



InMage Cloud Optimized Infrastructure

Raising the Bar for Both End User and
Cloud Provider Expectations

Introduction

Cloud computing is fast becoming an accepted way of doing business. While to date it has been used primarily by small and medium businesses, it is now being seriously considered by organizations of all sizes. One of the primary reasons for this is that cloud computing opens up significant new opportunities for enterprises to leverage the benefits of elastic computing models to create more flexible and cost-effective information technology (IT) infrastructures. In the past, most IT organizations built their own infrastructure, typically oversized to accommodate infrequent or seasonal resource usage spikes. These peaks often varied significantly from average usage, resulting in a large percentage of a company's IT infrastructure lying underutilized for much of the time. This led to IT infrastructures that were highly inefficient and overly costly. In today's "high data growth," increasingly regulated, and budget-constrained environments, IT managers are being pushed to create efficient organizations that can flexibly accommodate the needs of the business while at the same time providing the protections necessary to meet evolving business and regulatory mandates.

This is where cloud providers come to the fore. While initial cloud services met the infrastructure needs of smaller organizations or the limited, non-critical needs of larger organizations, they now need to evolve to the next level where they become viable options for the business and mission-critical applications and functions of enterprises of all sizes.

A good example of this is storage and disaster recovery/business continuity (DR/BC). Cloud providers have traditionally provisioned storage resources on demand focused more on non-production or backup and archive data. With the recent introduction of more "intelligent" cloud storage solutions, companies can now use the cloud for production data as well while maintaining the performance and data availability they require.

With respect to DR/BC, the primary focus of most cloud providers has been on traditional backup. While these services may have been billed as "disaster recovery" or "business continuity" solutions, the fact is that most of them limited themselves to restoring data, typically with recovery point objectives (RPOs) of 12 hours or more. In addition, the solutions didn't include application protection, leaving the recovery of the application services necessary to drive the business to the manual efforts of one or more administrators. This further extended the time required to get an application environment functional in the event of a disaster.

As server virtualization and wide area network (WAN) optimization technologies advanced, however, it became possible to think about recovering application services in the cloud, not only reducing the time required to get critical applications back up and running, but extending the concepts of IT elasticity to a broader set of use cases. When applications can potentially be a part of the cloud services mix, end users have new options for expanding and contracting IT infrastructure on demand to meet varying resource requirements without incurring any capital costs.

These developments are driving a new set of requirements which end users should use to evaluate the suitability of various cloud-based services offerings. When thinking about how to strategically leverage cloud computing to drive efficiencies in the IT infrastructure, it is necessary to look at a three to five year roadmap. High data growth rates, increasingly stringent recovery requirements, the increasing penetration of server virtualization technology, and expanding heterogeneity in hardware and software platforms are all market trends that need to be taken into account. Organizations will need to evaluate current and future applications and determine if and how to migrate them to the cloud in a way that makes sense over time. Companies will want a cloud provider that can accommodate their environments and support a set of tiered, solutions-oriented services, make it easy to migrate non-disruptively between services and other clouds, meet requirements for performance, scalability, security, and control, and offer the professional services that allow them to get the most value out of private, public, or hybrid cloud offerings over the long term.

What is needed is a set of infrastructure technologies to form the foundation for an IT infrastructure that is optimized for the requirements of cloud computing. A concept called “cloudbursting” plays a key role in expanding the utility that cloud-based computing offers to the enterprise.

Defining Cloudbursting

“Cloudbursting” is the rapid, on-demand provisioning/de-provisioning of cloud-based virtual servers pre-configured with the right CPU, memory, storage, and operating systems resources. While this capability has been available, what makes this really interesting are technologies which can migrate and/or expand application services at the touch of a button, either into the cloud or out of it, while maintaining the most current data states. This is enabled by several next generation recovery technologies, including continuous data protection (CDP), heterogeneous asynchronous replication, and application failover/failback.

Using these technologies, cloud providers can support true business continuity solutions that can host customers’ application services, enabling very rapid recovery of business services in the event of customer site disasters. A key part of this solution is the support of simple, non-disruptive DR testing, as often as monthly if required.

But more importantly, cloud providers are now in a position to offer a host of other services to customers using the same tools. For example, they can bring platform and application resources on-line quickly to build out test harnesses, enabling applications to be tested at scale without capital expenditure. Test beds can also be brought on-line for patch validation, performance tuning, and other maintenance purposes to meet short term needs without impacting production operations in any way. And, in cases where cloud providers are making the necessary investments to meet client-side performance, reliability, and security requirements, they may offer the use of cloudbursting to meet seasonal demands for additional production capacity with a model that does not result in underutilized in-house IT infrastructure for much of the rest of the year.

