419 Sargon Way
Horsham, PA 19044
215.394.1906
215.394.1908
sales@upsafety.net
support@upsafety.net

UNITED PUBLIC SAFETY

www.upsafety.net

Where is my Data?

We don't compromise with any of our customers data. When transactions are complete, your data is sent to the Microsoft Global Foundation Services data center in Chicago Illinois. Your data is triplicated among virtual and hardware servers.

All of our customer data is encrypted and backed up each day at 9:00 pm EST to our servers in Horsham, Pennsylvania. For a small additional charge, we can also send your encrypted data to an FTP site in **your server** at **your facility**.

The following is a redacted copy of an executive overview of Microsoft's data centers.

Cloud-Scale Data Centers

Delivering services at cloud-scale requires a radically different approach to designing, building, deploying, and operating data centers. When software applications are built as distributed systems, every aspect of the physical environment – from the server design to the building itself, creates an opportunity to drive systems integration for greater reliability, scalability, efficiency, and sustainability.

This strategy brief will explore how cloud workloads have changed the way data centers are designed and operated by Microsoft. It will also share information on how to approach developing more resilient online services that deliver higher availability while lowering overall costs.

Early Microsoft Data Centers

Microsoft built its first data center in 1989, and we have been delivering online services at scale since 1994 with the launch of MSN. The company has invested over $15 billion in building a highly scalable, reliable, secure, and efficient globally distributed data center infrastructure. These data centers support over 200 online services, including Bing, MSN, Office 365, Skype, Windows Live, Xbox Live, and the **Windows Azure platform**.

We are continuously evolving our strategies for how to provision and operate the underlying infrastructures for cloud services. Key learning gained from operating at huge scale are integrated to help us meet the performance and availability expectations of customers for our growing services portfolio, while driving greater cost efficiency.

419 Sargon Way
Horsham, PA 19044
215.394.1906
215.394.1908
sales@upsafety.net
support@upsafety.net

In our early data center designs we applied very conservative practices still used by most of the industry today. Availability was engineered to meet the highest common denominator – mission critical support for intranet, extranet, and online services.

Typical of the state of the art facility designs at the time, these facilities were fault tolerant and concurrently maintainable and populated with 2N servers with redundant power supplies and myriad of hot-swappable components. We made capital investments in redundancy to protect against any imaginable failure condition, including the loss of utility electrical and cooling water services, processor and hard drive failures, and network interruptions. By enabling highly reliable hardware, services' developers were free to scale-up their applications on relatively few expensive servers.

However, hardware redundancy could not protect against the non-hardware failures that result from software configuration errors, lax change management processes, network performance, and code bugs. These human errors and uncontrollable issues can impact service availability – no matter how many "9's" were engineered into the data center infrastructure.

Microsoft maintained the traditional approach of delivering availability through hardware redundancy until 2008. We had delivered highly-available services by supporting this model with rote process and procedures, but we quickly saw that the level of investment and complexity required to stay this course would prove untenable as we scaled out our cloud services.

At cloud-scale, equipment failure is an expected operating condition – whether it be servers, circuit breakers, power interruption, lightning strikes, earthquakes, or human error – no matter what happens, the service should gracefully failover to another cluster or data center while maintaining end-user service level agreements (SLAs).

Resilient Software

In many companies today, data center capacity is consumed through a rigid series of processes where each element of the stack is designed and optimized in a silo. The software is developed assuming 100 percent available hardware and scale-up performance. The hardware is optimized for maximum reliability and premium performance, and the operations team does their best to deliver service quality through the software application's lifecycle.

At Microsoft, we've begun to follow a different model, with a strategic focus on resilient software. We work to drive communications that are more inclusive between developers, operators, and the business.

419 Sargon Way
Horsham, PA 19044
215.394.1906
215.394.1908
sales@upsafety.net
support@upsafety.net

By sharing common business goals and key performance indicators, it has allowed us to more deeply measure the holistic quality and availability of our applications. As developers create new software features, they interact with the data center and network teams through a development operations model. This enables everyone to participate in the day-to-day incident triage and bug fixes, while also leveraging chaos-type scenario testing events to determine what is likely going to fail in the future.

