Oracle® Reference Architecture

Cloud Infrastructure

Release 3.0

E25633-03

November 2011

Cloud Infrastructure, Release 3.0

E25633-03

Primary Author: Anbu Krishnaswamy Anbarasu

Contributing Author: Dr. James Baty, Stephen Bennett

Contributor: Mark Wilkins, Scott Mattoon, Dave Chappelle, Bob Hensle, Cliff Booth, Martin Cookson, Constantine Steriadis, Dhruv Gupta, Honglin Su, Teju Ajani, Vincent Ho, Arvind Jain, James Bayer

Contents

3 Cloud Management Infrastructure

4 Cloud Implementation Perspectives

5 Product Mapping View

6 Deployment View

7 Summary

A References and Resources

List of Figures

Preface

Oracle Reference Architecture (ORA) is a product-agnostic reference architecture based on architecture principles and best practices that are widely applicable and that can be implemented using a wide variety of products and technologies. ORA does not include any implementation artifacts for the prescribed architecture. Rather, ORA addresses the building of a modern, consistent IT architecture while minimizing the risk of product incompatibilities and obsolescence.

ORA is an extensible reference architecture that describes many facets of IT. It is comprised of several documents that cover core concepts of technology, along with other documents that build upon these core concepts to describe more complex technology strategies.

The ORA Cloud documents present the ORA architectural concepts from the perspective of Cloud, highlighting the specific details of Cloud as an elaboration of the ORA core concepts with respect to this technological approach. This ORA Cloud perspective comprises two documents:

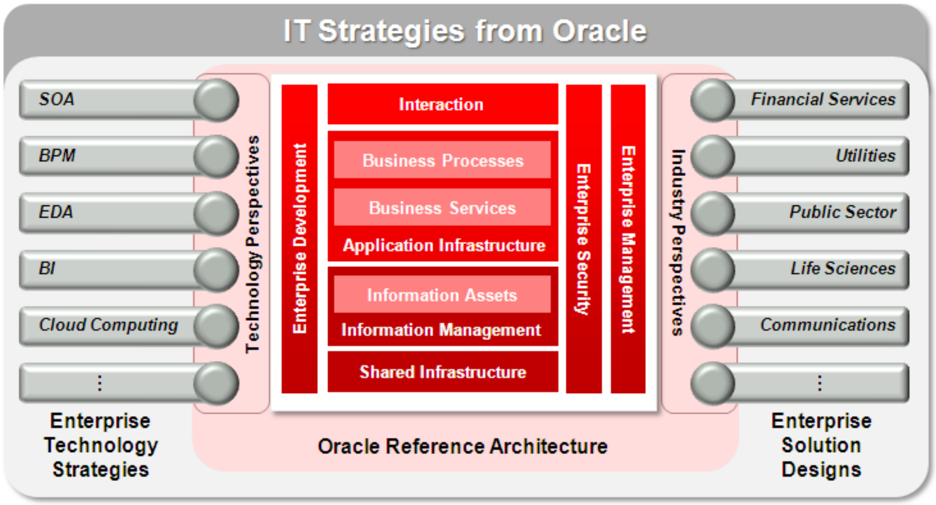
- Cloud Foundation: primarily a conceptual architecture for Cloud, specifying architectural characteristics and expectations of Cloud at a business and operational level. Also included in this document are architectural principles, standards, concepts, a conceptual view for Cloud architecture, and its relationship to ORA.
- Cloud Infrastructure: relates the Cloud characteristics and requirements, as defined by the conceptual architecture, to the Oracle infrastructure and provides a number of architecture views to help architects and developers who are focusing on Cloud.

This document is part of a series of documents that describe the Oracle Reference Architecture (ORA). This "Cloud Infrastructure" document provides the underlying architectural definitions of associated Cloud concepts. The Cloud Infrastructure document presents the key infrastructure capabilities and architecture views for Cloud providers to build and manage an Enterprise Clouds.

This document is part of the body of knowledge, IT Strategies from Oracle (ITSO). Please consult the ITSO web site for documents pertaining to Cloud and other technologies.

Introduction to IT Strategies from Oracle (ITSO)

IT Strategies from Oracle (ITSO) is a series of documentation and supporting material designed to enable organizations to develop an architecture-centric approach to enterprise-class IT initiatives. ITSO presents successful technology strategies and solution designs by defining universally adopted architecture concepts, principles, guidelines, standards, and patterns.



ITSO is made up of three primary elements:

- **Oracle Reference Architecture (ORA)** defines a detailed and consistent architecture for developing and integrating solutions based on Oracle technologies. The reference architecture offers architecture principles and guidance based on recommendations from technical experts across Oracle. It covers a broad spectrum of concerns pertaining to technology architecture, including middleware, database, hardware, processes, and services.
- **Enterprise Technology Strategies (ETS)** offer valuable guidance on the adoption of horizontal technologies for the enterprise. They explain how to successfully execute on a strategy by addressing concerns pertaining to architecture, technology, engineering, strategy, and governance. An organization can use this material to measure their maturity, develop their strategy, and achieve greater levels of adoption and success. In addition, each ETS extends the Oracle Reference Architecture by adding the unique capabilities and components provided by that particular technology. It offers a horizontal technology-based perspective of ORA.
- **Enterprise Solution Designs (ESD)** are industry specific solution perspectives based on ORA. They define the high level business processes and functions, and the software capabilities in an underlying technology infrastructure that are required to build enterprise-wide industry solutions. ESDs also map the relevant application and technology products against solutions to illustrate how capabilities in Oracle's complete integrated stack can best meet the business, technical, and quality of service requirements within a particular industry.

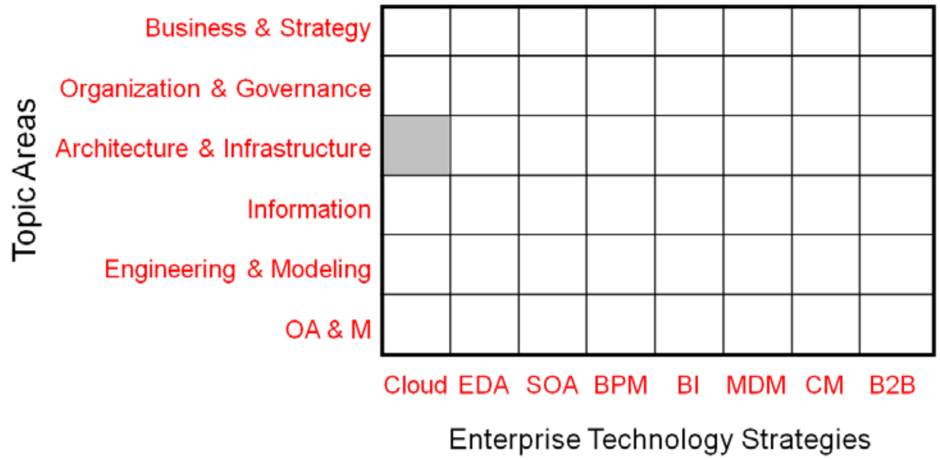
This document is part of a series of documents that comprise the Cloud Enterprise Technology Strategy, which is included in the IT Strategies from Oracle collection.

Please consult the ITSO web site for a complete listing of Cloud and ORA documents as well as other materials in the ITSO series.

Document Purpose

This document provides a Cloud perspective to the Oracle Reference Architecture. It offers several architecture views including logical, product mapping, and deployment views. It also describes the role of infrastructure in the reference architecture.

ORA documents are mapped by technology concern to areas of interest in order to represent document purpose. This mapping for the Cloud Infrastructure document is shown in the diagram below:



Audience

This document is intended for enterprise architects, application architects, project managers and developers. The material is designed for a technical audience interested in learning about the Cloud Architecture and how to prepare to develop an enterprise class Cloud infrastructure.

Document Structure

This document is organized in chapters spanning introduction, logical architecture, product mapping, deployment, implementation scenarios, and appendices. The main document chapters are as follows:

Chapter 1 provides brief introduction to Cloud infrastructure.

Chapter 2 provides a logical view of the Cloud infrastructure.

Chapter 3 describes the Cloud management capabilities.

Chapter 4 presents implementation perspectives that include use cases, business scenarios, and design patterns.

Chapter 5 provides product mapping to the logical and use case views along with the detailed product information.

Chapter 6 discusses some deployment views of the Cloud infrastructure.

Chapter 7 provides a summary of this document.

Appendix A provide lists of relevant supplementary reading and references.

How to Use This Document

This document is designed to be read from beginning to end. However, each section is relatively self contained and could be read independently from the other sections.

Conventions

The following typeface conventions are used in this document:

Convention	Meaning
boldface text	Boldface type in text indicates a term defined in the text, the ***ORA Master Glossary***, or in both locations.
italic text	Italics type in text indicates the name of a document or external reference.
underline text	Underline text indicates a hypertext link.

1 Introduction

Cloud computing is a significant advancement in the delivery of information technology and services. By providing on demand access to a shared pool of computing resources in a self-service, dynamically scaled and metered manner, Cloud computing offers compelling advantages in cost, speed and efficiency.

Cloud computing has emerged as one of the most important new computing strategies in the enterprise. A combination of new technologies and processes has led to a revolution in the way that computing is developed and delivered to end users. It promises the ability to deliver applications at a lower cost. However, today's Cloud computing environment is highly fragmented and standards are evolving to accelerate the adoption of Cloud technologies. Oracle is actively involved in developing and promoting key industry standards around Cloud computing.

Traditionally deployments require applications to be bound to a particular infrastructure. This results in low utilization, diminished efficiency and inflexibility. Cloud brings in capabilities to allow applications to be dynamically deployed onto the appropriate infrastructure at runtime. This elastic aspect of Cloud computing allows applications to scale and grow on demand without needing traditional patches or upgrades.

IT departments and infrastructure providers are under increasing pressure to provide computing infrastructure at the lowest possible cost. There are also a growing number of public Cloud providers that are looking for ways to build a versatile Cloud Infrastructure to support their clientele. To do this, the concepts of resource pooling, virtualization, dynamic provisioning, utility and commodity computing could be leveraged within the public Cloud or create a private Cloud that meets these needs. Customers driven by concerns over security, regulatory compliance, control over Quality of Service (QoS), and long-term costs, will build internal private Clouds. Private Clouds allow internal IT providers and application development teams more control of data security and meet their governance regulations. However, as the technology matures and these concerns ease, we will see more customers adopting a hybrid Cloud model that makes use of both private and public Clouds using the most suitable of the hybrid strategies.

This document focuses on the Cloud provider view. It covers the capabilities for public and private Clouds, a discussion of Cloud architectures and key architecture views to start your Cloud architecture initiative. This document was written for Enterprise Architects and experienced technologists to understand the Cloud computing architectures.

1.1 Definition of Cloud Computing

It is important to understand the definition of Cloud Computing before we look at the Cloud Infrastructure. The *ORA Cloud Foundation Architecture document* describes the definition of Cloud Computing in detail. The National Institute of Standards and Technology, Information Technology Laboratory (NIST) definition is repeated here for quick reference.

> **Definition:** "Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This Cloud model promotes availability and is composed of five essential characteristics, three service models, and four deployment models."

It should be noted that these characteristics and models do not apply equally to all 'Clouds' e.g., a self-service developer public Cloud likely has a completely different implementation from a government agency 'Community Cloud'.

1.2 Cloud Infrastructure Vs. Infrastructure as a Service (IaaS)

Before we get into the details, it is important to clarify the terminology we use in this document. This document uses the term "Cloud Infrastructure" to denote the control plane that manages and monitors the Cloud resources. Cloud Management Infrastructure (CMI) and Cloud Infrastructure are used interchangeably in this document.

Cloud Infrastructure is not be confused with Infrastructure as a Service (IaaS). IaaS is described in the *ORA Foundation Architecture document*. It is a type of service used to deliver infrastructure resources such as Compute, Network, and Storage over the Cloud.

1.3 Cloud Infrastructure Implications

Let us begin with the NIST definition of the Cloud and see what the Cloud Infrastructure must provide in order to generally support the key elements of the definition.

The table below illustrates the implication of the five essential characteristics on the Cloud Infrastructure. Although these characteristics may sound simple at first glance, a closer look would reveal that much deeper infrastructure capabilities are required to support these basic requirements. This is especially true for large Cloud providers that needs to support multiple large shared deployments.

Table 1–1 Implications of the five essential characteristics

Characteristics	Infrastructure Implications
On-Demand Self-Service	■ Support Self-service portals ■ Support APIs for self-service ■ Automation everywhere - registration to provisioning to management

Table 1–1 (Cont.) Implications of the five essential characteristics

Characteristics	Infrastructure Implications
Rapid Elasticity	■ Support for capacity on-demand ■ Right level of granularity of deployable entities for rapid change ■ Grids and Clusters to support elastic infrastructure ■ Engineered Systems offer great levels of coordination between pooled resources to enable rapid elasticity
Resource Pooling	■ The ability to estimate resource sizing to make appropriate number of resources available in the pool ■ Pool management capabilities to manage the active instances in the pool. ■ Multi-tenancy support to ensure that the pool resources can be assigned to multiple tenants ■ Abstraction/virtualization capabilities are required to divide the resources into manageable and measurable chunks that can be pooled. ■ Location transparency
Broad Network Access	■ High speed, high bandwidth network access capabilities need to be made available. ■ Multichannel support is required to support a variety of mobile devices and appications. ■ Standards based access should be supported.
Measured Service	■ Definition of metrics - Metrics most suitable for measuring the service consumption should be defined. ■ Metrics capture - Infrastructure should support the collection of the usage metrics. ■ Support for SLA and policy management to ensure that the Cloud services meet the SLA requirements and any violations are captured. ■ Billing management - Use the metrics collected to rate the charges and create invoices. ■ Revenue management - Manage the payment and accounting. ■ Customer management - Manage the customer accounts and associated service levels.

Now, let us see how the service models drive the requirements of a Cloud Infrastructure. The service infrastructure exposes the resources and services to the Cloud Consumer using appropriate interfaces. These service models may be deployed independently or in any possible combinations. A cascaded deployment may make it challenging to measure and manage the consumption and drive up the complexity of the infrastructure. The table below captures the key implications of these service models.

Table 1–2 Implications of the three Service Models

Service Model	Infrastructure Implications
SaaS	■ Support for running multi-tenant, shared business processes and services ■ Monitor and measure the services for an given customer ■ Protection and isolation of customer data in multi-tenant environments ■ Support for business processes, business services, applications, etc.
PaaS	■ Scalable platforms, ready to be deployed on-demand ■ Flexible interface to deploy customer applications/services ■ Monitor and measure platform usage for a given customer ■ Process model that supports isolation in multi-tenant environments ■ Support for a variety of platforms including containers, application servers, middleware, databases, queues, etc.
IaaS	■ Ability to expose the infrastructure resources through abstraction ■ Logical separation of underlying resources ■ Support for infrastructure isolation in multi-tenant environments ■ Support for infrastructure resources such as compute (hardware/OS/VM), storage, network, etc.

The Cloud Infrastructure must also support the four deployment models mentioned in the NIST definition. These deployment models bring in their own set of requirements that impose some interesting architectural constraints on the Cloud Infrastructure. The interests of an IT organization that implements a Private Cloud may be totally different from that of a Cloud Provider or a Cloud Broker that implements a Public Cloud. The capabilities of the Cloud Infrastructure for each of these cases may vary and the architecture should ensure that the infrastructure capabilities are prioritized based on the motivation and goals of the organization. The table below summarizes the key infrastructure implications of the deployment models.

Table 1–3 Implications of the four Deployment Models

Deployment Model	Infrastructure Implications
Private	■ IT as a business ■ Internal cost allocation and chargeback. Not all enterprises use chargeback but those that choose to use chargeback require infrastructure capabilities that provide chargeback and showback ■ Separation of Cloud resources from traditional IT resources ■ Internal economies of scale
Community	■ Inter-community economies of scale ■ Community interaction ■ Sharing of common resources ■ Security implications

Table 1–3 (Cont.) Implications of the four Deployment Models

Deployment Model	Infrastructure Implications
Public	■ Security ■ Network access ■ Subscription management ■ Contracts and SLA management ■ Multi-tenancy ■ Metering, billing, and revenue management - public Clouds need to be able to meter the services they provide, bill to the customers, and manage revneue.
Hybrid	■ Inter-Cloud interaction ■ Security propagation and synchronization ■ Cloud bursting support ■ Virtual Private Cloud (VPC) support ■ Synchronization support (Data and Services) - for Business Continuity/Disaster Recovery

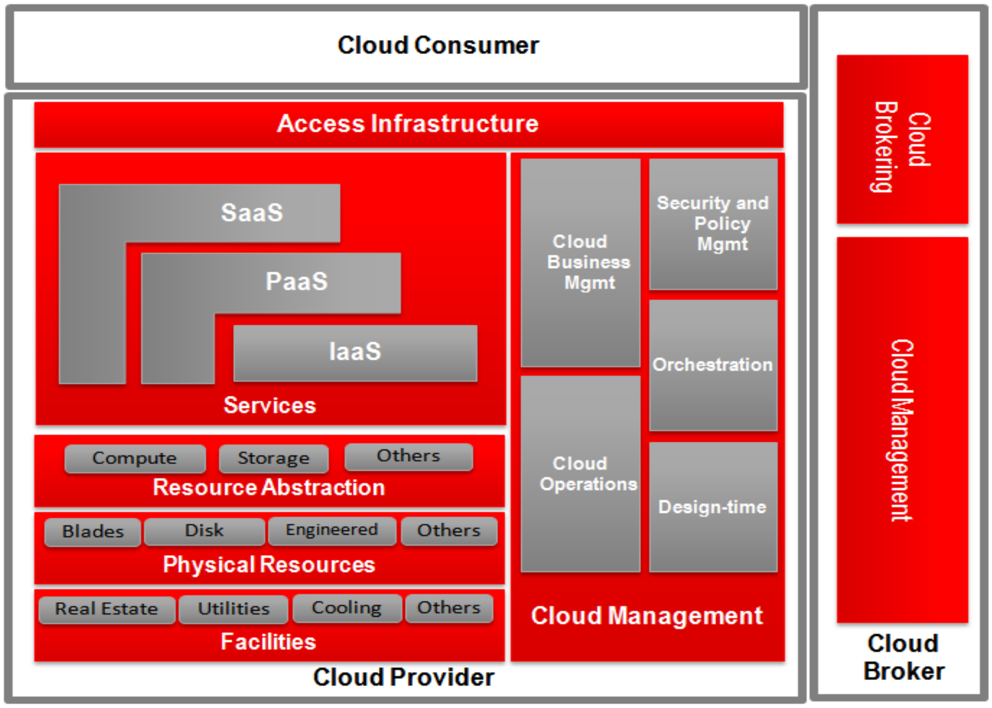
1.4 Need for a common Cloud Management Infrastructure

Cloud management responsibilities should be architecturally isolated and grouped in order to achieve the economies of scale required by Cloud. Regardless of the kind of service used and type of resource requested, the management responsibilities generally remain the same. The cross cutting concerns addressed by this layer, aka control plane, include security and policy management, model management, orchestration, business management, and runtime services. Capabilities such as metering of services, billing, customer management, and revenue management are cross cutting across the various service models, and hence should be designed to be versatile.

1.5 Cloud Infrastructure Overview

Figure 1–1 shows a high level overview of the Cloud infrastructure components from different perspectives. These perspectives are described in detail below.

Figure 1–1 Cloud Infrastructure Overview



1.5.1 Cloud Consumer Perspective

The infrastructure on the Cloud consumer side depends on the degree of sophistication required by the consuming organization. The types of components required in a consuming infrastructure include but not limited to:

- Security infrastructure for security management, encryption, and Virtual Private Networks (VPN).
- Network components that include routers, switches, load balancers
- Monitoring components that may include agents, gateways, and collectors

For example, a Cloud Consumer that uses a Hybrid Cloud Architecture may require components to monitor the capacity of the local resources and to burst into a public Cloud as needed.

1.5.2 Cloud Provider Perspective

The role of the Cloud provider is the most important and most complex of all. Infrastructure for the Cloud is usually of unprecedented scale and stringent requirements. Implementing the Cloud and maintaining it to satisfy the SLAs of the consumers requires extensive planning and precise execution. The provider should explore and plan several aspects of the Cloud infrastructure, which are briefly described below.

1.5.2.1 Facilities

A key part of the infrastructure expenses that is often overlooked is facilities. Especially, for small organizations, it is not economically feasible to establish a full data center that supports their SLAs. A Cloud provider can spread the costs of facilities, among other expenses, across consumers to achieve economies of scale that are otherwise impossible. Facilities may include real estate, cooling, utilities, and rack space among other things.

1.5.2.2 Physical Resources

The physical infrastructure components may include blades, networks, engineered systems, and disks. These resources must be pooled and provisioned through grid technologies in order to support the elasticity and scalability demands of Cloud infrastructure.

Definition: Engineered systems are a hardware and software stack designed and built as a single product. This allows both the hardware and software to be configured and even modified to provide the very best compatibility; thereby providing improved reliability, availability, security, and performance.

As companies move to cloud computing environments, the introduction of engineered systems into cloud architecture has an important role. The adoption of engineered systems shifts the architectural strategy away from a monolithic, shared-server architecture to a service-oriented, highly standardized, highly modularized architecture; thereby facilitating the transition to a dynamic cloud environment. Services provided by engineered systems may be considered as infrastructure or platform services provided within a dynamic cloud implementation.

