The Need for Service Catalog Design in Cloud Services Development

The purpose of this document:

- Provide an overview of the cloud service catalog and show how the service catalog design is an fundamental part of any cloud services definition and development effort
- Discuss the methodology and framework for defining, designing, and deploying a cloud service catalog in service provider and enterprise environments
- Illustrate how Cisco Services can assist organizations through this complex process

Introduction

In the context of cloud computing, the service catalog is an integral and critical component of the cloud computing architecture. Most cloud computing projects will invariably begin with a discussion of "what IT services does an enterprise need?"

Helping companies devise their service catalog strategy, design a service catalog, and design and implement a service catalog portal that supports the underlying cloud infrastructure are primary components of Cisco® Cloud Enablement Services.

The Context: Cloud Computing

Cloud computing is a service delivery model that abstracts the setup and management of IT resources from the underlying infrastructure, providing compute environments in a self-service mode, on demand, and at scale.

An enterprise can deploy cloud computing within its private network. This is commonly referred to as a *private cloud*, as it is restricted in access to the private network. A service provider can provide cloud-based services to its customers over the public Internet. This is commonly referred to as a *public cloud*. An environment that transparently combines both a private cloud and a public cloud is commonly referred to as a *hybrid cloud*.

A cloud can provide IT infrastructure (for example, machines and storage, including the base operating system), an application deployment platform (for example, machines and storage, including the base operating system plus standard enterprise middleware), or subscription-based software. These different types of cloud computing services delivery models are called *infrastructure as a service* (IaaS), *platform as a service* (PaaS), and *software as a service* (SaaS).

IT services that are delivered as cloud services typically have the following attributes:

- Pay as you go: minimal or no initial costs as well as self-service request capability
- Usage based pricing: end-user costs are based on actual resource consumption
- Elasticity: end customers can dynamically consume more or less resources

The NIST definition of cloud computing is a good reference for additional information about cloud computing deployment models and definitions.

The Front End: Service Catalog

Information Technology Infrastructure Library (ITIL® v3) service design defines a service catalog as a list of technology-enabled services that an organization provides, often to its employees or customers. More specifically, the service catalog is an expression of the operational capability of a service provider or enterprise within the context of an end customer, a market space, or an internal business unit stakeholder.

In the context of cloud computing, the service catalog is an integral and critical component of the cloud computing architecture. A cloud service catalog:

- Contains a set of cloud services that an end user can request (usually through a web self-service portal).
- Acts as the ordering portal for cloud end users, including pricing and service-level commitments and the terms and conditions for service provisioning.
- Can also be used as a demand management mechanism, directing or incenting customers toward particular services or service configurations or away from legacy or declining services, as well as making sure of alignment with governance and standards through default configurations and service options.
- Has a self-service look and feel; that is, it provides the ability to select service offerings from the cloud service catalog and generate service requests to have instances of those offerings fulfilled.
- Is useful in developing suitable cloud-based solutions, thus enabling other IT and business services, which in turn create the value propositions for the investments in cloud architectures.
- Contains features and characteristics (atomic items[1]) that can be configured (and preferably priced based upon a "cloud chargeback" mechanism) to fulfill a particular need.
- Serves as the provisioning interface to automated service fulfillment using a cloud orchestration subsystem.

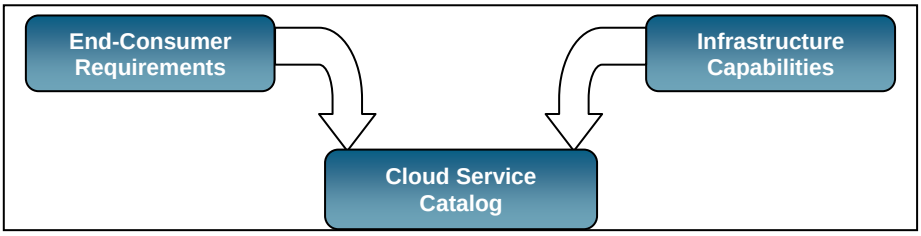
[1] Atomic items/units: Refers to the smallest unit of each of the billable items that will be available to the customers as a service. Atomic units therefore drive data collection, billing, and reporting functions in cloud architecture.

Developing an Optimum Service Catalog

An optimum catalog is one that maximizes the alignment of infrastructure capabilities with business requirements while delivering the best value for the end consumer. The service catalog can be used as an effective tool by IT organizations to implement enterprise standards, introduce new technologies, and enforce default regulatory requirements. The enterprise architect is responsible for the service catalog's alignment with the business architecture, thereby helping to maximize the return on investment in cloud and service catalog development.

