White paper

Four approaches to better data center availability

"Mission critical" is a much-used adjective for various hardware components and processes, referring to any element of a system whose failure will result in the failure of business operations. While it's a popular buzzword, it has little real meaning in the data center. Here, without redundancy, the failure of almost any piece of equipment or software could result in a shutdown.

It's very much like building a tower from blocks; no block more critical than another. In the same way, every functional element in the data center – power distribution, HVAC, BMS, IT network – depends on the rest. The failure of any component, such as a pump, transformer or operating software, could bring the center to a crashing halt.

Loss of service is the #1 fear that keeps data center managers awake at night. In a recent survey, they cited "availability" as their top priority.[1] Their fear is justified. The direct costs of an outage are estimated at approximately $260,000[2] for a partial shutdown, plus the less-tangible but possibly more-damaging indirect cost to the center's reputation.

While 95% of data centers experience one or more unplanned outages over a typical 24-month period, the good news is that 80% of these outages are preventable.[3] Given the tremendous demands placed on data centers and the complexity of technology applied to meet those demands, there is much that can go wrong. But there is also much that data center managers can do to ensure an uninterrupted flow of transactions.

Availability Versus Reliability

"People in our industry often talk about the 'reliability' of data centers," explains Mietek Glinkowski, Director of Technology at ABB. "What they are really talking about is availability. The terms aren't interchangeable, and it's useful to understand the difference."

Reliability generally applies to a component or subsystem. It is the probability that a device or system will perform its intended function over a specific period of time. The reliability can be statistically calculated.

"You can determine an actual value for reliability as mean time between failures or mean time between repairs," Glinkowski says. "For example, if you take a sample of 1,000 cars of a certain model, you can look at the rate of breakdowns over a certain period, say one year, in order to calculate the reliability of that model."

1. 2010 Ponemon Institute study
2. ibid
3. Ibid

Within a data center, there are three broad types or classes of reliability:

- **Component reliability** – the reliability of each individual, physical piece of hardware in the center.
- **Power system reliability** – the overall reliability of all the components and their interconnections that manage electricity from the street to the server.
- **Process reliability** – relates to the processes in data centers, including HVAC and other infrastructure-related systems.
-

The availability of the data center refers to meeting the uptime expectations. The goal, of course, is 100% availability, but both humans and technology stand in the way of achieving that goal. In the Uptime Institute's Tier Classification System, a Tier 1-certified data center is expected to provide 99.671% availability. That sounds like an impressive number until you do the calculations and realize that this equates to 28.8 hours of downtime per year.

Center managers determined to elevate their uptime metrics have several key areas where they can focus their attention.

Consider the Weakest Link

"Facility designers, surprisingly, often don't realize that the availability of the entire data center is only as good as the weakest point," Mietek says. "If you look at the power system and how the power is distributed throughout the data center, you have to look at all possible components and the flow of power in different conditions to make sure that in all of the scenarios you have the reliability that you are looking for."

To keep costs under control, many data centers are constructed using commercial-grade instrumentation. While cost-effective, these instruments are not engineered or built for the high reliability needed in data centers. By mistakenly assuming, for example, that an instrument is a commodity item, they might choose to save money by selecting a lower-cost model. The entire system is no better, no more reliable, than that instrument.

Using industrial-grade rather than commercial-grade instrumentation is one way to increase your availability. However, achieving higher availability is not as simple as spending more money on more reliable hardware.

"That may be true to some degree," Glinkowski says, "but the point is to have a balance of reliable equipment in all places combined with the infrastructure or control systems to knit them together into the overall system with high reliability. You must understand the interrelatedness of every element in the data center. They are all important. They are all part of the overall performance of the data center."

This fact makes the architecture of the system critical to reliability. You can mathematically calculate an overall process or facility reliability by considering the reliability of each component, but you must also consider how those components are connected and combined.

Mountain climbing provides a good analogy. Two climbers could be outfitted with exactly the same equipment. The climber who places that hardware in the most logical and effective arrangement has a much greater chance of returning from the summit safely than the climber who doesn't use the equipment properly.

Even the most reliable piece of equipment or software will fail eventually. Data center designers and managers must analyze their facilities to identify the points where the threat of breakdown warrants redundancy. In some cases, those decisions can be made intuitively. It is also possible to identify third-party service providers with the analytical tools and broad industry experience to do a more comprehensive analysis.

"You create redundancy selectively, in the areas where it is most important," says Mark Reed, ABB's Director for Data Centers in North America. "Look at the risk- or cost-benefit of redundancy in each area of the system. Go through that exercise and select the best places to make investments in reliability."

