

Accelerating Academic Research with Cloud Computing

Published: September 2014

For the latest information, please see www.microsoft.com/education

Overview.....	1
What Is the Cloud?.....	1
Cloud Service Models	1
Public, Private, Hybrid Clouds.....	3
Cloud Storage.....	4
Microsoft Azure	4
Enabling Next-Generation Research	6
Support for CIOs	6
Research as a Service	7
Microsoft's Commitment to Research	9
Conclusion	9
Resources.....	11

Overview

Universities today face increasing demands and constraints from many directions. Creating a successful academic research environment in this hyper-competitive and requirements-filled world requires powerful, efficient, and responsive systems.

The cloud delivers flexibility, scalability, security, and cost savings, as well as the efficiency, responsiveness, and speed that Principal Investigators and post-doctoral researchers need to achieve an even higher level of productivity and innovation. University Chief Information Officers (CIOs) and others responsible for delivering technologies that enable advanced research and science in higher education also benefit from cloud computing.

What Is the Cloud?

The cloud delivers on-demand computing resources—everything from operating systems to applications to servers to massive storage—over the Internet on a pay-for-use basis. Software and data are hosted on remote computers “in the cloud,” and users connect through the Internet via a browser or program interface from their own devices. Servers are secured, deployed, maintained, and updated by expert staff. Worries about data loss from disasters (natural disasters, flooding, power outages) are greatly reduced because data is automatically and securely protected and stored in the cloud designed for fast data recovery. Universities no longer have to provide and maintain their own costly computing infrastructure on-premises. In both the private and public sectors, the cloud is dramatically changing the economics and time-to-value for IT resources.

Cloud Service Models

The cloud typically has three service models: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS).

Infrastructure as a Service (IaaS). The cloud provides the underlying infrastructure; the university controls the operating systems and deployed applications. IaaS can be thought of as “hardware as a service.”

Platform as a Service (PaaS). The cloud hosts applications or user-created applications and provides the infrastructure. The university controls the deployed applications. Because the hardware and operational aspects of the cloud are managed by the cloud provider, applications can be developed quickly.

Software as a Service (SaaS). The infrastructure (network, servers, operating system, storage) and applications are managed by the cloud provider, which also manages, maintains, and updates the system. Applications run in the cloud. Users access applications from a web browser or a program interface on their devices.

Each different cloud service comes with some overhead to IT. It's useful to compare using on-premises equipment with cloud services from Microsoft Azure:

	ON PREMISES	INFRASTRUCTURE (as a service)	PLATFORM (as a service)	SOFTWARE (as a service)
Applications	Applications	Applications	Applications	Applications
Data	Data	Data	Data	Data
Runtime	Runtime	Runtime	Runtime	Runtime
Middleware	Middleware	Middleware	Middleware	Middleware
O/S	O/S	O/S	O/S	O/S
Virtualization	Virtualization	Virtualization	Virtualization	Virtualization
Servers	Servers	Servers	Servers	Servers
Storage	Storage	Storage	Storage	Storage
Networking	Networking	Networking	Networking	Networking

Compare cloud service models to on premises.

The Microsoft Azure cloud platform provides an open, flexible, global platform that supports multiple programming languages, tools, and frameworks. Not limited to the Windows platform, it also has the capability of including Linux, Python, Java, Hadoop, and the Microsoft .NET Framework, and it can provide backend services to iOS with Azure Mobile Services.

With Microsoft Azure and IaaS, the university manages applications, data, runtime, middleware, and operating systems. Microsoft deploys, maintains, and upgrades the rest, including virtualization, which enables the creation and management of virtual machines (VMs).