It is important to note that an optimized cloud service catalog can only be built when both the business perspective (for example, which services does the business need to deploy?) and the IT perspective (for example, what services can be provided?) are taken into consideration at the same time.(See Figure 1.)

Figure 1. Cloud Service Catalog Inputs



The cloud service catalog development methodology should be:

- **Repeatable**: When a service catalog is built for a customer, the process could be taken and repeated for multiple customers.
- **Measurable**: A service catalog's items should also be measureable in order to be priced for chargeback, as well as managed for availability and performance.
- **Comprehensive:** A service catalog should encompass all the possible combinations of infrastructure capabilities as well as different deployment requirements.

As a result, the cloud service catalog development framework should be:

- **Scalable**: To enable services provided to scale up or down according to market and end-user requirements. It should enable horizontal and vertical scaling requirements of the services provided through transparent integrated automation.
- **Flexible**: To accommodate new and changing service requirements for end consumers and implications on the IT service catalog.

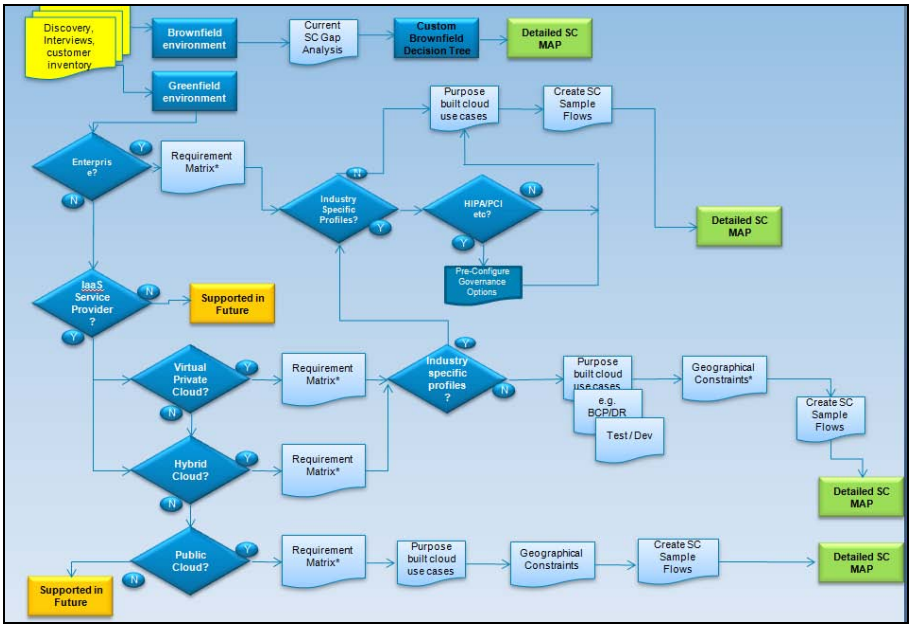
Service Catalog Development Methodology and Framework

The initial framework upon which the service catalog will develop depends in part upon the relative maturity of your IT organization. More advanced organizations will have an enterprise architecture (EA) practice, often at times reporting to the CIO, or an IT executive reporting to the CIO. One of the four pillars of the EA practice is the generation of the enterprise business architecture. In an IT organization with this practice in place, the service catalog will align exactly with the business architecture practice's artifacts and will be used as *the* one tool to manage change into the enterprise. The ability to use this group's work in the creation of the cloud service catalog will add significant velocity to the effort and greatly simplify the work. For the purposes of this paper, therefore, we will focus on organizations that have not yet reached this level of maturity. The following steps outline this methodology and framework for designing a cloud service catalog:

1. Capture initial requirements based on your environment (brownfield/greenfield).
2. Analyze and identify requirements from following perspectives:
   - Business
   - Services capabilities
   - Role-based access
   - Governance and compliance
   - Purpose-built cloud use cases
   - Geographical constraints
3. Create a template of the cloud service catalog based on the distilled requirements.
4. Create sample service catalog work flows for a self-service portal.
5. Review the cloud service catalog design with the customer and incorporate feedback.
6. Iterate through distilled requirements and finalize the design.

These steps are illustrated in the service catalog decision tree (Figure 2) and are discussed in detail in the remainder of this document. Note that this illustration shows an example for IaaS.

