, Ephraim Feig, Robert Grossman，Keynote Panel, Business Cloud: Bringing The Power of SOA and Cloud Computing，pp.xix, 2008 IEEE International Conference on Services Computing (SCC 2008), July 2008

[2] Amazon Elastic Compute Cloud (Amazon EC2), http://aws.amazon.com/ec2/, 2009

[3] XDriver Box service, http://www.box.net/xdrive, 2009

[4] Microsoft Skydrive service, http://skydrive.live.com/, 2009

[5] Hadoop Open Source Project, http://hadoop.apache.org/core/, 2009

[6] OSI Model, http://www.osi.org, 2009

[7] John Y. Sayah, Liang-Jie Zhang, On-demand business collaboration enablement with web services, Decision Support System, 40 (2005), pp.107-127.

[8] Ali Arsanjani, Liang-Jie Zhang, Michael Ellis, Abdul Allam, Kishore Channabasavaiah, "S3: A Service-Oriented Reference Architecture," IT Professional, vol. 9, no. 3, pp. 10-17, May/Jun, 2007

[9] Liang-Jie Zhang, Henry Chang, Tian Chao, Jen-Yao Chung, Zhong Tian, Jingmin Xu, Yingnan Zuo, Shunxiang Yang, Qingyun Ao, A Manageable Web Services Hub Framework and Enabling Technologies for e-Sourcing, IEEE Conference on System, Man, and Cybernetics (SMC'02), 2002.

[10] Shun Xiang Yang, Liang-Jie Zhang, Tian Chao, Jing Min Xu, Ying Nan Zuo, Zhong Tian, and Henry Chang: Adaptive profiling framework and system for service provisioning in e-business solutions. Electronic Commerce Research and Applications 3(2): 139-151 (2004)

[11] Amazon Mechanical Turk, http://aws.amazon.com/mturk/, 2009

[12] Liang-Jie Zhang, Jia Zhang, Hong Cai, Services Computing, Springer and Tsinghua University Press, 2007, ISBN: 978-3-540-38281-2, July 2007

[13] L.-J. Zhang,N. Zhou,Y.-M. Chee,A. Jalaldeen,K. Ponnalagu,R. R. Sindhgatta,A. Arsanjani,and F. Bernardini, SOMA-ME: A platform for the model-driven design of SOA solutions , IBM SYSTEMS JOURNAL, VOL 47, NO 3, 2008, pp.397-413.

[14] Liang-Jie Zhang, Jen-Yao Chung, Qun Zhou, Discover Grid Computing, IBM developerWorks Journal, February 2003, pp.14-19.

[15] Liang-Jie Zhang, Qun Zhou, Jen-Yao Chung, Developing Grid Computing Applications, IBM developerWorks Journal, May 2003, pp.10-15.

[16] Jing Min Xu, Ying Nan Zuo, Shun Xiang Yang, Zhong Tian, Henry Chang, Liang-Jie Zhang, Tian Chao: Membership Portal and Service Provisioning System for an Infrastructure of Hubs: Managed E-Hub. ICEIS (4) 2003: pp. 143-150.

[17] Liang-Jie Zhang, Haifei Li, Herman Lam, Toward a Business Process Grid for Utility Computing, IT Professional ,Volume: 6 , Issue: 5 , Sept.-Oct. 2004, Pages:64 – 63

[18] IBM Research, 2008 Global Technology Outlook, http://www-03.ibm.com/procurement/proweb.nsf/objectdocswebview/file3+-+ibm+gto+overview+-+agerwala/$file/3+-+ibm+gtooverview+-+agerwala.pdf

[19] Google Web Applications for Communication and Collaborations. http://www.google.com/apps