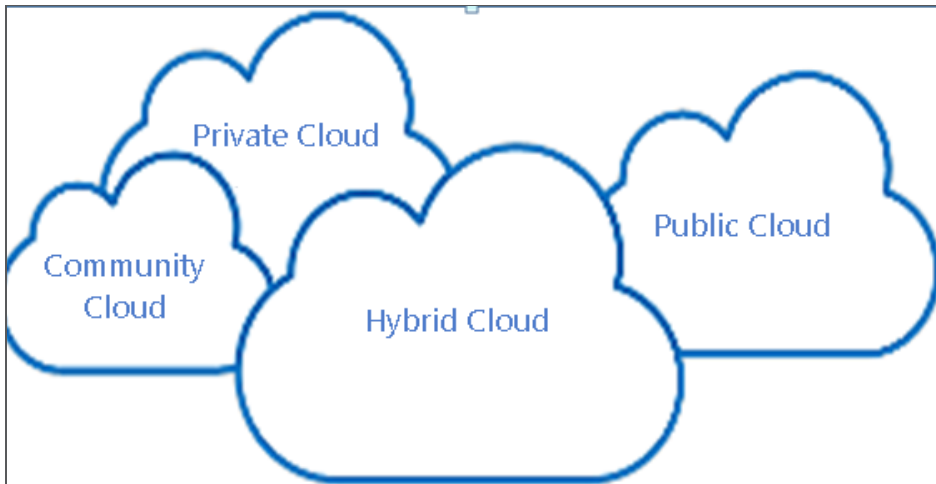
With PaaS, the university manages applications and data, and Microsoft manages the rest. In this case, users may host their own websites optimized to run directly on Microsoft Azure with a VM operating system layer in between (resulting in better performance, less operating system management, and lower-cost computing). The increases in productivity can be considerable, and because the hardware and operational aspects of the cloud platform are also managed by Microsoft, applications can be created and deployed quickly.

With SaaS, users can take full advantage of the cloud, connecting via their browsers or program interfaces, leaving Microsoft to manage all the underlying components.

Real innovation occurs in the algorithms and business logic that researchers develop to analyze their data—that is, at the top, or application, level. When researchers can invest more of their resources at this level, they have a better opportunity of delivering impactful outcomes and innovations from their research. Because Microsoft focuses on the platform, researchers are freed to focus on innovation at the application level.

Public, Private, Hybrid Clouds

A common misconception is that using the cloud is “all or nothing.” Although cloud technology is becoming ubiquitous, Microsoft Azure can be configured in the way that best meets the university’s and researchers’ needs.



Choose the cloud configuration that meets your needs.

Private clouds take advantage of many of cloud’s efficiencies, while providing control of resources and restricting which organizations can use the same servers. Private clouds can exist on- or off-premises for exclusive use by the university (and can be made accessible as appropriate to students, faculty, researchers, staff, and alumni). The infrastructure is owned and operated by a single provider that controls the way virtualized resources and automated services are customized and used.

Community clouds can exist on- or off-premises and are intended for exclusive use by a specific community that has shared concerns. For example, in the U.K., the Ministry of Education (MoE) might operate a private cloud for all its universities.

Public clouds are owned and operated off-premises by companies that use them to offer academic institutions rapid access. With public cloud services, hardware, software, and the supporting infrastructure is owned and managed by providers. This can lower universities’ costs and management overhead, but the university has less control.

Hybrid clouds use a private cloud foundation combined with the strategic use of public cloud services. For example, some universities use private clouds, but manage workloads and store data across data centers, private clouds, and public clouds—thereby creating hybrid clouds.

Cloud Storage

Research data requires vast, reliable, secure storage, and the cloud makes storing massive amounts of data feasible and cost-effective.

On-premises storage with cloud backup stores everything on local machine hard drives at the university (or elsewhere), and backups are stored in the cloud and accessible if ever needed.

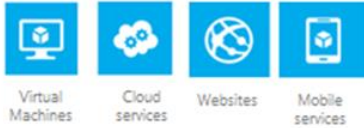
With a hybrid on-premises and cloud data storage solution, critical data can be stored both on premises and on the cloud with Microsoft Azure, satisfying many disaster recovery requirements. Because Azure stores data redundantly (RAID array), data is protected from hardware failure.