The ORA core document "*ORA Engineered Systems*" provides more details on Engineered Systems and how they influence Cloud architectures.

1.5.2.3 Resource Abstraction

The physical resources may not be suitable to be served to the consumer directly. They need to be logically partitioned and secured in order to support multi-tenancy. Rapid elasticity requires the resources to be quickly deployed and undeployed at granular levels. Traditional deployments require downtime for scaling and maintenance but Cloud infrastructure does not have that luxury. The approach is to create and deploy new instances on the fly to grow, shrink, or fix the existing deployments. The resource abstraction layer provides the capabilities to logically abstract the physical resources.

1.5.2.4 Services

The Service infrastructure exposes the *aaS services to the consumers through the access infrastructure. Many resources can be offered as services but the ORA conceptual model defines three broad categories of services listed below.

- Infrastructure as a Service (IaaS)
- Platform as a Service (PaaS)
- Software as a Service (SaaS)

The diagram also illustrates a core concept that there is no particular dependency between these three layers. PaaS and SaaS may exist independently without an IaaS service. In this case, they would use dedicated infrastructure for running the PaaS and SaaS services. Alternatively, they may use the any of the underlying layers. For example, a SaaS service may leverage a PaaS service that is built on an IaaS service. It should also be noted that the consumers may access any of these services through the access layer, as indicated by the placement of the layer bars.

1.5.2.5 Access Infrastructure

The access infrastructure controls access to the services and provides the appropriate interfaces to the consumers to access the services.

1.5.2.6 Cloud Management

The Cloud management infrastructure provides the control plane that was discussed in the previous section. It provides the key capabilities to manage the resources, control access, and govern the infrastructure. Cloud Management Infrastructure is explained in great depth in Chapter 3, "Cloud Management Infrastructure".

1.6 Cloud Broker Perspective

Cloud Broker infrastructure is a mix of Cloud Provider infrastructure and Cloud Consumer infrastructure. Cloud Brokers act both as Cloud Providers and Cloud Consumers at the same time by providing value added services that are built on top of the services offered by the providers to the consumers. So, Cloud Broker requires additional capabilities to the typical provider Cloud Management Infrastructure. Cloud Brokers also require both the Access and Consumption infrastructures since they perform the functions of both providers and consumers.

The other unique piece of the infrastructure the Cloud Broker requires is the Cloud Brokering infrastructure that provides capabilities to add value to the provider services. According to Gartner, Cloud broker capabilities may include aggregation, arbitrage, and intermediation.

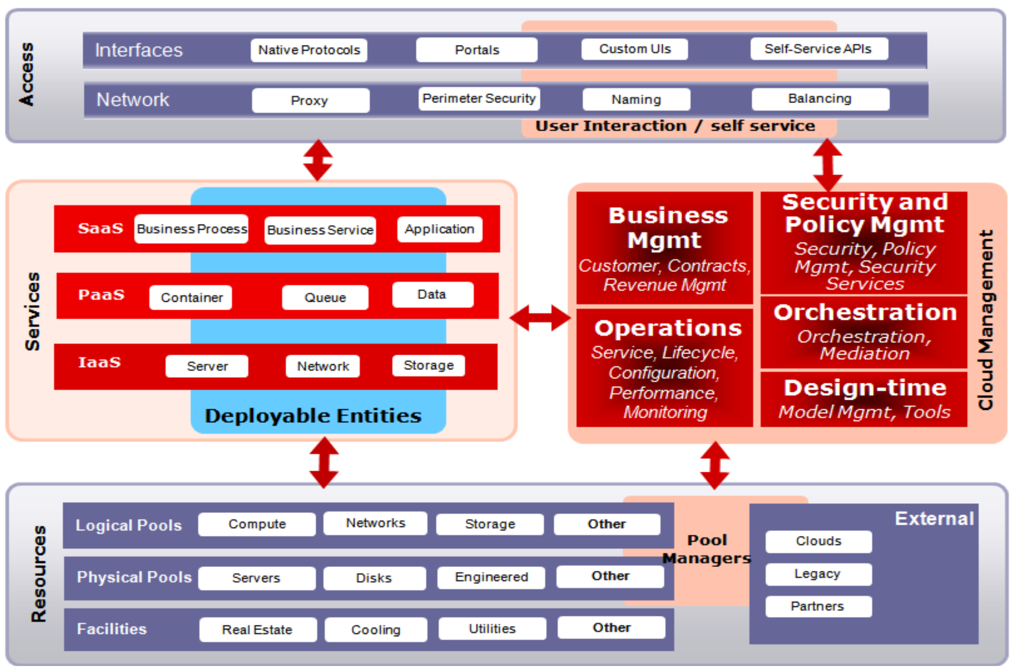
Cloud Broker perspective is out of scope for this document. However, it is an important role and one should understand how it fits in the Cloud eco-system.

2

Logical View

This section describes the logical view of the Cloud infrastructure. Figure 2–1 shows the logical layers and key logical components of the Cloud infrastructure.

Figure 2–1 Logical View



The logical architecture view shows the four major layers of Cloud architecture listed below.

- Access Layer
- Services Layer
- Resources Layer
- Cloud Management Layer

These layers are described in detail below.

2.1 Access Layer

The access layer decomposes into two important sets of functionality: interfaces and facilities.

2.1.1 Interfaces

The Cloud needs a variety of interfaces to allow for access to the underlying services as well as its management capabilities. While these will vary depending on the Cloud implementation, most Clouds will include the ability to access the services and management capabilities via:

- Native protocols (mostly for the services tier)
- Customer portals or self-service applications for Cloud application management
- APIs for Cloud application management

2.1.2 Network

In addition to the end-user facing interfaces present in the access tier, there will be several operator specific capabilities present. Most of these will be familiar to IT departments, as they cover infrastructure requirements similar to most enterprise networks:

- Naming (such as DNS)
- Perimeter security (such as firewalls)
- Proxies (for bridging or tunneling traffic)
- Balancing (load balancing for applications)

While most of these capabilities will be invisible to users other than the Cloud operator, some may be exposed as capabilities within services. For example, load balancing may be a feature exposed to IaaS developers to allow them to present highly available and scalable presentations tier within their applications. It is important that the provisioning process assigns and configures such network resources to support the services provided.

2.2 Cloud Management Layer

A common Cloud management layer provides the management capabilities required for all types of services. A unified management framework provides a holistic view of the infrastructure and end-to-end visibility. The management layer provides support for both build-time and run-time activities.

Figure 2–2 Cloud Management Layer

The top-level categories of capabilities are listed below:

- Business Management
- Operations
- Security and Policy Management
- Orchestration
- Design-time

Within each of these groupings within the management layer, a number of sub-capabilities exist. Chapter 3, "Cloud Management Infrastructure" describes these capabilities in greater detail.

2.3 Resources Layer

The resources layer represents the logical and physical resources provided through the Cloud infrastructure.

2.3.1 Logical & Physical Pools

Logical and physical pools of resources contain virtualized or physical resources managed by the pool manager. These resources may be fairly simple in nature (as the single virtualized server example above) or complex (such as a highly-available or clustered java application platform).

Examples of physical resources may include:

- Computer, Blades
- Storage Disks
- Network Components (Routers, Switches, Load Balancers etc.)
- JEE containers
- Queues

- Databases

The logical resources layer provides a layer of abstraction on top of the physical resources to ensure that the resources can be divided across the consumers. Server virtualization, service virtualization, and database partitioning are some of the examples of the technologies used for creating the logical resources. The list below shows some examples of logical resources.

- Compute (e.g. Virtual Server)
- Storage
- Virtual IP
- Database Schemas

2.3.2 Pool Managers

Pool Managers aggregate, coordinate, and expose virtual or physical resources to the Cloud management layer for deployment into the service layers for consumption. In addition to this coordination, pool managers will also provide other management functions similar to those found in the Cloud management tier, but localized in scope to the resources under their control.

Pool managers will also provide other run-time functionality similar to the service management capabilities of the Cloud management layer, but scoped to only the resources under their control. For example, most pool managers will provide basic logging and metering capabilities so that they can be relayed and then aggregated by the Cloud management layer for chargeback.

2.3.3 Facilities

Facilities refer to the underlying physical infrastructure such as real estate, utilities, cooling infrastructure, and other shared facility infrastructure. It may sound trivial but for a large Cloud provider the cost of facilities may prove out to be very expensive. Many organizations strategically locate their Cloud data centers to reduce the real estate and cooling expenses. For example, some organizations choose to build their data centers near the Arctic Circle to reduce cooling costs.

2.3.4 External Resources

External resources may include third party Cloud services, legacy systems, and partner services.

2.4 Services Layer

The services layer is not technically part of the operator's view of the Cloud. It contains the end user services deployed in the Cloud (the developers and application owners who deploy services into the Cloud infrastructure), which are combinations of IaaS, PaaS or SaaS presented by Cloud management and underpinned by Cloud resources.

The services layer also shows the concept of deployable entities that are logically grouped and packaged service resources that can be used to rapidly provision the services. The deployment entities are templatized to automate and accelerate the provisioning of services.

3 Cloud Management Infrastructure

This section describes the key capabilities of the Cloud Management infrastructure.

3.1 Cloud Management Capabilities

Cloud management is the core of the Cloud Provider infrastructure. It encompasses a wide range of technologies to manage the end-to-end Cloud lifecycle.

Figure 3–1 Cloud Management Infrastructure Capabilities

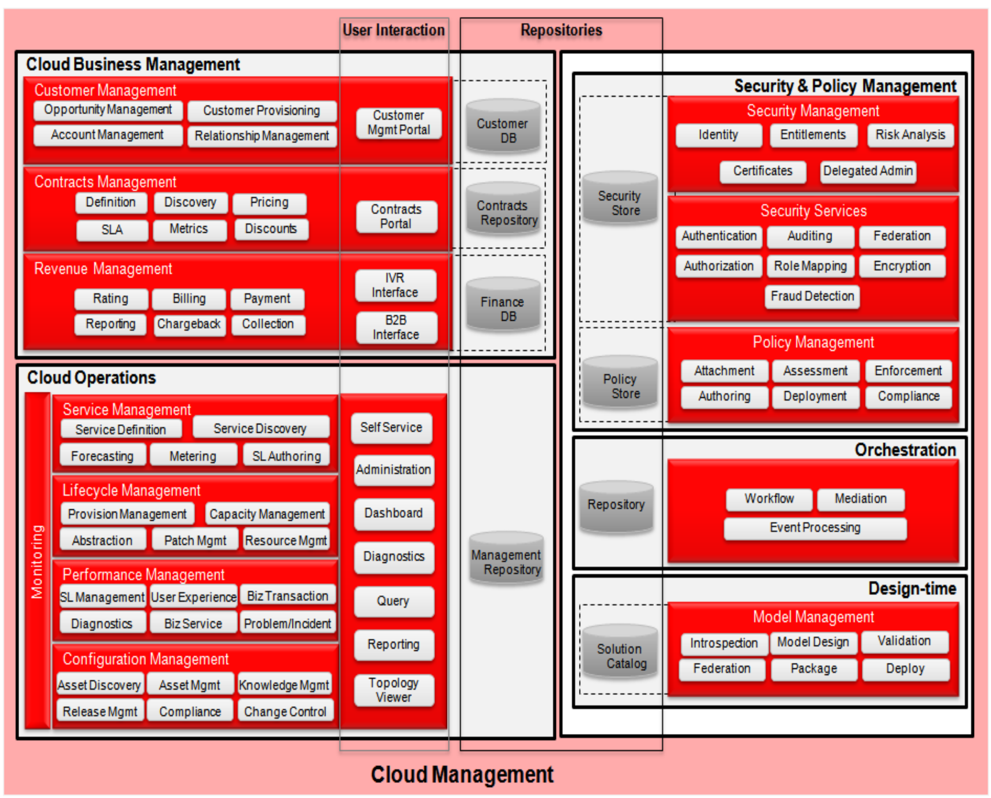


Figure 3–1 lists the key capabilities of the Cloud management layer. These are described in the section below.

3.2 User Interaction and Repositories

A core trait of the Cloud architecture is automation and the ability to self-service most the consumer needs. To that affect, access interfaces of various forms should be provided to the consumer to access the self management capabilities of the Cloud. The type of interfaces may range from sophisticated self-service portals to low level APIs to automate the management of the Cloud resources. The user interaction components show the logical view of the interface that is most appropriate for the task in hand.

Another cross cutting concern is the management and sharing of information across various management activities. For example, the service usage information is captured by the runtime/operations infrastructure components but need to be shared with the revenue management components for billing and revenue management. The repositories shown in the diagram are shared across the management components but their primary affinity is shown with the dotted lines.

3.3 Cloud Business Management

The primary goal of the Cloud providers is to build and offer services to the consumers at a lower cost and faster time-to-market than traditional IT. Providing the service infrastructure is only a part of the solution. Customer acquisition, customer management, account management, billing, and revenue management are critical functions that providers will need to support their business.

It may first seem like the business management function is only applicable to public Clouds. But even private enterprise Clouds need to be run as a business in order to achieve the benefits of Cloud. Some of the capabilities in this section may be used less so than others by private Clouds but the idea of monitoring usage, tracking users, charging back the appropriate business unit, and settling the revenue (or cost allocation) are examples of activities very much applicable to private Clouds as well.

Another interesting aspect of the Cloud business management is the possibility of using external Cloud services (SaaS) for common, non-core functionalities. Most customer management functions can be implemented and used as on-demand, SaaS services. This approach reduces the footprint of the custom management infrastructure and increases reuse by availing the services for both internal management as well as external pay-as-you-go offering.

3.3.1 Customer Management

Customer management capabilities in a Cloud environment are fairly standard but to support the scale and velocity requirements of the Cloud, self registration and self service should be supported.

Capability	Description
Opportunity Management	Manage the opportunities and prospects. Trial subscriptions and self registration may be used for capturing the information on prospects for creating leads and follow up.

Customer Provisioning	Provision customers and the ability to assign different levels of access. For example, a customer user may be designated as a delegated administrator with additional security access to manage the customer users. The key is to provide a self service interface for managing the customer provisioning and providing the ability to the customers to choose the services that fit their needs.
Account Management	Managing the account status, account information, account planning, and forecasts. Account profile is used to manage multiple services for the customer under a unique account.
Relationship Management	Managing the ongoing relationship with the individual and institutional customers. The ability to upsell and cross sell products and services.
Customer Management Portal	Ability to self register and self service. Customers will be able to sign in and order services using this self service portal.
Customer Database	Stores the customer and account information

3.3.2 Contracts Management

Contracts protect both Cloud Providers as well as Cloud Consumers, in terms of protecting both parties against risk and legal issues. Defining contracts is a good idea not only for public Clouds but also for private clouds, since the organizations providing the Cloud services are different from those consuming them. Private Cloud may choose to use informal contracts to define the service level agreements and usage criteria.

It is important to agree upon how to handle the data stewarded by the provider. More specifically the parties need to agree upon what happens to the data, how it would be handed back to the consumer (the process), what technology will be used for the data transfer (e.g. CD, flat file, CSV etc.), the time frame for the transfer, process for settling the outstanding balance (which may be a requirement for the cloud provider to hand over the data), and arbitration process to be followed if either party fails. The technology components (database, hardware, middleware, apps etc.) are fairly easy to mimic in a new environment, but the data is impossible to reproduce and is vital to the business. In the interest of the end-users (customers), both parties should ensure that the data is protected and is returned to the rightful owner promptly.

Contract management should consider several areas including the following:

- Data protection and data security
- Portability and interoperability requirements
- Legal process and protocols
- SLAs and metrics
- Pricing and Discounts

The table below summarizes the key capabilities of contracts management.

Capability	Description

Definition	Ability to define contracts and reuse master templates as applicable.
Discovery	Discover contracts and service agreements for reference and to ensure reuse when possible.
Pricing	Pricing for the services or at the least the pricing structure should be agreed upon. Pricing models such as monthly subscription-based or pay per use should be determined up front.
Discounts	Support for target promotions and discounts to customer segments or specific customers. Capabilities such as trial subscriptions allow providers to get prospects and later convert them to paid consumers.
SLA	Define SLAs as part of contracts and services
Metrics	Define metrics to be captured for tracking usage and assessing charge backs. In many cases, a combination of metrics will be used to assess chargeback. To avoid any surprises, the metrics to be used and the method of calculation should be agreed up front.
Contracts Portal	User interface for contract creation and management.
Contracts Repository	Enterprise repository to store and maintain the contract assets. The contract assets must be linked to the related runtime assets and metrics to provide closed-loop governance.

3.3.3 Revenue Management

Most Cloud services are offered in a basic subscription or one-time charge basis, limiting the creative possibilities of pricing and billing and future service profitability. Many Cloud providers rely upon billing and revenue collection mechanisms that cannot create dynamic pricing, nor provide the ability to leverage complex business models to react to market conditions or to fend off competitive offers. Maximizing real-time profitability of services is highly dependent on setting the right price points for the right customers at the right time.

Today's Cloud providers must address the technology and business challenges that currently limit the potential revenue upside of a successful cloud-based service. They need a billing and revenue management solution that offers:

- Flexibility - infrastructure that supports usage-based and subscription pricing models, including non-traditional price algorithms
- Scale - the ability to rate and price tens of millions of transactions per day
- Account Management - support for real-time balance management; discounting; and enterprise, consumer and hybrid accounts
- Integration - pre-built integrations to enterprise applications based upon repeatable business processes and an open service-oriended architecture
- Automation - Automated revenue management through standards based interfaces (e.g. EDI)

Monetizing all layers of the cloud requires an enterprise IT infrastructure that offers enterprise-grade performance, the ability to launch, deliver, revise and retire services quickly, and ultimately one that ensures an exceptional customer experience.

Capability	Description
Rating	Rating is the process of applying the pricing rules to the usage information to calculate the charges. Advanced capabilities such as batch rating (high performance processing of high volume usage events), real-time rating (rating real-time usage transactions using event processing), and re-rating (balance impacting the rate and requiring a recalculation) are required for more complex pricing rules.
Billing	Billing is the process of calculating bills and partner statements, applying additional discounting and taxes, and finally invoicing the customers. Should be able apply federal and local tax rules specific to the geographic region to ensure compliance to local laws. Support complex non-cyclic billing scenarios such as bill now, on-demand billing, and in-advance billing.
Payment	Ability to accept and process the payments using a payment engine of some kind.
Reporting (and BI)	Tailored and out of the box revenue analysis and reporting to capture business and revenue visibility, analyze process efficiencies, and revenue flows.
Chargeback	The ability to chargeback to internal organization to obtain cost relief. A related term is "show-back", which refers to the ability to provide cost information and accounting usage patterns. This helps the cost conscious consumers to be consumption aware.
Collection	Support collection activities by generating audit reports and payment violation reports.
IVR Interface	Interactive Voice Response interface for billing support and payments through telephone.
B2B Interface	Support for standards based interface for billing and payment.
Finance Database	Repository for billing and revenue information. Some may choose to use an existing revenue management instance for managing the customer revenue accounts.

3.4 Cloud Operations

Cloud Operations groups the operational capabilities of Cloud Management. The capabilities are further categorized into the following groups:

- Service Management
- Lifecycle Management
- Performance Management
- Configuration Management

- Monitoring
- User Interaction
- Management Repository

3.4.1 Service Management

Service management refers to the capabilities required to define and manage the Cloud services.:

Capability	Description
Service Definition	The ability to define the service and service characteristics. In the context of Cloud, various types of services may be offered with different sizing options and SLAs.
Service Discovery	Cloud services must be discoverable and the Cloud infrastructure should allow the consumer to discover the services they are interested in. Assets and metadata associated with the services need to be managed to provide additional information as required.
Forecasting	The readiness of the Cloud provider depends on their ability to forecast the demand and size the resources. Demand should be forecast based on historical demand and market changes anticipated.Resource requirements should be forecast based on the peak demand expectations and utilization projections.
Metering	Metering data from various resources, portals and runtimes are aggregated within this area so that it can be rated and ultimately billed. This would include metering data from internal resources as well as billing data from external resources (i.e. public Clouds).
Service Level Authoring	Service Level authoring tools allow the creation of executable SLAs for defining and enforcing policies around SLA thresholds and violations.

Service management may also contain a number of other capabilities specific to vendors' implementation.

3.4.2 Lifecycle Management

IT operations have long acknowledged the difficulty in deploying and maintaining new software, in provisioning and maintaining new servers with a variety of configurations, and the difficulty in adapting to changes in workload of the environment in a timely and consistent manner. This is especially true in Cloud computing environments. Cloud brings in several benefits to the enterprise but unless managed effectively, those benefits won't be realized. The infrastructure components must be constantly monitored and automatically provisioned based on the current demand conditions.

Capability	Description

Provision Management	Cloud requires resources to be provisioned dynamically in an on-demand and elastic fashion. Provisioning may require orchestration multiple activities and resources to ensure that the resources required for the service are allocated and organized for the consumer. Provisioning process should take the location preferences into account as many consumers have either a preference or a legal requirement to be physically located in a certain geography.
Capacity Management	The Cloud infrastructure should be able to grow or shrink the resources available dynamically based on proactive or reactive strategies to support the demand variations. Generally the resources span across multiple tiers and need to be orchestrated to ensure that any possible bottlenecks are eliminated.
Abstraction/Virtualization	Provide a layer of resource abstraction to create deployable entities that can rapidly be deployed and undeployed. Virtualization is an example of abstraction that is used at multiple levels to provide layers of abstraction.
Patch Management	Managing patches could get very complex in a Cloud environment. In an ideal Cloud environment, patches are never individually applied but rather, updated images/assemblies are deployed as a whole. The infrastructure should be capable of automatically identifying patching needs and provide the workflow to manage the patches efficiently.
Resource Management	Resources and resource pools must be managed independently to be able to support the scalability and elasticity requirements. Pool management aggregates, coordinates, and exposes virtual or physical resources to the Cloud management layer for deployment to the service layers for consumption.

