IT@Intel Brief

Intel IT
IT Best Practices
Cloud Computing
October 2010

Implementing On-Demand Services Inside the Intel IT Private Cloud

- On-demand self-service improves business agility by unlocking the full business value of our cloud computing environment.
- Virtualization has already reduced service provisioning time from 90 days to 14 days.
- On-demand self-service has the potential ability to further reduce our provisioning time to hours and then to minutes.

To meet changing business requirements at a faster pace, Intel IT has shifted from a traditional static enterprise computing environment to a service-oriented environment. We have embarked on a core business strategy to build an enterprise private cloud to improve infrastructure efficiency along with IT service level agility, availability, and security.

On-demand self-service is a critical aspect of a complete cloud environment; however, without underlying business logic, controls, and transparency, an unconstrained on-demand enterprise private cloud will quickly exceed its capacity by doling out allocations beyond its supply. By instituting a hosting automation framework that includes entitlement, quotas, transparent measured services, and data-driven business logic, we are establishing a true enterprise private cloud that provides a consumer-focused self-service portal.

This model, shown in Figure 1, allows Intel IT to provide capacity to our business users when they need it, removing IT from the critical launch path for business services—a key to creating a more agile enterprise infrastructure to support a dynamic and ever-changing set of business requirements.

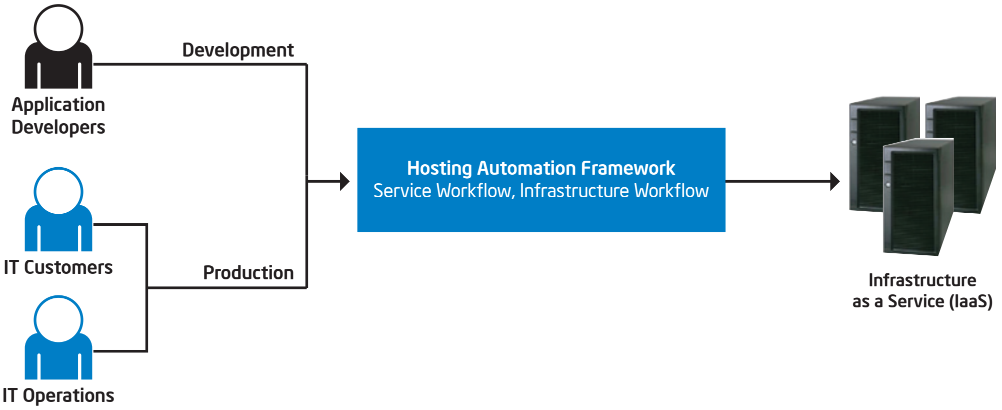
Application Developers
Development
IT Customers
IT Operations
Production
Hosting Automation Framework
Service Workflow, Infrastructure Workflow
Infrastructure as a Service (IaaS)

Figure 1. With Intel IT on-demand self-service, users request and consume services through a self-service portal while IT manages and measures service consumption on a highly utilized resource pool of virtualized assets.

Defining On-Demand Self Service

The National Institute of Standards and Technology (NIST) indentifies on-demand self-service as one of five essential characteristics of cloud computing. On-demand self-service makes IT resource capacity within a cloud infrastructure appear infinite to users. A consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service's provider. For more on NIST definition of cloud computing, visit http://csrc.nist.gov/groups/SNS/cloud-computing.

Background

Most enterprise environments deliver capacity to application owners, software developers, and service owners through cumbersome business processes that require a significant number of manual touch-points before a single server may land. These business processes were often developed as data center space, power, and cooling capacity were becoming constrained.

The emergence of x86 virtualization helped us alleviate physical capacity constraints by increasing server utilization and enabling consolidation of more applications onto fewer servers. However, resource virtualization only solves part of the challenge for IT and business partners. While a virtual machine (VM) can be granted in a significantly shorter period of time than a physical server, the length of time from service request to service provisioning remains significantly higher than is desirable.

Service delivery lead times remain tied to legacy business processes for IT fulfillment: We need to implement a process of checks and balances to justify the business demand before allocating capacity. Additionally, holistic operational management evaluations across various infrastructure components (compute, storage, backup and recovery, network, facility, and power and cooling) are needed to help ensure that there are no capacity constraints before landing a virtualized application on a physical server.

As virtual infrastructure becomes the norm, we have realized that business processes need to change significantly. For example, Intel IT has a 90-day service-level agreement (SLA) for provisioning a new physical server compared to 14 days for provisioning a virtual server. While a significant reduction in server provisioning time is beneficial to us, we realized that provisioning the server is only one of the steps required for delivering business services and that additional improvements were necessary.