With cloud-only storage, it's true that data can be lost if the data center has a disaster. To remedy this, Azure provides additional protection with the offer of geo-redundancy, which means that data is copied and stored in two widely separate data center locations, ensuring that any one disaster won't cause data loss.

Microsoft Azure

The following diagram illustrates the compute, data services, application services, and network services components of Microsoft Azure that enable researchers to build highly available, infinitely scalable applications and to quickly deploy and manage applications and services. Azure manages the deployment details, such as provisioning and load balancing to health monitoring for continuous availability, enabling researchers to focus on the application and not on the infrastructure.

compute services



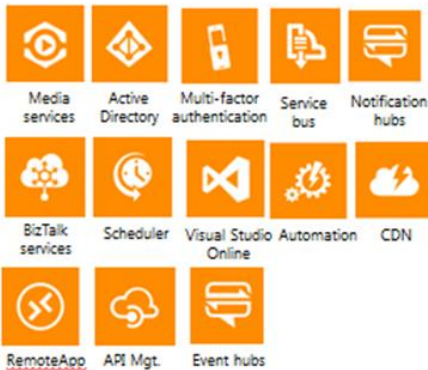
- Deploy a Windows Server or Linux image in the cloud with Virtual Machines (VMs).
- Quickly build, deploy, and manage applications, websites, and mobile applications. Azure Cloud Services manages the server infrastructure for you.

data services



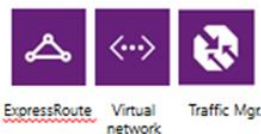
- Store and access your data in a SQL database, or store non-relational data in blob, table, queue, and drive form.
- Back up your data to the cloud.
- Build highly scalable and responsive applications and gain fast access to your cached data.
- Work with data of any type and size.
- Be assured of simple, cost-effective recovery in case of a disaster.
- Easily design, test, operationalize, and manage predictive analytics in the cloud.

app services



- Enjoy cloud-based media solutions, including on-demand and live streaming.
- Manage identities and control access. Use single sign-on capability to simplify user access, or use multi-factor authentication for an additional layer of protection.
- Schedule jobs and manage traffic.
- Automate the creation, deployment, monitoring, and maintenance of resources.
- Deliver high-bandwidth content to users around the world.
- Enable users to work from anywhere on the devices they choose.
- Publish APIs to developers, partners, and others securely and at scale.

network services



- Create private connections between Azure data centers and on-premises or colocated environments.
- Create virtual private networks (VPNs) and link these with your on-premises IT infrastructure.
- Load balance incoming traffic no matter where services are running.

Enabling Next-Generation Research

The cloud is changing higher education and is already enabling leading academic researchers to conduct groundbreaking research faster and more cost-effectively than ever before. Principal Investigators and post-doc researchers are able to provision supercomputing-magnitude resources quickly, access massive data sets, perform compute-intensive analyses, share data for collaboration, and share results for peer review and publishing. The cloud enables them to access massive technical resources rapidly and pay only for what they use.

The cloud is fast, delivering speeds not available before. Servers can be procured and provisioned within minutes instead of days or weeks. Applications can be developed and deployed quickly, dramatically increasing the pace of innovation. The result is that less time and money are needed for technology and more time is available for core research.

The cloud also provides real-time access to a broad set of tools, languages, and frameworks, delivering direct access to virtually unlimited computing resources. And because Microsoft Azure interoperates with Linux, OpenStack, and more, and at the application level with iOS, researchers can work with the tools they're already familiar with and comfortable using.