3.4.3 Performance Management

Because of the size, complexity, and business criticality of Cloud operations, the challenge for providers is to be able to maintain the levels of availability and performance required for both services and infrastructure components in order to ensure that business operations are not impacted. This requires a business context based performance, availability, and usage monitoring approach, whereby a proactive approach to correcting problems is achieved.

Performance Management provides a comprehensive, flexible, easy-to-use business context based monitoring and drill down analysis functionality, which supports the timely detection and notification of impending IT problems across the Cloud environment. To obtain a comprehensive picture, IT organizations must monitor end-user experience, understand service/infrastructure component dependencies, monitor infrastructure component health, and trace business transactions all in conjunction.

Capability	**Description**

Service Level Management	Services levels need to be monitored and checked against the policies defined. Maintaining the promised service levels is an important aspect of establishing trust between the provider and consumers. Policy enforcement and escalations must be applied as appropriate.
User Experience Management	Cloud providers need to maximize the value of consumers' business-critical applications by delivering insight into real end-user experiences. Cloud management should provide insight into business trends and user preferences. This requires the ability to monitor the business activities from the user perspective and correlate them to the the underlying system activities.
Business Transaction Management	Business transactions are often executed by arranging, or "orchestrating," existing applications and infrastructure in the cloud to implement business processes. They incorporate a wide variety of technologies, deployed across many platforms and organizational boundaries such as ESBs, Process Engines, middleware, legacy and packaged applications. As applications grow in capability and complexity, real-time visibility into business transactions supported by these applications become imperative. Cloud management requires capabilities such as transaction visibility, business KPI monitoring, and proactive exception management.
Business Service Management	Business Service Management deals with the management of business services underlying the business transactions. Business Service Management requires capabilities such as: service topology analysis, service implementation drill downs, and service dependency analysis. Business services must be monitored and managed holistically to be able to serve the customers better.
Diagnostics	Complex environments require performance and utilization diagnostic information at multiple levels to troubleshoot and resolve issues. Cloud management infrastructure must provide diagnostics capabilities for the full stack from application layer to hardware layer, and the ability to trace the dependencies across.

Problem/Incident Management	Incident Management aims to restore normal Cloud operation as quickly as possible and minimize the adverse effect on business operations, thus ensuring that the best possible levels of service-quality and -availability are maintained. Problem Management aims to resolve the root causes of incidents and thus to minimize the adverse impact of incidents and problems on business that are caused by errors within the Cloud infrastructure, and to prevent recurrence of incidents related to these errors. Centralized Cloud incident/problem management should be able to detect issues as they occur and take remedial actions such as self-healing or notifications to manage the incident. Capabilities such as issue reporting and patch advisories help find a long term problem resolution.

3.4.4 Configuration Management

One of the problems of Cloud operations is the difficulty in managing consistency and compatibility across the entire stack. This can lead to infrastructure component configuration drifts and security vulnerabilities that lead to lack of compliance.

Using configuration management, Cloud providers can rely upon automation to ensure that all infrastructure components are deployed following specified practices and rules. This way, only pre-tested, pre-certified configurations enter the Cloud environment.

Capability	Description
Asset Discovery	Many companies choose to build private Clouds with existing infrastructure. When building Cloud infrastructure with existing resources, it is useful to have the capability to automatically discover the assets in the network and baseline the resources to be able to plan the migration to the Cloud.
Asset Management	Assets, asset configuration, and asset metadata need to be managed in a controlled fashion. Asset management may require the use of a variety of techniques such as version control systems, CMDB, and enterprise metadata repositories.
Knowledge Management	Manage the knowledge base around the configuration, history, and issues. Ability to interface with vendor knowledge bases to facilitate quicker analysis and problem resolution.
Release Management	Release Management is used for platform-independent and automated distribution of software and hardware, including license controls across the entire Cloud infrastructure.

Configuration Compliance	Ability to track the configuration information and perform configuration comparisons and drift analysis to identify any violations to the baseline configuration. Cloud and compliance are often treated as opposing forces. While Cloud encourages dynamism, compliance enforces caution and control. This capability allows the use and enforcement of OOTB and user-defined configuration policies that are evaluated automatically on an on-going basis.
Change Control	Administrators can track changes to individual configuration properties in almost real-time, get alerts for out-of-policy changes, detect authorized versus unauthorized changes, and prevent IT audit problems. Change detection capabilities, as part of a change request management process, can be used to automate provisioning changes as well as setting up new environments throughout the application lifecycle, from development, testing and through production.

3.4.5 Monitoring

Monitoring has been covered to a certain extent under various specific topics such as metering and performance management. Conventional monitoring focuses on individual resources, but Cloud requires the ability to set a performance metrics on particular services, and then provide correlation down to the infrastructure components supporting that service. This correlation provides Cloud providers the ability to both diagnose and optimize the performance and availability of their services. In addition, Cloud monitoring brings context based end user and business transaction visibility by discovering how long an entire business transaction takes.

Cloud resources and services must be monitored across the stack and a unified view of everything must be presented to achieve "manage many-as-one". Some of the sample monitoring capabilities are listed below:

- View summary status of services and resources (targets)
- Monitor outstanding alerts and policy violations for a group or "Cloud zone" collectively
- Monitor the overall performance of a group or "Cloud zone" through performance charts
- Support monitoring templates, which help define the monitoring settings once and apply them to the monitored targets. They are also used for propagating user-defined metrics across a large number of targets

3.4.6 User Interaction

The functionality that interacts with the user will always vary from one Cloud implementation to another, so it is important that any user interaction framework have a fully customizable interface that can also support multiple devices such as browser, mobile, and portal. Please refer to the "*ORA Management and Monitoring* document" for more details on the Management and Monitoring user interaction components.

Below are a number of key architecture capabilities that are commonly provided:

3.4.6.1 Self Service

Cloud automation requires self service interfaces that are both GUI-based and API-based. Most provisioning and management tasks should be made accessible through self service interfaces to enable the scale and velocity requirements demanded by the Cloud.

3.4.6.2 Administration

Administration enables the ability though a single console to manage and monitor the entire Cloud environment, including all infrastructure components such as applications, services, virtual machines, operating systems, and hardware components. As well as managing all infrastructure components, it enables administration tasks to be applied to logically related infrastructure components.

The console has the built-in intelligence to understand the characteristics of each infrastructure component and allow the appropriate administrative tasks. This approach allows it to support new infrastructure component types in the future.

3.4.6.3 Dashboard

Dashboards provide an "at-a-glance" monitoring of all critical indicators for services and other infrastructure components. They offer access to a series of rich real-time customizable and consolidated views of the IT eco-system with the ability to drill down. Administrators are able to spot recent changes or issues by presenting actionable information using intuitive icons and graphics, which assist in identifying trends, patterns, and anomalies.

3.4.6.4 Troubleshooting & Diagnostic Analysis

As part of an overall approach to quality management, Troubleshooting and Diagnostic Analysis enables the ability to analyze collected metrics for the purpose of investigating and resolving application and service issues. Examples include:

- The diagnoses of the root cause of a performance problem, such as Services crashing and hanging in the production environment.
- The rapid detection of memory leaks using real-time heap and garbage collection metrics.
- The analysis and comparison of one or more memory heap dumps over a customized period of time to find the object that is causing a memory leak.
- Drill down to view the performance of a specific method call and even track the details of JDBC/SQL calls obtain via instrumentation.
- Diagnostics presented via an architecture view showing the call path.

3.4.6.5 Query

Query enables the searching of the management and monitoring repository using pre-defined or ad-hoc queries. For example, an administrator can use this capability to find all resources with a given configuration. Commonly used user-defined queries could be stored within the monitoring repository for future use.

3.4.6.6 Reporting

Reporting and publishing capabilities allow the definition of custom reports, that can be produced as needed or on a defined schedule. The reports present an intuitive interface to critical decision-making information stored in the Management Repository, which should be able to be distributed via several means, email, portal access, etc. For example, a report could be defined that reports on actual service levels achieved, helping IT and business to find out whether their Services indeed function as expected to support business activities.

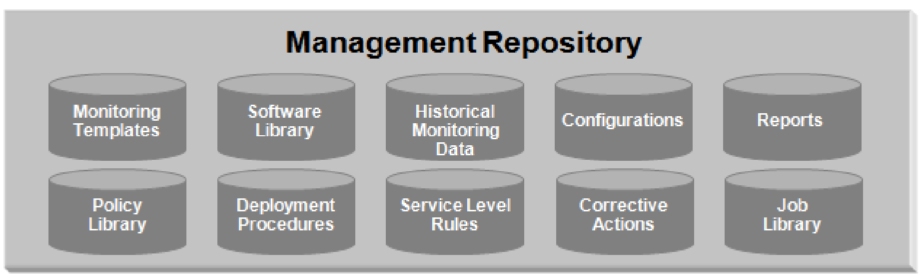
3.4.6.7 Topology Viewer

A topology viewer provides the ability to depict a graphical representation of the infrastructure, infrastructure components, services, and their dependencies. The viewer displays all the determinants for the service's availability in a graphical form and allows the understanding of how requests are routed through different layers of the infrastructure. In addition, the topology viewer can allow users to drill down to detail pages to get more information on the key infrastructure components, alerts and policy violations, possible root causes and services impacted.

3.4.7 Management Repository

The data required to manage and monitor in the Cloud infrastructure can be quite extensive, complex, and distributed in nature. Figure 3–2 shows the key information stores that are commonly required. Note that one should not infer that all data be centrally located. Please refer to the "*ORA Management and Monitoring* document" for the description on each of the stores shown in the figure.

Figure 3–2 Management Repository



3.5 Security and Policy Management

Security is one of the key elements of Cloud architecture. Cloud consumers are always concerned about security of their business information in a multi-tenant environment. *ORA Security* document provides the details of general security concepts in great depth.

In a Cloud, due to the self-service, automatic provision and dynamic elasticity, there is an expectation that IT resources get provisioned more quickly. Hence the normal defense mechanisms such as hardening, policy checking, governance etc. may be skipped which result in risk.

Hence one of the important aspects for security in Cloud is automation to ensure these essential steps are not missed. However, automation is just one part of the solution. There must be governance in terms of who can provision environments, what data can be loaded into an environment (think data classifications), what security controls need to be put in place (proper risk assessment), standardization of security services, etc.

Recklessness is a major concern with Cloud, as with all "new strategies". With most new technologies, security lags behind early adoption and most people just avoid it. With Cloud you can be reckless just by allowing more people to do things faster and with fewer hurdles. For example, it can be easy to stand up a development or test environment and load up a copy of real production data in a public Cloud. It is great for time to market, but bad for security, and nothing to stop people from doing it except common sense.

There are several other factors that make Cloud security unique and challenging. The following list summarizes these factors:

- Multi-tenancy – particularly with public Clouds, the issue of security comes from multi-tenancy and an issue or breach with one customer can easily impact or infect other customers.
- Elasticity – as instances are added, security must be applied to the new instances and the security infrastructure should also scale accordingly to support the new demand.
- Sensitive data is maintained by third party systems and people.
- Cloud bursting may raise questions on where the data is going and how it can be kept synchronized.
- Possible loss of audit trail
- The need for synchronization across enterprise boundaries.
- The need for self-service, from security administration standpoint.
- Some businesses are willing to accept certain risks to save money by using public Clouds. This may put a twist on how strong the security is and how well it is handled in the Cloud.

Security and policy management provides services common to both the build-time and runtime functions of the Cloud management layer. Commonly, these will include:

- Security Management
- Security Services (Runtime)
- Policy Management

3.5.1 Security Management

Security Management deals with the management of security information. Although the concepts described in this section are very generic, Cloud brings in interesting issues related to security. The key question is how to implement these security management functions in a Cloud environment to support the unique needs of the consumers, given the constraints of the Cloud. Also, given the scale and elasticity requirements, it is important to automate and self-service-enable as much of these function as possible.

Capability	Description
Identity Management (Users and Roles)	Identity Management (IdM) includes the capability to store, manage, synchronize, provision, administer, and audit security data related to user identity information in the Cloud.

Entitlements	Entitlements management pertains to the way low level security rules are handled. Entitlements may be hard coded into applications or may be centrally managed through an entitlements engine. Since important goals of Cloud include agility, consistency, and multi-tenancy, an entitlements engine is a much more favorable approach. It provides a way to manage this aspect of security policy from a central point in a consistent way.
Delegated Administration	The ability to partition administrative capabilities such that consumers can perform administrative function based on their role or group. Although security information and services are centrally managed, administration can be partitioned and delegated to provide a certain amount of control to groups and/or individuals within the Cloud consumers.
Certificates	A Cloud provider may decide to centrally manage a set of common encryption keys and certificates. It may also want to issue keys and act as its own internal certificate authority (CA).
Risk Analysis	This function involves the processing of fraud detection logs and/or audit logs in order to recognize potential risks to the systems. It may involve automated processing, manual analysis, or some combination of both. The output of risk analysis is an indication of the likelihood of malicious activity, which can influence ongoing authentication and access policy management.

3.5.2 Security Services

This category addresses the runtime services providing security functions. Implementing and utilizing these security functions in a Cloud environment might be quite challenging, nevertheless it plays a key role in establishing the trust between the Cloud provider and consumers. Application of a security framework that provides these services is a first step in offering them to the consumers through self-service portals and APIs.

Capability	Description

Authentication	Authentication is the process of verifying that a user or resource consumer is who he/she claims to be. This is generally accomplished by providing an identifier along with a password, token, or signature that is unique to the user and trusted to be secret.
Authorization	Authorization is the process of granting or denying requests by a party (e.g. client or user) to perform actions on a resource. It is generally performed by comparing rights Authorization (Access Control) granted to the user with actions the user is attempting to perform. In a Cloud scenario, it is important to limit the user to that particular consumer's resources and services.
Auditing	Auditing provides a record of activities or transactions that have occurred in the system. It may be used for many purposes such as retracing steps of a transaction, quantifying activities in the system, identifying erroneous behavior, or attempting to determine if a specific action did or did not occur. It is an essential tool in building trust between the Cloud provider and consumer.
Federation	Federation, in this context, refers to interoperation between entities in different security domains that may potentially be in different Clouds. Security infrastructure must provide identity mapping, credential mapping, SSO, and identity propagation necessary to support federation across security domains.
Encryption	Encryption is often handled by the platform transparently to the user. Each platform will often manage its own key and credential store, although some might support the use of a common store. In addition, a common encryption and decryption service could be useful in order to provide end-to-end encryption.
Fraud Detection	This service operates behind the scenes monitoring other services with the intent of detecting possible fraudulent activity. It is mostly associated with the authentication service since that is where hacking will first be evident. It can monitor events such as failed logins and various inconsistent behaviors in the Cloud as it occurs. It can send alerts and suspend accounts and sessions when suspicious activity is detected.
Role Mapping	Role Mapping is the ability to determine which roles a user is associated with. It is a necessary step for role-based access control. The Cloud consumer roles may need to be mapped to the Cloud provider application roles, if they are different.

Security Store	Security store is the logical repository that maintains the security information including identities, entitlements, role associations, etc.

3.5.3 Policy Management

Policy management pertains to the management of rules and the software and infrastructure that supports those rules. Security policy governs the way various aspects of security need to be handled.

Capability	Description
Authoring	Deals with the definition and creation of the policies. Authoring tools should allow policy rules to be created in a user friendly, non-intuitive fashion.
Attachment	Policy Attachment defines mechanisms for associating policies with the subjects to which they apply. They may be defined as part of existing metadata about the subject or defined independently and associated externally.
Deployment	The ability to deploy policies to specific services/components.
Assessment	The ability to monitor/assess the policies for compliance.
Enforcement	Policies are distributed to the appropriate policy enforcement points using common approaches such as gateways and agents and are enforced at these points.
Compliance Reporting	If infrastructure components fall out of compliance, remedial action can bring them back into compliance. Detailed compliance reporting highlights the infrastructure components that are in and out of compliance and details any deviations. This enables administrators to take action quickly and address the high impact items to improve the compliance score.
Policy Store	The logical repository that stores the policy information.

3.6 Orchestration

Orchestration is one of the key capabilities of a Cloud infrastructure. Many Cloud management requests require a group of resources to be coordinated. For example, a scale up provisioning request may require several resources such as hardware, VM, middleware, database, firewalls, proxy servers, and load balancers to be allocated, provisioned, and configured to fulfill the service level requirements. In addition, internal management, security, and revenue subsystems may need to be configured as well. Orachestration capability allows the automation of Cloud resource provisioning workflows by coordinating between various Cloud resources. This directly translates into scale and elasticity since orchestration supports resource provisioning through self-service or automated channels.

Capability	Description
Workflow	The ability to orchestrate processes, workflows, and tasks to achieve process automation and component coordination.
Mediation	Mediation can be broadly defined as the ability to resolve the differences between two or more systems in order to integrate them seamlessly. Mediation can happen at different levels such as security, transport, message, API, and protocol. Orchestration layer is required to interface with several disparate systems in order to coordinate activities. Mediation is a handy technique to create a loosely coupled architecture for interfacing with these heterogeneous backend resources. Mediation capabilities such as "Routing" are essential to route the incoming traffic to appropriate service instances.
Event Processing	Although not a requirement, event processing adds sophistication to the orchestration layer by adding asynchronous resource coordination and automated policy based orchestration. For example, event processing may be used to detect patterns in the resource utilization, and execute the orchestration process to scale up or scale down the capacity based on the policies defined.

3.7 Design-time

This category includes the capabilities related to design & build time. Cloud introduces the concept of DevOps that blurs the line between Development and Operations. This makes it all the more important to have the right infrastructure and tools to support the transition into the new paradigm.

Model management provides capabilities for developers and deployers to create virtual data centers (vDCs). It supports this build-time functionality by exposing Cloud resources, validating vDCs and then these vDCs for later deployment. Model management is the main point of interaction for developers with the Cloud infrastructure.

Capability	Description

Introspection	The ability to introspect a reference configuration to capture metadata and binaries to automatically generate the deployable entities. Support for introspecting components of various types (possibly plug-in based) on local and remote hosts should be provided.
Model Design	Ability to quickly configure and provision entire multi-tier application topologies (including templates, cloud models, and deployment models).
Validation	Validate the models and targets for correctness before provisioning and reconcile the them across the various nodes in the Cloud (multiple locations, data centers, private/public Clouds etc.).
Assemble/Package	The ability to combine components and their relationships to internal and external resources and create deployable packages with support for deploy-time customization.
Federation	Ability to manage and provision the models to multiple nodes across Clouds, data centers, servers etc.
Deploy	The ability to customize and deploy the deployable entities to the resource pools. Also be able to scale instances after initial deployment of the deployable entities and automatically wire the newly deployed instances into the existing deployable entity.
Solution Catalog	Solution Catalog is the repository for storing the solutions, services, and models for developer discovery, use, re-use, manipulation and eventual deployment. The solution catalogue is the repository for all deployable entities (a.k.a. vDCs).

4 Cloud Implementation Perspectives

Cloud infrastructure is very complex and has several moving parts. In most cases it is not possible to define the implementation of a Cloud infrastructure using a simple blueprint. It requires multiple perspectives based on business scenarios, deployment patterns, and use cases. This section classifies the implementation perspectives into these three categories.

4.1 Business Scenarios

This section describes some basic business scenarios around Cloud computing.

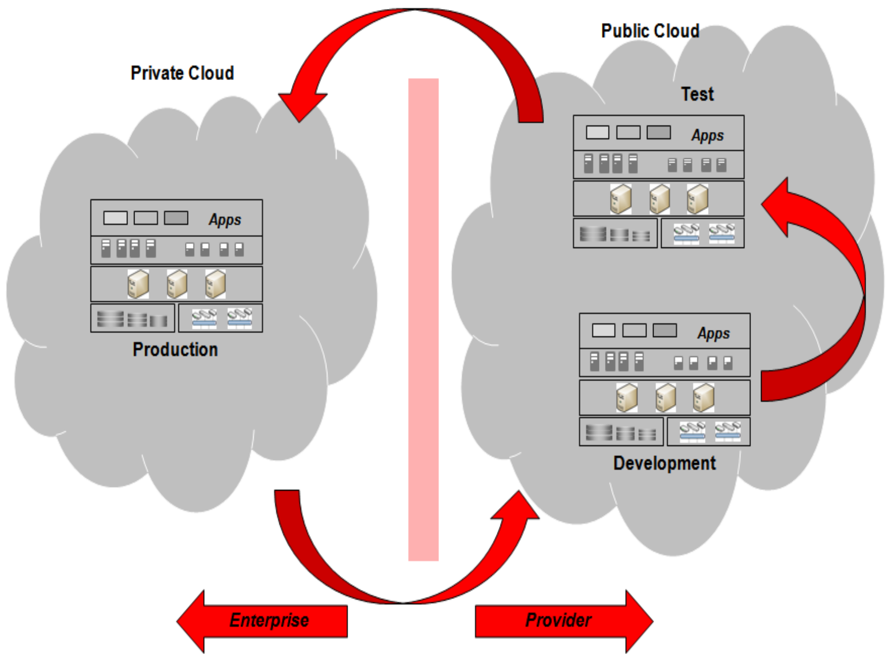
4.1.1 Hybrid Cloud Scenarios

There are three primary scenarios for the Hybrid Cloud. They are explained in this section using some examples. The examples assume a private Cloud and a public Cloud but the scenarios are equally applicable to any Cloud to any Cloud.