Deploy DCIM

Compared to a busy manufacturing plant or bustling distribution facility, the typical data center appears to be a fairly sedentary operation. In truth, there is an around-the-clock whirl of activity, the bulk of it hidden from view inside racks, fans and conduits. To monitor and manage that activity, data center managers are increasingly turning to Data Center Infrastructure Management systems (DCIMs). Earlier versions of this software were little more than building management system monitoring, collecting top-level metrics regarding temperatures of cooling fluid in and out of chillers, total voltage coming into the facility, and the operating state of pumps, fans and motors.

"More advanced systems followed, but there were different systems each focused on different center functions," says Reed. "In addition to power and HVAC, there were also security and fire systems, electrical/power monitoring systems and asset management systems. These systems weren't always integrated, making it difficult or impossible to get a true picture of the overall center status. As a reliability tool, they didn't do the job."

Today, there are DCIMs available that provide a unified view that allows data center operators and managers to see information across multiple disciplines and make well-informed decisions. Synthesizing a clear, timely picture of the operation requires data from sensors deployed throughout the facility.

The current crop of data center devices is increasingly "smart," equipped with instrumentation that enables them to communicate data about their condition. The DCIM eagerly accepts this data and uses it to paint a picture of overall process status.

"The DCIM can use inputs from all types of devices," Reed explains. "Inputs can come from field devices like pressure transmitters, temperature sensors and valve actuators. You can

get diagnostics from equipment like pumps, drives, motors, gensets, switchgear, batteries, UPSs, chillers and water towers. It's possible to get information from almost every piece of facility equipment that is "intelligent" today. The DCIM collects, synthesizes and analyzes the raw data and turns it into useful business information for making decisions in real time. In addition, the information can be used by the DCIM system to tell you what needs to be fixed and how to fix it, and then can schedule a technician to fix it."

Of course, all that instrumentation comes at a price, so managers need to identify which component or sub-system failures would be most catastrophic and/or where they are most likely to occur. Those are the areas where instrumentation should be deployed first. However, the more fully the center is instrumented, the more comprehensive and accurate is the DCIM output.

At this point, it might be better to replace "management" with "monitoring" in Data Center Infrastructure Management. DCIMs typically provide a comprehensive data stream representing operating parameters throughout the process. Analytics in the software also provide guidance and decision support for process improvements. But the actions required to implement those improvements remain, for the most part, a manual activity. The loop hasn't been closed to create automated management control except in a very few products now on the market.

In manufacturing, automation efforts started with Computer-Aided Design (CAD) that created part designs in a three-dimensional model. CAD provided the foundation for the follow-on application, computer-aided manufacturing. With CAD/CAM (Computer-Aided Manufacturing), parts could be designed electronically and then transmitted to machines for automatic manufacturing. That same evolution is now underway with DCIMs. Soon they will become true management systems.

As managers increasingly rely on DCIMs, their role to centralize operation becomes more critical. It therefore becomes more important to include redundancy in the DCIM itself.

"A capable DCIM is increasingly becoming a must-have piece of software," says Reed. "It's one of the fastest growing parts of what people are buying for data centers right now. The current generation of DCIMs harmonizes facilities and energy management in a unified control environment within and across multiple data centers. That helps drive both efficiency and reliability to higher levels."

Support the Staff

Half of unplanned outages is the result of operator error. While the size of that number may be disheartening, the good news is that those types of errors are 100% preventable.

Reducing operator error begins with ensuring that there are clearly defined procedures in place for both routine duties and for serious issues and emergencies. As in many industries, this type of knowledge is frequently conveyed simply by having the new person shadow the incumbent.

While easy, quick and cost-free, this training method isn't grounded in best practices. To ensure consistency and availability, operators need to define their work processes and document them. These documents are frequently provided to employees or made available in a binder kept somewhere in the work area.

A far more effective approach is to incorporate these work instructions in the DCIM. The current generation of DCIM software includes the ability to house work instructions and other documents in an online repository.

"This is the next step in reducing operator error – embedding operator procedures into the control software so that each action is prompted, checked, and documented," Reed says. "That provides consistent and safe operation of the processes based on predefined and tested procedures."

Many data center employees experience a work situation similar to jetliner pilots. Both are responsible for technology-laden assets, but both spend most of their time not doing very much. Once a pilot has the plane to altitude and cruising, it basically flies itself. When the data center is running well, the operator is mainly on watch, with little to do.

"There are hundreds of events in a typical shift that might trigger an alert or alarm," says Reed. "Many are routine and can be easily addressed or will self-correct. These alarms can become white noise to the operator and can possibly be missed. With intelligent alarm management, the DCIM identifies those that require urgent intervention. It can also synthesize apparently minor alerts from several points and conclude that, taken together, they indicate a potentially serious impending problem.

