What You Need to Know Before Implementing Infrastructure as a Service (IaaS)
What You Need to Know Before Implementing Infrastructure as a Service (IaaS)

John Hales, Global Knowledge VMware Instructor, A+, Network+, CTT+, MCSE, MCDBA, MOUS, VCP, VCAP, VCI, EMCSA

Review of IaaS Clouds

IaaS clouds provide the infrastructure (physical or virtual servers, networking, and storage) in a manner very similar to what was and is done in a typical data center deployment with traditional applications. The user has complete control over all aspects of the infrastructure (subject to the capabilities of the provider), including network speed, number and speed of CPUs, amount of RAM, type of storage, etc. The user can fully configure the operating system, applications, etc., and can tune it as desired. In short, it is much like deploying a physical or virtual server on premises today, except it may not be on premises and you don’t pay for it all up front. You pay for it as you use it.

The advantage—and disadvantage—of this cloud type is complete control. The user is responsible for sizing, installing, and maintaining operating systems and applications, backing up the systems, etc. This enables the user to configure everything in an optimal way for the workloads that need to be accomplished, but it requires time and effort to determine how it should be set up, secured, etc. One advantage of using an IaaS cloud instead of a traditional deployment, however, is that if conditions change (for example, there are many more or fewer users than expected or latency is higher than desired), the configuration can be rapidly changed to meet the changed conditions without wasting the large capital expenditures already made or requiring new capital expenditures. The design can literally change hour by hour, month by month as needed.

In this model, networking is always shared in some fashion on the cloud provider’s infrastructure with other consumers, and storage often is (though doesn’t have to be) shared. In many designs, compute may also be shared via the use of virtual machines (VMs), but some providers offer the ability to dedicate CPU and/or memory to specific VMs (not over-provisioning them) and/or a bare metal server, dedicating physical servers to customers.

Services that are Similar to IaaS, but for Specific Purposes

Some aaS categories don't fall neatly into one of the three broad categories (IaaS, PaaS, and SaaS) but seem to fit best here as IaaS providers.

Disaster Recovery as a Service

Disaster Recovery as a Service (DRaaS) is a great option for many companies that only have a single data center. They may be happy with their existing operations but just want to back up to the cloud (or they may want to have the ability to resume operations in the cloud in the event of an outage in their data center. Typically, the cost for this is the storage cost of keeping the data (and usually VMs) available in the cloud and the network cost to replicate the data from the existing data center to the DR location in the cloud. Usually there is no, or very little, compute cost, as the VMs are usually left in a powered-off state, but this isn’t always true, depending on the replication mechanism chosen, the platform being replicated from/to, etc. If the VMs are powered on and it costs to have them idle in the cloud, consider doing development, QA, or other functions on those servers to make use of them. In any case, this is typically much, much cheaper than opening an entire data center just for DR purposes that will be used only rarely (if ever) in most cases.
Companies that offer products in this area include HotLink DR Express for backing up vSphere to Amazon; Zerto for backup/DR of vSphere and/or Hyper-V to vSphere, Hyper-V, AWS, and a large range of cloud providers; and Sunguard Availability Services, which can support not only VMs, but AIX and other physical platforms as well.

**Desktop aaS**

Desktop as a Service (DaaS) is similar to the typical server deployments done in IaaS, but it is specialized to offer desktop operating systems only. It often involves specialized graphics cards for better video performance; management tools to deploy, configure, and allocate the desktops; and clients or web browser access from a variety of devices, including PCs, Macs, tablets, and even phones. The desktops that providers offer are often Windows-based and sometimes Linux-based. They are almost always VMs, as this service grew out of an on-premises capability known as Virtual Desktop Infrastructure (VDI). There are several advantages and disadvantages to using VDI instead of providing everyone their own desktop, including:

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Company owns a standard image and software, but users can bring any device they want to access that desktop (BYOD), whether subsidized by the company or not</td>
<td>Company may pay for multiple operating system licenses (virtual desktop and on endpoint devices, such as tablets, PCs, and/or MACs)</td>
</tr>
<tr>
<td>Company can ensure that backups and patching are completed</td>
<td>Extra storage space required for backups</td>
</tr>
<tr>
<td>Access from any device, anywhere, anytime</td>
<td>Extra corporate Internet bandwidth required to support all those who access the corporate network off site (e.g., from home)</td>
</tr>
<tr>
<td>Centralized, locked-down desktops for greater security, especially when used with thin clients, reducing endpoint management costs</td>
<td>Reduced flexibility to handle special circumstances for specialized software needs (though for security, this may not be an issue anyway)</td>
</tr>
<tr>
<td>All data is stored in the data center for higher security and to meet compliance requirements</td>
<td>No access to that data for those without an Internet connectivity</td>
</tr>
</tbody>
</table>

Using DaaS instead of VDI offers the same benefits as using IaaS instead of on-premises deployments with additional benefits, including:

- No need to purchase expensive servers or storage (usually the largest part of a VDI deployment); just rent what you need, when you need it
- No licenses to keep up with (at least OS licenses, as they are typically handled by the cloud provider, but application software may or may not be available from them or monitored by them)
- No need for staff to manage all the virtual infrastructure that VDI requires (though end user, helpdesk staff is still usually needed)
- Data centers around the world, potentially putting desktops closer to the users who use them (for remote workers), reducing latency, and improving the end user experience

The biggest differences between IaaS and DaaS are:

- IaaS server deployments usually involve (relatively) fewer, larger servers, while DaaS deployments usually involve more, smaller desktops
- IaaS server deployments usually don’t offer management tools for the servers beyond those natively provided by the installed operating system, such as RDP or SSH, while DaaS deployments usually provide custom (usually web-based) tools to manage the entire deployment, installed applications, data access, etc.
• Storage for the desktops is typically on local disks only—at least as far as the VM is concerned; the provider may put the VMs on shared storage such as iSCSI to facilitate quicker recovery of the VMs if a host server fails—and any data to be accessed and moved outside of the VM is via email or a cloud storage provider, such as Dropbox or Syncplicity.

• Networking is also usually much simpler—just connect and use the desktop—vs. the work involved in designing and deploying global and local load balancers, firewalls, VPNs, etc., that are needed in most server deployments.

Companies that offer products in this area include Horizon Air (the cloud version of the on-premises Horizon suite from VMware), Amazon Work Spaces, and Citrix XenDesktop from a Citrix Service Provider.

What You Need to Know Before You Implement IaaS

In this section, we are not talking about step-by-step directions or anything vendor specific, but rather a high-level review of what is involved in any IaaS deployment from any vendor. You can compare this with PaaS and SaaS and what they require in terms of setup and configuration, as well as how to implement them on an IaaS infrastructure. These areas will, in fact, be covered in more detail in the other white papers in this series.

It is worth noting that you should also check other nontechnical issues as well, such as how to get pricing discounts, whether long-term contracts are needed for those discounts, what SLA they offer and what it covers, and other such business issues.

Networking

Most cloud deployments begin with networking, for without that, there is no way to access any of the provisioned infrastructure in the cloud. You will need to consider the following questions when selecting a provider and deploying the necessary components for a good, functional design:

• What type(s) of access will be needed and for what purpose? There are a lot of things that may require network access and for many reasons. Ask yourself the following:
  – Do you need management access? The answer is usually yes. What do you need it for? Installing/upgrading software? Access to hardware in case of emergency? BIOS-level access to configure storage or other cards?
  – Do you need public access over the Internet? For whom? Your employees? Your customers? Both?
  – Do you need private, secure access between your existing corporate site(s) and the cloud infrastructure you’ve deployed?

• How much bandwidth will be required for the workloads you’ll be deploying? For management access? For communication between the various components you’ve deployed, including storage and other servers? Raw bandwidth is important, as is an understanding of what latency is acceptable. These two parameters may be related (for example, queuing requests during periods of peak demand), or they may not be (for example, how far away users are from the infrastructure they are accessing, with the laws of physics playing a role).

• What is the cost of accessing that bandwidth? With enough money, virtually anything is possible, but you will need to budget for this and ideally have a predictable bill each month. This can be a large portion of your bill (in many deployments, 30 percent or more of the total monthly bill), so check carefully to see what, if anything, is included for free and what the cost of everything else is.

