Microsoft

Taking the cloud to your datacenter

Microsoft Azure Stack

Version 1.0

1/29/2016

CONTENTS

The Challenge	The Goal	The Solution
Many of the competitive differentiators within organizations (e.g. supply chain, distribution model, customer data, operations, R&D, etc.) have been optimized by IT investments over many years. But these assets sit behind a corporate firewall, may be regulated, and are very proprietary. For many of these assets, participation in the cloud computing revolution remains elusive. All the while, new start-ups without these "assets" accelerate their businesses quickly.	*Enable the power of the cloud computing paradigm across your corporate assets. Unlock the agility that comes from application modernization, regardless of where that application runs. Do this all in a way that doesn't risk lack of functionality or stagnation from a solution 'brokered' between clouds and doesn't requires investment of limited resources in customizing the deployment rather than adding business value.*	*A hybrid cloud platform that is truly consistent with a major public cloud provider unlocks this enterprise value. This solution provides a continuous stream of innovation to enables developers and IT to quickly address business needs without limiting them by location.* *This is the promise of Microsoft Azure Stack.*

CLOUD IS A PARADIGM, NOT A PLACE

As you think of the IT investments throughout your company, you can see that over the course of time they have become inextricably interwoven into the very foundation of your business. Almost every asset in your company, from brick and mortar, to supply chains, to distribution networks, all the way through your personnel and processes, have organically grown with your IT investments. That organic growth pattern has led to tremendous value in the enterprise – and companies that do it well enjoy a strategic advantage over their competition. Most of these important assets are kept behind the corporate firewall, where companies feel they can best control performance, customization, and compliance.

As the realization of the disruption of cloud has spread from the server room to the board room, businesses and technologists alike are looking for how new innovation born of the cloud can be used to modernize existing investments so that they can continue to provide a strategic advantage to their company. Microsoft believes that a well formed modernization strategy begins with a modern application platform. The modern application platform today is being created in hyper-scale clouds. These vendors are setting the pace when it comes to innovation which has resulted in most organizations considering a public cloud as they evaluate modernizing their application portfolio.

Given its strategic significance, application modernization must be seen at an organizational level, not an application by application level. But what of existing investments, assets and business realities? One-off cloud solutions may lead to a modern application, but enterprise wide application modernization requires the coherent and continuously evolving platform similar to those created by hyper-scale cloud providers. And this is fundamentally the problem: existing business assets that cannot currently transition to public cloud risk being alienated from agility and innovation promised by the cloud computing era.

We believe this is not the way things must be and that the next evolution of application modernization for enterprise is enabled by leveraging the value of Microsoft Azure not just in our datacenters, but also in enterprise and service provider datacenters. In this way, you can modernize your portfolio in a way aligned with your existing business realities while not missing out on the best the hyper-scale cloud can offer – knowing that your investments can take advantage of the power of a true hybrid cloud platform.

CLOUD COMPUTING ON YOUR TERMS

The disparate nature of public clouds from one another and from each unique private cloud means that for most organizations, full participation in the cloud paradigm has not been possible. The discrepancies between environments ultimately compromise the promise of innovation and agility that is fundamental to cloud computing. As we work with our customers, we recognize three key areas that led us to see the need for a consistent hybrid cloud platform.

Business and technical considerations	There are important business and technical considerations for organizations that should all be factored when developing an overall cloud strategy. Adoption of public cloud services is constrained in regulated industries. Data sovereignty requirements may prevent information from leaving national borders. Latency between regions can be an issue when serving customers in remote locations. Finally, there can simply be custom requirements that may be considered niche by global public cloud providers, but are vital for an individual organization.
There is a need for new application patterns that enable sound business decisions and flexibility. Hybrid application patterns that run across clouds, taking advantage of the characteristics of both private and public cloud when appropriate. By designing applications to the platform, mobility between clouds becomes a runtime decision based on business rules. In this way, development work can be done once with the original design of being able to run in the public or private cloud environments. Similarly, when applications moves from dev-test to production, the code is the same (all the way to the declarative approach to the infrastructure) and so functionality is the same.	**Application flexibility**
Inadequate alternatives	The next generation of application modernization is coming from the incredible innovation that only a hyper-scale cloud provider can develop. Virtualization alone is certainly not enough. Infrastructure as a Service is only the beginning of the cloud transformation, not the end. However, a vendor who is only interested in a public cloud strategy is not meeting the full set of challenges of their customers as outlined above. By the same token, one-off deployments of highly customized on-premises solutions lead to fragmented and dissimilar technologies – preventing the development of the ecosystem as well as investment in people and processes required to realize the promise of cloud.

