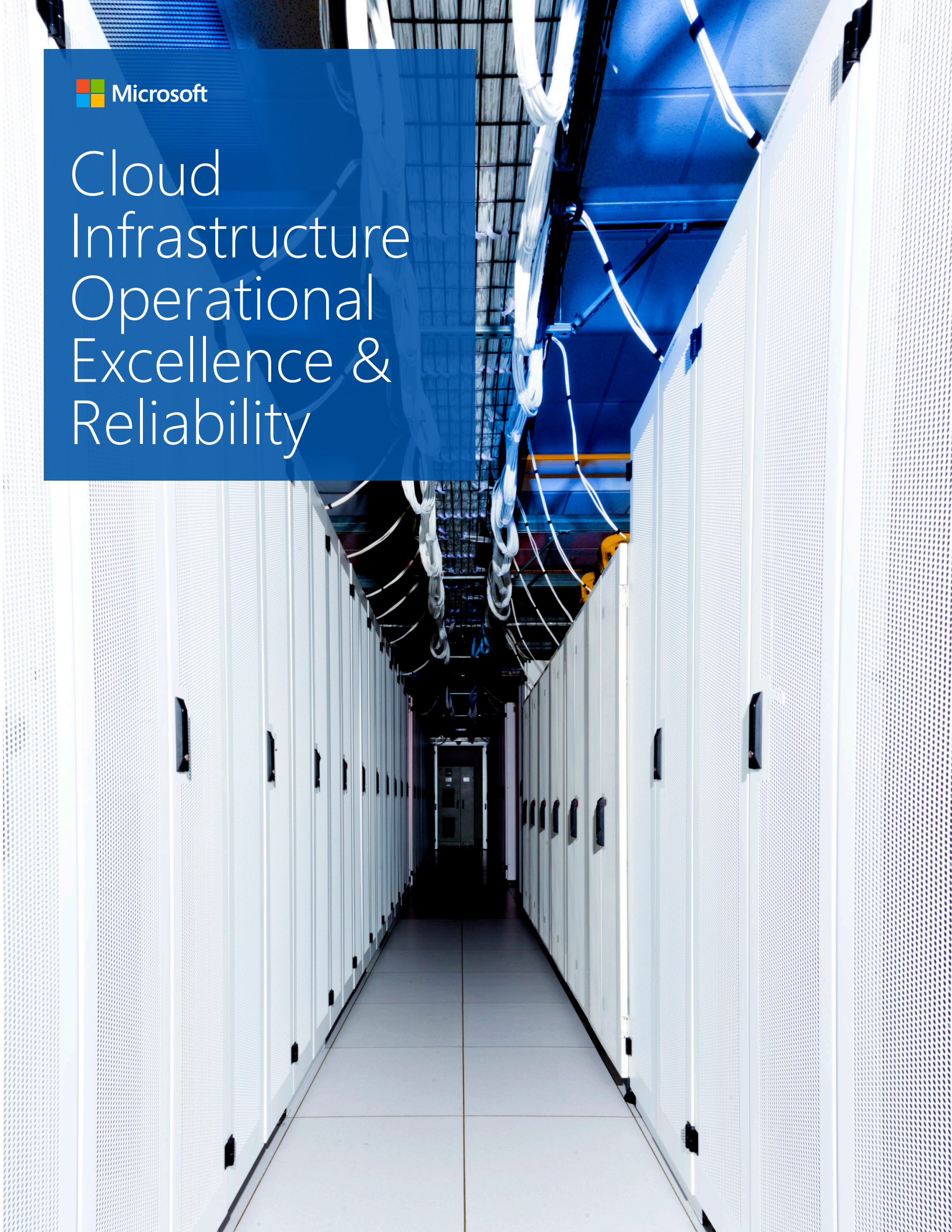
Microsoft

Cloud Infrastructure Operational Excellence & Reliability



Operational Excellence & Reliability

Microsoft has invested over $15 billion in building one of the world's largest global cloud infrastructures. Today, our cloud hosts over 200 cloud services, including Bing, Office 365, OneDrive, Microsoft Azure, Skype, and Xbox Live.