Figure 2. Service Catalog Methodology Decision Tree



Current Environment (Brownfield vs. Greenfield)

In a situation where you might already have an existing service catatog in place, the service cataloging process will encompass analyzing the requirements, assessing and identifying any particular gaps in your existing catalog with respect to best practices, and providing recommendations to mitigate those gaps. This is a highly customized effort and is, therefore, not covered in this white paper. This white paper only discusses environments where a service catalog does not exist.

Requirement Analysis Aspects

Business Requirements

The services business requirement analysis shown in Table 1 can be used to distill your business requirements.

Table 1. Services Business Requirements Analysis

Business Service	Market Readiness	Infrastructure Capability	Strategic Fit	Implementation Priority
Managed server hosting	**High**	**High**	**High**	**High**
Managed desktop hosting	**Low**	**High**	**Low**	**Low**
Application hosting	**High**	**High**	**High**	**High**
Disaster recovery	**Low**	**High**	**High**	**Medium**

Service Capabilities

In developing a cloud service catalog it is also important to understand the capabilities of your infrastructure that is intended to be used in the cloud architecture. If you are planning on providing, for example, a generic IaaS, for example, your infrastructure should be able to deliver that capability. The service catalog should describe and enable the allowed permutations of IaaS presented to the consumer as options. This can entail providing a selected set of compute profiles that can be provisioned collectively on a single hypervisor level, while excluding options to provide CPU pinning of the virtual image.

In creating and maintaining a service catalog, attention must be paid toward the alignment of the offered capabilities with the business processes they are supposed to support. In the context of continuous improvement, a regular validation process should be applied to the service catalog's offered capabilities that make sure of alignment to the broadest set of business process requirements. In the initial design of a service catalog, therefore, the "requirements" generation should be performed in the context of assessing and analyzing the "technology-enabled business processes" to distill down to the service catalog's content.

Role-Based Access

Different actors/consumer roles will be accessing the cloud catalog. Best practices recommend associating the permissions for each of the privilege levels of different types of roles to the minimum necessary to make sure of security, access, and compliance requirements.

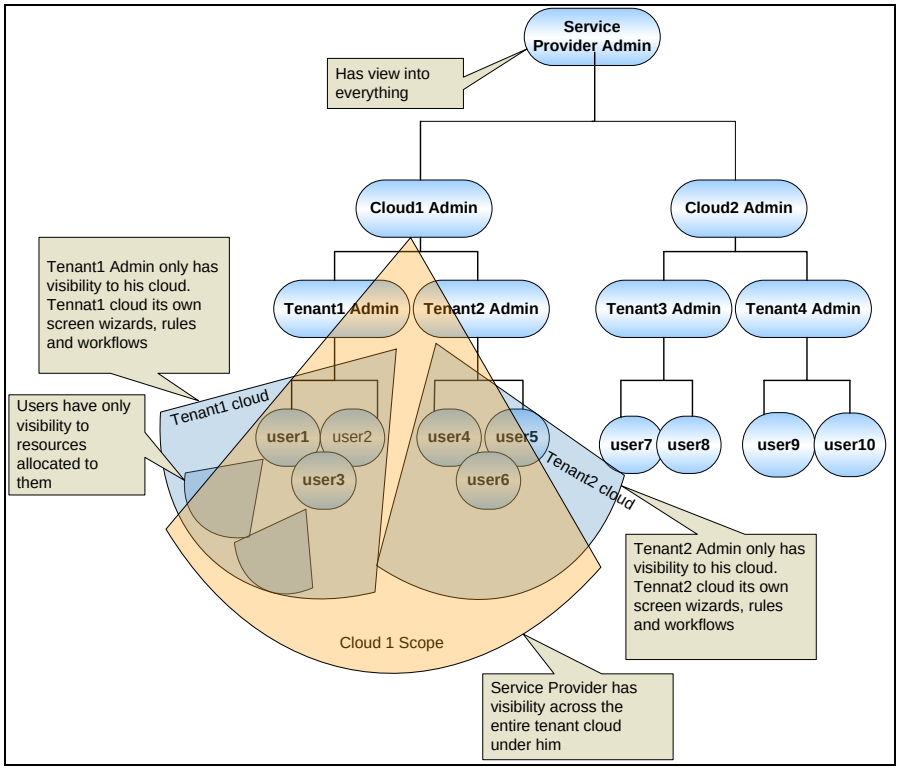
It is also recommended for the sake of efficient and effective cloud management that different roles and their scopes are well documented and standardized. This is crucial since these roles are going to play a part in reporting, monitoring, provisioning, and more in the service catalog portal.

