

UNOVIECLOUD

Cloud Computing Infrastructure Technical Architecture Document for Hexagrid's Unovie Cloud

By: David Rokita, VP of Technology Operations

November 2011



TABLE OF CONTENTS

Who is this paper for?	3
Executive Summary	3
Infrastructure as a Service, What does it really mean?.....	1
Design Principles	5
System Requirements.....	6
Hexagrid’s Unovie	7
Logical Design.....	8
Physical Design	8
Physical Unovie Specifications	11
Unovie Eco-system	12
Scalable Design.....	13
Licensing Model.....	13
Unovie Essentials.....	14
Unovie Pro	14
Unovie Elite	14
Ready to Deploy	14
Seamless Migration	15
Data centers Deployment	15
Cloud Public Service Interface.....	15
Private Intra-VLAN communication	16
Optional Private VLAN Network	17
Storage Architecture.....	17
Customer Interface.....	17
Summary.....	18

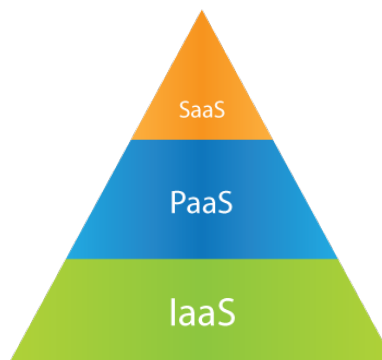
Who Is This Paper For?

Cloud computing, specifically Infrastructure as a Service (IaaS), brings efficiency, agility and value to the enterprise. Both enterprises tremendously benefit from these infrastructure models. Hexagrid Computing is a thought leader in the practice of transforming traditional IT delivery models into highly tuned service delivery platforms. By coupling the very best in hardware solutions with state-of-the-art management software, any data center can take advantage of IaaS. This paper documents the specific use-case for building IaaS clouds using Hexagrid's Unovie, including logical designs, hardware specs and best practices. Unovie creates a unified approach to implementing private and hosted private clouds.

Executive Summary

The business value of cloud computing is no longer disputed. Cloud computing decreases costs, increases asset value, and facilitates on-demand delivery that improves the dynamic of business computing. The industry recognizes three levels of cloud computing, Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS). IaaS is the foundation on which the others are built. Whether for cost-savings or to drive new revenue models, companies with successful IaaS strategies will maintain a material advantage over their competition. Business organizations are looking to adopt these technologies now. Some of these organizations will build services in-house, others will look to purchase these services from others. A majority of these organizations will adopt a combination of both models. The question that businesses should be asking: "how do I build IaaS to meet the growing demand?".

This document examines the requirements of IaaS and makes concrete recommendations for achieving IaaS as a consumer-facing service. The requirements in this document are gleaned from Hexagrid customer use cases and accurately reflect the demands of many service providers and enterprise end-users. In this work, Hexagrid will deconstruct the hardware model, logical and physical design. Unovie is a state-of-the-art approach, purpose built for consistently and repeatedly meeting the demands of IaaS consumers. When properly architected, Unovie serves as the keystone that completes the cloud and truly delivers your data-center as a Service to your end-users.



Cloud Computing Stack

Infrastructure as a Service, What Does it Really Mean?

Any truly compelling technology is driven by hype. Although hype performs wonders when driving industry growth, media outlets often become saturated to the point of confusion. The hype surrounding cloud computing maintains this tradition. For the purpose of this document, it is important to identify the real problems that cloud computing tries to solve. It is Hexagrid's strong belief that the majority of 'cloud computing' environments do not meet the criteria for a true IaaS cloud.

Many IaaS products are built around the concept of managing hypervisors and building virtual machines. This concept aggregates around a bundle of tools fit for only system administrators and technical elite. To only the most basic definition, these solutions are IaaS. These offerings fail to define the actual delivery models and supply chains that are required to truly offer IaaS. To be clear, Hexagrid certainly believes that the practical application of server virtualization can provide tremendous value to any organization, but the ability to provision a VM from a management portal is simply not enough to meet the criteria for cloud.

Cloud computing is a delivery and consumption model. In its purest form, computing resources should flow downward through the supply chain to be ultimately consumed by users that are completely abstracted from the technical implementation. Cloud computing implies self-service. Of course this doesn't mean that normal policies and protocols don't apply to the utilization of resources, but traditional challenges of finding hardware, capacity planning three to five years in advance, and racking servers may be relegated to the past. This becomes a very important point to understand because there are many tools available that claim 'IaaS' while ignoring the most important part: 'service'.