In this strategy brief, you will learn how Microsoft ensures high reliability through robust processes for incident management, service support, security and compliance, and change management—with speed, efficiency, and trust.

Operating at hyper-scale

Microsoft's cloud is one of the largest in the world and powers over 200 services for our customers and businesses 24x7x365 via an infrastructure of globally distributed datacenters, networks, and applications and tools.

Microsoft has been managing online service delivery at a global scale since the launch of our first cloud offering in 1995 – MSN. Today our cloud infrastructure supports more than one billion customers in 90 global marketplaces. With an investment of more than $9.5 billion in Research & Development, we continue to evolve our cloud infrastructure, while continually improving the operational processes to maximize reliability, performance, and efficiency.

The foundation of our cloud infrastructure operational management comes down to two key principles—keeping the sites up and costs down for our customers.

Microsoft's cloud strategy

When we look at our cloud strategy there are three differentiating factors that make Microsoft unique from the competition:

Enterprise grade – Microsoft has a long history of working with enterprise customers. We have a deep understanding of the capabilities and requirements enterprises require, and our infrastructure and services are designed to meet their needs for security, compliance, interoperability, reliability, and performance.

Hybrid cloud – Customers who adopt cloud services will retain some workloads on premises, but require consistent capabilities, a consistent management experience, and secure connectivity. We designed the Microsoft Azure cloud platform to seamlessly span from on-premises to the public cloud, allowing customers to run their services where and how they prefer.

Hyper-scale – With more than 100 datacenters worldwide, we have datacenters located in every region, connected by one of the largest cloud networks in the world. There are only a few cloud providers who are individually enterprise grade, hybrid, or hyper-scale, but Microsoft's cloud is the only one that offers all three.

Software-defined reliability

No matter how well engineered the physical infrastructure, service incidents do occur. We accept the three laws of operations – servers will fail, software will have bugs, and humans will make errors. At Microsoft, we employ sophisticated software-defined service instrumentation and monitoring that integrates at the component or server level, the datacenter edge, our network backbone, Internet exchange sites, and at the real or simulated user level, providing visibility when a service disruption is occurring and pinpointing its cause.

More importantly, we are continuously investing in developing greater application resiliency in our software so it will instantly recognize a disruption and gracefully fail over to a different set of servers or even a different datacenter, without interrupting the availability of the service.

Driving costs down

To respond with the right resource allocations while optimizing operating costs, Microsoft has moved away from the standard industry practice of allocating costs by use of space to a more relevant measure of allocating against key cost drivers. These drivers include energy consumption, bandwidth consumption, incident response, and server capacity to provide transparent and accurate billing across our resource pool.

This comprehensive billing approach helps our cloud businesses make the right decisions necessary for their services and allows the datacenter operations teams to optimize the placement of compute and bandwidth capacity with appropriate scale.

Microsoft also uses our own proprietary tool that tracks power usage, server utilization, and carbon output, along with the environmental conditions inside and outside the datacenter. This tool enables us to provide real time monitoring of all workloads and to manage costs and resources in a fully rationalized manner.

Migrating from enterprise to hyper-scale

From our early datacenters in 1989 until 2008, we followed the traditional enterprise IT approach to operations—to deliver high hardware availability through multiple levels of redundancy and maintain an active/back-up workload profile. In the enterprise IT model, we would have one active workload, with multiple back-ups in the datacenter or within the network for that service. If an active node were disrupted, the operators could take the back-up, copy it over, and bring the new workload node online. In this situation, the service could experience an outage.

As we move to a hyper-scale, software resilient design, our workloads become geo-redundant or what we call multi-mastered, where we have many active nodes in the network and software to automate the failover. In this situation, if we happen to lose an active node or we lose an entire datacenter on that network, our customers can still find that service on the network. There may be some latency, but the service will still be available. This approach allows us to remove redundant hardware provisioning, saving both capital and operating costs.

In a traditional enterprise IT model where a company may have an active and back-up workload, they could face a situation where hours would pass before they could restore their service. In a hyper-scale, software-defined model, when we lose an active node, the time the service can be restored is seconds.