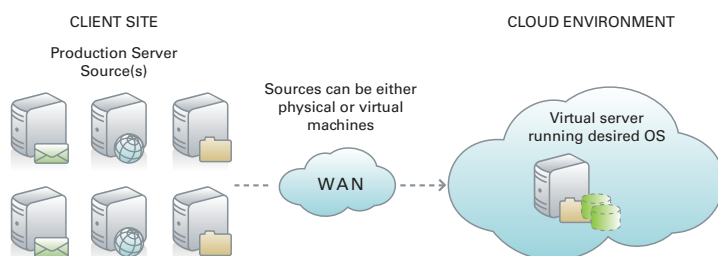


Figure 1. Cloudbursting allows virtual server provisioning on-demand, immediately creating a viable target within the cloud on which to host application services.

Capabilities of a Cloud Optimized Infrastructure

Evolving End User Requirements

To understand what a “cloud optimized infrastructure” should look like, it is important to know how end users want to use that infrastructure. The business requirements from the end user side include non-disruptive IT elasticity on demand that does not impact capital budgets, both for expansion and contraction of services and/or resources, low overhead on customer-side resources like production servers, networks, and storage capacity, ability to meet performance and security requirements, and ease of use.

Cloud providers have traditionally been very flexible about provisioning new services on demand, but have limited end users’ abilities to shut those services off with the same amount of urgency. This is in large part because the licensing models of the vendors that offer application and infrastructure products to the cloud providers do not support ready contraction – they only support expansion. The whole point to the utility computing model enabled by cloud services is to bring services online when they are needed, and turn them off (and stop paying for them) when they are not. Licensing models that support this flexibility provide significant value to end users, and they support significant differentiation for a cloud provider. Because the barriers to exit are lower in this more flexible model, it is reasonable to assume that end users will be more likely to provision cloud-based resources to address short-term needs, increasing the value proposition they can potentially offer to enterprises looking to harness additional services without incurring any capital expense.

Being able to provision additional resources and take advantage of them to meet business requirements has more value the more transparent it is. When evaluating service offerings, end users should consider what the potential impact is on production operations of not only the initial installation of the service, but also what the impacts are of accommodating application and data growth, handling increased I/O rates, provisioning additional server and/or storage resources, migrating application services back and forth between the cloud and in-house infrastructure, and adding tiered services (e.g. enabling encryption and compression or moving from an online backup-only service to a service that can restart primary site application services at a cloud provider’s location). Ideally, end users will want to be able to expand and contract the services they are using with the least amount of production disruption. This includes terminating a service and restoring all data to another location – something that has traditionally been difficult if not impossible.

To enable cloudbursting, a cloud provider must have a way to get both a copy of production data and the application service to the cloud. Traditional solutions have used either some type of modified “backup agent” to periodically capture production data and migrate it to the cloud, or some form of replication which maintains a more updated copy of production data. Both approaches can put anywhere from 10% - 20% overhead on production servers, depending on deployment configurations and the exact functionality they support. Additionally, approaches based on point-in-time backups can impose data loss that is unacceptable for certain environments that have low RPO requirements, while replication technologies do not address data corruption problems unless they are somehow paired with snapshot technologies at the cloud provider site.

A way to resolve both of these “data” problems is through the use of CDP technology. CDP captures data continuously and in real time from production servers while imposing very little (1%-2% per server) overhead. CDP makes data recoverable the instant it is created, not just once it’s backed up, and because it can re-create an almost unlimited number of recovery points within a defined span of time (called the retention window), it offers the industry’s best solution for recovering from data corruption while minimizing data loss. CDP is also the only way to get the data to cloud provider locations that is compatible with virtual server environments and the way they are generally provisioned. Unlike physical servers, most virtual servers are provisioned at utilization rates approaching 90% or above. In these high utilization environments, backup and replication agents that take up 10% - 20% of a server’s resources are just not workable.

In terms of getting application services to the cloud, the cloud provider must also support a simple way to bring an application service up on a target virtual server. Combining application failover/failback technology and shared nothing clustering can enable the launch (or recovery) of application services automatically. In cases where an application is being recovered due to a client-side failure, network-based users of that service can also be automatically re-directed to the new server providing the application service. In both cases, selecting the appropriate data set (i.e. recovery point, etc.) with which to start the application can be either automated or selected manually. The bottom line for end users is that it is easy to migrate application services around onto target server platforms as necessary to meet recovery, administrative, and or production use cases.