4.1.1.1 Lifecycle Distribution

Perhaps the most often cited one of these is on-demand development and test environments as shown in Figure 4–1. This example shows a set of applications deployed on a PaaS with the development and test environments running in the public Cloud and production running in the internal private Cloud.

Figure 4–1 Hybrid Cloud - Lifecycle Distribution

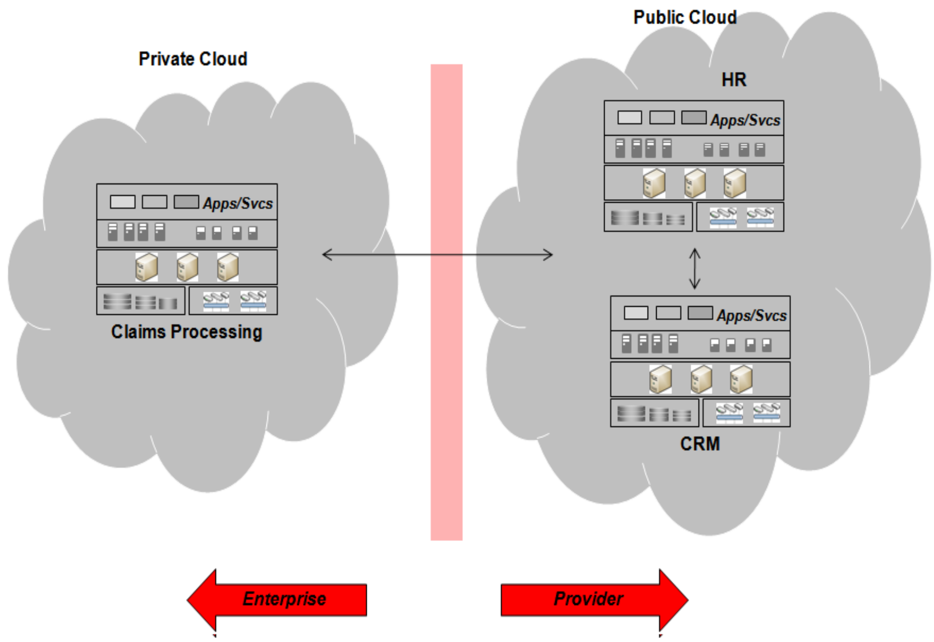


The simple act of moving from more dedicated and constantly re-wired development environments to one where pooled resources are dynamically managed, provides several levels of TTM (time to market) impact. This certainly reduces the time to use from the potential months to order and deploy new systems from weeks or months to days or even hours.

4.1.1.2 Functional Distribution

An enterprise may choose to split functionally distinct business applications and deploy them across the private and public clouds. These applications generally don't have much in common and the frequency of interaction is minimal. An example is shown in Figure 4–2, where Claims Processing, HR, and CRM functions are deployed separately, and as shown by the thin arrows between them, the interaction between them is very minimum.

Figure 4–2 Hybrid Cloud - Functional Distribution



When building your enterprise Cloud roadmap, you should identify such applications and plan to migrate them to the public Cloud at the early phases of the implementation.

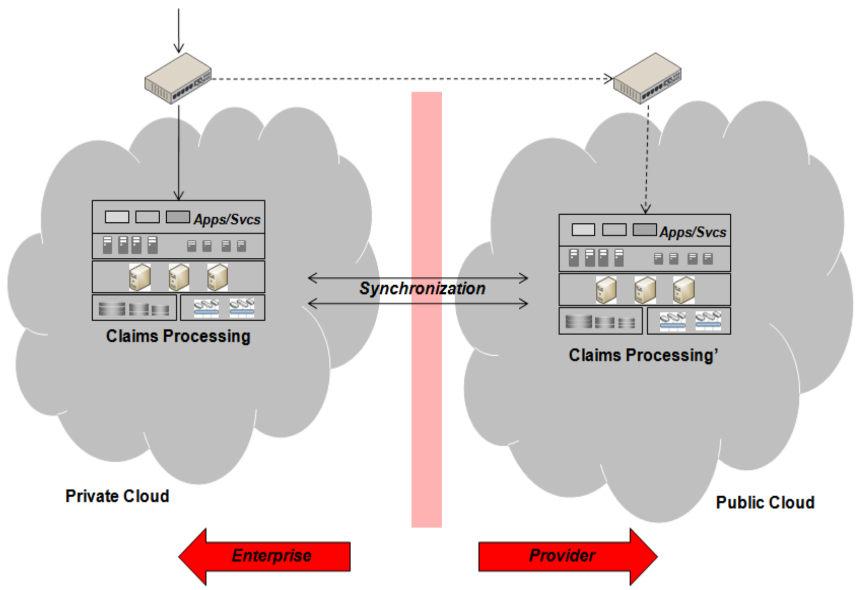
4.1.1.3 Load Distribution (Cloud Bursting)

Cloud bursting is a very popular use case for the Hybrid Cloud and technologically quite challenging to implement. Cloud bursting happens when the private Cloud runs out of capacity and the load is directed to the public Cloud for further processing.

Cloud bursting is a great option for handling unusual and spiky workloads. Enterprises avoid buying and maintaining the extra capacity required for these occasional spikes. The challenge, though are the synchronization and potential differences between the private/public environments. It also requires sophisticated hardware and software to identify the moment of bursting and to route the workload to the public Cloud.

Figure 4–3 shows an example of Cloud Bursting scenario using the Claims Processing application deployed in both the private Cloud as well as the public Cloud. The workload is routed to the public Cloud during peak load times to handle the additional load. This also means that the services needed to be kept synchronized to maintain the consistency and integrity of the business.

Figure 4–3 Hybrid Cloud - Load Distribution



An enterprise may potentially be able to build out for dominant average loads and utilize a Hybrid Cloud model for peak load. This would be especially attractive where the peak loads are unpredictable and potentially large. As in the case of Disaster Recovery explained later in this section, this likely requires pre-positioning of applications and data and resulting synchronization challenges. Off course, some peak loads (e.g. annual financial close) may be more predictable. In addition, Cloud resources, whether public or private could make an ideal platform for high load associated with temporary projects.

The decision to burst may use a proactive approach or a reactive approach. The reactive approach decides to burst into the public cloud on failure to handle the workload, whereas, the proactive approach uses pre-defined policies to failover to the public cloud ahead of the max out.

4.1.2 Business Continuity

Mission critical applications require contingency plans for disaster recovery, often not just for business purposes, but also to meet regulatory requirements. These architectures require not only the additional resources but plans to synchronize data, migrate control, and eventually restore original production. Even where costly hot-standby facilities can be justified, there are still some failure events that may jeopardize both the production and disaster recovery sites. A Cloud strategy may provide a disaster recovery strategy that not only results in lower costs, but potentially even higher overall reliability levels.

Figure 4–4 Cloud for Business Continuity

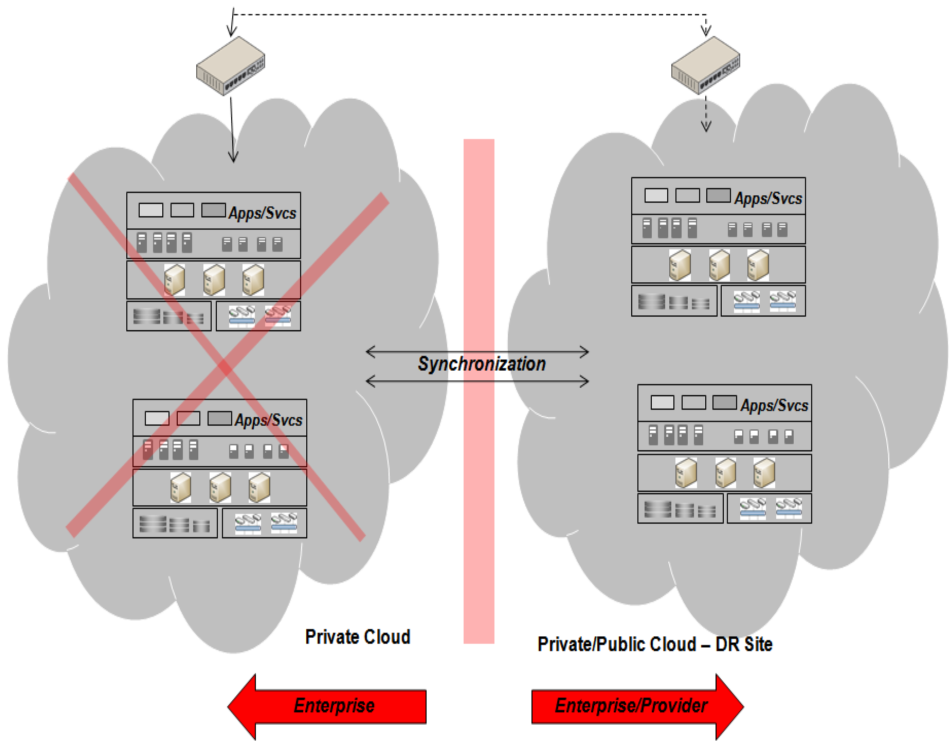


Figure 4–4 illustrates a simplified business continuity scenario where the private/public cloud continues after the primary site goes down. Cloud based disaster recovery could include more shared use of dynamically provisioned private Cloud resources, or could rely on contracts with public Cloud providers, or utilize a hybrid mix of public and private Cloud resources, with the potential of lower cost and more scalable failover. Of course using a public Cloud for disaster recovery would imply some significant baseline costs for pre-provisioning of data and application resources.

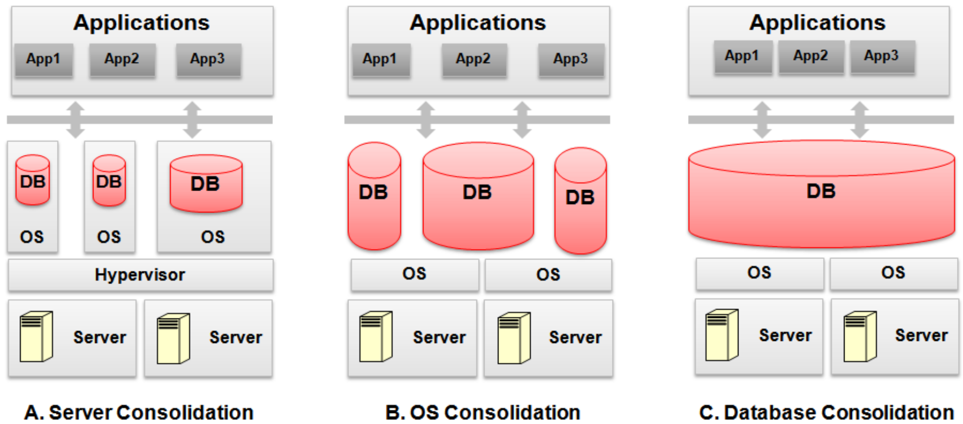
4.2 Deployment Patterns

This section describes some patterns around Platform as a Service, more specifically Database as a Service.

4.2.1 Database as a Service (DBaaS) Architectures

A popular resource that is often consumed over the Cloud is the database. Providing on-demand access to the database over the Cloud is a classic case of Platform as a Service (PaaS). This can be done in several ways. Figure 4–5 illustrates three options, A, B, and C that use different levels of consolidation to achieve this.

Figure 4–5 Database as a Service Cloud Architecture Options



4.2.1.1 Server Consolidation

This deployment model as shown in Option A, is enabled through server virtualization, where VM servers are added to a server pool, and Virtual Machine (VM) guests are created from the servers in this pool. There is generally a one-to-many relationship between VM server and VM guests, with the guest density driven by the server's physical resources (CPU and memory). In this deployment model, when a database service is requested, the entire OS stack is built and provisioned as part of VM guest creation.

The databases deployed in this model will generally be single instance databases that are instantiated in a VM guest. VM guests will leverage native VM high availability capabilities. In this model, tenant isolation is at the VM guest level. Elasticity and scalability in this model is provided through scale-up, i.e. by adding more virtual resources (CPU or memory) to the VM guest. For this reason, this option is called the Infrastructure Cloud.

4.2.1.2 OS Consolidation

In this deployment model as shown in Option B, a database is provisioned in physical servers that are part of a database cluster. This configuration enables OS and database instance consolidation.

This model employs a "multi-tenancy database" approach, where many databases share resources across a single clustered environment. In this model any one server from the cluster can host one or more database instances. Clients access the database instances via database services.

Elasticity and scalability are provided by adding additional nodes to the database server pool (scaling out) or by adding more physical resources such as CPU, memory, or I/O cards to an existing node (scaling up).

4.2.1.3 Database Consolidation

In this deployment model shown in Option C, the database essentially consists of one or more application-schemas. When a tenant requests a database to be provisioned, a schema is created within the database, with its own set of tablespaces and a database service.

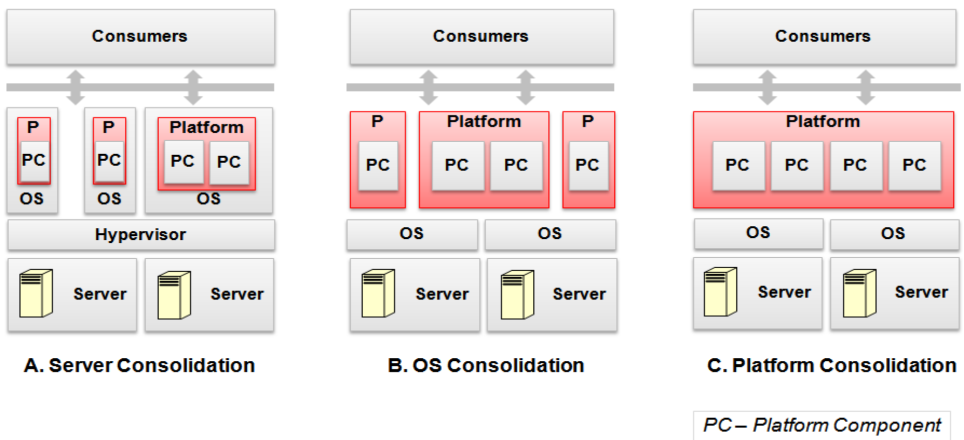
Schema name collisions can be mitigated by dynamically assigning schema names. It is also highly recommended to configure this database as a cluster, as this provides high availability and scalability.

Additionally, because all applications in this pool share the same data dictionary and database attributes, tenant data isolation and security may be a concern and should be addressed.

4.2.2 Generalized PaaS Architectures

The concept of DBaaS explained earlier in this section can be generalized to apply to any platform such as application servers or messaging infrastructure. Platforms can be built with resources providing isolation at several different levels. Figure 4–6 below illustrates the consolidation options for a generalized platform.

Figure 4–6 Generalized Platform as a Service Architectures



The platforms run platform components (PC). Option A illustrates server consolidation through server virtualization. The platforms run in their own host OS environments of the virtual machine. Isolation is provided at the virtual machine/OS level in this case.

The fact that a hypervisor layer prevents the direct access of the Platform Component to the bare metal may in some cases lead to a compromise in the quality of the provisioned PaaS service. For example in the case of the Database as PC, the hypervisor layer may affect the I/O communication of the Database instance to the storage, which may have overall impact to the DBaaS throughput or even availability (due to potential unequal distribution of I/O resources via the hypervisor). However business demands and the need for automation in many cases warrant this architecture.

Option B illustrates OS consolidation where the platform capabilities are used to span across multiple OS environments. As depicted, isolation is provided at the platform level in this case.

Option C shows platform consolidation where a common platform is used to host various consumer platform components and the platform is logically partitioned to isolate the consumer assets.

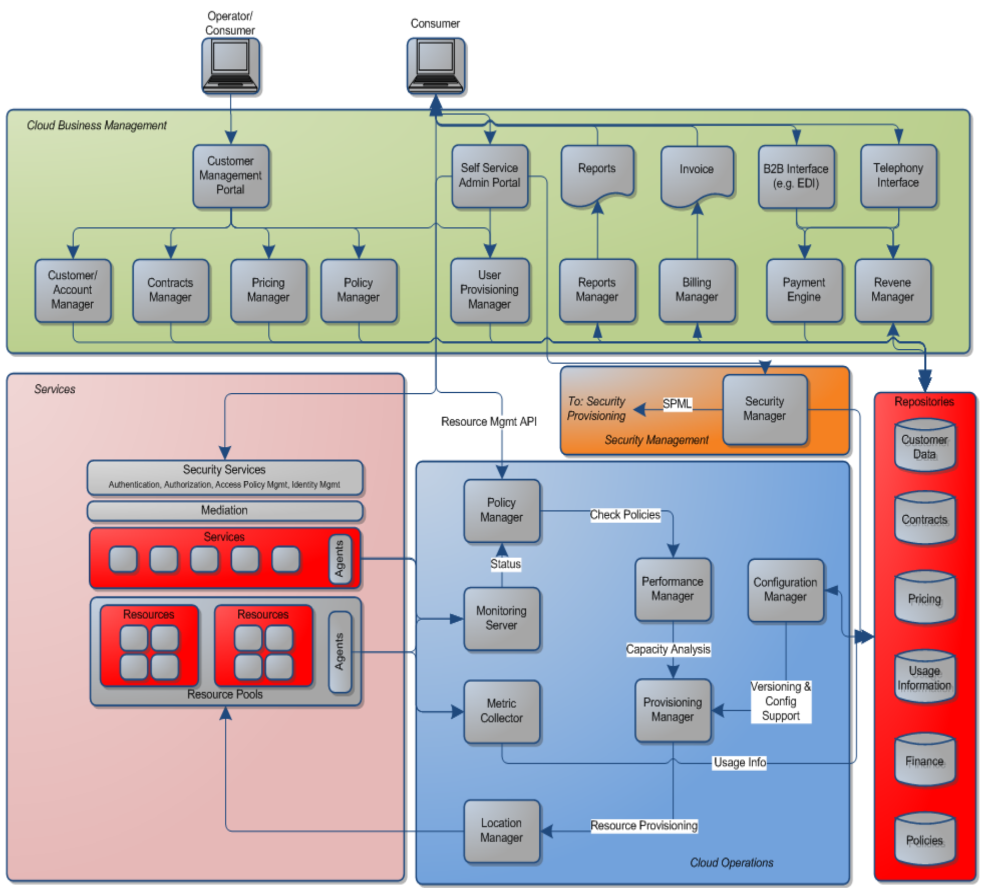
4.3 Cloud Use Cases

Cloud management is clearly a very broad category that encompasses a multitude of important use cases. This section explores the following scenarios using a logical model illustrated in Figure 4–7. This figure uses logical components to show the functions and relationships.

- Cloud Business Management
- Cloud Operations
- Security Management

These are further explained in detail in following subsections.

Figure 4–7 Cloud Management Logical Use Case View



4.3.1 Cloud Business Management

Let us look at two of the business management use cases, consumer establishment and revenue management.

4.3.1.1 Consumer Establishment

Business Management is as important to the Cloud provider as Service Management. As discussed earlier, Cloud introduces certain nuances to business management, from customer provisioning to revenue management.

Customer acquisition, relationship management, and customer provisioning are shown in the logical diagram. Customer (Management) Portal is used to add new customers to the Cloud. Consumers may be businesses, individuals, or multi-account hierarchies. Customer relationship management is a continuous process of ensuring that customers are satisfied and taking advantage of any up-sell and cross-sell opportunities. Initially customer account information is established along with common pricing, policies, and contracts.

Pricing Manager is responsible for managing the pricing of services, which could get very complex in a Cloud environment. Most consumers prefer flexible pricing, promotion, and payment options. A variety of pricing models such as one-time, real-time, recurring, and usage-based should be supported by the business management infrastructure.

The Cloud provider and consumer should agree upon suitable metrics for pricing. The metrics depend on the services used by the consumer. For example, named hosts, number of CPUs, storage, and thresholds are some of the metrics used by the providers to measure the usage. Metrics are covered in more detail in the revenue management topic. Cloud providers may also benefit from sophisticated pricing policies such as pricing validity periods, support for renewable terms, trial periods, and cross-product discounts.

Contracts Manager provides support for creating and managing contracts. Contracts should capture key concerns such as SLAs, mode of communication, standards, level of service, break-up agreements, and security agreements.

Policy Manager is responsible for the creation, association, and management of rules based policies. Policies may dictate sophisticated rules such as temporal rules, resource usage rules, and automated capacity management rules.

Once the consumer is established as an active customer, they can provision users through self-service channels or through the Cloud operator. Delegated administration is very commonly used in Cloud situations to enable the consumers to manage their user base. Both interactive and batch modes should be supported for user provisioning. User provisioning manager manages the provisioning of users.

4.3.1.2 Revenue Management

Revenue Management deals with the monetization of service usage. The consumer needs to be aware every important detail of usage and pricing.

Billing Manager often relies on Metric Collector (explained later in this chapter) to gather the usage data so that it can issue applicable billing charges to the consumer based upon the tenant-level aggregated usage of virtual network and resource metering information. It will also help to settle charges with external service providers that provide additional services as part of virtual networks. The Billing Manager manages the generation of an invoice for a given billing cycle and conveying charges and payments made/due. The Billing Manager supports the following capabilities:

- Highly configurable billing that supports flexible billing cycles, variable billing dates, on-demand, bill now, and in-advance billing
- Collections management to improve cash flow
- Complex revenue recognition rules
- Robust invoicing with support for simple or complex, single or multi-account invoicing
- Taxation, General Ledger, and reporting tools

Reports Manager provides revenue reporting and business intelligence capabilities. Tailored and out-of-the-box reports must be provided to gain business visibility. It should also provide views into current and expected revenue flows. Consumers would also like to see resource usage trends, demand history, and capacity management behavior.

