Accelerate scientific breakthroughs

Microsoft Azure for research

Microsoft

A science-led recovery

As the world recovers from the COVID-19 pandemic, economic disruption is colliding with wicked problems like climate change. To map the way forward in a post-carbon future, we need smart minds working together – and time is of the essence.

Extend your capabilities with ours

Microsoft Azure lets your research harness one of the world's most powerful public clouds for research.

- Infinitely scalable compute resources available globally
- 99.95% monthly availability with 24/7/365 support
- Secure, on-demand access to vast storage and supercomputing resources
- No additional investment in capital assets or IT infrastructure, freeing research funding
- Sustainable zero emissions technology

A new dawn for science

Our planet is at a historic inflection point. By investing in a science-led economic recovery, we can build a stronger and more resilient future for all. With the right strategy to guide us, we can transition faster to a knowledge nation built on big ideas and innovative technologies – one that's ready to tackle wicked problems like climate change, food and energy security, global competitiveness, our aging population and disease.

This strategy must focus on future-proofing our research capability to support excellent research at scale. It must also empower our entire innovation sector to work as one to conceive and commercialize our best ideas.

We have the smarts

The pandemic has shown the talent and the potential of great minds working to solve imminent challenges. Worldwide, scientists worked together with health professionals, policy makers and advocates to make significant progress tackling one of the greatest challenges facing our global public health.

Our health and that of our planet need innovative ideas to propel us to new prosperity in future. But to support their inception, we need to equip our researchers for how science is done now.

How science is done now

Sparking intense collaboration among researchers globally, the coronavirus outbreak accelerated a new research culture characterised by open datasets, rapid-fire exchange of knowledge on preprint servers and an explosion in public papers. Scientists talk of a 'firehose' of information. To contribute to the myriad incremental increases in knowledge that nourish the greatest innovations, our researchers need to surface important findings, analyse massive datasets and share them so they can reproduce each other's experiments. To spearhead a science-led recovery, they need sustainable technology infrastructure that supports open, collaborative science – plus the skills to use it to discover and commercialise valuable ideas faster.

Diving into data

More and more, today's academic research teams are using high-performance computing (HPC), artificial intelligence (AI) and machine learning (ML) to delve into complex questions and power scientific enquiry across huge datasets. Scaling from desktops to supercomputers, they require big data storage, processing and analysis. Many are now using cloud computing to help them meet these needs in a cost-effective way.

The public cloud

The public cloud provides modern research infrastructure to extend your own. Far more flexible than investing in your own technology, it allows researchers to dial up storage and compute when needed and dial down when they're done. With appropriate security to protect intellectual property (IP), academic research teams can share immense datasets with research partners and industry, accelerating game-changing discoveries.

"Modern science is being called upon to solve global challenges at a rapidly increasing rate. The academic model that has driven research outputs in the past is struggling to keep up with the pace of change and a new focus on converting research into action … The scientific community needs to embrace collaborative practises and harness the power of globally connected IT infrastructure to make solutions available when and where they are needed most."

Dr Justin Perry, Research Scientist and Project Leader, CSIRO

Help your research team achieve more using the cloud

Researchers around the world are using Microsoft Azure to accelerate their research. It provides an open, flexible, global platform that supports multiple programming languages, tools, and frameworks to enable faster results.
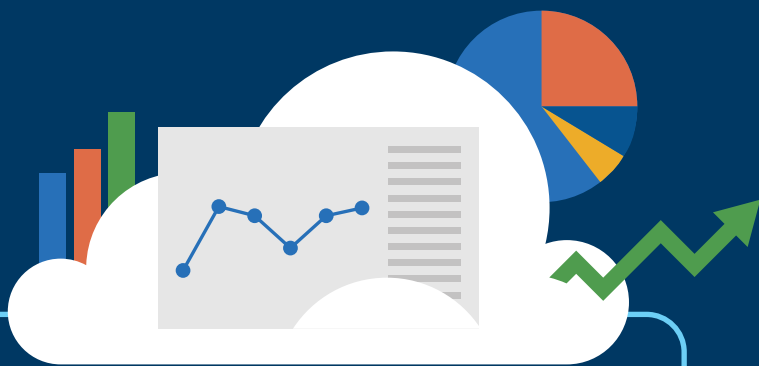
Data management and governance

Make research discoverable, apply data policy and protect university IP with a best practice platform automating governance.