Software-defined networks

To ensure fast and reliable connectivity to cloud services and data, Microsoft combines globally distributed datacenters and edge computing nodes with our network, which is one of the world's largest fiber backbones that provides multiple terabit connectivity with more than 70 points of presence.

This network provides multiple paths to many providers, allowing instantaneous re-routes around internet failures to maintain high reliability. It also maintains sufficient capacity to handle large-scale network interruptions without degradation of performance.

To improve flexibility and accelerate the adoption of advanced technologies into our network, we have broadly adopted a software-defined networking (SDN).

SDN provides the ability to dynamically modify our network using automated management tools to move data and resources to an area where it is best served. In an SDN environment, we are able to extract and separate the application, the control plane, and the transport of the data. This allows us to insert our own APIs to gain visibility of how the data flows and gain better control, and allows us to upgrade network performance outside of the hardware refresh cycle.

Our large, geographically distributed footprint of datacenters and networks enables us to be located close to our customers to reduce network latency and allow for geo-redundant back-up and failover.

Microsoft's software-defined network overview

Geo-Redundant Service/ Application Design	• All nodes active
Top 2 Most Connected Networks in the World	• Peer with over 3000 ISPs globally
DC-to-Internet Backbone	• Multiple terabits • Over 70 points of presence globally • Global backbone connecting MS Datacenter to the Internet
DC-to-DC Backbone	• Multiple Terabits of capacity • Dark fiber based DC-DC backbone to enable high bandwidth between datacenters
Dark Fiber	• 1.4 million+ route kilometers of 10G and 100G DWDM capacity deployed • Backbone consists of over 100,000 route miles of fiber
Cache Node	• Hosting Services collocated at user location (metro)
Edge Nodes	• Multiple Terabits of Edge Interconnect capacity • Interconnect over 1600 unique networks around the globe
Decoupled DCs	• Separation of CPU's Storage, SQL Services
IT Capacity Unit	• DC Capacity Unit or Workload Appliance

Microsoft's Open CloudServer

With more than one million servers hosted in over 100 datacenters, a major initiative involves designing our own server specifications that are optimized for a massive, hyper-scale cloud infrastructure environment. These specifications evolved into a new server design that significantly reduces power consumption and improves operations efficiency. We also shared this design with the industry through our contributions to the Open Compute Project Foundation in 2014.

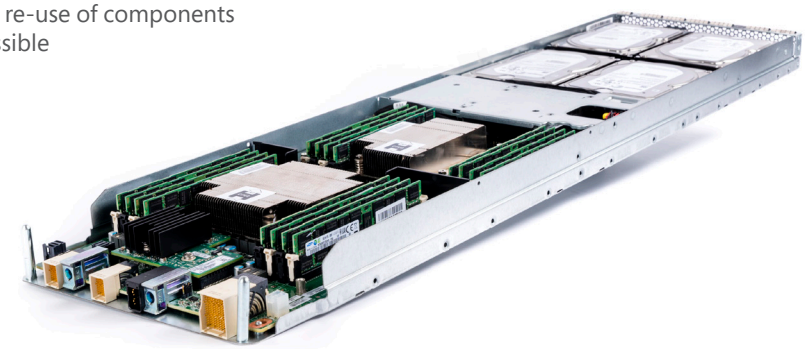
Microsoft's Open CloudServer was built around a number of key principles including:

- Simplicity of design
- Efficiency across cost, power, and performance vectors
- Modularity to accommodate evolving workload requirements
- Environmental sustainability that involves minimizing material use and ensuring re-use of components wherever possible

When operating a hyper-scale cloud, we are faced with unprecedented challenges associated with how to design, build, and operate a massive cloud infrastructure. We must also consider how to decommission and refresh hundreds of thousands of components every year.

This involves having a structured rotation plan that allows for efficient server refreshes. This also entails recovering, reusing, and recycling server components, as well as electrical and mechanical rack infrastructure, which in addition to reducing costs and waste, can also accelerate the rate of deployments. Finally, it is critical that any decommissioning plan evolve active wiping and destruction of hard drives to ensure the protection of customer data.