The Revenue Manager provides revenue assurance solution and value chain management in addition to typical account payable/receivable management. Revenue assurance refers to the process used to verify the end-to-end completeness, accuracy, and integrity of capturing, recording, rating, and billing all revenue generating events as they flow through the system. Value chain management allows automated revenue sharing rules with support for multiple partners. It is a very important feature for Cloud providers, as many providers do not get paid until the services are consumed. Hence, they prefer to go with a revenue sharing arrangement with the partners and vendors that a one-time initial payment arrangement. This allows them to share the revenues only when the customer is billed.

B2B interface is an optional capability to support standards based communication such as EDI between the provider and institutional consumers for revenue management.

The Payment Manager ensures that the payments are received, validated, and processed in a timely manner. A variety of payment methods such as invoice, credit card, direct debit, and voucher top-up should be supported by the Payment Manager. Payment Manager also supports dunning, reversals, and write offs.

4.3.2 Cloud Operations

This scenario explains a very simplified logical view of the Cloud operations management use case. Operation management is a key part of the Cloud runtime management. Several Cloud essential characteristics such as metering, rapid elasticity, on-demand self-service, and resource pooling are fulfilled by operations management functions.

At a high level, this use case illustrates the following:

- Capturing the resource usage metrics
- Determine if there is a need to adjust the capacity automatically using the pre-defined policies
- If required, provision or decommission capacity based on the configuration and location preferences.

Metrics Collector tracks consumer charges for the usage of the resources from aggregated data including usage statistics and number of network devices deployed. It also provides aggregation of metering and multi-tenancy-aware measurement metering information on different levels of aggregates. The provider and consumer must first agree on the metrics to be used to measure the usage. The type of metrics depends on the service being consumed. The table below summarizes some sample metrics that can be used in different service models.

Service Type	Category	Metric	Unit
SaaS	Frequency	# of service invocations	Number
	Complexity	Type of service/transaction	Small/ Medium/ Large
	Usage Cost	Transaction Cost	$$

Service Type	Category	Metric	Unit
PaaS	DB Usage	DML Operations	DML Statements
		DB Connections	Average/Max DB pool Size
		Data Transferred	GB
	Deployed Entities	Deployed Apps	# of .ear
		Exposed Services	# of services
	Service Consumption	Service Invocations	# of invocations
	Identities	Population of identities	# of users and/or groups
		Application Integration	# of application roles
	Authentication	Applications in SSO domain	# of applications/ services in SSO
		Usage	# of unique users authenticated
	Usage Cost	Transaction Cost	# of transactions
IaaS	CPU	CPU utilization	%
	CPU - config	CPU Count	#
	Memory	Memory Usage	GB
	Memory - config	Memory	GB
	Storage	Disk space	GB
	Bandwidth	Bandwidth	Mbps
	Other Costs	System	Count
	Facility	Base Facility charge	$$
	Facility	Base Utility Charge	$$
	HA	HA multiplier	Times X

These metrics are collected from various targets using appropriate collection agents. The monitoring agent architecture is described in the ORA core document, *ORA Monitoring and Management*. The collected metrics are saved in the repositories for billing and/or chargeback.

Monitoring Server monitors the services and resources, in many cases through agents, for specific events and trends of usage, detecting thresholds, conflicts, and failures within the infrastructure. They correlate to respective network and resource assets, ensuring compliance, and reporting problems for overall functional integrity of the infrastructure. Monitoring Server supports various targets through appropriate agents, much the same way as the Metrics Collector.

The status from the Monitoring Server is checked against the capacity management policies and SLAs by the Policy Manager. If determined appropriate, the Performance Manager takes it further through the process.

Performance Manager ensures that the optimal numbers of resources are available to satisfy the resource requests to support the performance specified by the consumer. It measures and reports the discrete resources within a Cloud environment to ensure optimal and committed performance levels. It also has the responsibility to manage performance of the tenants in accordance with the tenant's Service-Level Agreements (SLA) and track where resources are added or removed based on operating rules.

Configuration Manager provides configuration support within the infrastructure. It ensures that the right version of the deployment entity is provided to the Provisioning Manager.

Provisioning Manager ensures that the right number of resources is allocated to address fluctuating demands as well as interacts with the virtualization and platform management component to provision and deploy services within the infrastructure to optimize infrastructure resources as well as satisfying the customer's requirements for the requested Cloud services.

Location Manager helps to geographically allocate resources based on predefined rules and regulations. The infrastructure needs to provide location-awareness, such as indicating the state and location of the resources with which services are executing. This can be used when applications are moved from one location to another for various reasons including business continuity.

Although this use case explained the automated capacity management scenario, it is also possible to manage capacity through the self-service resource management APIs as indicated by the flow arrows in the figure.

4.3.3 Security Management

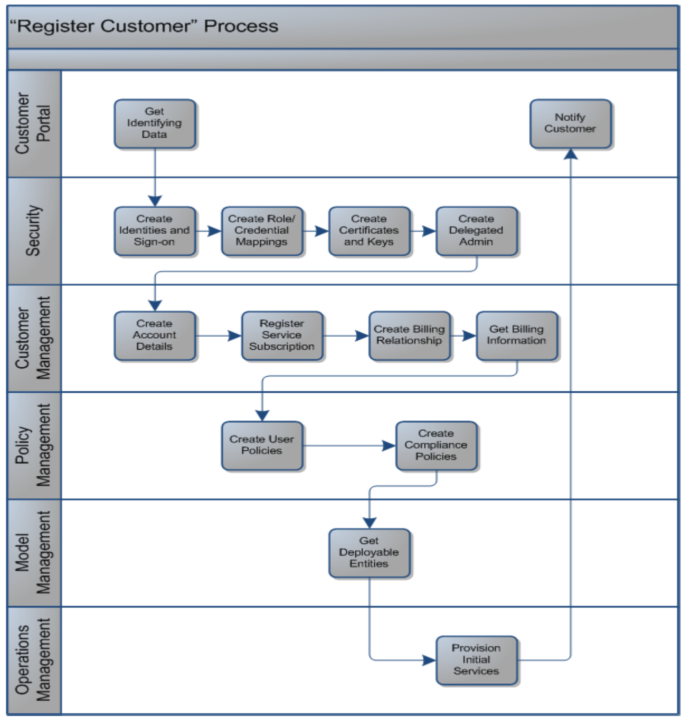
The last of the use cases is security management which is shown oversimplified in Figure 4–7. Security Manager is involved in a number of security related scenarios such as:

- Identity and Access Management - Managing the identity, roles, groups, certificates, role mappings, and credential mapping. What is important in a Cloud scenario is the ability to support multiple tenants. This may require creating multiple "security domains" that can be shared across tenants and Cloud service profiles. Another critical capability is Single Sign-On (SSO), which allows the users to access multiple services with the same credentials.
- Security provisioning - When consumers sign up for Cloud services, the security credentials must be created as part of the larger provisioning process. In a Hybrid Cloud environment, the security information needs to be propagated from one Cloud to another. The Security Manager may use standards like SPML (Service Provisioning Markup Language) to automate the provisioning of identity information.
- Provide additional value-added services such as risk analysis and fraud detection. These are optional services that the Cloud operator may provide. Considering that security is one of the key concerns for most consumers, these additional services provide a level of comfort to make the migration to the Cloud easier.

4.3.4 Register Customer Process

A new customer may register through self-service or customer management channels and the process touches several of the logical modules. Figure 4–8 shows a simplified version of a sample "Register Customer" process.

Figure 4–8 Register Customer Process



The following list summarizes the key steps in customer registration:

- Identifying data is obtained from the customer.
- After security authentication, the customer identities are created along with any additional security artifacts such as role mapping, credential mapping, and certificates, keys, and delegated administration. Customers may provision users at different levels for delegated administration. A "super user" should be able to manage the security credentials for the other users.
- Billing and account management information is created for the customer in the revenue management systems. Customer accounts are setup in such a way to be able to subscribe multiple services under the same account. Customers typically subscribe for one or more services when they sign up.
- User and compliance policies are created for the consumer based on the options customer chooses during registration.
- Initial resources are provisioned per the contract using the deployable entities (aka Virtual Data Centers - vDC). This process is explained in more details in Section 4.3.5.
- Consumer is notified of the completion of the registration process.

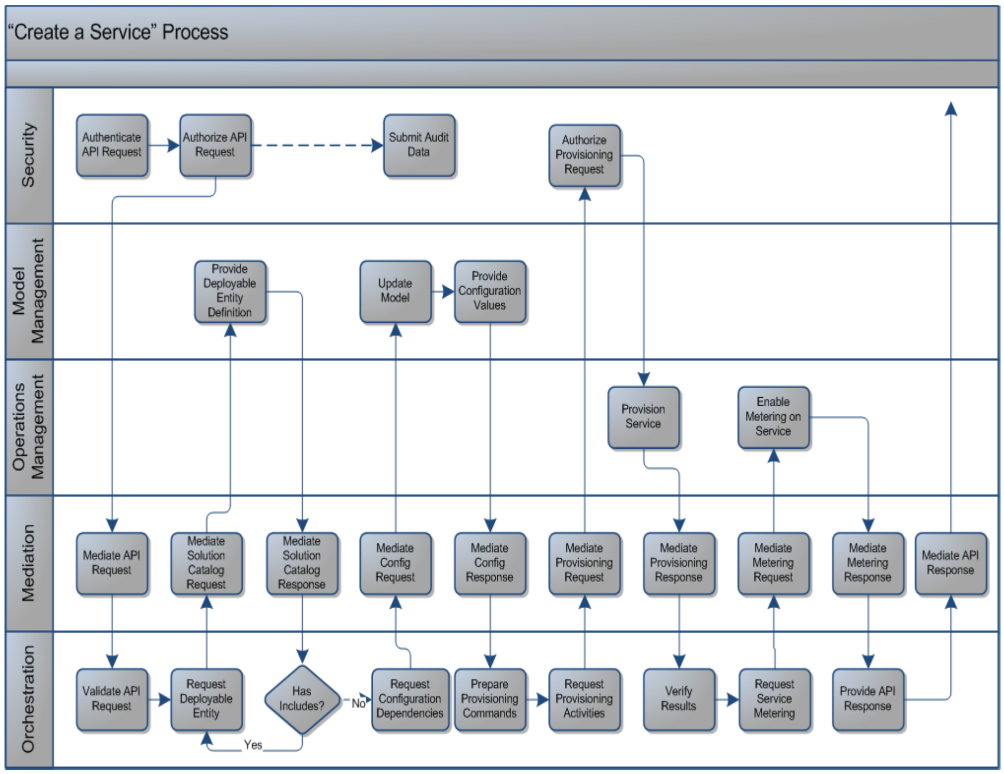
4.3.5 Create Service Process

As discussed before, Cloud requires most provisioning and deployment processes to be automatically orchestrated to support the scale and velocity requirements. Figure 4–9 shows the process to create a service (IaaS or PaaS) using the corresponding APIs. This process is applicable to several usage scenarios including:

- New customer registration : When a customer registers, he/she may subscribe for services that need to be provisioned.

- Scale up and Scale down: When a customer changes the capacity requirements through the APIs, a similar process is employed to adjust the capcity to the desired levels. Depending on the implementation, it may require a complete shut down of the service instances. However, a more efficient and real cloud-like approach is really to add capacity dynamically.

Figure 4–9 Create Service Process



The key steps illustrated in the figure are summarized below.

- The consumer makes a request to create (provision) a service through the IaaS or PaaS APIs.
- The user is authenticated and the request is authorized by the security module.
- The request and the subsequent ones are mediated by a mediation component such as a Service Bus.
- Orchestration layer validates the request and gets the deployable entity.
- If there are dependencies, all dependant deployable entities are requested from the solution catalog.
- The dependant configuration values are obtained and provisioning is initiated.
- After appropriate security checks, the service is provisioned. How the service is provisioned depends on the type of service and the architecture chosen.
- If applicable, metering is turned on the service. Most of the pricing models require the usage of the service to be monitored and captured for billing and revenue management purposes.
- The result of provisioning is sent back to the consumer.

5 Product Mapping View

Today, Cloud computing is at an early stage in its lifecycle, but it is also the evolution and convergence of technology trends that have been driving enterprise data centers and service providers over the last several years, such as grid computing, clustering, server virtualization, SOA shared services and large scale management automation. Oracle has been a leader in these areas with thousands of customer successes and a high level of investment. Today, Oracle powers the Cloud and is the foundation for many of the world's public and private Clouds.

Oracle's strategy is to offer a broad portfolio of software and hardware products and services to enable public, private and hybrid Clouds, enabling customers to choose the right approach for them. Oracle provides the broadest, most complete and integrated Cloud offerings in the industry.

For private Platform-as-a-Service (PaaS) Clouds, Oracle offers an extensive portfolio of horizontal and industry applications that run on a standards-based, shared services platform, leading middleware and database products, including Oracle Exadata Database Machine and Oracle Exalogic Elastic Cloud. For private Infrastructure-as-a-Service (IaaS) Clouds, Oracle offers leading hardware products. For public Clouds, Oracle On-Demand is a Cloud service provider delivering application and platform services to customers. Customers can also choose to run Oracle products in third party public Clouds. Many third party SaaS ISVs and other public Clouds are powered by Oracle technology. Finally, Oracle also offers software to integrate across public and private Clouds.

This chapter maps the Oracle products to the Cloud logical view, Cloud management capabilities, and the use case we discussed in the previous chapter.

5.1 Summary of Oracle Cloud Products

There are a number of products and product options from Oracle that can be used individually to satisfy specific the needs of the Cloud, or used in combination to establish a complete Cloud implementation. Note that product names, features, and options may evolve and change over time. Please consult your local sales representative or visit `http://www.oracle.com` for the latest product information.

The following products are included in the product mapping view:

- **Oracle Exalogic Elastic Cloud** - The Oracle Exalogic Elastic Cloud X2-2 is a one-size-fits-all datacenter building block, pre-integrating compute, storage and network components to provide an ideal out-of-the-box platform for the broadest range of typical enterprise application workloads, from middleware-based custom applications and packaged applications from Oracle to 3rd party applications and entire private clouds. Oracle Exalogic Elastic Cloud Software includes a number of optimizations and enhancements made to the core products within Oracle

WebLogic Suite. This includes Oracle WebLogic Server, Oracle Coherence, Oracle JRockit, and Oracle HotSpot. In addition to unique support for Java applications and Oracle Fusion Middleware, Exalogic also provides users with a choice of Oracle Linux or Oracle Solaris operating systems.

- **Oracle Exadata** - The Oracle Exadata Database Machine is the only database machine that provides extreme performance for both data warehousing and online transaction processing (OLTP) applications, making it the ideal platform for consolidating onto grids or private clouds. It is a complete package of servers, storage, networking, and software that is massively scalable, secure, and redundant. At the heart of every Oracle Exadata Database Machine are Oracle Exadata Storage Servers, which combine smart storage software and industry-standard hardware to deliver the industry's highest database storage performance. To overcome the limitations of conventional storage, Oracle Exadata Storage Servers use a massively parallel architecture to dramatically increase data bandwidth between the database server and storage.
- **Oracle Enterprise Manager (OEM)** is a family of management products, to manage Oracle environments. OEM enables centralized management functionality for the complete Oracle IT infrastructure, including systems running Oracle and non-Oracle technologies. OEM is a single, integrated solution for managing all aspects of the Cloud and the applications running on it. It delivers a top-down monitoring approach to delivering the highest quality of service for applications with a cost-effective, automated configuration management, provisioning, and administration solution. OEM includes several key capabilities required for Cloud management.
- **Oracle Billing and Revenue Management (OBRM) -** Oracle's BRM is the industry's leading billing and pricing solution for today's on-demand digital services. BRM is deployed by many of the world's largest and most innovative service providers, empowering the entire cloud computing ecosystem with scalable and flexible billing, pricing and revenue sharing capabilities. BRM helps today's next generation service providers to significantly improve time-to-market of new products and services, build stronger brands and lower operational costs. Oracle's Billing and Revenue Management is the only application that provides a product-based solution for revenue management within the Concept-to-Cash-to-Care lifecycle for digital goods and services available within the cloud computing ecosystem. It is built to industry standards on a highly available, real-time platform, and it is functionally rich enough to support all next generation service provider business processes across all lines of business. Oracle's Billing and Revenue Management solution to also includes the ongoing process of analyzing, evaluating, and optimizing each phase of the lifecycle, providing complete insight and intelligence into the revenue relationships that consumers have with the Cloud provider.
- **Oracle VM** - Oracle VM is server virtualization software that fully supports both Oracle and non-Oracle applications, and delivers more efficient performance. Oracle VM offers scalable, low-cost server virtualization. Consisting of open source server software and an integrated Web browser-based management console, Oracle VM provides an easy-to-use graphical interface for creating and managing virtual server pools, running on x86 and x86-64-based systems, across an enterprise.
- **Oracle Access Manager (OAM)** - OAM provides an identity management and access control system that is shared by all applications. It offers a centralized and automated single sign-on (SSO) solution for managing who has access to what information across IT infrastructure.

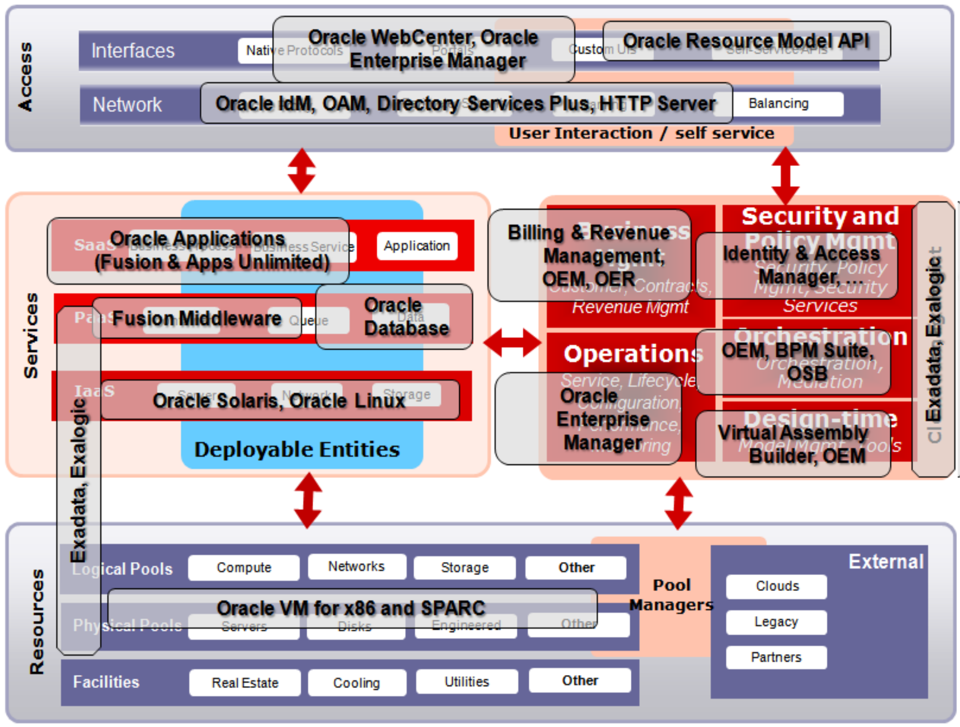
- **Oracle Adaptive Access Manager (OAAM)** - OAAM provides superior protection for businesses and their customers through strong yet easy-to-deploy multifactor authentication and proactive, real-time fraud prevention.
- **Oracle Identity Manager (OIM)** - OIM is a user provisioning and administration solution that automates the process of adding, updating, and deleting user accounts from applications and directories; and improves regulatory compliance by providing granular reports that attest to who has access to what resources.
- **Oracle Web Services Manager (OWSM)** - Oracle Web Services Manager is a JEE application designed to define and implement Web Services security in heterogeneous environments, provide tools to manage Web Services based on service-level agreements, and allow the user to monitor run-time activity in graphical charts. Oracle Web Services Manager is delivered both as a standalone product and as part of the Oracle SOA Suite.
- **Oracle Entitlements Server (OES)** - OES externalizes and centralizes fine-grained authorization policies for enterprise applications and Web Services.
- **Oracle Platform Security Services (OPSS)** - OPSS provides security APIs for Java SE and Java EE platforms that insulate application developers from underlying identity systems and allows them to focus on business logic. OPSS is an enterprise-grade, standards-based, and portable security platform used by all Oracle Fusion Middleware 11g components and Oracle Fusion applications. As the security platform used by Oracle for its own products, OPSS is now also available to customers of Oracle Fusion Middleware for writing custom applications.
- **Oracle Database (ODB)** - Oracle Database delivers industry leading performance, scalability, security and reliability. It provides comprehensive features to easily manage the most demanding transaction processing, business intelligence, security, and content management applications.
- **Oracle WebCenter (OWC)** - provides the foundation for delivering a modern user experience for Oracle Fusion Middleware as well as Oracle Fusion Applications. OWC is composed of four main components described below.
  - **Oracle WebCenter Framework (OWCF)** - provides the ability to create composite applications bringing together transactional applications, content management, and collaboration. OWCF enables developers to compose federated user interface components directly into their web applications by providing support for portal producers that can be accessed, deployed, and managed centrally.
  - **Oracle WebCenter Portal (OWCP)** - is a high performance portlet engine that provides the ability to layout a site structure, secure site resources, provide multi-level delegation model, and deliver personalized user experience built directly into Java Server Faces. OWCP is the convergence of the best capabilities from WebLogic Portal, Sun Java Portal Server, and Oracle Portal and is designed to provide an upgrade path for each of the portal products.
  - **Oracle WebCenter Spaces (OWCSp)** - provide a set of templates and site building and management tools to enable micro-sites tailored by business users to target a specific group of users for information sharing, collaborative team projects, and social connections.
  - **Oracle WebCenter Services (OWCS)** - provides ready-to-use components and portlets enriching applications with content, communication, and collaboration capabilities including document libraries, wikis, blogs, forums, presence, etc.