• What pieces of networking equipment can you control? Switches? Routers? Firewalls? Antivirus? Patching? VLANs? This can vary widely from provider to provider and have large implications for security, management overhead, etc.
• How does access work between the providers’ sites if you will be using more than one? This can have implications for cost of availability, as well as designing access and security between the sites.
• How will your existing network from your on-premises data center and/or other cloud providers’ networks (if any) integrate with theirs? What kind of charges are there for this? If the connections are across the Internet, what kind of VPN access will be required and what options are available?
• How does your IP scheme integrate with theirs? This is often overlooked until you try to connect two (or more) networks that have overlapping IP address ranges. Then you’ll need to look into how to handle that. Can the IP addresses be configured/selected when equipment is deployed? Are there supported NAT options? These have cost and management overhead implications.
• These, and many more questions, will need to be discussed with your cloud provider(s) or your potential providers. If networking is done poorly or incorrectly, it’s unlikely that anything will work as expected in the cloud.

Storage
Storage is a little simpler than networking but still presents questions and challenges. Among the questions you’ll need to get answers to are:

• What kind of storage is available? This question falls into a couple of categories that you may be able to explicitly get answers to, or it may be answered in terms of a performance target.
  – What kind of disks are available? Can you choose the number, type, and speed of each? Do they support SAS, SATA, and SSD? Are they available in RAID configurations or Just a Bunch of Disks (JBOD) or both? Can you select the RAID level?
  – What kind of shared storage is available? iSCSI? Fibre Channel? Fibre Channel over Ethernet (FCoE)? NFS? CIFS?
• Is storage performance guaranteed? If not, is it best effort?
• How is storage billed? Space consumed? IOPS used? IOPS provisioned?
• Are other options, such as object storage or Amazon’s Glacier storage, available to reduce the cost of static or slowly changing data, such as backups, to reduce the price of the storage used?

Compute
The last important component is the compute piece. In other words, servers—primarily CPU and memory. As with the previous categories, there are many questions to ask a cloud provider to determine what options are available and the cost of each, as well as to determine the suitability of their offerings for the workloads you want to run in the cloud. These questions include:

• Are physical (bare metal) servers available or just VMs?
• What are the CPU and memory options? Can you select any combination of CPU and memory or just predefined combinations? Can you select the CPU family? What is the process to upgrade CPU and/or memory? Is there downtime involved in upgrading either? If so, approximately how long? Can the time (or a time window) be specified?
• If VMs are available, are they multitenant (more than one company on the physical server) or single tenant?
• If VMs are used, can they automatically failover to another server if the underlying host fails? If not, what is the process?
• Are the CPU cores and memory that your VM uses dedicated to you or shared across VMs? Dedicated CPU and memory will lead to more consistent performance than when they are shared across VMs.
• For physical servers, do you have access to the BIOS? To the disk controller card to configure RAID?
• Can you choose the type and speed of storage you are connected to?
• Can you select the network card speed? If so, what are the options? Is there a cost associated with any of them?
• Are redundant power supplies an option? Are they included or an optional purchase? Is there any restriction on the servers that can have redundant power supplies?
• Is it possible to choose other cards for the servers, such as specialized graphics cards, InfiniBand cards, etc.?

Things to Consider When Implementing IaaS
That didn’t sound too bad, right? Very similar to what you currently do on premises, right? While there are a lot of similarities, there are also many differences and many things that must be considered beyond just the basics to get a secure, reliable, manageable infrastructure built that can scale up or down with demand, provide the requisite performance, and be cost effective. As in the previous section, the goal of this section is not to provide step-by-step directions, nor to imply that all vendors provide all these capabilities the same way or even that this list is exhaustive. The goal is to provide a good place to start and highlight some areas that should be planned for and designed for.

Security
One of the biggest impediments to moving to the cloud is the security of your data. You should ask the following questions to ascertain the security provided by the cloud providers that you are considering:

• How is your data separated and secured from all other customers hosted by the provider? How is it achieved at the network, storage, and compute levels?
• Is data encrypted at rest (on disk)? In motion (via the network)? If yes, who has access to the encryption keys? Can these keys be forced to be turned over to law enforcement under court order?
• What physical (data center) certifications does the provider hold (for example ISO 27001 or SOC 1, 2, and/or 3)?
• What kind of secure workloads can be hosted? Will the provider help you pass audits of those workloads? If so, how?
• Can you meet data privacy rules? For example, in the EU or in Russia personally identifiable information (PII) must be stored in the country or countries involved.
• What access do technicians have to the servers, VMs, and storage used by you? What kind of audit trail is available?
• If a breach were to occur, would you be notified? How about information on breaches or network attacks on other customers (in aggregate form)?