This document covers the entire spectrum of IaaS delivery, including logical integration points for supplemental value-added features. It entails technical implementation for building and creating cloud servers and infrastructure as well as logical hierarchies and delivery mechanisms to extend the management of these artifacts as a service to the end-users that require them. For enterprise, those end-users are divisions, departments, projects, or other company segments to which IT resources are allocated. For service providers, those end-users are direct customers or those of reseller channels.

Because this article includes providing IaaS as an end-consumer service, it is assumed that the trust relationship between the cloud provider and end-user is low to non-existent. This is important because many 'IaaS' tools simply do not have the controls in place to make this practical. Furthermore, it is a gap that needs to be filled by the administrators of the cloud. Since enterprises are consumers of technology rather than development houses, the prospect of building an IaaS cloud with these tools is difficult, if not impossible, to achieve at a price point that would allow them to obtain

Design Principles

When tackling the problem of IaaS, there are many possible paths. Hexagrid believes that the best solutions to problems are elegant in their design. Elegant solutions have common themes that provide a strong foundation. Unovie was designed on the following principles.

- The cloud will be built on a **simple and repeatable** deployment model. The single biggest barrier to scalability derives from escalating costs associated with growth. Hexagrid believes that this barrier to scale must be remedied in the business model as well as the technical model.
- The cloud must be built with **multi-tenancy and tiered delivery** in mind. Many believe that cloud computing relationships are flat. Hexagrid believes that IaaS supply chains are complex. The consumer of the cloud isn't always a person; instead it's typically an organization or a department within an organization. The deliverable isn't a virtual machine, but rather a virtual data-center. The cloud must support these complex multi-tenant models.
- Virtualization is a tool used for creating cloud, but by itself, is not cloud. Virtualization vendors are quickly racing towards parity as competitors assimilate feature sets. The majority of hypervisor technology already reaches lowest common denominator status. When the smoke clears, **clouds will be defined by the applications** that run in them, not by the hypervisor upon which it's built.
- The cloud should be open and should strive for **interoperability**. The cloud itself should never prevent a customer from transporting their workload to another cloud technology through a reasonable effort, including easy movement of virtual images between environments.
- **Fault tolerance** should be inherent in the system. The cloud should be a fluid engine for running workloads. When failure occurs, the user community should not be concerned about the workload failover location. Fault-tolerance for everyday data center failure should be transparent.
- Cloud computing implies **self-service**. The entire idea behind IaaS is that traditional IT artifacts like CPU, RAM, and network can be provisioned on-demand. Customers of the cloud should have the ability to build within the allotment of resources provided from the up-stream partner.
- Cloud computing should **conform to existing business models**, not the other way around. It's unreasonable to expect that business relationships and processes change because of new technology. The burden is on cloud technology to align with the relationships that exist today.

System Requirements

The requirements discussed in this section have been collected as a digest of real use cases. As such, the requirements detailed here accurately reflect the actual demands of enterprise and service provider customers as they design and build cloud computing infrastructure.

- **Public Internet Access**
Customers need to be able to define public interfaces so that Internet facing devices can be created in the cloud.
- **Wired vLAN Access**
Customers need to be able to create interfaces to privately maintained VLANs for the purpose of accessing shared MSP services or interfacing with non-cloud infrastructure across Layer 2 networks.
- **Virtual Private vLANs in the Cloud**
The customer should be able to build complex network topologies in the cloud. This ability should be exposed as a self-service function of the end-user.
- **High-Performance**
To meet the requirements of business, the cloud must maintain a minimum performance profile. This performance profile should be in-line with that of non-cloud virtual environments. This also means that the performance should be consistent and predictable.
- **High-Density**
Enterprise data-centers benefit from infrastructure consolidation to achieve maximum density.
- **Fault Tolerance**
Fault tolerance should be inherent in the cloud. The cloud should be elastic enough to handle the most common faults without resulting in an outage of service to the customer.
- **Complex Infrastructure**
The cloud should be able to support logically complex infrastructure to sustain a multitude of topologies and system designs. This requirement implies policy control over intra-vLAN communication for customers' cloud network segments.
- **Self-Administration**
Users of the cloud must have the ability to administer their environment to the extent that the cloud provider chooses. This may include operational interaction, design and build, and console access to the cloud servers.