Following is a sample set of roles and scopes that can be used when developing the catalog:

- **Service provider admin:** Super or root user for the entire infrastructure. Scope: Entire infrastructure. Access to all cloud instances within the service provider domain, access to all tenants.
- **Cloud admin:** Root user for a particular cloud within the service provider infrastructure. There could be multiple clouds in a service provider environment. For each cloud there should be a cloud admin. Scope: Has the visibility to own cloud infraresources, not the entire service provider infrastructure
- **End user:** End consumer. Regular user without any administration privileges. Can use resources, see utilization reports, but cannot select reports outside privilege scope. Scope: Very limited scope down to the virtual machine level access.

Figure 3 illustrates the scope of different roles within a multitenant environment and can be used to capture the role-based access requirements.

Figure 3. Role Definitions and Scopes



Governance and Compliance

The service catalog role in governance and compliance might be best understood by example. Any organization deploying cloud services for healthcare or the financial industry must make sure of compliance with relevant HIPAA and PCI regulations. In developing a cloud service catalog, it is important to capture regulatory requirements and predict any effects on the service catalog options as and when they are affected by these regulations. The regulatory or governance constraints are called out as features (or subfeatures) associated to services in the service catalog and facilitate adherence to regulation as well as enhancing auditability.

For example, the HIPAA Security Rules specify certain safeguards that are “required” (that is, must be implemented) and others that are "addressable" (that is, do not have to be implemented if the organization can document why the specification is not reasonable or appropriate to its circumstances). These include:

- Authentication: authenticating entities or individuals prior to data access (required)
- Data integrity: protecting against unauthorized modifications (addressable)

- Data access: controlling which users, applications, and devices can access patient information (addressable)
- Data confidentiality: encrypting data in transit or at rest (addressable)

Therefore, healthcare organizations/customers using a cloud service catalog will have some catalog options that will be “required” (as mandated by HIPAA) instead of just being “optional” for customers in other industry verticals, as applicable in the delivered service.

Purpose-Built Clouds

If the cloud service catalog is built for a specific purpose such as a dev/test environment or a business continuity planning and disaster recovery (BCP/DR) solution, then these specific use cases should be reflected in the requirements of the catalog.

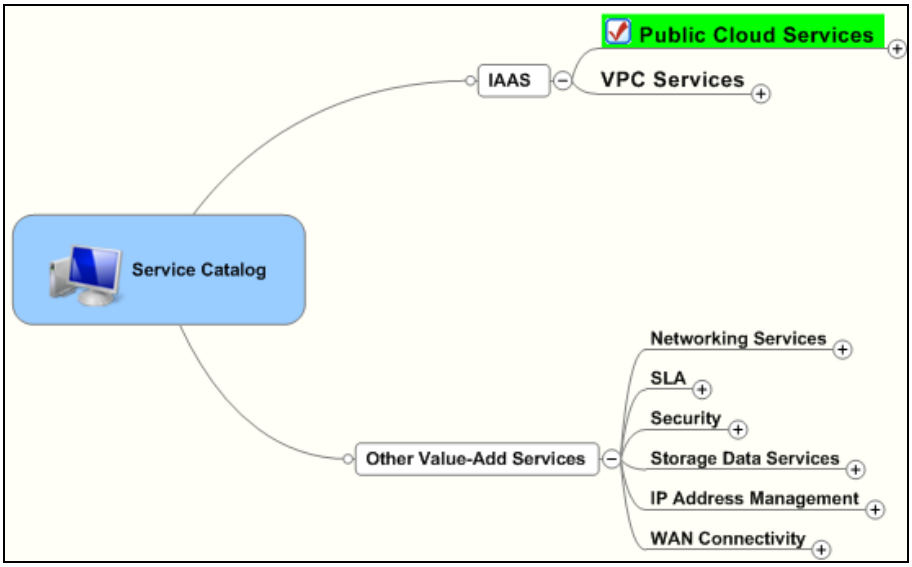
Geographical Constraints

Specific requirements relevant to access, security, load balancing, and the like might be required based on geographic considerations. These considerations could be different from one geographical location to another. For example, since cloud services could be ubiquitously accessed by network, local restrictions in Asia could be different than those in Europe or the Americas. Similarly, load balancing the infrastructure according to the geographical location of the data center can affect the SLA options provided in a cloud service catalog.