Cost control

Pay for computing resources only when needed, plus use Azure tools to optimise cloud workload costs, reducing total cost of ownership.

Quality control

Students can view project data, upload results and run analytics workloads with role-based access protecting data integrity.

Faster discoveries

Work easily with huge datasets and use AI, HPC and Azure discipline-specific tools to arrive at new insight faster.

Virtual research teams

Share and manage code across research communities using GitHub repositories. Integrate IoT for remote sensing data and enable hybrid working for your team in a crisis.
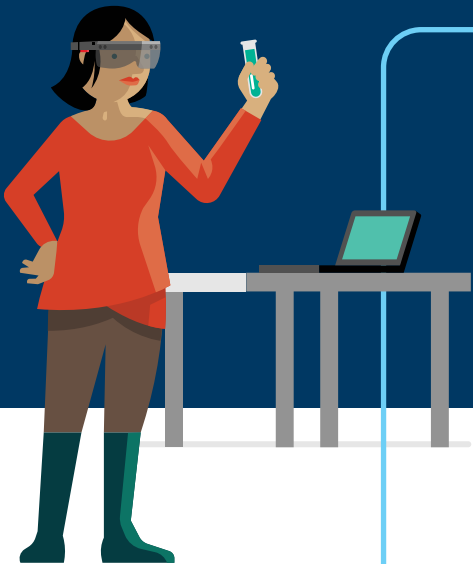
Best practice code

Software engineers and researchers can use Azure's best-practice coding toolsets to create custom research software to solve new scientific challenges.

Solve wicked problems at speed

To find answers to big challenges like climate change and pandemics, scientists in data-intensive fields like climatology and epidemiology are interrogating immense datasets to reveal the complex interplay of factors driving natural phenomena and disease. The Azure platform helps them turn evidence to conclusions more quickly.

Faster discoveries

Solve complex problems with big data

Make headway in data-intensive fields like astrophysics, geophysics and genomics. In minutes, researchers can set up cost-effective Azure infrastructure to ingest, store and share their biggest datasets. It's easy to add computing power to further investigate data using modelling, simulations and analysis. Or to import Azure Open Datasets to train ML models using public data on weather, city safety and more.

Build on what came before

With thousands of new findings coming out of the worldwide scientific community daily, Azure AI can help researchers digest the 'firehose' of information and stay on top of their fields by surfacing relevant finds and data in scientific literature. And to help researchers reproduce findings and take the next step, Azure makes information more discoverable with best-practice governance and collaboration.

See things differently

Arrive at new insight with big data tools built for the Azure platform and available on demand, including HPC, AI, ML and data analytics. Azure helps your team adapt code to the world of big data and the cloud, plus use our intelligent tools to reveal new patterns, relationships and trends in research data. There's no need for capital outlay – simply switch the tools you need on and off to suit research timelines.

Virtual research teams

Hive minds

Azure lets everyone working on a research challenge access the latest data so you can collaborate across entire disciplines and industries. Data integrity is protected by access controls with multiple layers of security and governance, so you can co-create knowledge with global experts and citizen scientists alike. On BRUVnet, they're populating the world's largest training dataset of fish species this way. With Trove, you can pay people for the photos you need to train ML models.

Easily share ideas, methodology and research

Startling innovations can result when researchers connect data points and insights across existing knowledge and scrutinise and reproduce each other's work. Azure enables this with an open, flexible, fast, efficient infrastructure for sharing large datasets and findings. With research data at their fingertips, researchers can collaborate easily, plus replicate experiment results on other researchers' datasets to strengthen the foundation for future enquiry.