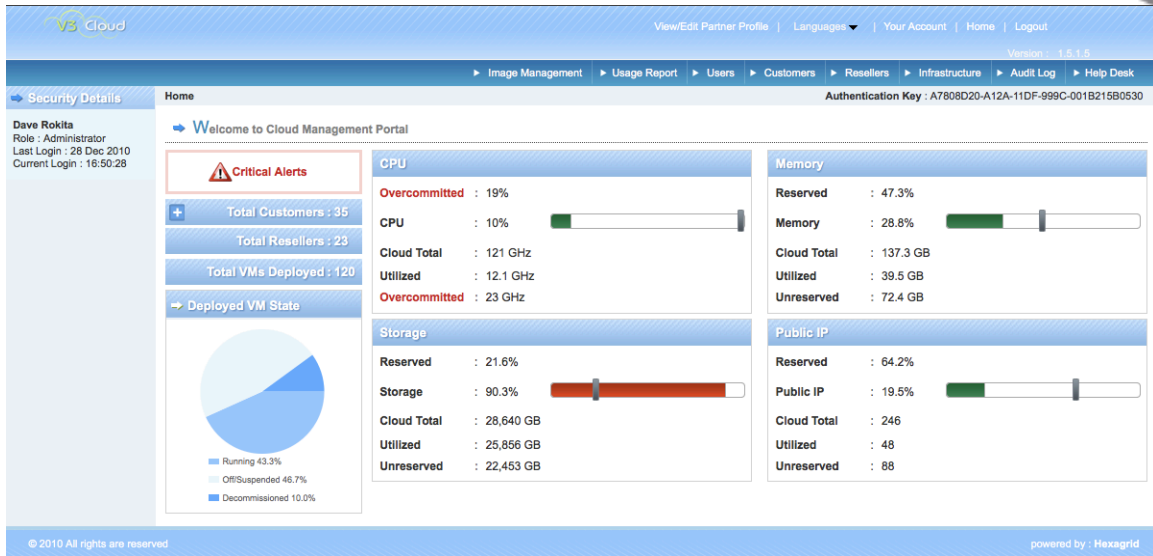
Hexagrid's Unovie

Unovie represents the most exciting development in cloud computing since the idea first captivated the industry. Unovie creates a unified approach to deploying scalable IaaS platforms with minimal cost and complexity. This gives rise to an easily reproducible model that maximizes opportunities for business value and revenue. With Unovie, Hexagrid controls the variance in the hardware deployment model to create a cloud experience that works every time. Using cutting edge technologies, Unovie further reduces the data center footprint and power consumption by unifying compute, network and storage resources. At no other time in the history of computing has more business value been available in a smaller package.

Unovie is purpose built to empower cloud providers with a single practice to implement a variety of service delivery designs. Hexagrid accomplishes this while maintaining the same dedication to service delivery through value-added channels. It is Hexagrid's belief that the delivery model of the cloud must conform and facilitate the execution of these relationships. It remains a key distinction between Hexagrid and other IaaS technology providers.

Unovie Features

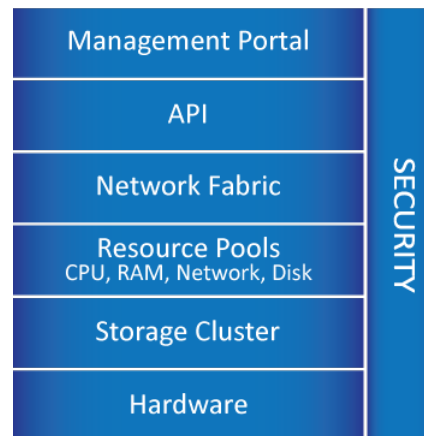
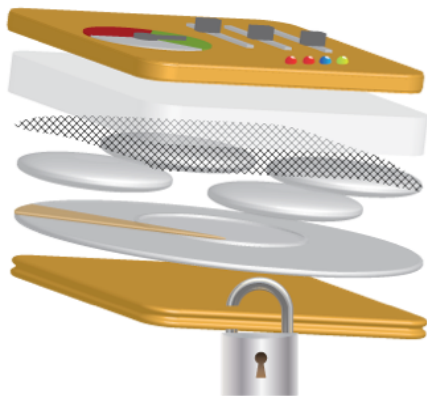
- Multi-tenant channel support model
- Powerful APIs for automation and customization
- Reseller co-branding
- Billing/chargeback integration
- Multiple users and ACL for each organization
- Multi-image support
- Support for direct customers
- Dedicated portal views at every level
- Resource metering and usage accounting
- Intelligent workload scheduling
- Application marketplace/image library
- Integrated secure console access



Unovie Management Portal

Logical Design

Unovie encapsulates all the value and functionality of a distributed cloud-computing environment within a single, easy to manage form-factor. This vastly simplifies the deployment of cloud services. The layer diagram illustrates the comprehensive approach that Unovie packages in the single box deployment.