Support for CIOs

The role of university CIO is evolving from being the gatekeeper of scarce IT resources to the champion enabling leading-edge research by connecting researchers with the best next-generation technologies. Gartner describes this evolving role for both the university CIO and IT organization:

*"Soon to be gone are the days when all solutions and other educational resources or services come from—or through—the institution's or the school district's data center or infrastructure. Time to deployment will decrease and become less dependent on the skill set or availability of IT staff. IT's role will change from development and maintenance to customer relationship and vendor management."*¹

University CIOs are strategic decision makers, shouldering the responsibility for delivering cost-effective computing resources, for securing data, and for meeting compliance and reporting requirements. They also work to advance the university by supporting and promoting vital research that benefits the university as well as society at

¹ Gartner, "Predicts 2014: Technology Drives Education Toward Transformation," Bill Rust, Allen Weiner, Marti Harris, Jan-Martin Lowendahl, Terri-Lynn B. Thayer, 20 November 2013.

large. They must be confident that data and services will be highly available and that data and confidential information will be protected, secure, and remain private. They must think about a wide range of issues—from scaling to security, to applications and virtual machines, to storage, data access, and sharing.

Universities benefit from the economy of scale that Microsoft delivers, with its hundreds of thousands of servers in data center locations designed for efficiency and specifically located where land is inexpensive and power costs are low. Expert data center staff perform upgrades and hardware replacements in bulk and manage deployment, maintenance, and security. CIOs can feel confident about the cloud because Azure delivers 99.95% monthly availability (SLA). No single university can come close to this scale, so costs are potentially lower with the cloud.



Microsoft has more than 100 data centers distributed globally.

The cloud helps CIOs supplement the university's infrastructure with secure, on-demand access to vast storage and supercomputing resources without any need for any additional investment in capital assets or IT infrastructure, freeing up funding for new projects and priorities.

Research as a Service

Researchers are increasingly burdened with demands for reporting and compliance as government and private funding agencies require them to make their data publicly available. Years of using internal infrastructure tools and applications have often made it difficult for researchers to share data beyond the university.

Just as the cloud provides IaaS, PaaS, and SaaS service models, Microsoft adds a fourth cloud service model, **Research as a Service (RaaS)**, which supports researchers in sharing data and meeting reporting requirements.

Researchers report that they can spend up to 90% of their time collecting and organizing data for analysis, including setting up servers and building applications that collect and organize data sets. When they have successfully collected their data sets, researchers then need to correlate data with prior research or to share data with collaborators around the world.

With the cloud, researchers do not need to repetitively build everything from scratch or re-create data collection and data organization solutions. Researchers can plug their data sources into an available prebuilt set of data collection and data organization solutions. Their data sets can then plug into a data repository that supports easy correlation of their findings with other research data sets. Sharing data and making data public is streamlined, advancing research and helping lead to increased innovation and practical application of results.

For example, the Wellcome Trust in Cambridge, England, researching the genetic causes of seven diseases (bipolar disease, coronary artery disease, hypertension, Crohn's disease, rheumatoid arthritis, and Type I and Type II diabetes), uses Microsoft Azure for a genome-wide association study (GWAS). The project involves searching for combinations of genomic information to gain insight into an individual's likelihood to develop one of these diseases.

They need both massive storage and powerful computing capacity for their database containing genetic information from 2,000 people and a shared set of approximately 13,000 controls for each of the seven diseases. They use the power of the cloud to not only store but also to analyze massive amounts of data. Accelerated by Microsoft Azure, they are able to make faster breakthroughs in critical genetic research. For example, using a 27,000-core run on Microsoft Azure, the nodes were busy for 72 hours. During that time, 1 million tasks were completed—equal to approximately 1.9 million compute hours. Had the same process been run on a typical 8-core system, the computation would have taken 25 years.

In this sort of research, it's also essential that even distant relations be accounted for so that associations can be attributed to genes rather than to common ancestry. To overcome this "confounding" factor, Microsoft Research developed the [Factored Spectrally Transformed Linear Mixed Model](#) (FaST-LMM), an algorithm that extends the ability to detect new biological relations by using data several orders of magnitude larger. FaST-LMM enables much larger data sets to be used, which can result in more subtle signals in the data being detected.

The Wellcome Trust has been able to analyze 63,524,915,020 pairs of genetic markers as they search for interactions among these markers for the seven targeted diseases. The new associations they are finding between the genome and these diseases can lead to potential breakthroughs in prevention and treatment.