The operations team on-boards the software applications and develop a playbook on how to operate it. Focus is placed on the capabilities that need to be provided by the underlying infrastructure, service health, compliance and service level agreements, incident and event management, and how to establish positive cost control around the software and service provided.

The software and the playbook then is layered on top of public, private, and hybrid cloud services that provide an abstraction layer where workloads are placed virtually, capacity is advertised, and real-time availability is communicated with the services running on top of the cloud infrastructure.

From a hardware standpoint, the focus is on smart physical placement of the hardware against infrastructure. We define physical and logical failure domains and recognize that workload placement within the data center is a multi-disciplined skillset. We manage our hardware against a full-stack total cost of ownership (TCO) model. And we consider performance per dollar per watt, not just cost per megawatt or transactions per second. At the data center layer, we are focused on efficient performance of these workloads – how do we maintain high availability of the service while making economic decisions around the hardware that is acquired to run them.

We automate events, processes, and telemetry; integrating those communications through the whole stack – the data center, network, server, operations, and back into the application to inform future software development activities. A tremendous amount of data analytics is available to provide decision support via runtime telemetry and machine learning that completes the loop back to the software developers, helping them write better code to keep service availability high.

The telemetry and tools available today to debug software are several orders of magnitude more advanced than even the best data center commissioning program or standard operating procedure. Software error handling routines can resolve an issue far faster than a human with a crash cart. For example during a major storm, smart algorithms can decide in the blink of an eye to migrate users to another data center because it is less expensive than starting the emergency back-up generators.

419 Sargon Way
Horsham, PA 19044
215.394.1906
215.394.1908
sales@upsafety.net
support@upsafety.net

Hardware will fail and as cloud providers and a new generation of application developers embrace this fact, service availability is increasingly being engineered at the software platform and application level rather than by focusing on hardware redundancy. By developing against compute, storage, and bandwidth resource pools, hardware failures are abstracted from the application and developers are incented to excel against constraints in latency, instance availability, and budget.

What Your Cloud Should Provide

In a hardware-abstracted environment, there is a lot of room for the data center to become an active participant in the real-time availability decisions made in the software applications.

Resilient software solves for problems beyond the physical world. However, to get there, the development of the software requires an intimate understanding of the physical in order to abstract it away.

In the cloud, software applications should be able to understand the context of their environment. Smartly engineered applications can migrate around different machines and different data centers almost at will, but the availability of the service is dependent on how that workload is placed on top of the physical infrastructure. Data centers, servers, and networks need to be engineered in a way that deeply understands failure and maintenance domains to eliminate the risk of broadly correlated failures within the system.

As you are looking at private, public, and hybrid cloud solutions, now is a good time to start thinking about how you present the abstraction layer of your data center infrastructure. How you place workloads on top of data center, server, and network infrastructure can make a significant difference in service resiliency and availability.

Virtualization technologies like Microsoft's Hyper-V allow you to move application workloads around machines and data centers. However, most folks are just extending a bin-packing exercise in an attempt to get more utilization out of the existing asset. Without recognizing the location of the compute or storage block in a broader ecosystem, packing more processes onto the same physical hardware can amplify the impact of a failure.

Understanding the failure and maintenance domains of your data center, server, network, and manageability infrastructure is critical to placing virtualized workloads for high availability. In most data centers, you are limited to one or a handful of failure domains at best. With cloud platforms like

Microsoft's Windows Azure, a developer or IT professional can choose from many physical hardware environments.

With public and hybrid clouds, virtualization when paired with resilient software not only reduces costs, but also improves availability. Rather than scaling up hardware with faster processors and more RAM, PaaS and IaaS cloud platforms such as Microsoft's Windows Azure enable seamless scale-out of applications across many commodity servers in multiple data centers.

Cloud-Scale Data Center Design

The shift to software resiliency often represents an inflection point to re- examine how data centers are designed and operated where software-based resiliency can be leveraged to improve reliability, scalability, efficiency, and lower costs, while simplifying the data center.

We achieve service resiliency in software, which allows us to start removing hardware redundancy from the data center environment. At cloud-scale, we rationalize our availability model and rethink our infrastructure.

