White Paper

Power Redundancy Schemes for Data Centers

ASCO Power Technologies™

Power Redundancy Schemes for Data Centers

Uptime Institute Tier Levels Explained

Data center downtime can have profound impacts on business operations and revenue, and power interruptions are a common cause of data center outages. The following narrative surveys data center power challenges and explains the Uptime Institute's tier levels for designing power systems to meet availability objectives.

TYPES OF DATA CENTERS

Since the advent of networked digital communications, Information Technology (IT) resources for connecting users and processing information have been aggregated into data centers located throughout the world. Over time, several types of data centers have emerged, each with a different purpose. Most data centers are described by the following types:

Enterprise Data Centers

Enterprise Data Centers are typically operated by the organizations they serve. They meet the networking and computing needs of the organization, such as those used in manufacturing, human resources, payroll, and other core services. Enterprise data centers also host business-to-business and/or business-to-consumer transactions via Internet. The primary focus of most enterprise data centers is to maintain high levels of availability.

Colocation Data Centers

Colocation facilities rent space for networking and data processing resources, including power and cooling, to multiple users. Some colocation centers manage equipment and systems for their customers; others merely provide an acceptable environment for customers to conduct their operations. The primary focus of these facilities varies with the types of customers they serve.

Cloud Data Centers

In this model, IT equipment and services are provided by the data center operator and are managed by the data center's staff. IT resources are standardized across the facility to promote rapid scalability. Cloud data centers focus on deployment speed.

Hyperscale Data Centers:

Hyperscale facilities can encompass more than 500,000 square feet of IT space. Operated by the world's largest computing, retail, and social media firms, these data centers leverage trends such as virtualization, software-defined networking, and software-defined storage.

The content of this document is provided for informational purposes only. Redundant power systems should be designed by a qualified professional. Assessment and certification of redundancy levels for data center systems should be performed by professional qualified by the Uptime Institute.

Edge Data Centers

Edge data centers are smaller facilities built to move time-sensitive computing resources closer to the user. They serve as extensions of larger data centers and provide low-latency connections for cloud computing resources, cached streaming, and data transmission for IoT devices.

Regardless of type, size, and focus, data centers strive to provide reliable and cost-effective service. However, data center reliability and availability are only as good on the systems that make IT equipment operation possible, including electrical power distribution systems.

DATA CENTER CHALLENGES

The Uptime Institute is an "...advisory organization focused on improving the performance, efficiency, and reliability of business-critical infrastructure..."[1] In 2018, the Institute reviewed information from multiple sources to evaluate challenges to data center availability. Their sources of information included (1) data collected from more than 100 major publicly recorded service outages since 2016, and (2) survey responses from more than 1,100 data center operators and IT practitioners.[2] The highlights of the findings are as follows:

- IT service and data center outages around the world are common and may be increasing
- The biggest cause of IT service disruption is a data center power outage, closely followed by network problems, and then by an IT system failure
- Failures at third-party cloud, colocation, and hosting providers are the second most common reason for IT service failures
- Most IT service outages are preventable. Eighty percent of respondents said that their most recent service outage could have been prevented

Data center operators have historically promoted high-availability designs and careful operating procedures to reduce service disruption. Nevertheless, the Uptime Institute study demonstrates that outages remain common.

[1] Uptime Institute, About Uptime Institute. https://uptimeinstitute.com/about-ui. Viewed 2/26.
[2] Lawrence A. Uptime Institute Data Shows Outages are Common, Costly, and Preventable. Uptime Institute. June, 2018. p. 2.

UPTIME INSTITUTE PRINCIPLES

Because utility power outages remain a common risk, power professionals strive to design facilities that can tolerate them and still serve their intended mission. The task of designing systems to provide the appropriate availability levels becomes clearer when the advantages and disadvantages of both utility and onsite power sources are considered. In its *Tier Requirements for Power*, the Uptime Institute states the following precepts should be observed when designing power systems:[3]

- Disruptions to utility power are not considered failures, but as an anticipated operational condition for which any site must be prepared
- Onsite power generation is the only source of truly reliable power
- Utility-provided power is the most economical source when compared to onsite-generated power
- Data centers can combine onsite and utility sources for cost efficiency if there is autonomous failover capability to another power source

To provide appropriate systems, power industry professionals must understand how to achieve the levels of availability required for common applications. For this reason, the Uptime Institute maintains a framework for configuring systems according to acceptable levels of power outage risk.

UPTIME INSTITUTE TIER CLASSIFICATIONS

The Uptime Institute classifies data center power systems according to level of redundancy for both power equipment and power pathways. The four classifications are summarized in the table below and are further detailed in the subsequent narrative.