For more information, see [FaST-LMM and Windows Azure Accelerate Genetics Research](#) and [Supercomputing on Demand with Windows Azure](#), which explain more about how Microsoft Azure is essential to this groundbreaking research.

Microsoft's approach to supporting academic researchers (and the CIOs who support them) is quite simple:

The Goal: Let Researchers Be Researchers...



- Most researchers don't want to be system administrators. They don't want to have to learn how to use supercomputers.
- They want to focus on their research.
- They want to use standard spreadsheets, statistical packages, desktop visualization tools. Programming for them means modifying a few parameters in a trusted scripting language.
- They want to share their research with collaborators.

We help researchers do what they do best.

Microsoft's Commitment to Research

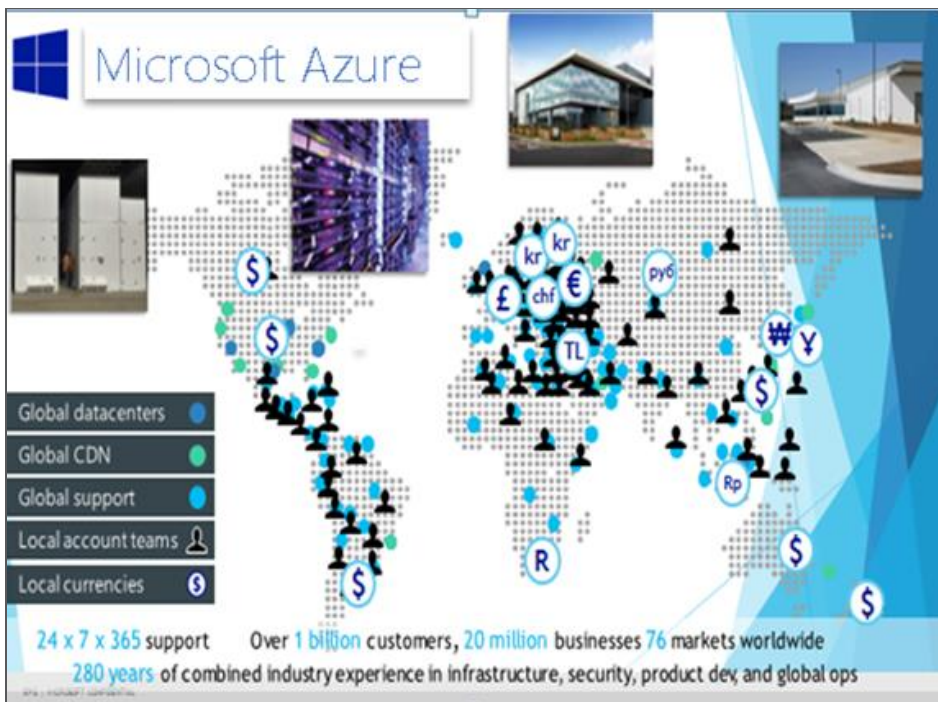
In addition to supporting university researchers with leading-edge technology, Microsoft has also made a huge commitment to research with Microsoft Research, backed by more than \$10.4B in funding (2013). The more than 125 Microsoft Research teams worldwide use Microsoft Azure cloud services in their own research, leading to innovations in computing, breakthroughs in social science research, powerful new products, and more.

Conclusion

Microsoft Azure is the foundation for twenty-first century research and provides:

- A global-scale data center foundation. Data centers are distributed all around the world, and geo-redundancy for critical data and disaster recovery is available.
- Virtualized, fault-tolerant computing.
- Highly secure, highly reliable (24/7x365, 99.95% availability) computing.

- High energy efficiency.
- A scalable platform for web data and analysis services.
- Cost-effective storage, with 60 to 80% reduction in total cost of ownership (TCO).
- Services that support from a few up to millions of concurrent users.
- Pay-by-use cost-effectiveness.
- Simple purchase process—by credit card or P.O., or work with a Microsoft Account Executive to add it to your Campus Licensing Agreement.