We therefore believe that the notion of "cloud-first" needs to be fully enabled across on-premises and public environments. For organizations looking for the agility and innovation of cloud computing in their datacenter, Microsoft Azure Stack offers the only hybrid cloud platform that is truly consistent with a leading public cloud. Only Microsoft can bring proven innovation – including higher level PaaS services – from hyper-scale datacenters to enterprise-scale environments to flexibly meet customers' business requirements. Ultimately, we want organizations to be able to embrace the notion of cloud-first on their terms – every journey to cloud computing is different and we want to support all of them. In this way, we believe customers can realize the value of the cloud paradigm across their organization, regardless of location, and thereby refocus their resources on the investments that competitively differentiate them in the marketplace.

MICROSOFT AZURE STACK VISION

Azure Stack extends the Azure vision by bringing the cloud model of computing to every datacenter. Azure Stack is a new *hybrid cloud platform* product that enables organizations to deliver Azure services from their own datacenter in a way that is consistent with Azure. In this way customers can focus on business innovation rather than spend time building their own cloud computing platform.

Organizations can create these Azure services from datacenter resources - enabling developers and IT professionals to quickly provision and scale services using the same self-service experience found in Azure. The product also allows IT organizations to leverage the same management and automation tools used with Azure to customize the service delivery experience to the business units they serve.

This all adds up to an environment in which application developers can maximize their productivity using a 'write once, deploy to Azure or Azure Stack' approach, because the Azure APIs are consistent regardless of where the resources are provisioned - Azure Stack is simply an extension of Azure. Part of the value of this approach is bringing the large ecosystem of operating systems, frameworks, languages, tools, and applications we are building in Azure to individual datacenters. Developers can create applications based on a variety of technologies such as Windows, Linux, .NET, PHP, Ruby or Java that can be deployed and operated the same way on-premises or in Microsoft Azure datacenters. They are also able to leverage the rich Azure ecosystem of templates, tools, and applications to jumpstart their Azure Stack development and operational efforts.

REINVENTING DATACENTER AGILITY THROUGH CLOUD SERVICE DELIVERY

In order to understand how Azure Stack works, it's important to take a step back and think about the people and processes involved in running and operating a cloud. The cloud era of computing more closely aligns the economic principles of supply and demand with IT operations. There are people, **Cloud Consumers**, such as developers, IT pros, DevOps, who need resources to do their job, building and running business applications, and there are organizations, **Cloud Providers,** who provide those resources to these customers. In Azure, Microsoft acts as the Cloud Provider and anyone who has a subscription to Azure is a Cloud Consumer. Cloud Providers can be any company, as well.

Cloud Consumers need a variety of resources to accomplish their business objectives. Those **Cloud Resources** are provided by **Cloud Services** on demand. For example, some Cloud Consumers just want a virtual machine. This is one kind of resource provided by Azure. Specifically, it is a form of Infrastructure as a Service (IaaS). On the other hand, if a Cloud Consumer does not need an entire virtual machine, but instead just a website resource, this discrete resource can be supplied to them. This would be a type of resource delivered through Platform as a Service (PaaS) – where the infrastructure is abstracted away and the Cloud Consumer need not worry about it. Different types of customers use different kinds of resources and tools to accomplish their goals. Developers, IT, and DevOps users are all welcome customers for Azure. The thing they all have in common is that they need resources and are looking for someone to supply them.