Uptime Institute Data Center Tier Levels			
Tier	Description	Power Pathways	Redundancy
1	Basic Capacity	Single	N
2	Redundant Components	Single	N+1 Components
3	Concurrently Maintainable	1 Active 1 Passive	N+1 Systems
4	Fault Tolerant	2 Active	System + System

Tier I Power – Basic Capacity

Tier I backup power systems offer a single set of components and a single pathway for connecting utility and generator power sources. At Tier I, a power distribution system can switch loads to an onsite standby generator when utility outages occur, avoiding downtime. Nevertheless, because the system offers only a single power pathway and does not offer redundant components, the failure of any single component or pathway element will result in downtime. Figure 1 represents a Tier I power system configuration.

[3] Uptime Institute. Tier Requirements for Power. Undated. p. 1.

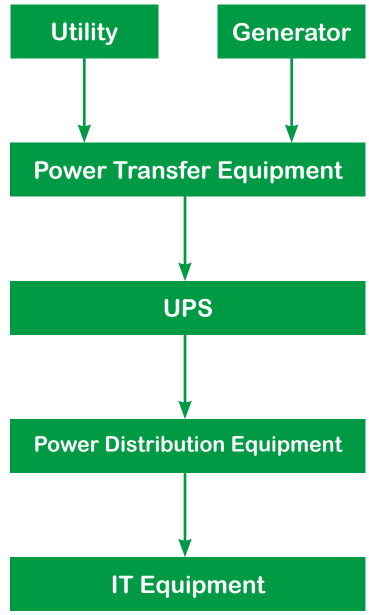

Figure 1: Tier 1 (N) Systems Offer a Single Power Pathway Without Redundant Components.

The expected availability of a well-designed and well-run Tier I power system is 99.671 percent. Assuming 24x7 operation, this equates to nearly 29 hours of expected downtime per year. For this reason, Tier I power is typically deployed where business operations can tolerate occasional service outages without experiencing significant impacts to overall operations or revenues. In these applications, Tier I architecture offers the simplest and most cost-effective backup power solution.

Tier II Power – Component Redundancy

Like Tier I systems, Tier II backup power systems offer a single pathway for connecting utility and generator power sources to loads. However, Tier II systems offer redundant units for one or more key components, typically those that pose the highest risk to reaching uptime goals. Figure 2 shows a Tier II system with redundant generators and uninterruptible power supplies (UPS). This system will insure that power remains available if either a generator or UPS becomes inoperable. Nevertheless, the loss of any single power path element will still result in an outage.

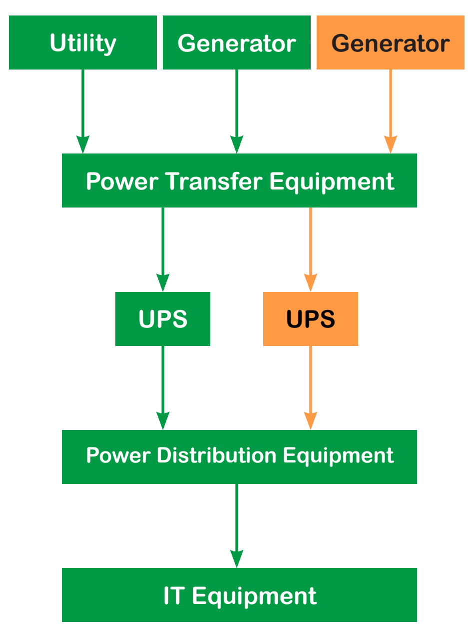

Figure 2: Tier II (N+1 Components) offers redundant components with a single power pathway.

For comparison purposes, the expected availability of a well-designed and well-run Tier II power system is 99.741 percent, which equates to nearly 23 hours downtime annually. For this reason, Tier II power is typically deployed in enterprise data centers to improve availability. When compared to Tier I systems, Tier II systems present only an incremental increase in complexity and capital cost.

Tier III Power – Concurrent Maintainability

Tier III systems provide a single redundant set of components arranged along a separate power pathway. This allows one system to be taken offline for service without disrupting power to loads.

One example of a Tier III system uses a *Dedicated Reserve* design, which provides a redundant pathway and redundant equipment that can functionally replace another power system. For example, Figure 3 shows a Tier III system comprised of *N* primary systems (green and blue) and one reserve system (orange). During normal operation, the primary systems provide up to 100 percent of their respective capacities while the secondary system remains idle. The secondary system supplies up to 100 percent of its capacity when one of the other systems fails or is taken offline for service. This state is shown in Figure 4.