End user customers hesitant to experiment with cloud-based services are often also concerned about either performance, availability, security, and/or control issues. These concerns are even more pronounced if they will potentially be running

application services in production within the cloud for certain periods. In these cases, customers should determine through pre-deployment testing with the cloud provider what sort of performance their own end users will experience when accessing application services from the cloud. What provisions does the cloud provider make for high availability and/or disaster recovery within their own infrastructure, and are they willing to commit to enforceable service level agreements with real penalties? Does the cloud provider support role-based access control within their multi tenant management tools that give their own clients the control they want? In terms of security, customers want encryption options and want to understand the multi tenant management capabilities their cloud provider will be using to keep their data safely separated from other cloud provider clients hosted on the same infrastructure. Lastly, what sort of control does the end user customer retain over their applications and data when they are hosted in the cloud?

Evolving Cloud Provider Requirements

The business requirements from the cloud provider side include flexible licensing models, low overhead in deployment and use, non-disruptive scalability, enterprise multi tenancy, and broad heterogeneous support.

For cloud providers to offer a more flexible licensing model to their customers, their vendors need to enable that. Cloud providers would like to be able to start and/or stop services for their customers on demand with greater granularity than what currently exists. At best, most services today require a minimum purchase of a month's worth of services. With more flexible licensing from their own suppliers, cloud providers could differentiate themselves from their competition with more granular offerings while maintaining target profitability.

Because solutions with low overhead are important to end users, they are important to cloud providers as well. Technologies that can keep production server utilization low, like CDP, are important, as are technologies that can store data very efficiently and make very efficient use of network bandwidth.

Cloud providers want to be able to easily scale to accommodate growth without disrupting their customers' production environments, but they are also concerned about the impacts of scaling infrastructure on their own environments. Many of the end user concerns in this area are shared by cloud providers as well. How are existing customer deployments scaled? How do cloud providers scale their own infrastructures to add additional customers? Is it easy to enable tiered services for their customers? Does the management paradigm scale gracefully as well?

To help keep costs low, cloud providers need as much as possible to leverage multi tenant infrastructure on their back end. Multi tenancy can potentially cause end user concerns around security, and cloud providers need to be able to prove that they can meet their customers' requirements. Accurate ways to predictively gauge performance are a requirement. So are encryption options. Cloud providers also need the requisite management infrastructure to be able to manage multi tenant environments efficiently, tracking resource utilization by customer for billing purposes, managing large groups of servers through centrally administered policies, and collecting the appropriate statistics to be able to support their own internal capacity planning.

Finally, the ability to support a broad range of heterogeneous customer environments with a given cloud infrastructure increases a cloud provider's total available market for a given investment. It enables cloud providers to focus on the services they offer, not on what platforms a potential customer may have. To help keep costs low, cloud providers should look for as simple an infrastructure as possible that enables them to meet the requirements of their target markets. Their potential customers have a wide variety of different servers, operating systems, storage architectures and storage device types, and applications. The ability to support heterogeneity maximizes the cloud provider's market opportunity, while at the same time it offers benefits to potential end user customers for preserving existing investments and providing maximum flexibility in new purchases going forward that they may want to incorporate into their use of cloud-based services.

InMage Cloud Optimized Infrastructure

InMage offers an infrastructure solution for cloud providers that maximizes the value of cloud back ends that already support cloudbursting. Combining multiple next generation recovery technologies, including CDP, application-consistent snapshots, heterogeneous asynchronous replication, application failover/failback, storage capacity and WAN optimization, and disk-based recovery, InMage offers the foundation to support a broad range of use cases across recovery, administrative, and production scenarios. InMage has been selling software-based solutions leveraging this technology base to mid tier and larger enterprises for the last five years, and with the addition of the RX Dashboard, a scalable, multi tenant management interface introduced in 2009, began selling "cloud optimized infrastructure" solutions to support managed services offerings from cloud and managed service providers. This technology foundation is equally applicable to private, public, and hybrid cloud environments.

InMage defines “cloud optimized infrastructure” as follows:

- **Flexible licensing** which supports the elastic expansion/contraction of cloud-based services and accommodates the billing implications
- **Low overhead** in deployment and use across server, network, and storage resources, making it a great fit for virtual machine environments that are a key supporting technology in cloud-based computing
- **Non-disruptive scalability** which accommodates the need for server, storage and other infrastructure growth on both the end user and cloud provider sides without impacting client-side production servers
- **Enterprise multi tenancy** that provides for the secure delivery of reliable services to multiple customers with a scalable management model
- **Broad heterogeneous support** that maximizes cloud provider market opportunities by covering a wide range of server, storage, and application environments found in customer settings