Resilient in a crisis

The COVID-19 pandemic has underscored the importance of resilient research infrastructure that supports flexible work styles, including remote working and hybrid working. With Azure, you can keep researchers connected and productive wherever they are. You can easily integrate Teams, SharePoint and OneDrive so they can disseminate their results and communicate more effectively. Power BI helps them manage shifting priorities with up-to-the-minute reporting tools.
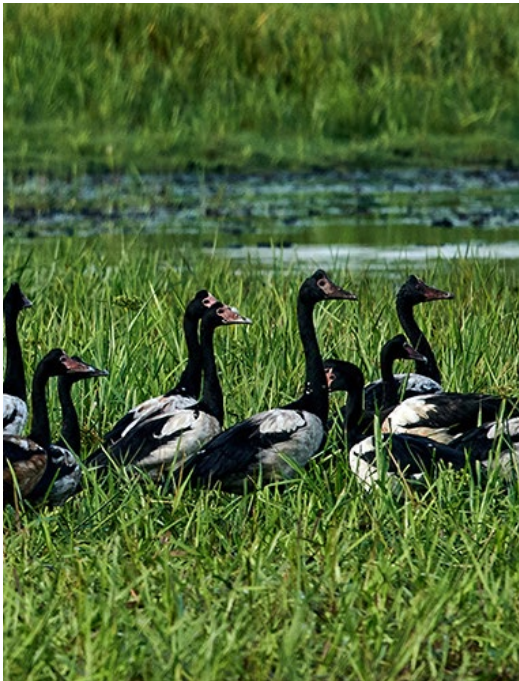
CASE STUDY
Prehistoric DNA

Analysing blood from around 300 Asia-Pacific islanders led a global research team to an amazing discovery: two previously unknown groups of prehistoric ancestors. Without Azure powering their statistical model, the team might still be waiting for compute resources. Instead, they're using Azure to dig deeper into the data to understand parasite resistance and disease and find treatments for islanders with autoimmune disorders.

"Science moves fast, and the questions would have been addressed by others if we hadn't got there at the speed we needed."

Professor Murray Cox, Computational Biologist
Massey University, New Zealand

CASE STUDY
Healthy country

In Kakadu National Park, drones are helping scientists and traditional owners control invasive species and bring back magpie geese, a sign of healthy country. In 2018, only 50 magpie geese were counted – nine months later, more than 1800. Combining Indigenous knowledge, Microsoft AI, data visualisation and scientific research, drone video footage data is collected, labelled and interpreted to guide environmental management.

"This is a unique project. We've allowed AI to sit next to Indigenous knowledge, to inform adaptive approaches to caring for this land."

Dr Cathy Robinson, Principal Research Scientist and Group Leader, CSIRO

Researchers

With Azure, it's easier for researchers to find patterns, test ideas and surface their data and results easily to collaborators globally. By providing access to all research data on the Azure platform, you can improve collaboration within your research community – and with research partners everywhere.

Andrea is an agritech researcher at a university facility. She's investigating how microorganisms in soil affect the productivity of crops and livestock and farm health. She's part of a collaborative project that brings together researchers from all over the world to improve global food security through sustainable farming.

Andrea is using drones and sensors to collect data from a network of farms all over the world to help her build a more complete picture of the factors that impact soil microbiology. She's particularly interested in what native animals and plants are beneficial to soil culture and the effects of different crops and livestock.

By combining data from her IoT devices with global agricultural datasets from Azure FarmBeats, she's training an Azure AI to help her dig more deeply into soil health and productivity.