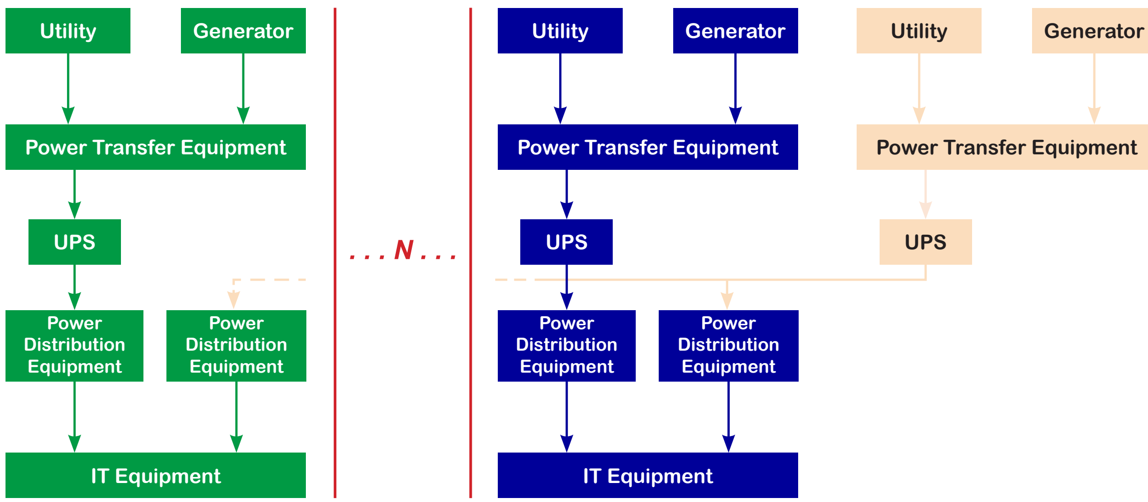

Figure 3: A Tier III (N+1 Systems) Dedicated Reserve System
Under Normal Operation

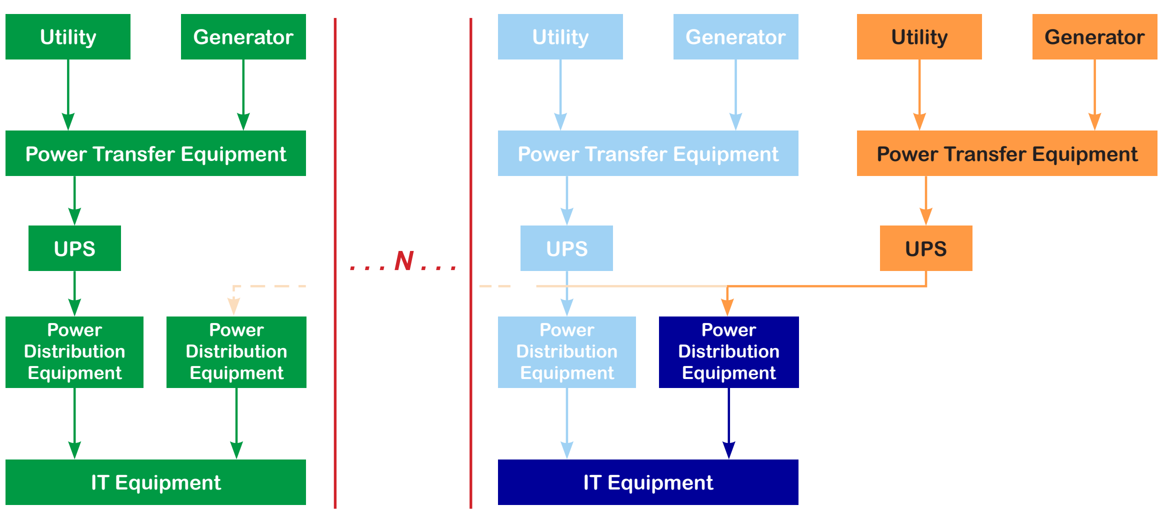

Figure 4: A Tier III (N+1 Systems) Dedicated Reserve System
Operating in Backup Mode

In these examples, the Tier III Dedicated Reserve system provides backup power to one of the other systems through the primary distribution equipment. Consequently, it can provide concurrent maintainability without significant additional investment in power distribution equipment.

The expected availability of a well-designed and well-run Tier III power system is 99.982 percent, which equates to approximately 1.5 hours of downtime annually. For this reason, Tier III power is typically deployed in colocation, cloud, and hyperscale data centers to provide very high-availability.

Medium Voltage Systems

In large facilities, multiple generators may be required to supply the necessary load capacity, typically through a medium voltage backup power system. When a utility outage occurs, these multiple engine-generators must be started, accelerated to their designed run speed, then synchronized and connected to a main bus. If a large system were to use a single main bus, no single generator could be serviced without taking all units offline.