A core aspect of our enterprise private cloud is the introduction of on-demand self-service for end users of various services within our enterprise computing environment. After analyzing our business processes and identifying bottlenecks, we successfully conducted a self-service proof of concept (PoC) to validate the model. This PoC demonstrated the ability to provision services in a matter of hours and encouraged us to proceed towards production implementation.

Automating Self-Service Delivery

Evaluating requests for demand and providing capacity in a more expedient fashion allows us to deploy a functioning and consumable business service more quickly. We see an opportunity to establish a VM time-to-provisioning goal of less than three hours for 80 percent of our enterprise computing environment.

Supporting this goal requires a high degree of business process automation so that customer-initiated provisioning of IT services does not increase the support burden for operations or impact other customer environments on shared infrastructure.

At the core of our self-service functionality is a hosting automation framework comprised of web services for receiving and responding to service requests, a database to track the status and progress of these requests, a scheduler to help ensure requests are being fulfilled, and an orchestration engine with a set of workflows to complete the tasks. The design goal of this framework is to automate the business processes that handle the majority of IT workflow.

Currently, we have deployed portals for both user self-service and IT measured services (discussed below) to aid in operational decision-making as well as to enable faster provisioning and better management of VMs across both development and production environments. Our strategy has been to release automation in phases to validate functionality and new architecture prior to releasing it universally across our business. In Phase 1, self–service was released only to the virtual development environments and incorporates measured services only for IT Operations. Operations-facing automation allows IT to provision VMs and helps ensure business processes and products are mature before extending these capabilities broadly. Later this year, in Phase 2, we plan to expand self-service to production virtual servers and expose measured services data to end users.

Measured Services: Unlocking Reliable On-Demand Self-Service

As we planned our self-service capability, early concerns included making the infrastructure appear infinite and meeting demands for capacity without significant pre-planning on the part of the consumer. We identified three primary elements—entitlement, quotas, and transparency—critical to achieving our goals. We designed our solution around these elements to help minimize sprawl while still allowing us to service ad-hoc capacity requests within a short timeframe and with minimal human intervention.

GRANTING ENTITLEMENT

We use an entitlement system to help ensure that users have access to data and services appropriate to their job roles. We have extended entitlement to include on-demand capacity requests and can link approval loops and training requirements to various job roles and use cases. For instance, Intel IT policy dictates that server users on the general network regularly complete data handling and classification training. Our entitlement system allows us to validate that users have completed required training and that users' managers have approved access to on-demand self-service infrastructure capacity. The entitlement system detects changes to users' job roles or classifications in Intel's human resources database and flags the need for regular training updates.

ESTABLISHING QUOTAS

Once users are granted entitlement, they are assigned a pre-determined quota of flexible capacity. This services business demand more quickly while helping ensure that resource pools are not drained by a single request. We distribute these quotas by giving each VM a SKU, which maps capacity to quota. For instance, a small SKU is worth half a quota point, a medium is worth one point, and a large is worth three points.

User are granted five quota points by default, which allows them to deploy a basic three-tier application environment consisting of a medium SKU database server, two or three small SKUs for web front-ends, and one or two small SKUs for middleware, with the remaining quota available in the event they need to scale the environment. Quota points are configurable and based upon users' entitlement. While most demands are satisfied with the default quota points, some users will require more. These requests can be handled on an exception basis.

IMPLEMENTING TRANSPARENCY

The third principle to maintaining a cost-efficient and equitable supply-and-demand landscape is exposing resource consumption transparently through measured services. This process is vital to minimizing waste and is a precursor to automated elasticity. Our measured services focus is based on three factors for minimizing waste and helping ensure that SLAs are met:

- **Utilization** compares resource allocation against actual consumption. Our primary focus is on storage, CPU, and memory capacity utilization.
- **Usage** shows whether or not users are taking advantage of server resources. This can be tracked either by the number of web hits or transactions.
- **Health** monitors Quality of Service by showing availability and that we are meeting key SLAs such as time to complete transactions.

Business Process Transformation

Transforming our business process was the biggest barrier to achieving the benefits of our on-demand self-service implementation, with bureaucracy and manual controls impeding change and agility. The number of control points, double-checks, and handoffs between teams created a very cumbersome process, which remained stagnant as business needs evolved (see Figure 2).

As we developed on-demand self-service, we started in areas of the business where there were fewer controls and where processes were less established. We introduced these new capabilities first into our development environment. This allowed us to test leasing, entitlement, and the overall business logic engine, building trust and valuable experience to help ensure that we could run it in our production environment.

Documenting business process, discovering valuable lessons, and repeating best practices allow us to move away from the as-is process to a new paradigm of self-service. Automation is able to fulfill the majority of demand requests, using exception-based logic to trigger a manual review as necessary.