9am – Analyse

Azure Logic Apps alerts Andrea that her drones, sensors and other IoT devices have sent new data from her study group of farms. It's time to run her data against Azure FarmBeats datasets to surface new insights. She starts her compute-heavy data analysis workload on Azure Batch. No need to hang around – her compute resources will scale up and shut down automatically – so she goes to check her soil samples in the wet lab. Back at her desk, she finds her data has already been sent to Azure Blob Storage, optimising storage cost for her large dataset. Using PowerBI to quickly visualise her results, she spots interesting correlations between soil microorganisms and native fauna. She shares the PowerBI dashboard with her research group on Teams.
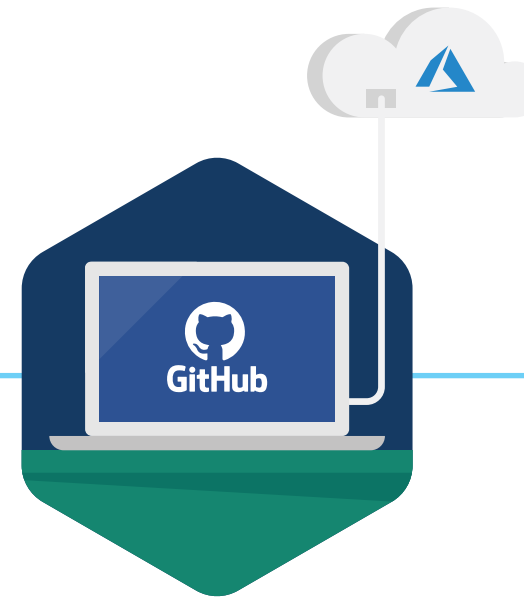
11am – Collaborate

Abroad at a conference, Andrea's supervisor is excited about her recent results and calls on Teams to review the data together. He suggests sharing the new dataset and results with international collaborators. With Azure Data Share, Andrea does so immediately and securely. To protect her unpublished research, she specifies conditional access to prevent collaborators from sharing her data further. The system records their consent before allowing access. They join Andrea in a Teams call to review her findings and contribute their own insights. During the call, they share notes and ideas in OneNote. Later, they'll write up their findings together with realtime co-authoring in Word.

1pm – Model

Andrea uses Azure ML to build a new model to better predict soil microorganisms based on the presence of wild flora and fauna, crops, livestock, fertilisers, pesticides, water and other variables. Azure ML leverages the latest graphics processing unit (GPU) resources, automates compute allocation and shutdown, and meticulously records all parameters to enable research reproducibility.

4pm – Learn

Andrea attends a Teams seminar hosting a renowned agritech researcher from another institution. She records it so colleagues who are busy in the lab can watch it later on Stream. By searching for their model organism name in the automatically generated video transcript, they can skip to relevant sections, saving time.

5pm – Build knowledge

- Good news! Her paper was accepted by a prestigious publication, so she shares her analysis code on GitHub and her data is stored securely in the cloud.

- She invites citizen scientists to tag her drone footage and submit images of wildlife and plants on Trove to expand her AI training dataset. Her aim is to train Azure AI to recognise more native flora and fauna to reveal how they impact microorganisms beneficial to farm health.

Sharp minds need powerful technology to solve wicked problems

Now even the smallest research teams have the computing firepower to investigate science's biggest questions. On Azure, there's no capital outlay, no on-premise hardware to maintain, and data controls and governance are built in. Simply pay for what you use – and scale up and down as needed. Plus spark new ways of seeing data with HPC, ML and other tools on demand.

Data management and governance

Manage rapid data growth easily

Scale your Azure research infrastructure easily to keep pace with scientific breakthroughs and collaboration opportunities. You can add on-demand storage and backup to suit any research requirement – all with best-practice security and governance. Data tags enable fine-grained life cycle management, curtailing data and budget bloat.

Make research findings discoverable

Data saved on hard drives and thumb drives doesn't fuel the future of collaborative science – and it's vulnerable to loss, theft and corruption. Azure provides a universal platform to ingest and share every researcher's data. Policy-based tags impose a common logic, helping to surface data to the right scientific and industry collaborators.

Protect your IP

There's a misconception that the cloud isn't for sensitive data. In fact, it can help prevent catastrophic loss and data leaks resulting from compromised university servers. Security and sound governance are built into Azure and constantly updated. Fine-grained access controls protect data safety, integrity and reproducibility – and your IP.

