

Inside Microsoft Datacenters

All Microsoft Azure services are provided by a globally-networked, highly-resilient group of datacenters that are distributed worldwide. The intent of Azure is to provide global computing power when and where it is needed, hence the distribution of datacenters across the planet to provide the horsepower Microsoft Azure requires.

As cloud services continue to gain momentum, the datacenters that support Azure deliver a robust, solid computing bedrock upon which to build new solutions. For many customers, the esoteric nature of what “cloud” computing is composed of (under the covers) can be brought down to earth by an analysis of the physical plant Microsoft has built and deployed to deliver Azure cloud computing services. These physical resources are complemented by a methodical approach to operations, providing optimized security and robust reliability by relying on standard operating procedures and best-of-breed support and management processes.

The Microsoft approach to building the Azure backbone has followed three key guiding principles:

- Provide global reach to the Azure services, to ensure optimal networking access based on the end-users' geography
- Provide massive computing and networking capacity to enable a new generation of cloud-based services and capabilities
- Use a specific set of criteria (beyond the above) for evaluating the potential locations for locating said datacenters

Note To learn more about the Microsoft Datacenters and see a video tour, go to <http://www.microsoft.com/en-us/server-cloud/cloud-os/global-datacenters.aspx>.

Global reach

The first compelling feature of the Microsoft Azure datacenters is that they are truly global in scale and reach, with more datacenters coming online across the globe on an ongoing basis. The Azure content delivery networks (CDNs) are distributed worldwide. Each CDN essentially acts as a connection point into Azure as a whole—they are the last stop a customer must connect to in order to access Azure services. CDNs are distributed worldwide to afford numerous, global points of connectivity into Azure so that a given end-user of the service is getting better overall performance by accessing resources through a local connection point.

Microsoft considers the global distribution of these CDNs to be a massive competitive advantage for Azure as a whole for this very reason— the resources a given user requires can effectively be co-located in the region that best suits their user base. Not all cloud services can

claim this capability. Again, this is a key tenet—any given administrator using Azure can choose the locations globally that best suit their target user environment. For example, this global reach allows a developer in China to build a service specifically for users in Brazil, or an infrastructure manager in the U.S. to bring a server online to support his business concern in Japan.

Note For more about the Azure regions and their global reach, go to

<http://azure.microsoft.com/en-us/regions/>.

The investment to achieve this level of global infrastructure is currently estimated to be \$15 billion (US). This money has been spent on the physical plant to build these datacenters, the network that connects them, and the resources to maintain and support them. In the end, the global nature of the Azure approach to datacenter build-out gives customers incredible flexibility in where and how to deploy services to support their business, all the while giving them a level of simplicity when it comes to deploying resources – that which once took hours (such as installing an operating system on a new piece of massive computing hardware), now only takes minutes.

More importantly, because Microsoft is maintaining the global infrastructure for Azure customers, the struggles of dealing with traditional hardware procurement and support (maintenance contracts, hardware obsolescence, configuration and upkeep, internal support resources, component failure, etc.) are obviated on a global scale. No longer does a given company have to evaluate the process, tax implications, and so on, of purchasing servers in a foreign country/region; Azure provides a simplified approach to delivering on capacity challenges when and where a given customer needs to take advantage of them.

The tectonic change for traditional IT is that this massive capacity can be exploited and used literally “on-the-fly,” without having to go through additional procurement cycles to bring a server, a service, a development project, etc. online and into production. Customers can now eschew the traditional challenges of procurement (what many would refer to as Capital Expenditure, or CapEx) and instead leverage the Operational Expenditure (OpEx) of using Azure when it is needed. Consider all the challenges of traditional CapEx:

- Procuring physical hardware

- Acquisition and approval processes

- Hardware service contracts

- Vendor hardware quotes

- Order-to-delivery timeline

- Networking logistics

- Physical switching

- Datacenter cabling

- Storage – SAN Capacity, Direct Attached Storage, and networking

Backup/Redundancy
Cooling/AC/Power
UPS management
Shipping/setup
Rack space/floor space (rental or owned)
Fire suppression

Azure datacenters have a global footprint to ensure that these CapEx components (there are many more, literally too many to print) will no longer factor into every single IT decision. Where it makes sense to go through traditional procurement practices and house services on-premises, customers will continue to leverage their current processes. Azure offers a new alternative in streamlining the setup and configuration process dramatically by taking these factors out of the overall scheme. It's up to customers to choose which path is the best for them, based on their needs.

This brings up another good point about Azure datacenters: they are built from the ground up to support Hybrid IT—that is, Microsoft recognizes that some services owned and maintained in traditional, on-premises datacenters will remain there based on customer needs. And while it will make sense for many customers to run key services in Microsoft Azure, not all services will be run in the Azure cloud. Customer choice is paramount in the overall consideration of when and where to take advantage of Microsoft Azure. In the end, the global scale of Azure provides new capability, ultimately delivering computing capacity when and where a customer needs it.

Capacity

As previously stated, the intent of Azure is to provide massive capacity on demand, with regional access to the Azure backbone through distributed CDNs. In the end, customers tap into Azure to leverage this capacity that sits essentially across a typical WAN link. The distributed Azure datacenters provide truly breathtaking capacity – as of this writing, Azure inherently maintains the capacity to support:¹

- More than 1 billion customers and 20 million businesses in 90 global marketplaces
- 200-plus online services delivered 24x7x365
- 15 trillion-plus data objects
- 1.5 million-plus network requests processed per second (on average)
- 50 million-plus users of Office Web Apps
- 50+ billion minutes of connections handled each day through Skype

¹ Microsoft's Cloud Infrastructure - Datacenters and Network Fact Sheet, June 2014

- 400 million+ active accounts on Outlook.com

It is important to note that Microsoft not only taps the Microsoft Azure environment for customers to use for platform as a service (PaaS), infrastructure as a service (IaaS), and software as a service (SaaS) applications, but also to run its own services and its own businesses. In this, Microsoft is its own best example of a company that has cloud-transformed its business. The opportunity for Azure customers is similar – the infrastructure built by Microsoft can be counted on to scale to support massive numbers of users.

There is perhaps no greater testament to the scale of Azure as when it was used to support the Sochi Winter Olympic Games in 2014. In this worldwide event, more than 100 million people tuned into streaming media that broadcasted over 6,000 hours of high-definition content. Streams were broadcast across 22 countries/regions on 4 continents, including 204 individual live streaming channels from the games themselves. Microsoft Azure provided the capacity to meet this demand, serving live streams and more than 500 billion transactions with over 100 TB of stored content.²

Locations

In order to provide the massive computing power that Azure brings to the table, the underlying physical environment must be accordingly massive and a true example of “big computing,” both in terms of how the environment is built as well as how it is operated. The physical underpinnings of Azure are irrelevant if operated haphazardly

The analogy one can easily draw is a massive, super-powerful rocket used in space missions. If not operated and maintained properly, this massive capacity (to send men to the moon) would be wasted.

² <http://blogs.technet.com/b/work/archive/2014/03/06/2014-winter-olympics-behind-the-scenes-with-microsoft-dynamics-and-windows-azure.aspx>