Disaster Recovery
Just as you need to make sure your data is secure, you need to ensure that it is accessible. Among the questions you should ask are:

• What are the options for local failover (if a VM fails)?
• Is failover automatic or manual?
• What if a site-level failure occurs? How do you provide high availability across sites?
• What options need to be purchased to make site availability possible? (For example, global load balancers, redundant infrastructure, storage-based replication, etc.).
• Is there a charge for replicating between data centers? If so, does it matter where in the world you are replicating between (i.e., are some locations free or cheaper than others)?
• Are there application design or deployment decisions that need to be made to make DR possible?
Server (VM) Sizing Options
In addition to the compute questions covered earlier, you should consider the following questions about the options for sizing your servers and/or VMs:

- How many CPUs can be placed in a server? Can you order multicore CPUs, or are they all single core? How many sockets are available? These questions have licensing implications.
- What CPU families are available (ex: Sandy Bridge vs. Haswell)? Do you get a choice?
- How much memory can be placed in a server?
- What is the memory speed and type? Do you have options to choose different types?
- Can memory and storage be sized independently, or are there only predefined combinations allowed?
- Are there redundant power supplies in the server? Are they connected to separate power supplies?

Network Throughput
Other network considerations relate to the throughput available between various components:

- What network speed options are available between VMs in the same data center?
- What network speed options are available between VMs in different data centers?
- What speed is possible between VMs and storage?
- What speed is possible to the Internet? From the Internet?
- How much bandwidth does the provider have between their own data centers?
- How much bandwidth is there within a rack (i.e., to top of rack switches)?
- How fast is the network with a data center?

Manageability
Finally, the best design in the world is not worth very much if it can’t be managed easily or requires many more people to manage. Among the questions to ask in this area are:

- What components in the infrastructure can you control?
- Do you have access to the native management apps, or must you access only through a portal or API the parts they want you to have access to?
- Can you purchase anything in a single-tenant manner (i.e., only you on the server, storage, and/or network device)?
- Can you manage with your own tools or only with those provided by the cloud provider?
- Can you manage both your on-site and your cloud-based infrastructure from a single pane of glass?
- Can you move workloads between on-site and the cloud easily? How much downtime (if any) is required to do this?
- If there are issues in the cloud infrastructure, how will they let you know?
- What is the SLA offered? What are the penalties for not meeting the stated SLA? What do you need to do to collect on the penalties?
Help to Get Started
Who are some of the big IaaS players? Here are just a few. Note that inclusion or exclusion from this list in no way reflects an endorsement to use or a caution to not use any company.

- Amazon Web Services
- Azure (Microsoft)
- Google Compute Engine
- Rackspace Cloud
- SoftLayer (IBM)
- VMware vCloud Air

Conclusion
While this list of questions may seem like a lot, going into any cloud deployment with this information will greatly increase your chances of success and enable you to more successfully articulate to management why they would want to move to the cloud, along with any risks.

Learn More
Learn more about how you can improve productivity, enhance efficiency, and sharpen your competitive edge through training.

- Architecting on AWS
- Systems Operations on AWS
- SoftLayer Fundamentals
- SoftLayer Solutions Design

Visit www.globalknowledge.com or call 1-800-COURSES (1-800-268-7737) to speak with a Global Knowledge training advisor.

About the Author
John Hales (A+, Network+, CCT+, MCSE, MCDBA, MOUS, MCT, VCA-DCV, VCA-Cloud, VCA-Workforce Mobility, VCP, VCP-DT, VCAP-DCA, VCI, EMCSA) is a VMware instructor at Global Knowledge, teaching all of the vSphere and View classes that Global Knowledge offers. He is the author of Administering vSphere 5: Planning, Implementing, and Troubleshooting, published by Cengage, as well as other technical books—from exam-preparation books to quick-reference guides. He has also authored custom courseware for individual customers. John lives with his wife and children in Sunrise, Florida.