Cost control

Scale out to a supercomputer

Empower researchers to take control of their technology budgets and eliminate waits for computing resources. Azure offers a cost-effective resource model with no capital outlay. You can start with what you need today and scale out to a supercomputer tomorrow. Switch on HPC, analytics, AI and ML instantly, plus access discipline-specific tools like Cromwell or FarmBeats on demand to accelerate insight.

Reduce technical overhead

It makes no sense to duplicate or under-utilise expensive equipment. Why not pay only for the time you need? With Azure, you can make the most of scarce funding by reducing on-premise hardware and, with it, total cost of ownership. You can also maximise the value of your infrastructure investment by storing operational, compliance, production and research data on the same Azure system.

Simplify your IT

Boost productivity with a cloud solution that's built to integrate with you. Azure automation makes it easy to ingest, manage and analyse data, saving time. Alerts help control cloud workload costs so researchers stay on budget. And by outsourcing technology maintenance, upgrades and IT resources, your IT team can deliver more value by working on your institution's strategic priorities instead.

CASE STUDY

Democratising data

The University of New South Wales (UNSW) can respond faster to challenges like the COVID-19 pandemic with Azure providing a single source of truth for reporting. UNSW used Azure Databricks to create a second Azure data lake to store curated data. Workers can create their own reports from this data in record time using Power BI. UNSW is now exploring using AI and ML to analyse the data to identify contract cheating and more.

"We want to identify what are the real issues that are impacting on student experience and student performance, and not mere correlates."

Kate Carruthers, Chief Data and Insights Officer, University of NSW

CASE STUDY

Azure, asteroids and AI

John Hefele used Azure Batch to spin up 200 virtual machines in about 15 minutes. He ultimately scaled to 500 VMs. He was able to generate a million asteroid trajectories in just nine hours – a task requiring at least three weeks on the university's supercomputer, a shared resource that's hard to access. He used the data to train a neural network that identified 11 asteroids that could impact Earth and weren't yet flagged as a threat.

"[People] assume that cloud computing is going to be really expensive, but I was able to generate a million simulations using Azure Batch for just $150."

John Hefele, PhD candidate, Leiden University

University systems managers

Azure enables research and data managers to keep up with ever expanding data demands from thousands of researchers, all with different needs.

Walter manages a university research facility whose genomic researchers have joined the race to find effective treatments for the coronavirus. Essential to this work is collaborating with multidisciplinary international teams, including clinicians and biotechnology companies.

Walter oversees the technology infrastructure that research teams need to test their ideas, analyse data and collaborate with peers and industry globally. A big part of this is helping them share immense datasets – everything from genomic, clinical and pharmacological data. This includes applying university data governance policy to outside datasets.

Azure helps Walter do all this and more. With highly scalable storage and compute, he can resource the team's many trials, experiments and studies on demand and in a cost-effective way. And it's easier to meet the specific needs of each research discipline because of Azure's advanced tools, which help data scientists tailor data analysis technologies to the research data, instead of the other way around.

Safeguarding the university's IP and important research data is also critical. Azure's security and access controls help to protect valuable and sensitive data. Governance is built in, helping Walter preserve valuable research data so it's reproducible years from now.

8:00am – Onboard

Walter onboards a new international clinical team. Using Azure Role Based Access Control (RBAC), he assigns permissions to collaborators by function/organisation. This protects sensitive patient and clinical data and the team's IP. He integrates the clinical team's patient data into the university cloud, including from thumb drives, mobile devices, desktops and local servers. He tags it to comply with university governance and to stop data silos forming. Now all collaborators can access project datasets on the same system with consistent security, governance and data integrity protections.
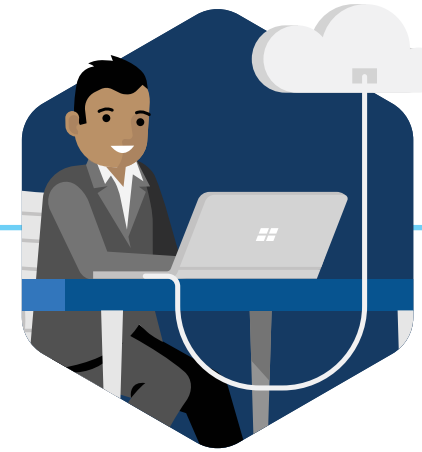
11:20am – Commercialise