Understanding InMage Foundation Technologies

CDP, deployed to capture the data from production servers at customer sites, provides a much less disruptive, lower overhead way to collect data than conventional, point-in-time oriented backups. InMage’s CDP is a game-changing technology that eliminates backup windows, can meet requirements for near zero RPOs, and interfaces with application snapshot APIs (e.g. Windows VSS, Oracle RMAN, etc.) to track application-consistent recovery points which can meet the need for rapid, reliable application recovery on (and/or migration to) target virtual servers in the cloud. InMage’s CDP operates at the block level, making it more efficient than file-based approaches. For those unfamiliar with CDP technology, think of it as “tivo” for your data, a process that logs every single change in real time as it occurs and enables the retroactive re-creation of any previous point in time (application-consistent or otherwise) on demand for recovery or other purposes. Simply turn a dial back to the selected point to generate a disk-based image of the application state at that point.

Because CDP just deals with change data (after an initial synchronization), it generates very little overhead on servers, networks, and cloud provider storage capacity. Think of CDP as a form of storage capacity optimization (it just stores changes) so that large data sets can be protected over time with very little additional storage capacity. Network requirements for InMage’s CDP implementation are very low because it operates

at the block level, only transmits changes across the network, and leverages WAN optimization technologies like compression, TCP optimization, and bandwidth shaping to keep WAN bandwidth requirements low. When combined with InMage’s integrated asynchronous replication, it can support very long distance configurations without impacting production application performance at customer sites.

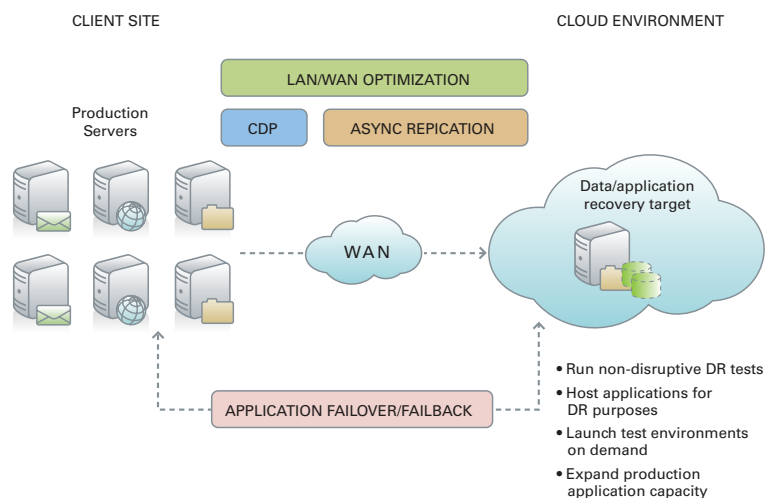


Figure 2. InMage’s “cloud optimized infrastructure” technologies leverage cloudbursting to cover recovery, administrative, and production use cases.

InMage’s use of automated application recovery technology, combined with shared nothing clustering, enables the simple deployment of configurations which can failover and failback to/from local (at the end user customer site) and/or remote (at the cloud provider site) targets. Shared nothing clustering allows servers that are connected by networks (LANs or WANs) to be paired for recovery purposes without the requirement for shared physical disks between them, and relies on some form of replication to keep data on the two servers in sync. Unlike other failover products, InMage’s solution includes productized, fully supported failover/failback for key enterprise applications, including Microsoft Exchange, SQL, and SharePoint, Oracle, MySQL, Blackberry Enterprise Server, SAP, and any file systems on Windows, Linux, or Unix. This makes implementing application migration and/or recovery solutions very easy for cloud providers, and delivers a proven application solution for end users.

Defining the Tiered Services Offerings

For cloud providers that can provision virtual servers and storage resources on demand, InMage provides the foundation technologies to leverage that capability to deliver a broad set of tiered, solutions-oriented services that span recovery, administration, and production use cases. Cloudbursting, made simple through the use of InMage’s application failover/failback technologies, are the key to the higher value services offerings that will differentiate cloud providers by expanding use cases beyond those just supported by resource provisioning on demand.