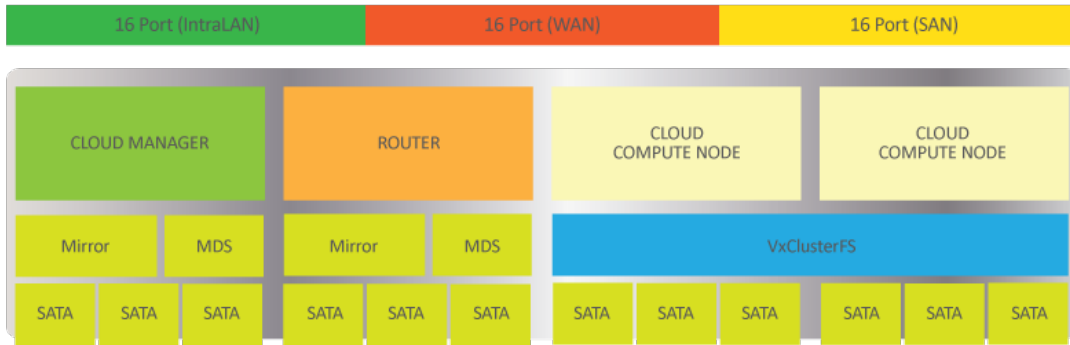


Physical Design

Unovie incorporates a simple blade design to deliver IaaS. Unovie comes in two possible packages. The first package is a head unit. All Unovie deployments require a head unit. The head unit contains the software required for management of the Unovie and supplying portal access to customers. The other package is a scale unit. A scale unit can be added to an existing head unit to increase scale of the Unovie solution. The scale unit does not require that the management functions be replicated, so it contributes 100% of available resources to scale.

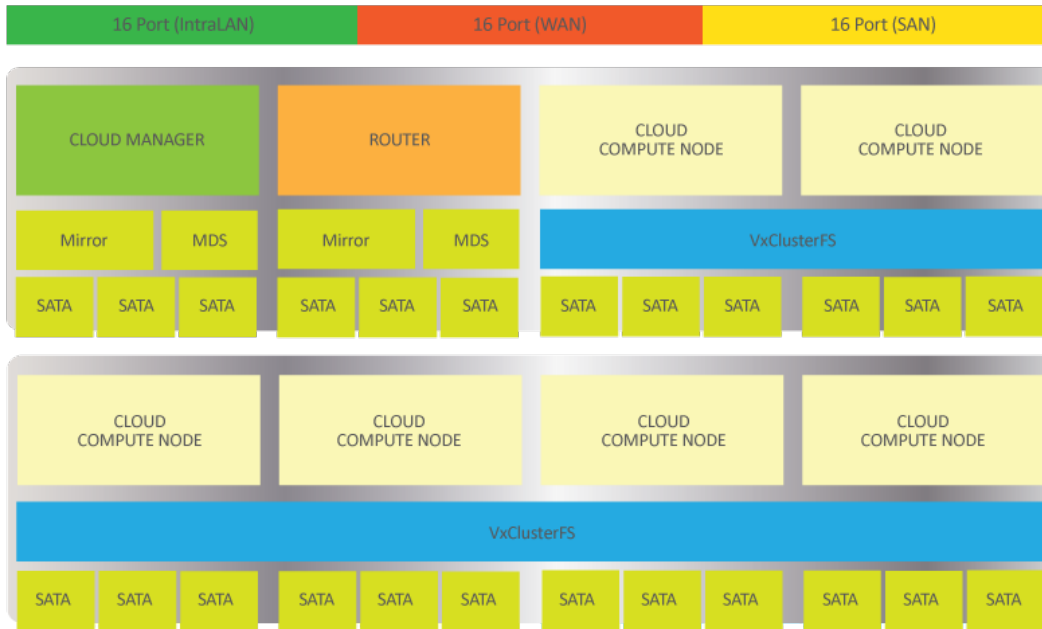


48 Port x 1 GB Ethernet Switch (Layer 2 / Layer 3 Intelligent Non-Blocking Switch)



Single Unovie Head Unit

48 Port x 1 GB Ethernet Switch (Layer 2 / Layer 3 Intelligent Non-Blocking Switch)



Unovie Head Unit with Scale Unit

Physical Unovie Specifications

Head Unit

4 Server Blades – 2U

Each Blade

- 2 - AMD Opteron 6238 2.6 Ghz 115W 12 Core Processors
- 64 GB DDR3 ECC Registered
- 3 - 2TB Enterprise Class SATA Drives
- 4 - 1GB NIC Ports

A Unovie Head Unit uses two blades for management and infrastructure services and allocates two blades for running workloads. The total amount of resources available on a Unovie Head Unit is 96 GB RAM and 24 CPU cores. Unovie Head Units may be deployed in a redundant fashion if the use-case demands that level of fault-tolerance.