- Walter receives a Teams message from an affiliate researcher working on new treatments for the coronavirus. She wants to replicate the clinical team's results on different and associated datasets.
- Walter works with IT to validate her identity. Once confirmed, he upgrades her access level so she can download the data.
- He also responds to a university IP lawyer to confirm that a new treatment that a pharmaceutical company wants to commercialise was pioneered by university researchers. He provides a link to the research data as proof.
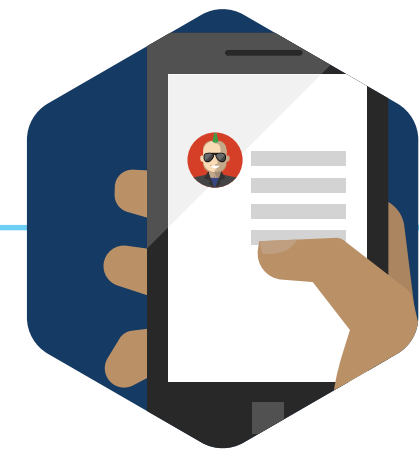
12:30pm – Optimise

- Using Azure Monitor, Walter analyses infrastructure performance. Azure Blob storage is close to capacity, so he deduplicates and compresses the team's legacy data and auto-archives all data not recently accessed in the university's Azure Data Lake.
- To further optimise storage, Walter asks researchers to review Azure Advisor's automated recommendations. They agree to archive several large datasets ahead of time, but want to retain access to an older dataset that's relevant to new research. Walter updates the policy.
- He also reduces time-to-archive by two weeks for infrequently accessed data, freeing more space. The team now has storage head room and will stay on budget.
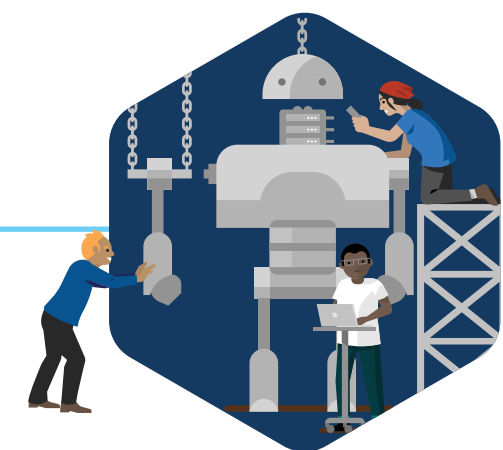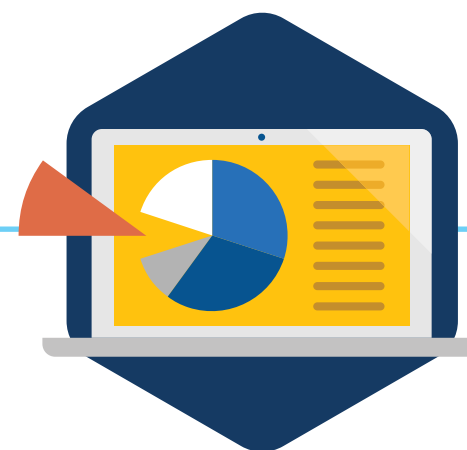
4:00pm – Provision

- A lead researcher needs more storage to spin up a data-intensive simulation. Walter decides how much to make available and lifts her limit on the Azure portal.
- Another researcher needs one-off infrastructure to run data workloads for a six-week vaccine trial. Walter checks the team's budget and messages back with costs. On go-ahead, he provisions the resources instantly. The trial can now start.
- When data processing is complete, the resources will be turned off and simulation and trial data transferred automatically to Azure Blob for permanent storage.

5:30am – Report

Walter uses Azure Cost Management + Billing to analyse usage by research team. To keep everyone on budget, he checks against previous periods to identify excess spend and switches off any underused resources. As the coronavirus team ramps up, they'll need more storage and compute, so he increases their spending threshold.