As Is

Manual Web and Database Provisioning Activities

- 12- to 16-day throughput time
- 38 steps
- Manual with workflow automation

To Be

Automated Web and Database Provisioning Activities

- 3-hour throughput time
- 13 steps
- Fully automated

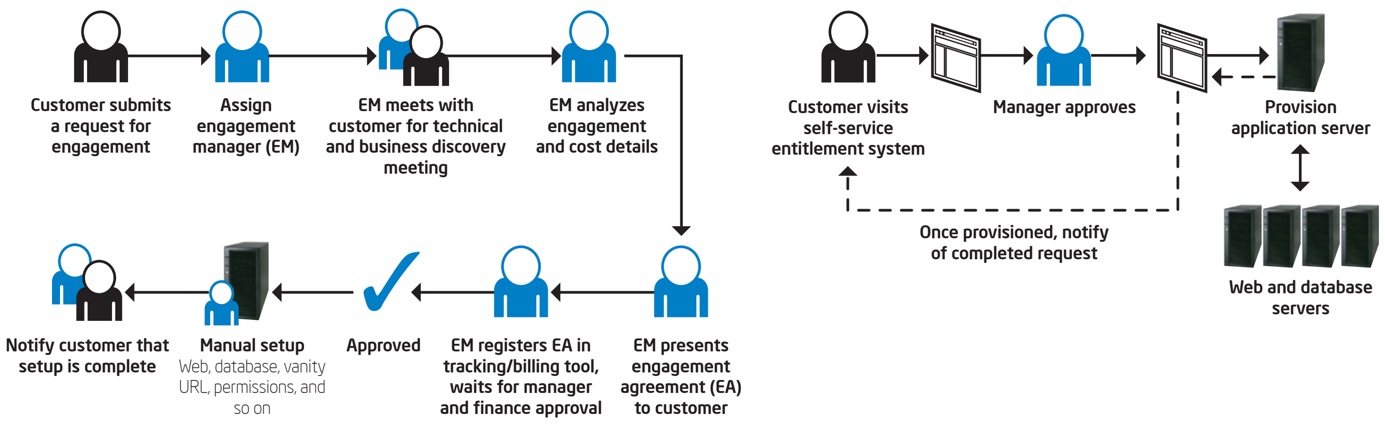
Customer submits a request for engagement
Assign engagement manager (EM)
EM meets with customer for technical and business discovery meeting
EM analyzes engagement and cost details
EM presents engagement agreement (EA) to customer
EM registers EA in tracking/billing tool, waits for manager and finance approval
Approved
Manual setup
Web, database, vanity URL, permissions, and so on
Notify customer that setup is complete
Customer visits self-service entitlement system
Manager approves
Provision application server
Web and database servers
Once provisioned, notify of completed request

Figure 2. Modifying our business processes to enable the complete realization of on-demand self-service required a holistic look at how we work today (as-is) and how we should be working tomorrow (to-be).

Results

By accelerating the virtualization of our infrastructure and deploying our initial self-service portals for infrastructure consumers, we have built a foundation for dramatically reducing the time it takes to provision services.

As we fully automate our service hosting framework into our entire path-to-production environment, we see the ability to provision IT services in only a few hours.

Conclusion

On-demand self-service is a critical aspect of our cloud environment; however, without underlying business logic, controls, and transparency, an unconstrained on-demand enterprise private cloud will quickly exceed its capacity by doling out allocations beyond its supply. By instituting entitlement, quotas, transparent measured services, and data-driven business logic, Intel IT is enabling a true enterprise private cloud providing a consumer-focused self-service portal. This allows us to provide capacity to our business users when they need it and to remove IT from the critical path for launching business services—a key to creating a more agile enterprise infrastructure to support a dynamic and ever-changing set of business requirements.

For more straight talk on current topics from Intel's IT leaders, visit www.intel.com/it.

AUTHORS

Das Kamhout
Cloud Engineering Lead, Intel IT

Greg Bunce
Automation Engineering Lead, Intel IT

Chris Peters
Industry Engagement Manager, Intel IT

This paper is for informational purposes only. THIS DOCUMENT IS PROVIDED "AS IS" WITH NO WARRANTIES WHATSOEVER, INCLUDING ANY WARRANTY OF MERCHANTABILITY, NONINFRINGEMENT, FITNESS FOR ANY PARTICULAR PURPOSE, OR ANY WARRANTY OTHERWISE ARISING OUT OF ANY PROPOSAL, SPECIFICATION OR SAMPLE. Intel disclaims all liability, including liability for infringement of any proprietary rights, relating to use of information in this specification. No license, express or implied, by estoppel or otherwise, to any intellectual property rights is granted herein.

Printed in USA
1010/IPKA/KC/PDF

♻ Please Recycle
324162-001US