- **Oracle Enterprise Repository (OER)** - OER serves as the core element to the Oracle SOA Governance and metadata management solution. Oracle Enterprise Repository provides a solid foundation for delivering governance throughout the development lifecycle by acting as the single source of truth for information surrounding assets and their dependencies. Oracle Enterprise Repository provides a common communication channel for the automated exchange of metadata and asset information between consumers, providers, policy decision points, and additional governance tooling. It provides the visibility, feedback, controls, and analytics to deliver business value.
- **Oracle Service Bus (OSB)** - OSB transforms complex and brittle architectures into agile integration networks by connecting, mediating, and managing interactions between services and applications. Oracle Service Bus delivers low-cost, standards-based integration for mission critical SOA environments where extreme performance and scalability are requirements. Service Bus represents many capabilities including mediation, routing, and composition.
- **Oracle Middleware -** Oracle's complete family of application infrastructure products-from the Java application server to SOA/BPM and enterprise portals-are integrated with Oracle Applications and technologies to speed implementation and lower the cost of management and change. Best-of-breed offerings and unique hot-pluggable capabilities provide a foundation for innovation and extend the business value of existing investments. Oracle middleware includes products such as SOA suite, BPM suite, and EDA suite.
- **Oracle database and Oracle RAC** - Oracle database is a proven and market leading database that offers sophisticated clustering and management features. Oracle Real Application Cluster (RAC) provides scalability and high availability at the database tier. The database tier can be horizontally scaled by adding database server instances that access the storage grid. The RAC instances can be configured for load balancing or failover based on the specific needs of the application.
- **Oracle Directory Services Plus** provides identity virtualization, storage, and synchronization services for high-performance enterprise and carrier-grade environments. It is the only integrated solution that provides a complete set of directory capabilities. Oracle Directory Services Plus is a single package that includes
    - Oracle Directory Server Enterprise Edition (formerly Sun Directory Server Enterprise Edition), a highly scalable and reliable directory server ideal for heterogeneous environments
    - Oracle Internet Directory, a highly scalable and reliable directory built upon Oracle Database and ideal for Oracle environments
    - Oracle Virtual Directory, the market-leading solution for unifying identity data from multiple data sources in real time without copying or synchronizing

5.2 Oracle Product Mapping to the Logical View

Figure 5–1 shows the mapping of Oracle products to the Cloud logical view discussed earlier in this document.

Figure 5–1 Oracle Product Mapping to Logical View



This view shows the groupings of products by the logical layers.

5.2.1 Access Layer

The self-service portals and customer portals can be implemented using Oracle Web Center. User interfaces and business components may also be developed using Oracle Application Development Framework (ADF). Oracle JDeveloper is the tool used for developing the portal and UI applications.

Oracle Enterprise Manager also provides a rich set of user interfaces for managing services in the Cloud platform. These user interfaces are not intended for system administrators but rather the users that provision and manage the services.

Perimeter security, proxy, and naming services are provided by the Oracle security products such as Oracle HTTP Server (OHS), Oracle Identity Management, Oracle Access Management, and Oracle Directory Services Plus.

Self-service APIs are exposed to manage the resources automatically using custom applications or management programs. Oracle has defined a Cloud Resource Model API or simply Oracle Cloud API. The Oracle Cloud API defines an Application Programming Interface (API) to consumers of IaaS clouds based on Oracle's solution stack. This API enables an infrastructure provider to service their customers by allowing them to

- Browse templates that contain definitions and metadata of a logical unit of service
- Deploy a template into the cloud and form an IT topology on demand
- Perform operations (such as ONLINE, OFFLINE) on the resources
- Take backups of the resources

5.2.2 Service Layer

The Software as a Service (SaaS) layer represents business processes, business services, and applications. Oracle has a number of products to implement processes and services. Oracle on-demand applications offer pre-built services in a Cloud model. Oracle Fusion applications implement best practice business processes and services.

Oracle middleware products including SOA suite, BPM suite, and EDA suite are used to build standards based platforms to develop and run services. Oracle database can also be made available as a platform for third party development. Oracle ExaData and ExaLogic engineered systems are used to build high performance, mission critical platforms offered as services.

Oracle Exalogic, Oracle ZFS Storage, Oracle Solaris and Oracle Linux are examples of compute and storage infrastructure resources that are made available as a service through the Cloud.

5.2.3 Resources Layer

Oracle's virtualization technology, Oracle VM, is a key technology mapped to the resources layer. It provides the abstraction and pool management capabilities for compute resources. Oracle SPARC, ZFS Storage server, Oracle Sun Super Cluster, and Sun Blade Ethernet Switch are examples of compute, storage, and network resources.

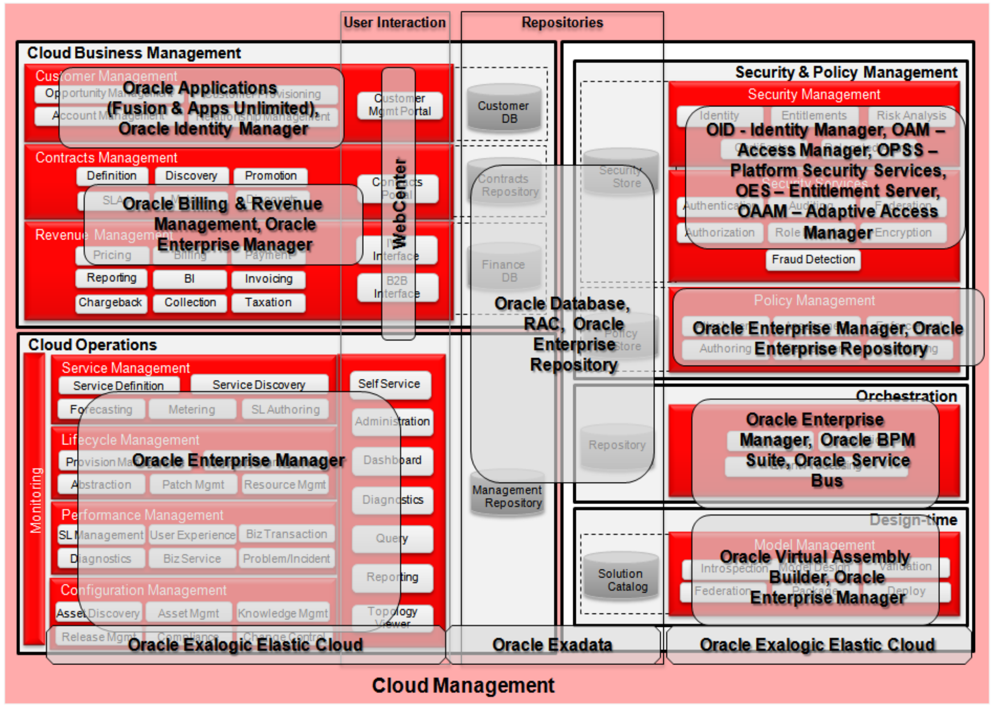
5.2.4 Cloud Management Layer

Product mapping to the Cloud Management Layer is discussed in the next section.

5.3 Oracle Product Mapping to the Cloud Management Layer

Figure 14 shows the mapping of Oracle products to the Cloud Management layer. Oracle products incorporate a rich set of capabilities to manage large scale Cloud implementations. Oracle Cloud Management layer may possibly run on an Oracle ExaLogic Elastic Cloud platform. Oracle ExaLogic is an optimized stack of Oracle products to implement Cloud Services and Management components. Alternatively, the Cloud management layer may be implemented on a open hardware/OS stack such as Oracle Sun Sparc servers and Oracle Solaris OS platform.

Figure 5–2 Oracle Product Mapping to the Cloud Management Layer



5.3.1 Cloud Business Management

The business management portion is clearly dominated by Oracle Billing and Revenue Management (BRM) and Oracle Applications. OBRM provides most revenue management capabilities including billing, payments, reports, and business intelligence. Customer Relationship Management is provided by Oracle Applications. Web Center provides the technology to implement self service portals and user interfaces. Oracle security products provide security and policy management.

Oracle B2B supports B2B integration using a variety of technologies that include EDI. Oracle B2B is an e-commerce gateway that enables the secure and reliable exchange of business documents between the Cloud provider and the consumers (trading partners). Oracle B2B supports business-to-business document standards, security, transports, messaging services, and trading partner management. The Oracle SOA Suite platform, of which Oracle B2B is a binding component, enables the implementation of e-commerce business processes.

5.3.2 Cloud Operations

Oracle has several products that map to the operations category. They are discussed in detail in this section.

5.3.2.1 Service and Lifecycle Management

The key product mapped to this category is Oracle Enterprise Manager (OEM). OEM is a modular product with a comprehensive set of plug-ins that support various parts of enterprise management. In this particular case, OEM supports metering, monitoring, and provisioning.

In a sever-virtualized environment, cloud infrastructure primarily consists of server pools, storage and network. A server pool is a grouping of Oracle VM servers that share common storage. Operations such as live migration of Guest VMs can be performed within a server pool. Oracle Enterprise Manager provides the ability to provision server pools and configure ISCSI or NFS/NAS storage areas for them. Servers in the pool can be used to provision Guest VMs using Oracle VM Templates or ISO images of operating systems. Administrators can view relationships among server pools, Oracle VM servers, Guest VMs and application workloads running in the cloud environment.

More details on how Oracle Enterprise Manager supports Cloud management is provided later in this chapter.

Oracle Enterprise Manager (OEM) provides a rich, comprehensive support for ITIL and with the integration with products like Oracle PeopleSoft Help Desk and seamlessly supports the ITIL processes.

5.3.2.2 Problem Management

To assist with the identification of the root cause of incidents, OEM offers a wealth of diagnostic features. These features range from identifying recent configuration changes that could have caused the incident to occur, to identifying where time was most spent in an application call stack to even identifying the SQL statement that caused poor performance in an application's transaction.

On the proactive side, Problem Management is also concerned with identifying and solving problems and known errors before incidents occur. One type of activity toward this goal is trend analysis, which is identifying recurring problems of a particular type. EM's Configuration Management Database (CMDB) stores historical information about all incidents detected, as well as configuration information about the component and environment on which these incidents have occurred. Using EM's Information Publisher features with the CMDB, administrators can generate trend analysis reports identifying the most common incidents by time or by type, and correlate these with their configuration information enabling them to perform root cause analysis and take preventive measures for these incidents.

5.3.2.3 Configuration Management

The OEM Configuration Management solution manages configuration discovery, enforces configuration changes and automates IT processes. Using agent and agent-less technologies, it discovers and collects detailed configuration information about all hardware and software resources in the enterprise, patches, the relationships and associations between them, topologies, systems and services, and availability and performance metrics. OEM also provides tools for comparing systems enterprise-wide at great detail, allowing an administrator to quickly and easily pinpoint any potential differences

5.3.2.4 Release Management

Oracle Enterprise Manager provides a deep solution that governs the coordinated rollout of software within an enterprise. By centralizing the release process around the Software Library, it allows for adequate planning and control. As a downstream discipline of the Change Management Process, it can perform automated determination of impact and conflict with other software releases. It also orchestrates the release rollout process by sequencing the blackout, shutdown, and startup of services. While rolling out the software, Enterprise Manager notifies interested parties of status changes, giving them the opportunity to take corrective action. Enterprise Manager has built-in capabilities to rollback and re-try a release operation in case the release process did not meet release goals. After the software has been rolled out, Enterprise Manager ensures that the configuration changes to the CI's are updated in the CMDB.

The Enterprise Manager Provisioning Pack ships out of the box best practices deployment procedures that comprise enumeration of a set of steps that are orchestrated by Enterprise Manager. Deployment procedures can be extended and customized for customer needs.

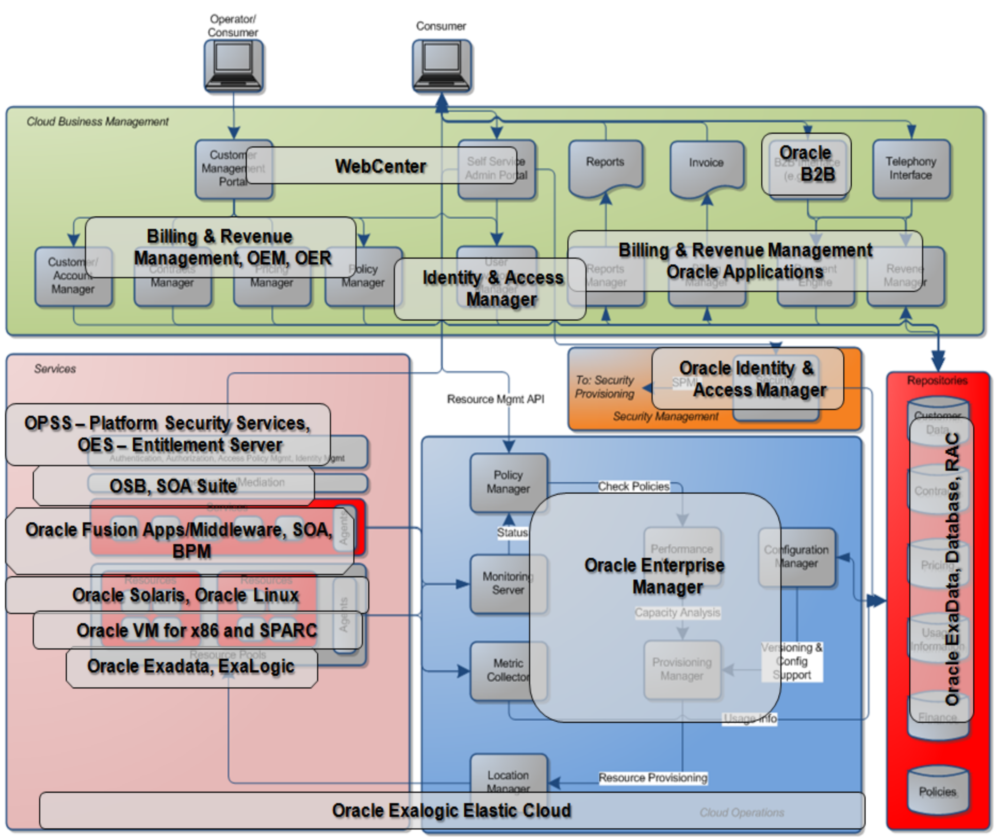
5.3.3 Security Management

Oracle Identity Manager (OIM), Oracle Access Manager (OAM) and Oracle Platform Security Services (OPSS) are some of the key products mapped in this view. Automated user provisioning, delegated administration, and security federation are some of the key issues of Cloud infrastructure and these Oracle products provide the capabilities to address them. Please refer to the "*ORA Security document*" for more information on security architecture and Oracle's security products.

5.3.4 Mapping to the Cloud Management Use Case

Figure 5–3 maps Oracle products to the Cloud Management use case discussed in the previous chapter. The products are not different from what has been discussed so far but it is just a different perspective on how the mapping looks like in the use case view.

Figure 5–3 Oracle Product Mapping to the Cloud Management Use Case



Cloud Business Management, Security Management, Service Management, and Capacity Management have been discussed before; this section describes the product mapping in Services and Repositories.

5.3.5 Services

Oracle security products such as OPSS and Oracle Entitlement Server (OES) provide the security services required for authentication and access control.

Mediation is implemented using products such as Oracle HTTP Server (Routing) and Oracle Service Bus. In this view, mediation refers to business process and business service mediation.

Oracle Fusion Applications are offered as SaaS services and Fusion Middleware, Oracle Exadata, and Oracle Exalogic products are offered as PaaS services.

Virtualization and resource management are provided by Oracle VM and Oracle Linux.

5.3.6 Repositories

A number of logical repositories are shown in the figure. Oracle Database and Oracle RAC are great choices for implementing the repositories. Oracle Exadata database machine is an optimized engineered system to manage the Cloud management data. The Cloud solution often requires data warehousing, BI, and predictive analytics that can best be implemented using an Exadata Solution.

5.4 Details of Key Oracle Cloud Products

Oracle offers products for every aspect of Cloud infrastructure. Although all these products are essential, some are noteworthy and play key roles in Cloud implementation. This section provides more details on the architecture of these Oracle products.

5.4.1 Oracle Enterprise Manager

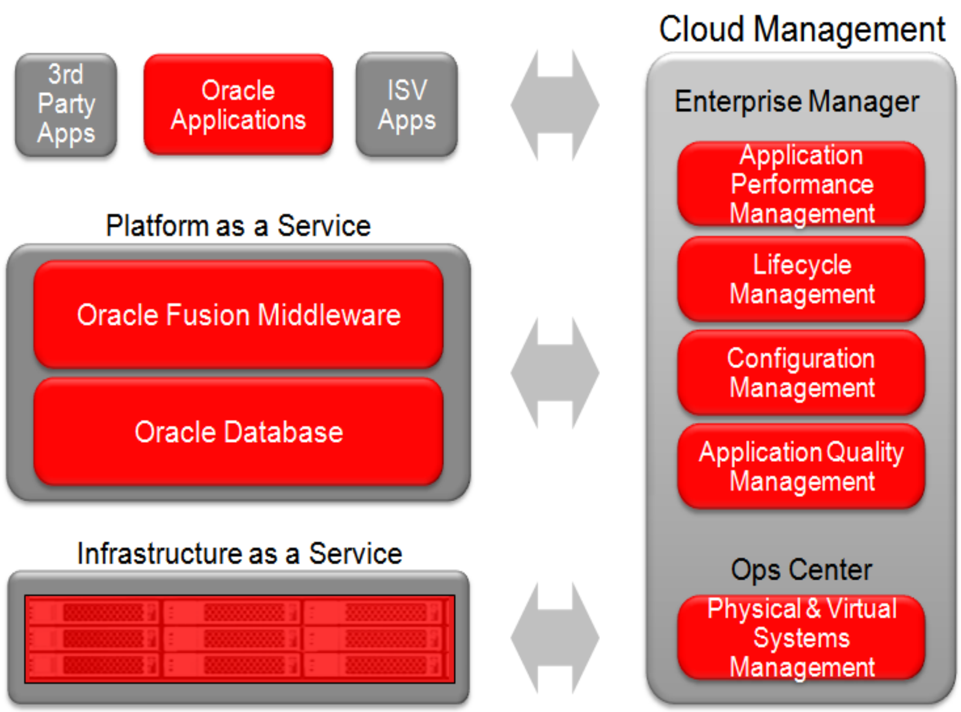
Several organizations have been exploring how to realize the economic benefits of Cloud computing within their own datacenter. Whether it is faster provisioning on demand, agile resource scheduling based on policies, chargeback rules to ensure optimal utilization of resources or more control over the environment, IT is moving from reactive to proactive to predictive approach for data center management.

Oracle Enterprise Manager, Oracle's flagship products for systems management provides single, integrated solution for testing, deploying, operating, monitoring, diagnosing, and resolving problems in today's complex IT environments. It offers a simple, scalable solution for managing the Oracle stack and other third party products, from applications to disk, in Cloud environments. It manages everything from the hypervisor to the operating system, database, and application tier. Oracle Enterprise Manager's broad range of capabilities includes performance and availability monitoring, configuration and compliance management, patching, provisioning, and performance diagnostics and tuning. Oracle Enterprise Manager provides the most comprehensive management solution for Oracle's Cloud platform for both private as well as public Cloud uses. Its key differentiators are:

- Complete coverage of entire Cloud Lifecycle
  - Integrated Cloud Setup, Delivery and Operations

- Support for Multiple implementation choices
  - Choice of IaaS, DBaaS, PaaS from single Self-Service interface
  - Choice of physical and virtual environments (x86, Sparc)
  - End-to-end Application to Disk Awareness
- Integrated management from end-user experience to Cloud infrastructure
  - Out-of-box management of Engineered (Exa*) systems

Figure 5–4 Oracle Enterprise Manager



Oracle EM has capabilities that will enable you to deploy out-of-box Cloud management solutions. Figure 5–4 illustrates these capabilities.