The Microsoft OpenCloud server design delivers advantages across the spectrum from acquisition cost, ongoing operations, to final decommissioning.



Security, Privacy, and Compliance

While the challenges of providing security and privacy are evolving along with the cloud, the underlying principles haven't changed. We work to build secure systems that will help protect an individual's privacy and we adhere to clear, responsible privacy policies in our business practices—from software development through service delivery, operations, and support.

Microsoft's Online Services Security and Compliance team operates a comprehensive security program and control framework that is evaluated regularly by external parties. The ISO standard is the foundation of our program. While the ISO/IEC 27001:2013 certification standard includes about 150 security controls for our scope, we adhere to more than 800 security controls. We choose to exceed the standard to manage risks that are unique to a cloud infrastructure. In addition, our security program and capabilities are subject to a SSAE16/ISAE3402 review.

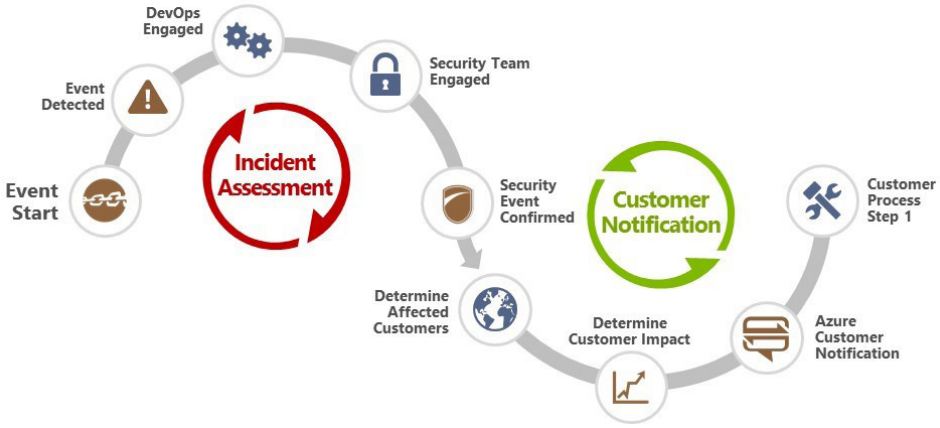
The FISMA certification and accreditation, ISO certification, and SSAE16 SOC 1, 2, and 3 Type I and Type II attestations demonstrate Microsoft's commitment to delivering a trustworthy cloud computing infrastructure.

Other audits and assessments that the Microsoft cloud infrastructure environment undergoes on a regular basis include: the Payment Card Industry Data Security Standard, Sarbanes-Oxley, Health Insurance Portability and Accountability Act (HIPAA), and Media Ratings Council for the integrity of advertising system data generation and processing.