Within the Cloud Provider, there is are two important roles of enabling Cloud Consumers to *request* and access Cloud Resources. The first is the Cloud **Service Administrators**. Cloud Service Administrators both support existing services and identify new Cloud Consumer needs that can be fulfilled through the cloud. For example, we have a group of people in the Azure team responsible for creating new services to meet the needs of Cloud Consumers as well as developing, delivering and managing the Cloud Service once it is deployed. The second role is the Cloud **Infrastructure Administrator** that works to ensure sufficient supply of physical infrastructure to *fulfill* the request from the Cloud Consumer that the Service Administrator has facilitated. Service Administrators and Infrastructure
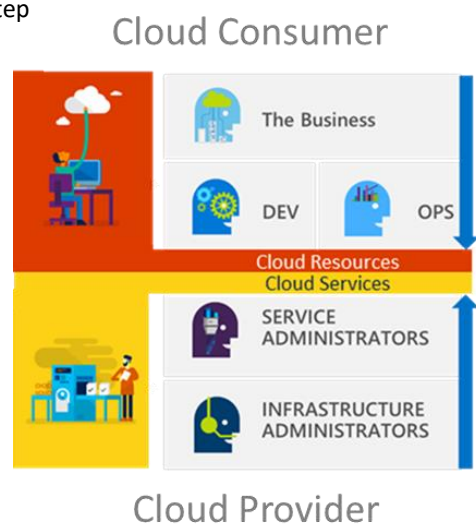
Administrators work together to ensure that there is a very large pool of physical infrastructure so that new Cloud Services developed by the Service Administrators have a place to run and fulfill customer Cloud Resource requests. In offering Azure, Microsoft is responsible for both providing Service Administrators and Infrastructure Administrators.

CLOUD SERVICE DELIVERY WITH AZURE STACK

Since Azure Stack is fundamentally born from Azure, the design challenge was clear. In Azure, Microsoft operates datacenters around the world each with tens of thousands of servers and a minimum "scale unit" of twenty racks. When we talk about 'hyper-scale' in Azure, this is what we mean. But designing services to also run effectively in the enterprise required creating a shared framework that allowed the services to be scaled and operated at 'enterprise-scale' – meaning significantly smaller scale point. The graphic below provides a simplified view of the Azure Stack product architecture.
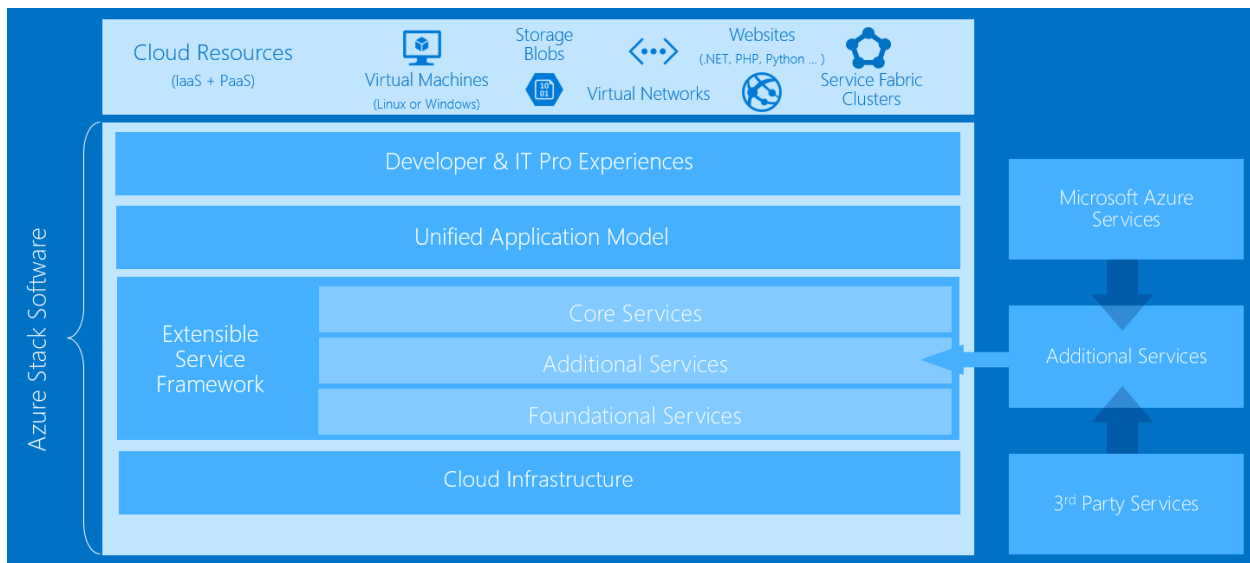
DEVELOPER AND IT PROFESSIONAL EXPERIENCES