Advanced scientific tools in easy reach

Now even the smallest research teams can use AI, HPC and discipline-specific tools created for the Azure platform. No huge overheads – simply pay for what you use.

Perform large-scale parallel and HPC batch jobs efficiently

Azure Batch and Batch Shipyard

Run large-scale applications efficiently in the cloud with cloud-scale job scheduling and compute management. Docker and Singularity containerised workflows don't require infrastructure setup. Add Azure Kubernetes Service (AKS) to help deploy and manage containerised applications.

## Build, train and deploy AI models collaboratively at scale

Azure Databricks

Work at scale with AI models using GPU-enabled clusters with this Apache Spark-based analytics service. Databricks Runtime comes preinstalled and preconfigured with deep learning frameworks and libraries e.g. TensorFlow, Keras and XGBoost.

Integrate data from field devices or leverage AI in smart farming

Azure IoT, including FarmBeats

Securely connect, monitor and manage billions of devices to develop Internet of Things (IoT) applications. Easily integrate meteorological, satellite and sensor data plus other IoT feeds.
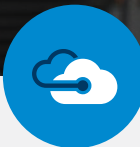
Analyse large genomic datasets

Cromwell on Azure

Orchestrate the computing tasks needed for genomic analysis with an open source scientific workflow management system that lets you run scripts on local machines or computing clusters plus the cloud.

Manage HPC clusters on the cloud cost effectively

CycleCloud

Dynamically provision HPC Azure clusters and orchestrate data and jobs for hybrid and cloud workflows. Compatible with Slum, Kafka, Zookeeper, PBS, Anaconda and more.

Simplify data science in the cloud

Data Science Virtual Machines

Get started quickly with Azure Virtual Machine images (Linux/Windows), pre-installed, configured and tested with tools widely used in data analytics, machine learning and AI training.

Achieve real high performance computing (HPC) on a public cloud

InfiniBand-enabled HPC

Enable MPI and RDMA workloads to run rapidly at scale with compute and GPU resources connected over a low-latency, high-throughput network.
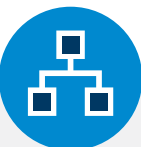
Build, train and deploy ML algorithms and enable reproducible ML research

Azure Machine Learning

Whatever your level of AI expertise, get up and running quickly with everything from automated machine learning (AutoML) to low-code drag-and-drop tools (Azure ML designer), to full code with Jupyter Notebooks. Azure ML automates deployment and downscale of compute clusters.

## Access the latest research

Microsoft Academic Graph (MAG)

Synchronise MAG to your Azure Storage accounts and see your searches in a heterogeneous knowledge graph – find scientific publication records and see citation relationships, authors, institutions, journals, conferences and fields of study.

Manage expanding data efficiently with a scalable system

Azure Storage (Blob, Disks, Data Lake)

Azure Blob storage provides massively scalable, secure storage for your unstructured data. For critical workloads, disk storage is dynamically scalable. Extend your current storage systems easily by integrating an Azure Data Lake.
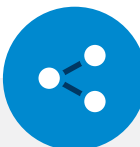
Share large datasets with research partners

Azure Data Share, Azure Industry Collaborative Service

Share research datasets with multiple collaborators with fine-grained controls. Enhance insights by easily combining data from third parties to enrich analytics and AI scenarios.

Manage low latency for realtime inference or model scoring requests

Field Programmable Gate Arrays (FPGAs)

Azure FPGAs enable ultra-low latency inferencing, even with a single batch size, using ResNet 50, ResNet 152, DenseNet-121, VGG-16.

Use off-the-shelf capabilities to train ML algorithms and keep AI costs down

Azure Cognitive Services

Tailor your AI capabilities to your specific needs using custom-developed models based on our broad range of AI services.

Make research reproducible

Azure Open Datasets

Improve the accuracy of your ML models and save time on data discovery and preparation by integrating ready-to-use public datasets straight from Azure services into your own ML workflows.