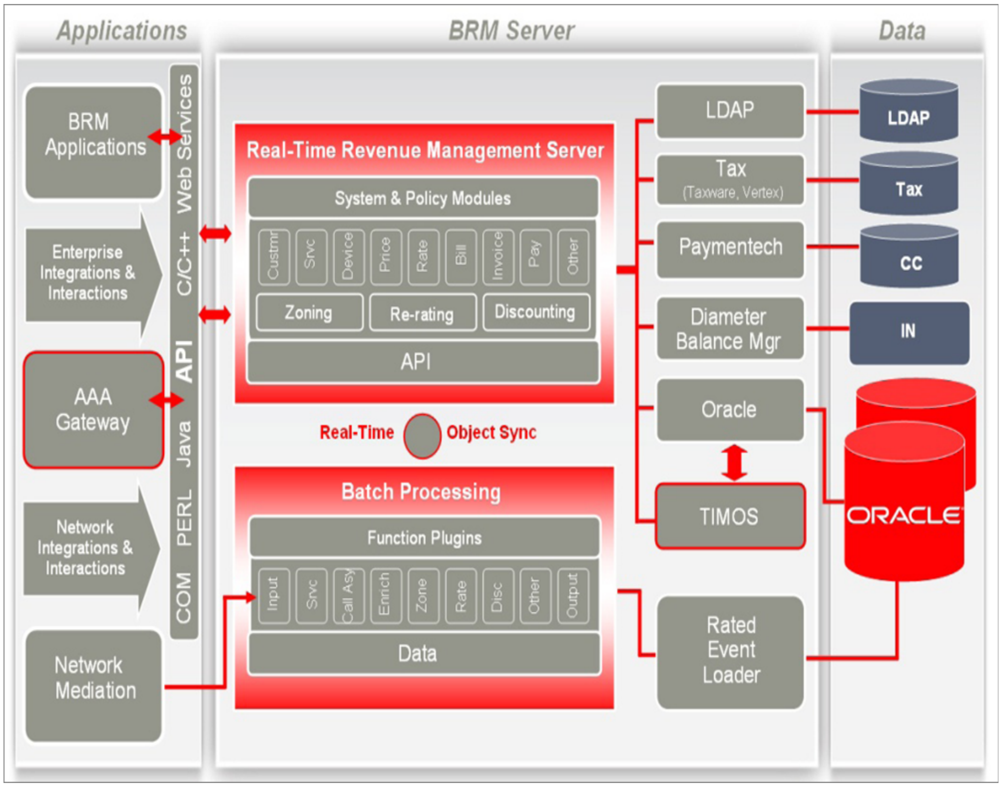
Oracle Enterprise Manager offers many capabilities that can help customers build, test, deploy, and operate Cloud solutions for their enterprise today. In addition, many of these capabilities are being significantly enhanced to further automate and simplify the planning, creation, deployment, and operations of Cloud environments. Whether an enterprise decides to deploy a IaaS or a PaaS, Oracle Enterprise Manager offers rich capabilities for both.

5.4.2 Oracle Billing and Revenue Management (BRM)

Oracle Communications Billing and Revenue Management uses a unified approach to address all of today's communications and media providers' services and business models. Oracle Communications Billing and Revenue Management helps service providers accelerate the launch of innovative new services, enrich global customer relationships, and reduce crippling costs associated with traditional approaches to billing. Figure 5–5 shows the architecture of BRM.

Billing & Revenue Management (BRM) solution addresses the end to end convergent revenue management processes right from Customer definition, mapping to complex customer hierarchies, customer maintenance, inventory management, collecting the usage information, guiding, rating, billing & invoicing, collections/dunning, payment processing, to finally General Ledger accounting and interface to GL.

Figure 5–5 Oracle BRM Architecture



Oracle Communications Billing and Revenue Management provides the world's first convergent revenue management platform that supports multiple services, business models, and customer types. Oracle BRM supports both online (real time) and off-line (batch processing) charging modes, any payment method (cash, credit card, direct debit etc.), any customer type (residential, business etc.), any network/service/product, any business model (retail, wholesale, MVNO etc.). With its productized solution based on industry standards, Oracle offers a single platform by which customers can dramatically accelerate the launch of innovative, profit-rich services including but not limited to wireline, wireless, broadband, cable, voice over IP, IPTV, music, and video.

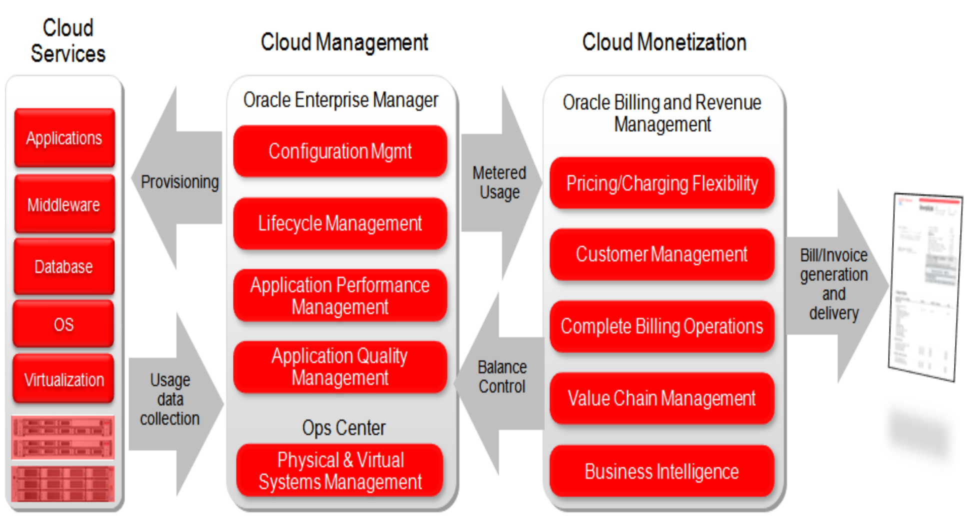
As communications, publishing, media, and entertainment services converge. Oracle can support companies with a proven offering for billing, customer interaction, and management of digital services and content. Oracle can deliver the first end-to-end packaged enterprise software suite for the communications and media industry including service creation, offer management, and order orchestration, through provisioning and service delivery, to billing, revenue assurance, and reporting. BRM has been deployed for all the above-mentioned networks/services and proven at various tier one customers having significant end customer base.

In addition to the above, Oracle offers Oracle Application Integration Architecture (AIA) which is an open standards based framework for creating cross-application business processes that support the way you run your business today, while paving the way for your long term, strategic, business transformation plans. Its application independent framework enables you to utilize the applications of your choice to create composite business processes unique to your business, on a flexible service-oriented architecture (SOA). For customers looking to quickly deploy integrations between Oracle applications, Application Integration Architecture also offers packaged integrations, or Process Integration Packs, allowing for rapid implementation of mission critical business processes which enables service providers to achieve instant value from three best-of-breed industry solutions, Oracle Communications Billing and Revenue Management, Siebel Customer Relationship Management and Oracle Financials. The integrated solution framework provides communications service providers with the following key business benefits:

- Enable faster time to market for new products and services
- Improve business agility and performance
- Streamline operations, while reducing implementation and operational cost

With a flexible, standards-based framework, the productized integrations are designed to work together, can be easily extended to meet business needs and are maintained and upgraded with new releases.

Figure 5–6 Cloud Management and Monetization using OEM and OBRM

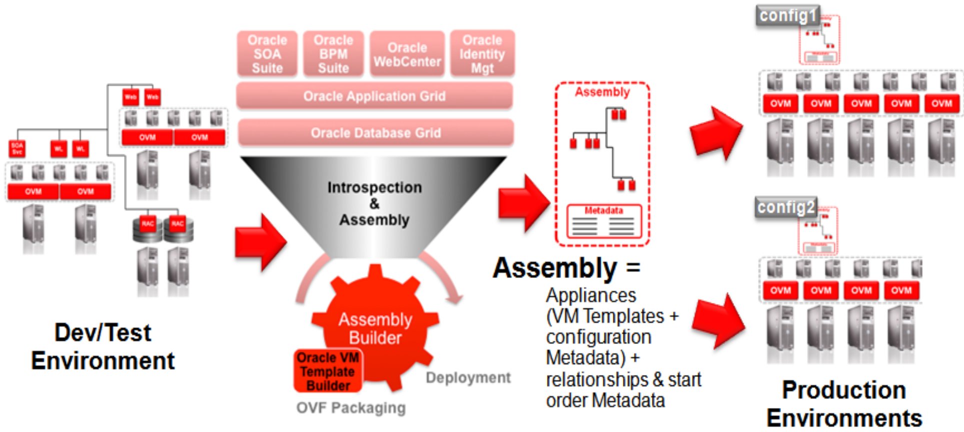


OEM and OBRM complement each other to provide a complete Cloud management and monetization solution as shown in Figure 5–6. OEM provides basic metering and chargeback capabilities with integration to OBRM for any advance billing and revenue management activities. A private Cloud chargeback solution can be implemented using OEM and a comprehensive public Cloud billing solution can be implemented using both OEM and OBRM.

5.4.3 Oracle Virtual Assembly Builder

A self service environment demands automated packaging and delivery. This is a key area of investment for Oracle. Oracle Assembly Builder is a tool for packaging multi-tier applications, including both middleware and database components and their "interconnections" in a way that is easily deployable, movable, etc. Once an application or service is packaged up, EM can consume it as a software image/configuration and deploy it automatically to many locations as needed.

Figure 5–7 Creating the deployable units with Oracle Assembly Builder



In order to ease deployment and management of Oracle products, a set of pre-configured virtual machine appliances are supplied and they can be provisioned on-demand. These could either be a single virtual machine image (for example for single instance database) - referred to as appliances - or multi-tier applications involving several virtual machine images (virtual appliance assembly). For example, an eBusiness Suite assembly consists of a database appliance, a mid-tier/application server appliance, and possibly a proxy/load balancer appliance. Users will then be able to provision a complete eBusiness suite environment as a single entity - with the proposed Cloud infrastructure automatically deploying all the required individual virtual appliances and knitting them together.

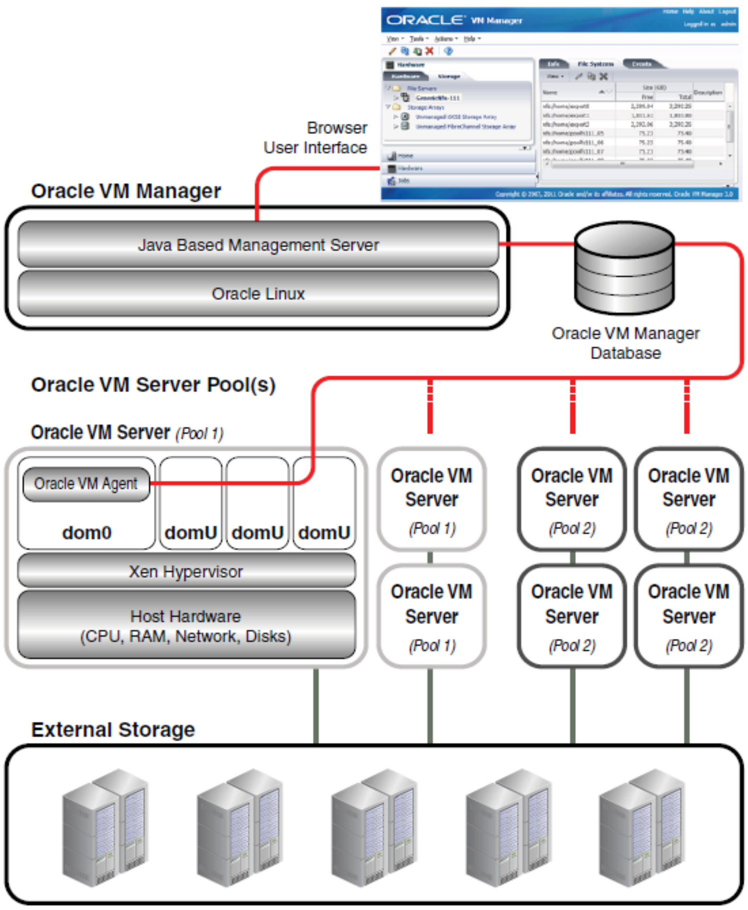
5.4.4 Oracle VM

Oracle VM is server virtualization software that fully supports both Oracle and non-Oracle applications, and delivers more efficient performance. Oracle VM offers scalable, low-cost server virtualization. Consisting of open source server software and an integrated Web browser-based management console, Oracle VM provides an easy-to-use graphical interface for creating and managing virtual server pools, running on x86 and x86-64-based systems, across an enterprise.

Oracle VM Templates deliver rapid software deployment and eliminate installation and configuration costs by providing pre-installed and pre-configured software images. Oracle VM and Fusion middleware products combine the benefits of server clustering and server virtualization technologies, delivering integrated clustering, virtualization, storage, and management for grid computing.

Users can create and manage Virtual Machines (VMs) that exist on the same physical server but behave like independent physical servers. Each virtual machine created with Oracle VM has its own virtual CPUs, network interfaces, storage and operating system. With Oracle VM, users have an easy-to-use browser-based tool for creating, cloning, sharing, configuring, booting and migrating VMs.

Figure 5–8 Oracle VM Architecture



5.4.4.1 Components of Oracle VM

Figure 5–8 shows the architecture of Oracle VM. The components of Oracle VM are described below:

5.4.4.1.1 Oracle VM Manager

Oracle VM Manager …

- Provides the user interface, which is a standards based JEE web application, to manage Oracle VM Servers. Oracle VM Manager 3 is an Oracle Fusion Middleware application, based on the Oracle Weblogic Server application server and Oracle Database.
- Manages virtual machine lifecycle, including creating virtual machines from installation media or from a virtual machine template, deleting, powering off, uploading, deployment and live migration of virtual machines.
- Manages resources, including ISO files, virtual machine templates, and sharable hard disks.

5.4.4.1.2 Oracle VM Server

A self-contained virtualization environment designed to provide a lightweight, secure, server-based platform for running virtual machines. Oracle VM Server is based upon an updated version of the underlying Xen hypervisor technology, and includes Oracle VM Agent.

5.4.4.1.3 Oracle VM Agent

Oracle VM Agent is installed with Oracle VM Server. It communicates with Oracle VM Manager for management of virtual machines.

5.4.4.1.4 Domains

Most of the responsibility of hardware detection in a Oracle VM Server environment is passed to the management domain, referred to as domain zero or dom0. Domains other than the management domain are referred to as domU. These domains are unprivileged domains with no direct access to the hardware or device drivers.

5.4.4.1.5 Hypervisor

Oracle VM Server is architected such that the hypervisor (or monitor, or Virtual Machine Manager) is the only fully privileged entity in the system, but is also extremely small and tightly written. It controls only the most basic resources of the system, including CPU and memory usage, privilege checks, and hardware interrupts.

5.4.4.1.6 Oracle VM Management Pack

The Oracle VM Management Pack is an Oracle Enterprise Manager plug-in that adds Oracle VM Manager functionality to Oracle Enterprise Manager. The Oracle VM Management Pack provides management at the Oracle VM Server, server pool and virtual machine layers, along with guest operating system monitoring, administration, provisioning and patch management.

Oracle VM Servers can be managed from only one of the two management options, either Oracle VM Manager or from the Oracle VM Management Pack. In a Cloud environment, Oracle VM is most likely to be managed using the Oracle VM Management Pack, since Oracle Enterprise Manager provides single point of management for all infrastructure components.

5.4.4.2 Dynamic Resource Management

Automating the virtualized environment to improve application quality of service and reduce power consumption are key enablers for improving operational efficiency. These capabilities are all the more important in Cloud infrastructures. Oracle VM offers Distributed Resource Scheduling (DRS) and Dynamic Power Management (DPM) for capacity management and power management.

5.4.4.2.1 Distributed Resource Scheduling (DRS)

DRS provides real-time monitoring of Oracle VM Server utilization with the goal to rebalance a server pool to provide consistent resources to the running virtual machines. DRS migrates load away from heavily loaded Oracle VM Servers to servers running lighter loads.

5.4.4.2.2 Dynamic Power Management (DPM)

DPM optimizes server pool to minimize power consumption. DPM complements DRS to reduce the servers in the pool when there are periods of low resource utilization. It can automatically add capacity as needed when resource utilization ramps up.

5.4.4.3 Oracle VM Templates

Oracle VM Templates provide an innovative approach to deploying a fully configured software stack by offering pre-installed and pre-configured software images. Use of Oracle VM Templates eliminates the installation and configuration costs, and reduces the ongoing maintenance costs helping organizations achieve faster time to market and lower cost of operations. Oracle VM Templates of many key Oracle products are available for download, including Oracle Database, Oracle Linux, Fusion Middleware, and many more. Oracle also provides tools like Oracle VM Template Builder to create VM templates for third party software and applications.

5.4.4.4 Virtual Assemblies

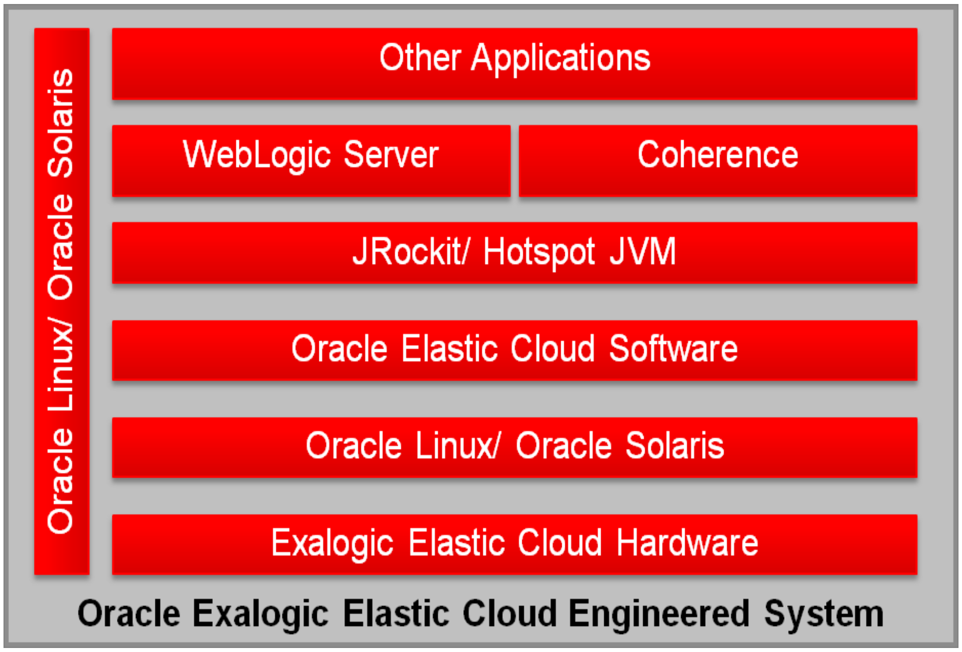
An application "assembly" is very similar to an Oracle VM Template except that an additional set of configuration information and management policies are packaged along with the set of multiple virtual machines, their virtual disks, and the inter-connectivity between them. These assemblies are packaged using the industry-standard Open Virtualization Format (OVF) for the configuration and policy information. Assemblies can be created as a set of .ovf Open Virtualization Format and disk image (.img) files, or may all be contained in a single Open Virtualization Format Archive (.ova) file.

Section 5.4.3, "Oracle Virtual Assembly Builder" provides more details on the Oracle Virtual Builder product that provides the capabilities to build the virtual assemblies.

5.4.5 Oracle Exalogic Elastic Cloud

Oracle Exalogic Elastic Cloud is the world's first engineered system specifically designed to provide enterprises with a foundation for secure, mission-critical private cloud capable of virtually unlimited scale, unbeatable performance, and previously unimagined management simplicity. Exalogic is the ideal platform for applications of all types, from small-scale departmental applications to the largest and most demanding ERP and mainframe applications. While Exalogic is optimized for enterprise Java, Oracle Fusion Middleware, and Oracle's Fusion Applications, it is also an outstanding environment for the thousands of third-party and custom Linux and Oracle Solaris applications widely deployed today.

Figure 5–9 Oracle Exalogic Elastic Cloud



The Exalogic Elastic Cloud Software encapsulates a set of enhancements made to Oracle WebLogic Suite, for optimized performance when running on Exalogic hardware.The list below captures the key benefits provided by the optimizations.

· *Increased WebLogic Scalability, Throughput and Responsiveness:* Improvements to WebLogic's networking, request handling and thread management mechanisms, which enable it to scale better on the high multi-core compute nodes that are connected to the fast InfiniBand fabric that ties all the compute nodes together. The net effect is that each WebLogic server can handle more client requests whilst also reducing the time taken to respond to each individual request.

· *Superior WebLogic Session Replication Performance:* WebLogic's session replication mechanism for Exalogic is improved to utilize the large InfiniBand bandwidth that is used between WebLogic servers for parallel connections over the network. The net effect is choke free communication between WebLogic servers, resulting in superior performance for Java web applications that require high availability.

· *Oracle RAC Integration for More Reliable Database Interaction:* A new technology component called Active GridLink has been added to WebLogic on Exalogic that provides optimized WebLogic to Oracle RAC database connectivity and provides intelligent load-balancing across RAC nodes with faster connection failover if a RAC node fails. The net effect is higher throughput and faster response time for enterprise Java applications that are involved in intensive database work.

· *Reduced Response Times for Exalogic to Exadata Communication:* In situations where an Exalogic system is directly connected to an Exadata system, using InfiniBand, a native InfiniBand networking protocol called SDP is used to interact with the Oracle RAC database on Exadata. This results in low latency for request-response times for calls between WebLogic and the database. The performance gain is most significant when large results sets are transferred from the database. The net effect is applications are able to respond to client requests faster, leading to overall performance gain for enterprise Java applications.

5.4.5.1 Oracle Exalogic Elastic Cloud Hardware

Exalogic hardware is pre-assembled and delivered in standard 19" 42U rack configurations. Each Exalogic configuration is a unit of elastic cloud capacity balanced for compute-intensive workloads. Each Exalogic configuration contains a number of hot-swappable compute nodes, a clustered, high-performance disk storage subsystem, and a high-bandwidth interconnect fabric comprising the switches needed to connect every individual component within the configuration as well as to externally connect additional Exalogic or Exadata Database Machine racks. In addition, each configuration includes multiple 10 Gigabit Ethernet ports for integration with the datacenter service network and Gigabit Ethernet ports used for integration with the datacenter's management network. All Exalogic configurations are fully redundant at every level and are designed with no single point of failure.