"But both the pilot and the operator need to be able to respond quickly and appropriately when problems arise," Reed says. "For the data center operator, the response often depends on whether or not they have immediate access to the right procedures or work instructions."

The DCIM can also be used for training by providing a continuous improvement tool. Some DCIM software includes the ability to record operator actions and system response. The information is a tremendously valuable training resource. You can review the way the situation was handled and the outcome in terms of how quickly the problem was resolved and whether other problems were created.

"The DCIM enhances operator effectiveness in two ways," Reed says. "On the one hand, it guides the operator's actions through context-sensitive information; and on the other, provides feedback and documentation of how successful the operator's actions were. This prevents mistakes in the first place and refines your processes to provide improved responses in the future."

Consider a Direct Current Conversion

"Everything old is new again."

The name of the person who first uttered that quote has been lost, but the wisdom of that observation is apparent in the recent revival of DC power for commercial applications.

"We see the DC comeback in our daily life, in your cell phone, tablets, laptops and most other electronic devices," says Glinkowski. "They are all DC powered. In the data center, at the end-level user load – the server – you mainly see DC. That's not apparent because most of the devices are powered or charged by AC. Within the server, the first thing that happens to the power is conversion to DC."

DC power isn't widely seen in industrial and commercial applications, but its reliability and portability have made it the current of choice in some very demanding applications. Mining and some other large industrial process equipment runs on DC. The military also prefers DC-powered devices. It is a proven technology with great potential for data centers. DC current, mainly 48V, has been used for decades in telephone company switching centers and PBXs.

The many advantages touted for DC power are the result of its simplicity compared to AC. Fewer devices are required to deliver power from the street to the server. The elimination of power management devices and simplification of the distribution system provide a long list of benefits.

In the current North American AC power world, line power (typically 480 or 208V) is rectified to DC to charge the data center UPS batteries and then inverted back to AC for distribution within the center. On the way to the racks, there is a power distribution unit (PDU) – essentially a transformer – that drops the 480 V power to 208 V or 120 V that can be used by the hardware in the racks. In many facilities, there are additional energy conversions.

With every power rectification, inversion and step-up or step-down energy is wasted, mainly in the form of heat that must be removed by the HVAC system, requiring more cooling capacity. "Anytime you can have less equipment, you're going to have a more reliable system," Reed says. "And that doesn't relate only to the power distribution system. When you make less heat, you need less HVAC, so you also reduce the complexity of that system. Greatly increased energy efficiency is a huge bonus benefit."

With everything DC has to offer, why isn't everyone making the rush to adopt it?

"One of the major stumbling blocks is that there's so much AC infrastructure," Glinkowski explains. "With DC, we have a great thing with remarkable potential, but it's different than what we have been building for the last 110 years. People are hesitant to quickly migrate from their existing, well-known infrastructure and to walk away from capital already invested."

There are other roadblocks as well.

"Putting it into practice will require many changes to things that we've long taken for granted, "Reed says. "It comes down to things as simple as receptacles. We've also been waiting for manufacturers to ramp up production of servers and other equipment with DC power supplies. They're already becoming more readily available.

"There's also the issue of standards. In the AC world, there are many, many standards for equipment, conductors, safety, monitoring, metering and so on. For DC, there are some standards, but very few."

The potential for significant improvements in reliability and efficiency have most data center professionals at least considering the DC alternative.

Conclusion

Overall application availability is more critical than the availability of each individual data center. That's why organizations typically have redundant centers. While Tier 4 data centers are very well protected, natural disasters and extended power outages still create downtime risks. Taking steps to enhance the availability of each data center ensures overall availability.

A slogan for a well-known luxury car brand was "The relentless pursuit of perfection." Every data center manager can relate to that sentiment.

The challenges faced in achieving that perfection is mathematically apparent when you consider the fact that the difference between the availability of the lowest-ranked Tier 1 facility and highest-ranked Tier 4 facility is .00324 … with two Tier increments in between.

Ratcheting up the availability of a data center requires tightly managing every aspect of operations. Highly reliable equipment must be selected and deployed in an architecture that not only capitalizes on the native device's reliability but enhances it through synergies in the system.

As data centers grow in size and complexity, it will become essential to rely on high-functioning DCIMs to collect and synthesize data and to present it to operators in a way that guides them toward, or simply implements, a correct response.

Finally, when considering a technology refresh or addition of new capacity, DC solutions should be considered instead of AC power distribution. The benefits are apparent, and the technical roadblocks are quickly falling.

Through these and other efforts, data center managers can make significant progress in their pursuit of a perfect record of availability.

ABB Data Centers
125 East County Line Road
Warminster, Pennsylvania, U.S.A.
Phone: +1 800 HELP 365

www.abb.com/datacenters

9AKK105408A8194

Power and productivity
for a better world™