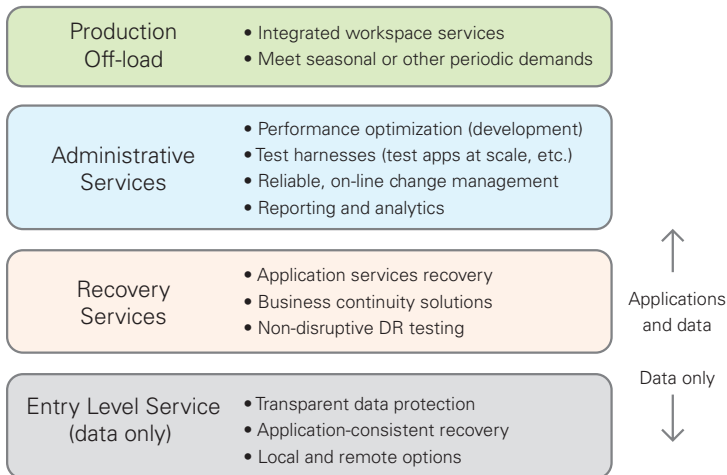


Figure 3. A sample menu of tiered services enabled by InMage Cloud Optimized Infrastructure.

The key for cloud providers that leverage InMage is that a single deployment model - that of installing the InMage data tap on a production server at a customer site - brings that server into the fold, supporting everything from basic, on-line backup services to more sophisticated services that leverage cloudbursting to provide higher value solutions around recovery, administration, and production. The InMage data tap is a filter driver residing on a client-side server that mirrors each write as it occurs to a network-based target which can be either local or remote, depending on which services the configuration must support. Moving to certain higher level services may require a reconfiguration of the cloud provider’s infrastructure, but those changes can be made without disrupting client-side production servers ever again. This type of flexibility gives end users an added incentive to consider tiered services that a cloud provider may offer – they can provision and/or de-commission these services easily and without impacting production.

InMage can cover a wide range of heterogeneous environments, including Windows, Linux, and Unix servers, running in

either clustered or non-clustered modes, and any type of storage subsystems (DAS, SAN, NAS) from any vendor. InMage supports physical and virtual servers as well as mixed physical/virtual server environments, and all the major server virtualization platforms (VMware, Microsoft Hyper-V, Citrix XenServer). A unique feature of cloud provider infrastructures based around InMage foundation technologies is that it can provide application-consistent recovery options for all environments – physical and virtual – using the same exact deployment model. This keeps things simple for the cloud provider, helping to keep costs down. Contrast this with approaches which only support virtual servers or, even more limiting, only support one type of virtual server platform.

The InMage Analyzer

Given that cloud-based services that use changing data sets require some sort of replication technology, end user customers generally have a range of questions they want answered for planning purposes. These questions include:

- What is the data protection load on each server?
- What is the additional load on the LAN?
- What sort of on-premises hardware (if any) is needed to meet performance requirements?
- How much WAN bandwidth do I need?
- Are there spikes in I/O rates over time that need to be managed to maintain quality of service on networks?
- What sort of RPO can I support with my existing bandwidth?
- If I increase or decrease my network bandwidth by 10%, how does that impact my RPO?
- What impact does the use of WAN optimization and encryption services have on WAN bandwidth requirements and recovery performance?
- How much extra storage capacity do I need to provision to meet my business requirements?

InMage is unique in the industry in providing the Analyzer tool which can be deployed to run transparently in the background on a customer’s production servers for several weeks to collect the quantitative data necessary to answer these questions. This provides definitive answers to the questions listed above based on accurate data, ensuring that there are no surprises in deployment configurations or costs for either the end user customers or the cloud providers. This feature has proven itself out as a strong differentiator when InMage has been deployed in in-house infrastructures as a software solution, and it offers the same value in planning for cloud-based services deployments.

Conclusion

InMage's Cloud Optimized Infrastructure solutions will open up new opportunities for cloud providers to differentiate themselves against their competition while at the same time offering new "IT elasticity" options to end user customers as cloud-based services. By bringing application services into the mix through the use of cloudbursting in a way that is simple for cloud provider customers to use, InMage allows cloud providers to go far beyond just provisioning resources to enable new use cases across recovery, administrative, and production off-load scenarios.

A unique combination of next generation recovery technologies defines "cloud optimized infrastructure," including CDP, application snapshot API integration, heterogeneous asynchronous replication, application failover/failback, storage capacity and WAN optimization, and disk-based recovery. InMage delivers these technologies, all controlled centrally with a secure, multi tenant management interface that scales well from small to large configurations, as well as a unique predictive capacity planning tool that results in the most accurately sized deployments available from cloud service providers in the industry.

InMage's Cloud Optimized Infrastructure is based around five key tenets that end users should look for:

- Flexible licensing
- Low overhead
- Non-disruptive scalability
- Enterprise multi tenancy
- Broad heterogeneous support

As the cloud services industry moves forward, cloudbursting will become a valuable usage model to help enterprises control costs while providing the flexible IT infrastructure needed to meet dynamic business requirements. InMage has defined the core technologies that will maximize the value of this model for end users going forward, and is making them available today through InMage Cloud Optimized Infrastructure solutions.