Developers and IT pros have an experience with Azure Stack that is consistent to that which they experience in Azure. This is fundamentally made possible because the Azure Stack portal environment is the same code as Azure. However, the real innovation of Azure Stack is the implementation of a consistent cloud API as Azure, so there is a consistent developer experience across clouds. Simply connecting to a portal to choose from preconfigured patterns is not enough; the definition of self-service has evolved to include programmatic access to the cloud API for the creation, deployment and operations of workloads in a cloud.

A consistent API surface area between Azure and Azure Stack is the path to a set of *experiences*, *tools*, *application patterns*, a*utomation capabilities, deployment and configuration,* and *operations* that work across clouds.

- **Experiences**: The first engagement with Azure and Azure Stack usually comes through the portal which provides a web-accessible conduit into the system. The portal is a graphical expression of the cloud API.
- **Tools**: Cloud Consumers can use the tools they use in Azure and know they will work in Azure Stack. Cloud Consumers can focus on solving business problems, rather than constant tooling and deployment transitions.

- **Application Patterns**: Programmatic and abundant access to Cloud Services is changing the way that applications are being designed, developed and operated. You can work with the resources in your application as a group – mixing resources *across* IaaS and PaaS services.
- **Automation Capabilities**: Having a consistent API means that Cloud Consumers can invest in automating development, deployment and operational activities knowing that they will not have to be rewritten to be used with a cloud supplier that offers Azure services.
- **Deployment and Configuration**: Deploy, update or delete all of the resources for your application in a single, coordinated operation. This can be done from the portal or programmatically through the SDK as code.
- **Operations**: Templated deployments work for different environments such as testing, staging and production. Role based access control, usage and audit capabilities are standardized across all cloud resources in the deployment. Updates made to application resources can be performed in an incremental and non-destructive manner.

These are all examples the breadth of impact enabled by this hybrid cloud platform. In each area we believe Azure customers should be confident that their investments in people, processes and technologies will be fully transferable between Azure and Azure Stack.

UNIFIED APPLICATION MODEL

As stated earlier, running and operating a cloud is fundamentally about Cloud Consumers requesting resources and the cloud fulfilling them on-demand. In order to do this, a key component of Azure and, accordingly, Azure Stack is **Azure Resource Manager**. Azure Resource Manager plays two important roles that enable Azure to operate at hyper-scale. The first role is focused on enabling Cloud Consumers to create, organize and control their cloud resources. The second is that it coordinates the fulfillment of that request in Azure datacenters. Each service that a Cloud Consumer sees in the portal is one, or a combination of several, **Resource Providers** (RPs). Azure Compute, for example, is a combination of what we call the Compute Resource Provider (CRP), Storage Resource Provider (SRP) and Network Resource Provider (NRP).

Resource Requests: First, Azure Resource Manager is the cloud endpoint to which your Cloud Consumers will connect. This is true for Azure Stack, just at it is true for Azure. Every installation of Azure Resource Manager is a cloud endpoint and the Azure APIs are the set of commands that can be implemented against that single system. Every command initiated by Cloud Consumers can be organized into groups that correspond to the individual services with which they work. So the service that provides virtual machine resources has commands like "Create VM" or "Delete VM". Similarly, the service to create website resources has commands like "Create Website" or "Scale Website".