Cloud Service Catalog Templates

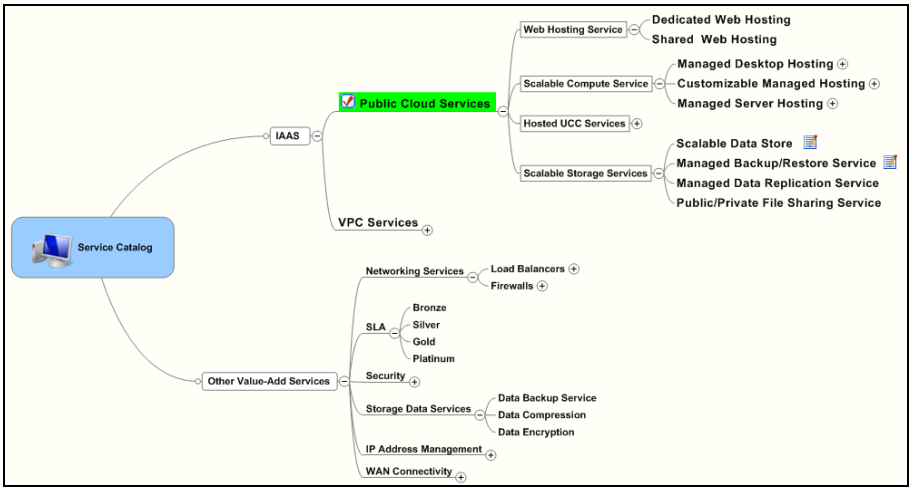
After all the requirements have been captured and distilled, a cloud service catalog template can be designed. This requires mapping all the business requirements into available infrastructure capability models. Figure 4 demonstrates a sample template where a service provider is required to provide a public cloud service along with multiple value-add services, including networking services, SLAs, WAN connectivity, and so on.

Figure 4. Cloud Catalog Services Template (Business-Level View)



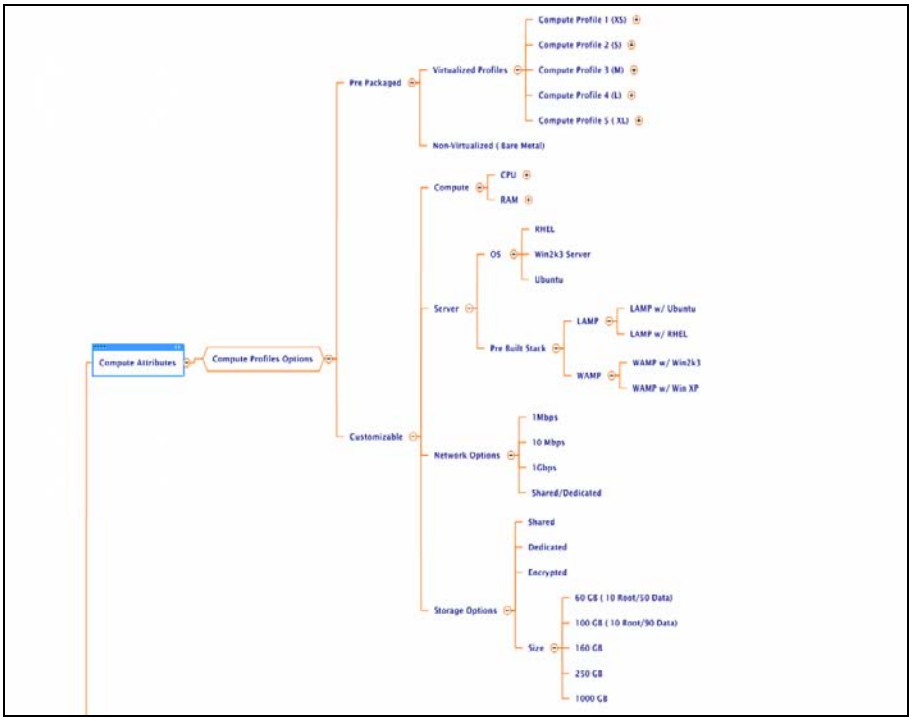
Based upon this business-level view of a service catalog template, it is now possible to start building out the options that will be required for creating a public cloud offering. Here, for example, the various classes of public cloud could be defined, along with the primary business variants requiring dedicated web hosting versus shared web hosting. This is illustrated in Figure 5.