Each Exalogic compute node is a fully self-contained unit of compute capacity with multi-core x86 Xeon processors, redundant power supplies, fast ECC DIMM memory, and redundant InfiniBand Host Channel Adapters. Each compute node also contains two solid-state disks (SSDs), which host the operating system images used to boot the node and act as high-performance local swap space and storage for diagnostic data generated by the system during fault management procedures.

InfiniBand is fundamental to the Exalogic Elastic Cloud system. In addition to providing an extremely fast, high-throughput interconnect between all of the hardware units within a deployment, it also provides extreme scale, application isolation, and elasticity.

5.4.5.2 Oracle Exalogic Elastic Cloud Software

Java is the most successful and pervasive application implementation technology in use by enterprises today. Exalogic has been designed from the ground up to provide the ideal environment for enterprise Java applications and Java-based infrastructure. Oracle's entire Fusion Middleware portfolio is designed for deployment on Exalogic. Ultimately, as the Oracle Fusion Applications portfolio is released, they too will be deployable on Exalogic. Oracle Exalogic Elastic Cloud Software includes a number of optimizations and enhancements made to the core products within Oracle WebLogic Suite, the essential Java foundation on which Oracle's next-generation applications are being developed. Oracle WebLogic Suite includes Oracle WebLogic Server, Oracle Coherence, Oracle JRockit, and Oracle HotSpot.

In addition to unique support for Java applications and Oracle Fusion Middleware, Exalogic also provides users with a choice of Oracle Linux or Oracle Solaris operating systems. Exalogic is 100% compatible with all standard Oracle Linux 5 and Solaris 11 applications, and no special certification for Exalogic is required - all Oracle applications that are certified for the appropriate releases of Oracle Linux and Solaris are supported on Exalogic.

5.4.5.3 Exalogic Elastic Cloud Management

Oracle Enterprise Manager provides application-to-disk management through Grid Control and OpsCenter. Enterprise Manager allows every individual hardware component within an Exalogic deployment to be monitored in real time and, at the customer's option, have system status automatically reported to Oracle Support for proactive system maintenance. Through integration with Oracle Support, Enterprise Manager can apply tested patch bundles tailored for Exalogic that cover every layer of the system, from device firmware and operating system to JVM, application server, upper-stack Fusion Middleware, and Oracle Applications.

5.4.6 Oracle Exadata Database Machine

The Oracle Exadata Database Machine is an extreme-performance data warehouse built using Exadata Storage Servers and state-of-the-art industry-standard hardware from Sun. Designed for large, multi-terabyte data warehouses with I/O-intensive workloads, the Oracle Exadata Database Machine is a complete, pre-optimized and preconfigured package of software, servers, and storage that is simple and fast to implement. The introduction of Oracle Exadata Database Machine X2-8, which comes complete with two 8-socket database servers, 14 Oracle Exadata Storage Servers and InfiniBand switches, extends the family of Exadata Database Machines, offering customers an ideal consolidation platform for very large OLTP and Data Warehousing applications.

The Oracle Exadata Storage Server is a database storage device powered by Exadata Storage Server Software running on Sun hardware. The hardware of the Exadata Storage Server was carefully chosen to match the needs of high performance query processing. The storage server comes preconfigured with: two Intel six-core processors, twelve disks, four flash cards, dual port InfiniBand connectivity, a management card for remote access, all the software preinstalled, and can be installed in a typical 19-inch rack.

Oracle Exadata Database Machine X2-2 (formerly known as V2) includes eight 6-socket database servers, 14 Oracle Exadata Storage Servers and InfiniBand switches, and is available in a choice of configurations that scale from a quarter rack to a full 42-unit rack

The Oracle Exadata Database Machine is built to accommodate up to 150TB of non-compressed user data and includes the following hardware:

- X2-2: Eight SunFire X4170 M2 Oracle Database Servers
- X2-8: Two SunFire X4800 Oracle Database Servers
- Both X2-2 and X2-8:
  - Fourteen Exadata Storage Servers (either SAS or SATA)
  - All the required InfiniBand infrastructure (HCAs, switches, and cables) for database server to Exadata Storage Server communication
  - Ethernet switch for communication from the Database Machine to clients or other computing systems
  - Keyboard, Video or Visual Display Unit Mouse (KVM) hardware

Utilizing a building-block methodology, the Oracle Exadata Database Machine provides a quick and easy way to scale. As new racks of Oracle Database Machines are incrementally added to a system, the storage capacity and performance of the system grows; a two-rack system is simply twice as powerful as a single rack. Scaling out is easy; the additional Oracle Database Machine is connected to the InfiniBand interconnect in existing racks, and Oracle automatically rebalances the database to fully utilize all of the storage and processing power of all racks.

For smaller configurations, Oracle provides Oracle Exadata Database Machine X2-2 Half Rack and Quarter Rack, the Half Rack being a system that is exactly half the size of full-rack Database Machine with four database server and seven Exadata Storage Servers, while the Quarter rack is a system with two database servers and three Exadata Storage Servers.

The Exadata family delivers the scalable hardware capabilities to provide the required bandwidth for high-end data warehousing applications. This solution complements the sophisticated software of Oracle Database 11g, with its broad set of query-processing algorithms, advanced analytics, and robust data integration capabilities that will now be discussed.

6

Deployment View

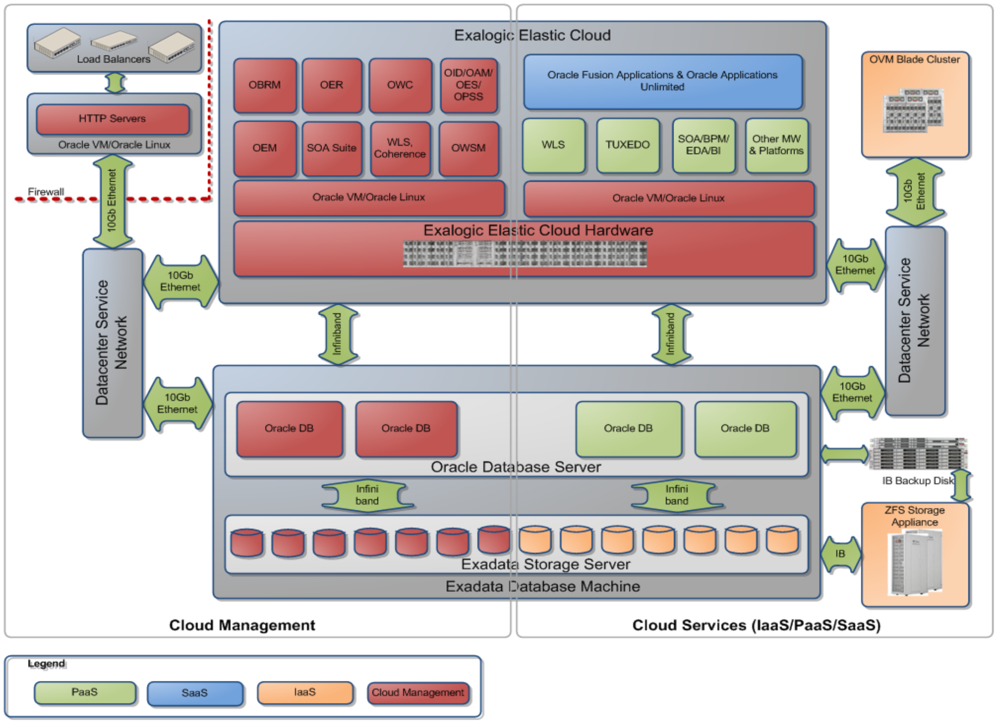
Cloud deployment architecture depends on a number of factors such as the architectural options chosen, the scale of deployment, and the type of services provided. Private Cloud implementations may be relatively simpler compared to large-scale, massive public Cloud infrastructures that support hundreds or thousands of consumers. Hence, it is not possible to create a boilerplate deployment view that can be applied to all Cloud deployments.

This document illustrates two deployment views that take two different approaches. One of them uses engineered systems to implement the Cloud and the other one uses traditional distributed deployment models.

6.1 Engineered Systems Deployment

Most Cloud deployments are massive and require mission critical infrastructure to achieve the performance, elasticity, and SLA goals. Cloud infrastructure can be best implemented using engineered systems. Figure 6–1 illustrates the use of Engineered systems to implement an enterprise Cloud infrastructure.

Figure 6–1 Cloud Infrastructure using Engineered Systems



As discussed in the logical view, Cloud infrastructure includes both the management components and service resources. Engineered systems are generally powerful enough to accommodate both of these layers of the Cloud infrastructure. Figure 6–1 shows a sample configuration in which the Cloud infrastructure is built on Exalogic and Exadata engineered systems. The Exalogic Elastic Cloud runs the software infrastructure required for managing the Cloud, such as OEM, OID, OAM, OBRM, SOA Suite, and others.

Exalogic also hosts the SaaS and PaaS services offered to the consumers. Figure 6–1 shows a sample set of SaaS and PaaS components that are hosted on Exalogic platform. A variety of platform products may be hosted on the Exalogic platform. Examples include Oracle WebLogic server (JEE application server), Oracle TUXEDO (OLTP monitor), SOA/BPM/EDA suite, and other middleware platforms. Oracle Fusion applications and Oracle application unlimited are examples of SaaS services that can be provided on Exalogic.

Note: The capabilities of the products may be different between versions. Please check to make sure the version of the products you have supports the capabilities before deciding on the architecture configuration. For example, this deployment view assumes that the version of Exalogic supports Oracle VM (e.g. Oracle Exalogic 2.1+).

The asset and model management are design-time aspects of Cloud management. The tools used for design-time activities may include Oracle Virtual Assembly Builder and Oracle Enterprise Repository. OER can also used for managing the service catalog for the consumers. OVAB is not shown in the figure but it can be deployed and run externally in a non-production environment.

Figure 6–1 also shows the load balancers, firewall, and HTTP servers that are essential capabilities of the Cloud infrastructure. These components need to be configured to support the multi-tenancy requirements of the Cloud. For example, the HTTP server needs to be configured with routing rules to direct the customer traffic to appropriate service instances.

Exalogic Elastic Cloud and Exadata Database Machine are connected through the high speed infiniband network and connected to the datacenter service network using 10Gb Ethernet connection.

Exadata Database machine provides data management capabilities for the Cloud management layer in addition to providing the resources for the database as a service (DBaaS). The Exadata Storage Server resources are offered as Storage as a Service. In addition Oracle ZFS storage can be connected to this platform with Infiniband to add additional storage capacity to be offered as a service. Computing capacity may be offered through Oracle SPARC SuperCluster or Oracle VM blade cluster, aka Oracle Optimized Solution for Enterprise Cloud Infrastructure.

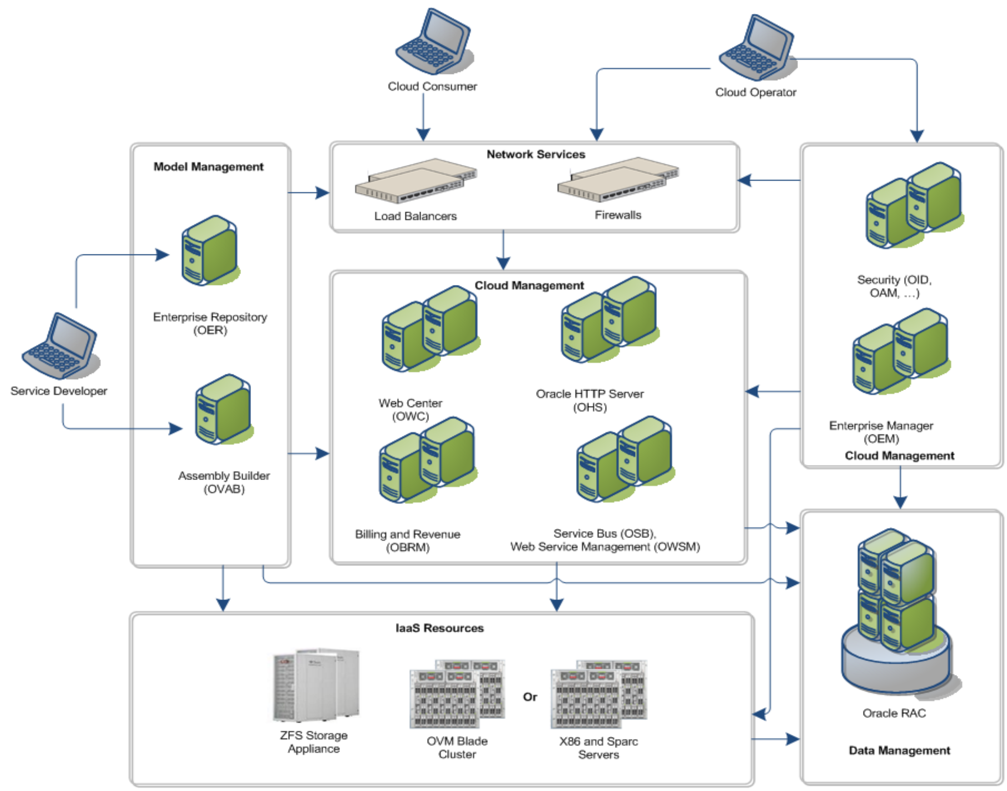
The ORA core document, *ORA Engineered Systems* provides more details on the benefits and architectural insights into Engineered Systems and how they can be deployed in a Cloud infrastructure.

6.2 Distributed Deployment

A holistic deployment view that includes every resource and service would be way too complex to represent in a single diagram. The key to successfully implementing a Cloud infrastructure is to isolate the common management components from the service components and ensure that the management components address the needs of all service models.

Figure 6–2 shows a simplified distributed architecture for Infrastructure as a Service (IaaS) using Oracle products. Service developer tools include Oracle Enterprise Repository (OER) and Oracle Virtual Assembly Builder (OVAB). OER provides asset management, model management, service/resource discovery, and governance capabilities. OVAB is used to model and manage the deployment entities that are deployed to the service layer or management layer. For example, a virtual image may be created for providing a pre-configured virtual machine instance to the consumer.

Figure 6–2 IaaS Simplified Deployment View



Network Services include devices such as load balancers, firewalls, and routers that provide access to rest of the Cloud infrastructure. Depending on the level of security, firewalls may be deployed at any tier. For example, for extra data protection, a firewall may be deployed in front of the database. Similarly a load balancer may be used in front of a web tier or application tier of a PaaS or SaaS service.

Cloud Management is broken down into two categories. The box on the right includes security and management infrastructure that apply to both the management components as well as service components. For example, both the Cloud management components and service resources need to be secured with appropriate authentication and authorization controls. *ORA Security* document provides detailed views into the deployment of the security components.

Oracle Enterprise Manager server is deployed in a HA configuration to manage and monitor the complete stack including server, storage, virtualization layer, JVM, management products, and platform components. Agents are deployed where applicable to monitor and manage the target components. *ORA Monitoring and Management* document provides more details on the agent-based architecture for monitoring and management.

The other part of Cloud Management shows the products that support the various management activities required in a Cloud. Web Center is deployed in a clustered environment to provide the portal functionality (e.g. customer self-service portal). *ORA User Interaction* document provides a detailed view of Web Center deployment.

Oracle Service Bus (OSB) provides mediation capabilities for the services deployed. OSB is clustered to meet the HA and SLA requirements. Oracle Web Service Manager (OWSM) uses an agent-based architecture to virtualize and manage the services. More information on the deployment of these products can be found in the *ORA SOA Infrastructure* document.

Data management for all the management components is provided by the Oracle RAC database shown in the Data Management box. The database is logically partitioned to provide the isolation required by the products. Please note that this database is distinct from the databases offered as service. The databases in the service layer are DBaaS resources dedicated to the use of the Cloud consumers. The *ORA Information Management* document provides an in-depth view of data management options.

Finally, the IaaS Resources box shows the resources offered as infrastructure as a service. ZFS storage appliance and Network Storage represent the storage resources available as services. Oracle VM Blade Cluster and X86/Sparc servers represent the "compute" resources available as a service. These resources are virtualized (not shown in the picture for simplicity) and offered to the consumer as measurable services.

7 Summary

Cloud is quickly becoming an key strategy for business and IT alignment and is starting to dominate architecture roadmap discussions. Most enterprises have either adopted or have plans to adopt Cloud as a strategic choice in support of their business and technology goals. Most Cloud implementations are going to involve some kind of a hybrid approach where enterprise private Clouds are integrated with either other private Clouds or public Clouds. Understanding both provider and consumer perspectives of the Cloud is necessary to successfully implement complex and highly scalable Cloud infrastructures that support the internal and external needs.

Cloud infrastructures are differentiated from traditional IT infrastructure by the scale, velocity, and the level of automation required. Building a successful Cloud infrastructure requires extensive planning and precise execution to ensure that resource capacity is not under- or over- estimated by a large margin. Automation and self-service should be built inherently into the infrastructure to support provisioning, elasticity, and management.

Oracle Reference Architecture (ORA) Cloud ETS provides a reference architecture for building enterprise-class, highly scalable Cloud infrastructure. Oracle provides comprehensive end-to-end solutions to implement and manage Cloud environments. Oracle products such as Oracle Enterprise Manager, Oracle VM, and Oracle Billing and Revenue Management help implement the Cloud management layer with their out of the box Cloud management capabilities. Oracle hardware solutions, engineered systems (Exadata and Exalogic), Oracle Applications, and Oracle Middleware products provide high performance hardware and software for the Cloud services layer. Oracle products help organizations build their Cloud architecture faster, better, and at a lower cost.

A

References and Resources

The ***IT Strategies from Oracle*** series contains a number of documents that offer insight and guidance on many aspects of technology. In particular, the following documents from the ORA series may be of interest:

ORA Service-Oriented Integration - This document examines the most popular and widely used forms of integration, putting them into perspective with current trends made possible by SOA standards and technologies. It offers guidance on how to integrate systems in the Oracle environment, bringing together modern techniques and legacy assets.

ORA Engineered Systems - This document outlines the core concepts around engineered systems and the Oracle engineered system technologies. Cloud infrastructure may be optimally implemented using engineered systems as described in the Cloud

ORA Security - This document describes important aspects of security including identity, role, and entitlement management; authentication, authorization, and auditing (AAA); and transport, message, and data security required to secure the modern IT environment.

ORA Monitoring & Management - A common thread running through many applications, services, and systems is the ability to monitor and manage assets in a consistent and efficient manner. ORA Monitoring and Management offers a framework for OA&M to rationalize these capabilities and help optimize the operational aspects of enterprise computing.

ORA Service Orientation - The promise of cost savings and agility derived from a service oriented approach to architecture has garnered widespread attention within the IT industry. This document describes how Oracle Reference Architecture embraces service orientation to connect disparate technologies into a unified reference architecture

ORA SOA Foundation - This document describes they key tenets for SOA design, development, and execution environments. Topics include: service definition, service layering, service types, the service model, composite applications, invocation patterns, and standards.

ORA SOA Infrastructure - Properly architected, SOA provides a robust and manageable infrastructure that enables faster solution delivery. This document describes the role of infrastructure and its capabilities. Topics include: logical architecture, deployment views, and Oracle product mapping.

ORA BPM Foundation - This document defines the core concepts of modern BPM, provides a conceptual architecture depicting the key capabilities required, and identifies the architectural principles for successful BPM.

ORA BPM Infrastructure - This document connects the conceptual architecture with a logical architectural view and includes the functional components necessary. Topics include: logical architecture, deployment considerations, and Oracle product mapping.

ORA EDA Foundation - This document describes the concepts and business benefits of EDA, provides a conceptual architecture depicting the key capabilities required, identifies the architectural principles for successful EDA, and identifies and describes the relevant industry standards.

ORA EDA Infrastructure - This document describes the infrastructure and its capabilities necessary to process complex events. Topics include: logical architecture, deployment views, and Oracle product mapping.

In addition, the following materials and sources of information relevant to Cloud Computing may be useful:

A.1 Other References

- "Database Consolidation onto Private Clouds", Oracle Whitepaper. By Vengurlekar et al.
- "ITIL Best Practices with Oracle Enterprise Manager 10g and Oracle PeopleSoft Help Desk", Oracle Whitepaper, By Sharma, et al.
- "Billing and Revenue Management for Cloud Computing", Oracle BRM datasheet.
- `http://www.opengroup.org/projects/soa-soi/`, Service Oriented Cloud Computing Infrastructure, The Open Group
- "Oracle Cloud Computing", June 2011 Oracle whitepaper, Rex Wang.
- `http://www.oracle.com/technetwork/topics/cloud/oracle-cloud-resource-model-api-154279.pdf`, Oracle Cloud Resource Model API
- "Oracle ExaLogic Elastic Cloud - A brief introduction", Oracle Whitepaper, By Piech, Palmeter, Lehman
- "http://www.gartner.com/it/page.jsp?id=1064712", Gartner Says Cloud Consumers Need Brokerages to Unlock the Potential of Cloud Services
- "http://www.gartner.com/DisplayDocument?ref=g_search&id=973412&subref=simplesearch", Three Types of Cloud Brokerages Will Enhance Cloud Services
- "Cloud Management using Oracle Enterprise Manager", Oracle Whitepaper, By Madhup Gulati, Sudip Datta