Scale Unit

4 Server Blades – 2U

Each Blade

- 2 - AMD Opteron 6238 2.6 Ghz 115W 12 Core Processors
- 64 GB DDR3 ECC Registered
- 3 - 2TB Enterprise Class SATA Drives
- 4 - 1GB NIC Ports

A Unovie Scale Unit utilizes all blades for additional capacity. The total amount of resources available on a Unovie Scale Unit is 192 GB RAM and 48 CPU cores.

Unovie Switch

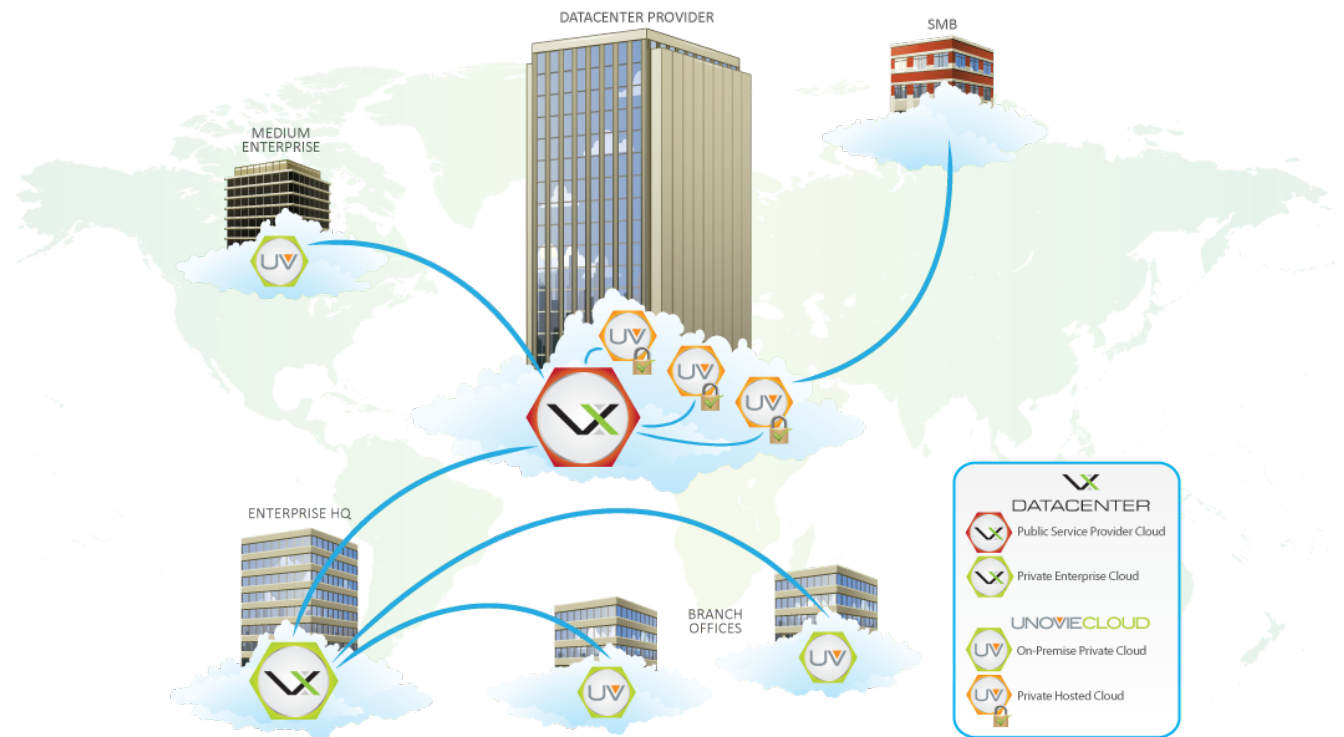
The Unovie Head Unit comes packaged with a single 1U 48 port switch. This switch is configured to work directly with the Unovie Head Unit. This configuration will accommodate up to four Unovie Units. A complete Unovie configuration will support:

- 672 GB RAM and 336 CPU Cores with a single Head Unit
- 576 GB RAM and 288 CPU Cores with redundant Head Units

The base Unovie installation requires only 3U of rack space with a full Unovie configuration requiring only 9U of rack space.