Resource Fulfillment: Second, Azure Resource Manager enables customers to describe their resource requirements of the available Azure Services (such as VMs for IaaS or Websites for PaaS), and then *actively* places them into the backend cloud infrastructure through the Resource Providers. As all cloud application resources are really just software (e.g. Virtual Machines, NICs, Databases, etc.), Azure Resource Manager breaks the description apart and manages them individually against the Azure Service to which they correspond, whether the resources reside in our datacenter or in yours. As each resource type is fundamentally unique (databases are very different from virtual machines), each individual Azure Resource Provider is responsible for integration with Azure Resource Manager to achieve the fulfillment of the Cloud Consumer's request.

EXTENSIBLE SERVICE FRAMEWORK

Microsoft Azure is composed of over fifty services today. As Microsoft continues to innovate and release new Azure services, we needed a way to deploy and manage their corresponding Resource Providers throughout our Azure datacenters. A key principle of operating a hyper-scale cloud is constant innovation - new capabilities, new

insights and new customer needs are the norm. Azure is fundamentally designed to enable the release of new innovation on a regular and on-going cadence. With Azure Stack, the framework we use to install and publish new capabilities is the same. The only difference is that we've tuned the size and automation requirements for enterprise-scale vs. hyper-scale. This installation of new Resource Providers and Cloud Services happens in a controlled and predictable manner all without disrupting the underlying infrastructure or the running services and doesn't require changes to the Cloud Consumer-facing experience and API.

CORE SERVICES

There are characteristics of Cloud Resources that are common across all types of IaaS and PaaS services and are essential enterprise features of a hybrid cloud platform. These include business level information, such as a subscription, which contain identity and quota information. Essentially this establishes financial responsibility for billing purposes. Another core service is role-based access control, which handles authorization rules describing what anyone can do with a particular resource. Additionally, all cloud resources need to be "tracked" for how much they are used and who made changes, so usage and audit are common requirements. Azure Resource Manager provides this support for all services, helping the Cloud Consumer focus on the resources they need to do their work, such as virtual machines and databases.

FOUNDATIONAL SERVICES

Foundational Services include both IaaS and PaaS services. These services are important for two reasons. First, they are available to be part of the Cloud Consumer-facing set of services in a given cloud. Second, they can be the backbone to higher order PaaS services. The Foundational Services can create Cloud Resources such as VMs and VM Extensions, VM images, Virtual Networks, Software Load Balancer, Gateway as well as Storage Disks and Blobs. They are also used by Cloud Providers to establish the foundation of additional services as they are layered into Azure Stack.

ADDITIONAL SERVICES

When a Cloud Provider chooses to add a service to their cloud, they are changing the API surface area of the cloud, in effect, growing the capabilities they can offer to their Cloud Consumers. As we move forward, we will grow the list of additional services that can be installed to Azure Stack. It is important to keep in mind that in Azure, in our datacenters, services take dependencies on other services – they layer. So installing an additional service might include also installing the services on which it depends. Each service in Azure is a candidate for being distributed through Azure Stack and we will listen to customer input and consider technical feasibility in determining the roadmap.

It is also worth noting that the extensible service framework serves not only to easily enable the onboarding of incremental Azure services, but it also provides a framework through which 3rd parties can also create resource providers to instantiate completely new services. This is a whole new world of extensibility for ISVs, SIs and Service Providers who want to create new services to offer to their customers.

Below is a table containing the services that we plan to have available when Azure Stack goes to general availability.

Category	Azure Stack Services at GA
Compute	Virtual Machines (including extensions & availability sets), Service Fabric*
Data & Storage	Blobs, Tables, Queues*
Networking	Virtual Network, Load Balancer, VPN Gateway
Mgmt. & Security	Microsoft Azure Portal, Key Vault*
Web & Mobile	App Service (Web Apps, Logic Apps*, Mobile Apps*, API Apps*)
Developer Services	Azure SDK