Rather than independently design a data center, a network, spec servers, and design management software, Microsoft has been working to define the entire environment as a converged ecosystem. In reexamining the data center, we look at elevated supply temperatures and delta T. We look at optimizing the size of our server clusters to match the failure domains that exist within in the physical world. Plus, we balance preventative maintenance against a "fail small" topology that allows us to compartmentalize failures. We rationalize the full stack availability, mapping end user SLAs to real-time availability performance of the facilities and the servers that run in them.

Additionally, we reduce the hardware redundancy in this space by focusing on TCO-driven metrics like performance per dollar per watt, and balancing that against risk and revenue. At cloud-scale, each software revision cycle is an opportunity to improve the infrastructure.

The tools available to the software developers – whether it is debuggers or coding environments – allow them to understand failures much more rapidly than we can model in the data center space.

Microsoft's Software-Resilient Data Centers

An example of a software resilient data center is our Chicago facility. This facility is populated with dozens of shipping containers that house thousands of servers. However, each container is a failure domain, and the logical grouping within a physical stack allows us to manage how we place workloads on

top of it. We purchase these containers through a competitive TCO-driven process where the server vendor is incented to provide the optimal performance per dollar per watt within a set of electrical, mechanical, and physical constraints.

Additionally, we focus on full stack integration – the software applications that operate in these containers are multi-way, Active/Active applications that can shed load and move it around the building to another container, or to other data centers around the globe. We have converged the electrical, mechanical, controls, networks, and servers into one management framework that allows us to have greater insight into the operation and performance of the system as a whole. With this approach, we attain a deeper understanding of Mean Time Between Failure (MTBF) and Mean Time To Recovery (MTTR), and can tailor support models that allow us to come in on a break-fix, next business day model.

Because these containers and the applications running in them are software resilient, there are no emergency back-up generators needed to support it. There is a short amount of ride-through, allowing us to do switching within the utility. However, if we have a broad utility-level failure within this space, we can move our workloads to another data center without impact on end-user performance.

More recently, Microsoft opened our Boydton data center where we use modular ITPACs (data centers) outside with no roofs or walls, and just an open- air electrical breezeway powering and supporting them. This has been one of our design evolutions, as resilient software ha improved. Consequently, we have been able to remove more inefficiencies, water use, and costs out of our data cent environment.

At cloud scale, we accept the three laws of operations – hardware will fail, software will have bugs, and humans will make mistakes. In moving to a model where resiliency is engineered into the service, we not only ensure better availability to the customer, but can make dramatic architectural and design decisions to improve reliability, scalability, efficiency and sustainability, simplify operations, an reduce acquisition costs.

419 Sargon Way
Horsham, PA 19044
215.394.1906
215.394.1908
sales@upsafety.net
support@upsafety.net

UNITED PUBLIC SAFETY

www.upsafety.net

Holistic Data Center Optimization.

The Microsoft data center operations team continually seeks ways to drive more efficient use of power and cooling in data centers. Our measurements of server performance under load and workload analysis has enabled us to right-size our server platforms. We have eliminated unnecessary components, use higher efficiency power supplies and voltage converters, and bounded the expandability of server platforms to achieve significant power savings.

Specific measures such as processor performance per dollar per watt help to determine the optimum trade-offs in processor selection. By using lower performance and lower watt processors we can deliver greater overall efficiency for both power utilization and deployment cost. For several years, we have shared our key learning for widening the operating range of our servers and using free air cooling and water economization to improve efficiencies.

This continual measurement and analysis allows us to push the boundaries of data center efficiency, helping reduce operating costs while improving environmental sustainability within our facilities and the industry.

Maintaining High Availability

Microsoft data centers are operated and managed according to rigorous compliance standards including ISO 27001, SSAE 16, and FISMA. Under these standards, all events are managed according to service impact (Outage, Impact, Incidents). For each event that impacts operations, an Abnormal Incident Report (AIR) is prepared that includes the root cause and recommended corrective action. A wide range of software-based tools for incident management, deployment, and scheduling help Microsoft’s service operations teams keep our data centers and services operating in a reliable, trustworthy, and cost effective manner.