Unovie Ecosystem

Now that the dust has settled and the smoke has cleared in the cloud computing market, it seems clear that some of the initial industry assumptions were a bit ambitious. The assumption that enterprises will ditch their data centers in favor of public clouds may be the most egregious of these. In reality, public clouds are a far shot from being comprehensive enough to solve the problems of most enterprises. Security and compliance remain the biggest obstacles. Due to this, enterprises are looking to gain the efficiencies of cloud, but retain ownership and control. A huge drive towards private, hosted-private and hybrid IaaS clouds dominates the market. Successful enterprises are seizing control of this market swing by taking advantage second generation IaaS product offerings. Unovie offers a unique opportunity for service providers to take advantage of these business opportunities.



The Unovie Ecosystem

Scalable Design

Unovie is designed to start small while meeting the needs of a super scaled environment. Hexagrid does not set artificial limits to scale. The management system is centralized, but out-of-band with respect to the operation of the cloud. This means that it can be scaled independently of the cloud, and if it ever fails, service availability continues unhindered.

One key aspect of the Unovie design allows virtual infrastructure to now be created with no regard to the physical layout of Unovie infrastructure. Customers creating virtual servers and networks are completely abstracted from the actual implementation. Workloads are completely portable within the cloud and can be migrated from physical to physical with no required downtime. Unovie's dynamic workload scheduler detects conflicts and race-conditions to ensure that any contentions are automatically handled with no impact. These facilities allow cloud providers to consolidate their infrastructure to the highest levels while maintaining peace-of-mind that any issues will be addressed. At the option of the cloud provider, quotas can be put in place to protect overall resource utilization from overuse whether intentional or unintentional.

In the Unovie design, one primary goal is self-service for a public/semi-public audience. The challenge is controlling policies so that a user can create their own public facing infrastructure (firewalls, load-balancers) as-required. This allows the cloud consumer to design and build without cloud provider assistance. The act of simply creating virtual machines across managed vLANs is insufficient to provide this capability.

Licensing Model

Unovie is designed to accommodate a wide variety of use cases ranging from simple infrastructure needs to complex deployment patterns through multiple levels of service provisioning. Unovie makes this simple by using a single hardware deployment practice to handle all of these scenarios. Unovie licensing is available in three varieties: Essential, Pro and Elite.



Unovie Essential

Unovie Essential licensing intends to provide everything required to manage a direct infrastructure relationship. This means that the purchaser of Unovie Essential does not require additional relationships to control access to the infrastructure. Unovie Essential is a great fit for flat organizations that manage infrastructure through a single department, still offering role-based access control.

Unovie Pro

Unovie Pro provides an additional layer of service provisioning to that of the Unovie Essential model. Unovie Pro includes a high level management tier that creates smaller organizations or projects (tenants) beneath it. Through this top tier the IaaS owner can allocate resources downstream from a single control point. The top-tier cloud provider in this model can view downstream to understand how cloud resources are being distributed. Depending on the relationship, the top-tier provider may also perform additional services for the lower-tier tenants. Unovie Pro excels at giving small service providers the ability to manage multiple customers. Unovie Pro also simplifies managing multiple projects/divisions within a single organization.

Unovie Elite

Unovie Elite comprehensively responds to the challenge of complete IaaS from the data center through multiple complex relationships to the end-user. Unovie Elite addresses the complexity of using a single, shared infrastructure to manage complex relationships and service delivery chains. With Unovie Elite, the top-tier infrastructure provider may onboard a variety of direct cloud consumers or intermediate IaaS brokers. These brokers either provide value added services or pass IaaS resources through to the organizational hierarchy. Either way, Unovie Elite establishes an IaaS infrastructure ready to conquer the demands of both complex enterprises and top-tier service providers.

Licensing Units

All Unovie licenses may be purchased in 24GB RAM and 1.5 TB Storage licensing units. This allows for scalability in both the licensing and the hardware model as the system grows. Each Unovie Head Unit is sold with four required license units. Each Scale Unit is sold with one required license unit and the capacity to accommodate seven additional licensing units before adding the next Scale Unit to further expand the system.

Ready to Deploy

As previously stated, a key advantage of Unovie is pre-integration to ensure that it works flawlessly straight out of the box. This means that Unovie can be deployed quickly and confidently with minimal configuration. This also makes scaling Unovie a lightweight process that can be completed without requiring the involvement of highly skilled technical resources. Additional units are pluggable to extend the overall physical profile of the cloud and software licensing also can be added incrementally. New customers can effectively jump on-board in less than 24 hours. This ultimately means that Unovie represents the most flexible and business ready IaaS platform on the market today.