Figure 5. Cloud Service Catalog Template (Expanded Business View)



After the initial high-level templates are built out, the next stage would be to start adding more details to the options, based on the capability of the infrastructure. These technical details could, for example, include the amount of storage, the CPU rate, and the operating system to be installed. This is illustrated in Figure 6.

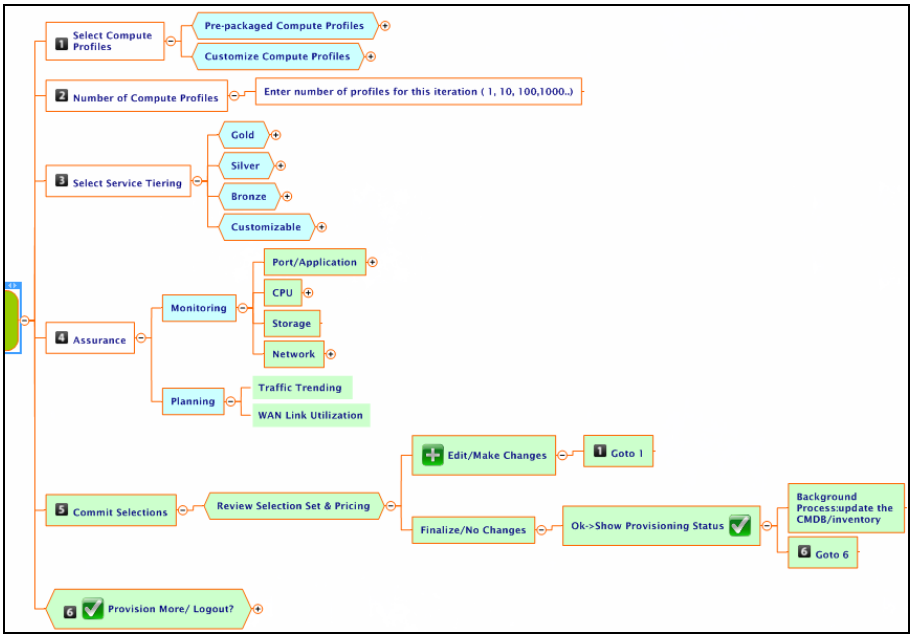
Figure 6. Adding Technical Specification Details to Cloud Service Catalog Template



Service Catalog Work Flows

Once all the levels of detail have been finalized for a cloud service catalog, the details of how the catalog will be utilized by end users can be outlined by providing sample work flows (effectively sequences of operational steps). The purpose of work flows is to provide insight and guidance regarding the self-service portal functionality, which can then pass on the provisioning/monitoring/management information to an orchestration layer below the service catalog to interact accordingly with the underlying infrastructure. Figure 7 illustrates one such example: a user can select a compute profile, select a service tiering (gold/silver/bronze and so on), define what level of monitoring is applied to the service instance, and commit activation of the service instance across the multiple pieces of equipment required to deliver this service.

Figure 7. Cloud Service Catalog Sample Work Flows



Designing the Solution: The Case for Professional Services

Defining and designing the right service catalog for your cloud computing solution is a demanding task. It requires detailed service catalog design experience and deep understanding of cloud computing models as well as detailed knowledge of how the service definitions interact with data center equipment, the external (cloud) environment, and the underlying infrastructure. It also requires experience integrating IT systems and cloud management and – a significant project in its own right – a structured approach to program management.

Cisco Services has significant expertise and experience in each of these areas, backed up by multiple years of helping customers transform their data centers. Cisco Services offers a range of Cloud Enablement Services, from strategy to planning and design to implementation. Application migration is an integral component in each of these services, and detailed application methodologies have already been developed by the cloud computing experts in Cisco Services. Cisco Services, therefore, is ideally placed to help you design and develop a state-of-the-art application migration approach for your cloud computing deployment.

Summary

A cloud service catalog is a critical component of a cloud architecture since it provides an abstraction for the underlying infrastructure. It enables other IT and business services, which in turn create the value propositions for the investment in cloud architectures.

A well-defined service catalog methodology and framework, therefore, are paramount before an investment in cloud architectures is made. Engaging Cisco Services to help you with this design exercise will provide you with additional expertise to accelerate your cloud services adoption.

For More Information

Data Center Services on Cisco.com
www.cisco.com/go/dcservices

Cisco Cloud Enablement Services on Cisco.com
www.cisco.com/go/cloudenablement

Email: ask-cloud-services@cisco.com

Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV Amsterdam,
The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at **www.cisco.com/go/offices.**

Printed in USA C11-669500-00 05/11