Microsoft Azure worldwide

Cloud computing powers research and makes innovations possible. Microsoft Azure enables academic researchers to:

- Use the **tools and technologies** most relevant for them.
- Get **cost-effective access to massive data storage and compute resources**. For example, DNA sequencing analysis research, which has the potential to lead to a wide range of medical breakthroughs, requires supercomputing resources and massive data storage. Working through a grant provided by the National Science Foundation in partnership with Microsoft, a team of computer scientists at Virginia Tech developed a cloud-computing model using the Microsoft Azure HDInsight service. By moving to an on-demand cloud computing model,

researchers have easier, more cost-effective access to DNA sequencing tools and resources, which may lead to even greater and faster advancements in medical research.

- **Have global capabilities and compliance.** Microsoft Azure makes it possible to share public data sets across the globe, which can facilitate large-scale collaboration. In places where infrastructure is poor, having massive data resources available on the cloud is especially valuable.

Many governments have strict privacy and security policy requirements for cloud infrastructure. To meet these various requirements, Microsoft Azure adheres to the highest standards. For example, Microsoft Azure is the only cloud platform that meets EU compliance standards.

- **Confidently store their data.** Because the cloud manages storage, less time and resources are required for storage infrastructure maintenance, and storage can scale seamlessly as needed. Confidential and highly sensitive data is secure and data is protected with faster and proven data recovery capabilities. CIOs and researchers can feel confident, knowing that their data is automatically and secure and protected, and that it can be stored for the long term.

Resources

Training: Microsoft offers free two-day technical training events. [Learn more, or sign up for Training.](#)

Webinars: One-hour webinars help researchers understand the value and capabilities of cloud computing and how they can use it for research in practical ways. [Participate in webinars as they happen, or view them later on demand.](#)

Online training: Videos provide a focused curriculum for scientists interested in using Microsoft Azure in their research. [Learn more.](#)

Technical Papers: Covering a range of topics—including Microsoft Azure virtual machines (Linux and Windows), storage, websites, big data, HPC, Microsoft business intelligence and other tools, and various open source tools—these [technical papers provide in-depth information that helps researchers get a fast start working with Azure.](#)

Microsoft Azure for Research Projects: For the past several years, Microsoft Research has collaborated with researchers to experiment with cloud computing on Microsoft Azure. These pioneering projects have cut across disciplines, from bioinformatics to ecology, social network analysis, civil engineering, mobile computing, natural language processing, and more. [View slides and information about some of these innovative and important projects.](#)

[Microsoft Azure for Research website](#)

The information contained in this document represents the current view of Microsoft Corporation on the issues discussed as of the date of publication. Because Microsoft must respond to changing market conditions, it should not be interpreted to be a commitment on the part of Microsoft, and Microsoft cannot guarantee the accuracy of any information presented after the date of publication.

This white paper is for informational purposes only. Microsoft makes no warranties, express or implied, in this document.

Complying with all applicable copyright laws is the responsibility of the user. Without limiting the rights under copyright, no part of this document may be reproduced, stored in, or introduced into a retrieval system, or transmitted in any form or by any means (electronic, mechanical, photocopying, recording, or otherwise), or for any purpose, without the express written permission of Microsoft Corporation.

Microsoft may have patents, patent applications, trademarks, copyrights, or other intellectual property rights covering subject matter in this document. Except as expressly provided in any written license agreement from Microsoft, the furnishing of this document does not give you any license to these patents, trademarks, copyrights, or other intellectual property.

© 2014 Microsoft Corporation. All rights reserved.

Microsoft, Active Directory, Azure, HDInsight, Visual Studio, Windows, and Windows Azure are either registered trademarks or trademarks of Microsoft Corporation in the United States and/or other countries.

The names of actual companies and products mentioned herein may be the trademarks of their respective owners.



This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](https://creativecommons.org/licenses/by-nc-nd/4.0/).