Seamless Migration

The term cloud-computing invokes a picture of an environment not bound by hardware. Hexagrid shares this belief. Workload fluidity should be inherent in the cloud. The migration of a workload should be transparent to the end-user. At the same time, the consumers of the cloud should be completely abstracted from the implementation. Unovie dynamic workload scheduler can migrate virtual machines to other machines while maintaining business continuity and requiring no intervention.

Image-Agnostic

Interoperability with other virtualization technologies is inherent within Unovie by allowing cloud consumers to bring in those diverse file formats currently used in other virtualized environments. With Unovie you can run VMDK Fixed (VMware), VHD (Hyper-V), QED/QCOW2 (KVM), XEN Containers and RAW images. This is all accomplished without requiring any conversion, which allows cloud consumers to easily move between virtualized environments and eases concerns about vendor lock-in.

Data Center Deployment

Unovie supports multiple network integrations for a variety of use cases. Unovie manages most of the network integrations of a single switch. This switch is managed by the Unovie software and requires no independent intervention.

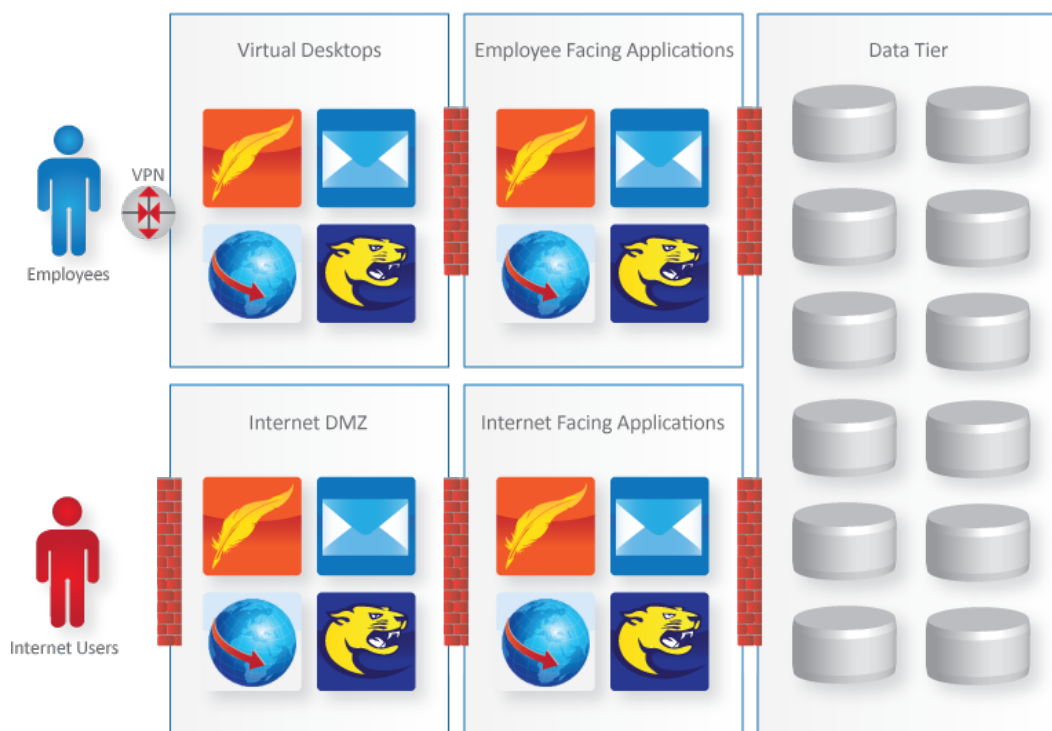
Cloud Public Service Interface

Service providers wanting to provide an Internet interface to the cloud must have some method of providing public facing servers to customers on request. Many of the existing public clouds are flat in their hierarchy. This means that every server created is a public server—perfectly fine for some models, but in many cases, an insufficient design. It is Hexagrid's belief that by default, each server should be a private server on a private vLAN; unless someone explicitly specifies that a server has public interface, it will not be accessible from the Internet. With Unovie, cloud users can create their own vLANs and control the policies between them. In essence, all of the facilities required to create complex network topologies are provided by the platform. Full control of the IP pool is maintained so that public IP addresses are not overused. These IP addresses can be assigned when creating tenant/consumer accounts and can be changed at any time. Finally, bandwidth can be capped for any tenant/consumer through this public facing interface.

For the purpose of providing public cloud services, a Class C block of public IP addresses would be used to create the public pool. As virtual machines are created, the individual building the server has the option to create a public interface (this can of course be edited after server creation). After creating the server, this IP address is statically mapped until the user releases the IP address back into the pool.