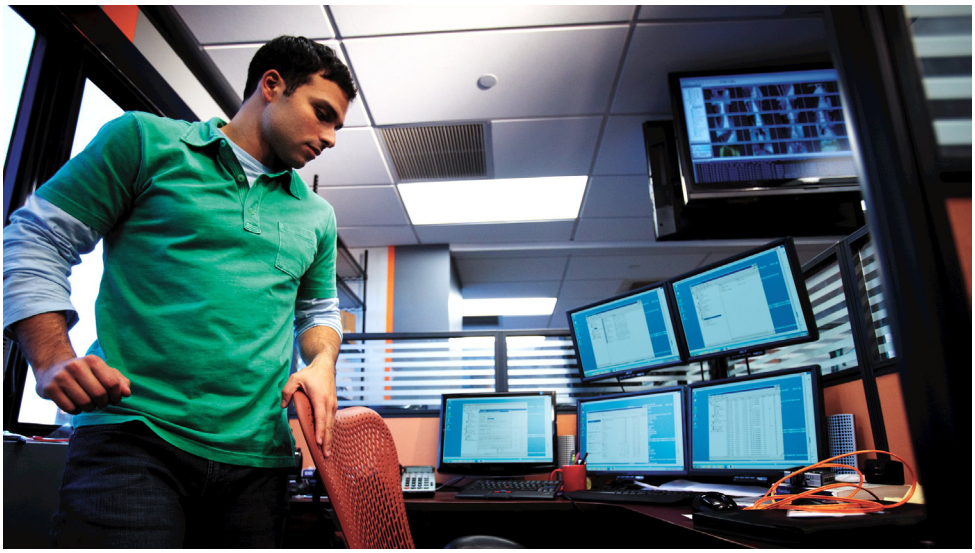
Incident management model

Our Microsoft Operations Centers (MOCs) are globally distributed and work around the clock in a "follow-the-sun" model to ensure our cloud services are persistently available. Each MOC is staffed with a team of incident management professionals and collectively they are responsible for monitoring service health, process automation, infrastructure operations, event and crisis management, and communications across the business. They are responsible for more than five hundred service components and monitor the servers and devices for the services we provide. Most critically, this is the team that identifies and resolves service incidents and outages when things go wrong.

In an environment with so many services, not all incidents are the same. Over half of all incidents are handled through automation and 90 percent of all incidents are handled at first touch. The most severe issues are resolved by a highly trained and qualified crisis management team, working on technical resolution, escalations, and communications to the impacted business groups.



Microsoft Azure Incident Response Model



Cloud sustainability

The IT industry is reportedly responsible for two percent of all worldwide emissions, about the same as the airline industry. However, IT can have a significant impact on the other 98 percent by enabling more sustainable business practices, creating greater efficiencies, and inserting smarter controls into everyday operations. We are not just concerned about efficiency inside the datacenter, but rather we are focused on how to create systems, from the power plant to chip, that are dramatically more efficient than anything that's been done before.

Microsoft focuses on two core priorities to reduce the environmental impact of our cloud infrastructure:

- Building from the ground up with sustainability in mind, such as choosing sustainable materials and finding locations where renewable energy is available and accessible.
- Operating our datacenters to optimize for efficiency, to increase utilization, and to better monitor and manage power consumption. Data and analytics are playing a critical role in allowing us to tune for efficiency in our datacenters.

At Microsoft, we look at data as just a refined form of energy. As a result, our path for delivering power to supply Microsoft's cloud infrastructure is focused both on how we optimize for efficiency inside our footprint and how we integrate and invest in driving greater efficiencies across the broader energy supply chain.

We maintain three key initiatives in power generation and consumption to improve our efficiency and realize our efficiency goals:

1. We distribute hyper-efficient power generation to the datacenter that radically reduces the amount of energy required to deliver cloud services.
2. We deliver to the grid low-cost and efficient energy through participation in utility-scale generation projects.
3. We foster the development of the next generation of energy technologies that will make future distributed and grid-connected projects radically more efficient.

Microsoft is committed to driving software and hardware innovations that help people and organizations reduce their impact on the environment. To learn more about our datacenter sustainability research projects and renewable energy investments, please visit our engineering blog site.

The benefit for our customers

We've invested heavily in our infrastructure, technologies, and processes to insure your cloud services will be available and protected. We back-up our commitment to customer service in the cloud with the industry's most extensive, financially-backed SLAs. In the event of a service outage, we have programs, procedures, engineers, and operational experts in place to contain issues or rapidly recover with minimal impact. We are continually evolving our infrastructure, tools, and methods to improve the reliability and availability of our cloud services. To ensure our operations are environmentally sustainable, we continually innovate to improve efficiency and strive to power our datacenters with clean, reliable energy.

Microsoft's has staked our future on being the company that drives productivity in a cloud-first, mobile-first world. The innovations and investments we are making in our global cloud infrastructure today will provide a strong foundation for the next generation of cloud services and beyond.

Microsoft has extensive experience operating a cloud services infrastructure since 1995. As Microsoft's cloud services portfolio and infrastructure continues to grow we are making thoughtful investments to answer customer needs for greater availability, improved performance, increased security, and lower costs.

Contributors:

Rick Bakken

Monica Drake

For more information, please visit www.microsoft.com/datacenters