* PLANNED TO BE IN PREVIEW AT AZURE STACK GENERAL AVAILABILITY

CLOUD INFRASTRUCTURE

As stated earlier, in Azure the Azure Resource Manager works with each service Resource Provider to fulfill the customer request onto the back end cloud infrastructure we run in our datacenters. For the purposes of this paper you can think of the **Cloud Infrastructure** as the underlying physical capacity of the cloud. Every resource requested by a Cloud Consumer and fulfilled by Azure Resource Manager and the Resource Providers, ultimately takes up capacity in the Cloud Infrastructure. As that capacity fills, Microsoft accepts the responsibility for growing the back end capacity. When we grow the footprint of Azure capacity, we think in terms of scale units. For Azure, our minimum scale unit is twenty racks of servers. We enjoy many benefits from that operational model, but one of the most important is the purpose-built nature of the datacenters. They are not general purpose. They serve only one purpose: to provide the backend capacity for Azure. That is the only way to run a hyper-scale cloud. Additionally, in order to operate at hyper-scale, a software-defined approach is required across the basic infrastructure components of Compute, Storage, and Network. A software-defined approach helps free us from the "physics" of scaling. Each component of the Cloud Infrastructure can be automated, provisioned and managed. In this way, we can create a standardized abstraction layer that helps us grow our cloud capacity.

A software-defined approach is equally important when thinking about the Cloud Infrastructure for Azure Stack. For enterprise-scale, planning in terms of twenty rack increments is neither realistic nor desirable. Accordingly, Azure Stack includes Cloud Infrastructure management technologies that are purpose built to supply Azure Service capacity and to do it at enterprise-scale. The foundational IaaS RPs of Compute, Storage and Network (CRP, SRP and NRP) are integrated with their corresponding platform at an infrastructure level: Hyper-V Clusters, Scale out File Server, and Network Controller. The software installed to the physical systems is part of Azure Stack.

DELIVERING CONTINUOUS INNOVATION

Based on all of this context, it should be readily apparent that true cloud computing is fundamentally different than a highly virtualized and automated traditional datacenter. With this understanding it is important to know that in Azure we are continuously innovating and deploying new technologies to our datacenters. The services listed in this paper are indicative of the beginning of a stream of continuous innovation and not a final list. Additionally, in order to maintain the consistent API, Azure Stack operators should expect to implement updates more frequently than with traditional software.

TAKING THE FIRST STEPS WITH AZURE IN YOUR DATACENTER

A key to success with the cloud model are the people and processes around it. Getting ready for Azure Stack may require change from the way you currently administer applications and systems, so it is recommended you begin evaluating your existing approach.

Consuming cloud services is something you can do today in Azure. You can use Azure today to learn what you can do with the services that will be made available with the release of Azure Stack. A great place to start is by learning how Azure Resource Manager works and trying some of the templates available for Azure Stack in both Azure and Azure Stack.

You can also begin prioritizing the services you want to offer from your datacenter and what kind of scale and SLA you will require. Download the single server deployment of Azure Stack and set it up. Once you do, you can accomplish several things.

- First, work with your developers and IT teams to use the Azure SDK to develop and create automations, deployments and configurations that work against Azure and Azure Stack.
- Second, work with your datacenter team to learn and understand the Service Administrator concepts in the single server deployment of Azure Stack (e.g. Plans, Subscriptions, RBAC, etc.).
- Finally, all Service Administrators should begin familiarizing themselves with the Foundational Services in Azure and Azure Stack that they will use to run the additional services.

CONCLUSION

The cloud computing paradigm is a transformational shift in technology that happens only once in a generation. The potential for agility and innovation is clear, but the vast majority of organizations still grapple with questions of how to fully enable cloud computing in their organization … on their terms. When focusing on how to most readily empower developers and IT Pros to do their best work, working within the global ecosystem of a leading public cloud provider is critical. This allows limited resources to be focused on innovation and not simply creating and supporting an isolated private cloud environment. It also allows for a unified, declarative approach to dealing with applications and the resources that power those applications. Microsoft is the only company taking the innovations from their hyper-scale cloud environment and enabling enterprises and service providers to fully participate in that ecosystem with a truly consistent hybrid cloud platform. With Microsoft Azure Stack, organizations can stop asking how their applications and datacenters can move to the cloud and start thinking about how they can also bring the cloud to their datacenter.