Private Intra-vLAN communication

Using Unovie, cloud consumers may create multiple vLANs per user. Furthermore, through the management portal, this functionality can be extended to the customer for self-service. The vLANs created by any user are isolated from any other user's vLAN(s). This creates a tremendous advantage for Unovie customers giving them the ability to recreate complex network topologies in the cloud. Unovie commands the Layer 2 network fabric at the hypervisor level so that virtual machines are free to move and migrate from server to server without having to reconfigure the network implementation.



Example Infrastructure Architecture

The network topology above can be self-provisioned from the Unovie portal without assistance from a cloud provider administrator. Additional segments can be added or removed based on demand. All of the policies are controlled through the Unovie portal interface. With the power of Unovie, users of the cloud no longer have to settle for *just* virtual machines. Virtual data centers can be created to meet the demands of any business need.

Optional Private vLAN Network

The combination of Unovie's managed public IP addresses and virtual private vLANs meet the needs of most public or semi-public cloud computing installations. For those situations where this does not meet the need, Unovie provides an optional network interface to interact with shared services or Layer 2 integration with non-cloud infrastructure. Unovie does this by extending its vLAN tagging implementation to include consumer-provided vLAN designations. The designations can be allocated to a tenant/user allowing them to tag virtual machines as required to meet the business need. This enables the cloud provider to create and sell non-cloud shared services to their customer community as well as interconnect physical servers that the cloud provider manages on their behalf. Implementing the optional managed vLANs also provides another option to ease physical to cloud migrations.

Storage Architecture

Designing the storage tier of a cloud is one of the most challenging tasks with many issues to address. Probably the most important is bandwidth because overall cloud performance will be determined by access to the storage layer. Understanding this, Hexagrid took the initiative to design a purpose-built solution to address the demands of scale and performance in the cloud.

To accomplish this, Unovie establishes two distinct storage pools upon installation. The first pool is a storage cluster for static data and workload templates. The second pool is a high-performance storage cluster on which running workloads reside. Unovie incorporates the latest in storage technology to create a scalable, fault-tolerant and self-balancing storage sub-system. As Unovie scales, the clustered storage automatically scales across the physical hardware. Unovie's physical storage may be augmented by attaching iSCSI LUNs. The LUNs will be incorporated into the cluster and data will be balanced across them. This storage sub-system supports the base requirement that the system be able to migrate a workload by simply re-pointing to the disk images from another cloud server. Unovie's proprietary access controls ensure that no user may access another users secure volumes.

Customer Interface

Hexagrid designed Unovie around the concept of supporting several consumer levels to align with existing supply chains and value-added IT delivery channels. Unovie effectively supports multi-level business relationships. Each level of the relationship has a dedicated portal that exposes the required functionality and information. Resources flow down through the relationship while consumption data flows up for accounting purposes. All of the traditional relationship boundaries between these entities are maintained.

In addition to the portal interface, Unovie also supports a fully featured set of APIs that can be used to further automate the cloud or expose additional services and functions to cloud users. One of the most common uses of the API is to integrate the cloud with an existing service catalogs or show-back systems

Summary

The advantages of adopting an IaaS strategy are clear. For the consumers of IaaS, it decreases the cost of doing business while facilitating agile and flexible IT delivery that empowers business. Unovie's IaaS model represents a new paradigm for addressing customer demands. Service providers build supply chains for computing resources, and likewise require a tool that will empower their existing relationships as well as allow them to define new ones.

Unovie is purpose-built to address the needs of organizations ranging from small business to large, multi-level, distributed enterprises. It was designed to facilitate the smallest environments while maintaining the ability to massively scale. Fault tolerance is not an option nor a configuration, but rather a baseline demand for any IaaS service. Unovie makes this inherent in the system.

For enterprises, Unovie provides not only the tools to deliver IaaS, but also the tools to manage the business that IaaS generates. By defining multiple service interfaces, the service provider may build the business they want, free from technology limitations.

Delivery creates the single biggest barrier to IaaS delivery. Unovie was designed to streamline the process into a simple, repeatable and cost-effective way to deploy cloud services. With Unovie, business can launch multiple cloud computing options with a single delivery practice.

There is no longer any doubt that IaaS and cloud computing have merit. The next generation of computing will be defined by these technologies. Unovie revolutionizes how these services are offered. Unovie is a technology that eases the adoption of the IaaS and offers a long-term strategic vision. The technology is here and the time is now. With Hexagrid and Unovie, start building your cloud today.