By dividing bus into two or more segments connected by tie breakers, generators can be separated into smaller groups that can be isolated to facilitate maintenance. By placing tie breakers between bus segments, faults can be isolated to a single bus, leaving the remaining units online to provide standby power. In additional, isolating one bus segment would remove only a subset of generators from service. Nevertheless, linear bus arrangements can leave operable generators without a connection to load, negating their backup function and leaving important loads unpowered.

Ring bus arrangements can be formed to connect generators to loads, as shown in Figure 5. Under normal operation where all bus segments are connected, any power source will have two pathways to any load. Should any one generator, tie breaker, or bus segment fail, power from all remaining sources will still have an available path to all loads.

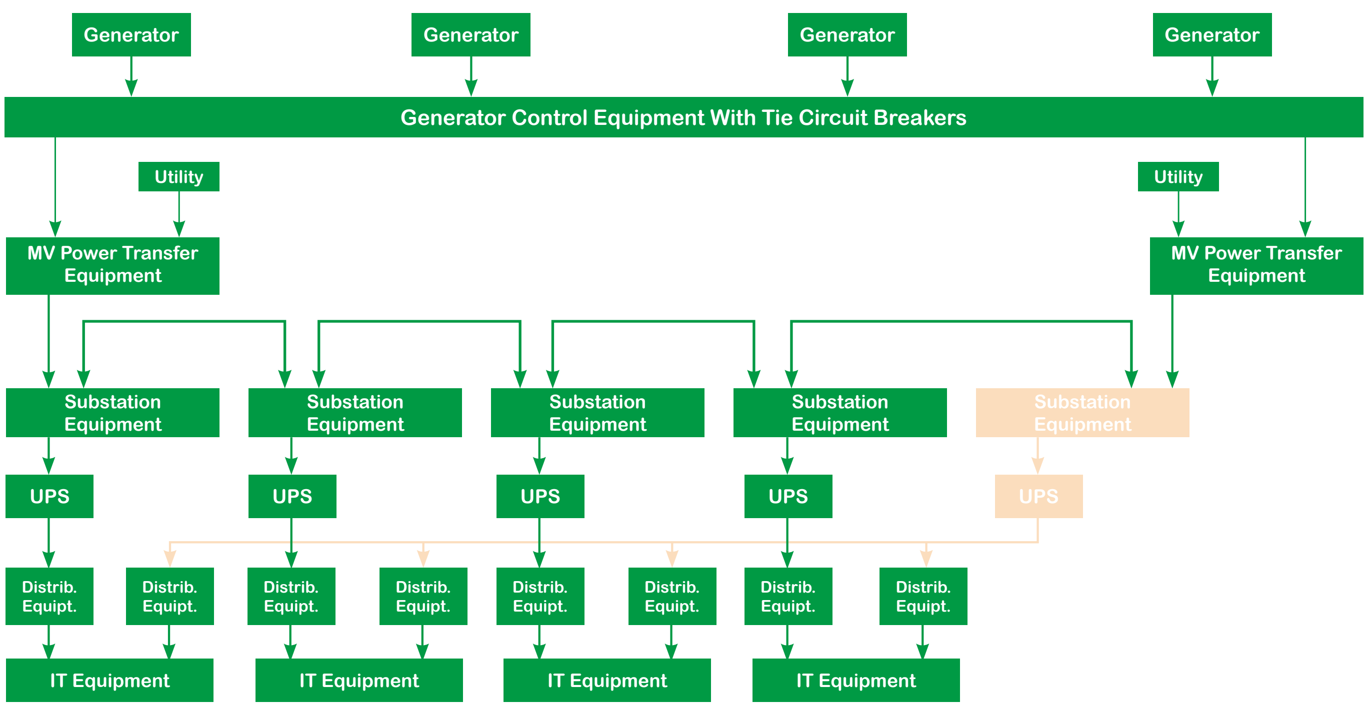

Figure 5: A Tier III Ring Bus Configuration Provides Inherent Pathway Redundancy

Tier IV Power - Fault Tolerant

Tier II and Tier III systems offer redundancy by providing supplemental power through one of two pathways to ensure that capacity needs are met. Tier IV power systems achieve full fault tolerance by providing two entire stand-alone power distribution systems that can each provide the full amount of power needed for all downstream IT loads. In normal operation, the systems operate in parallel at a fraction of their respective total capacities, and each serves a proportional amount of the IT load. This is illustrated in Figure 6. When a power outage or anomaly occurs in one system, the alternate system instantly assumes operation and provides the total amount of power required for all the loads. This state is shown in Figure 7.

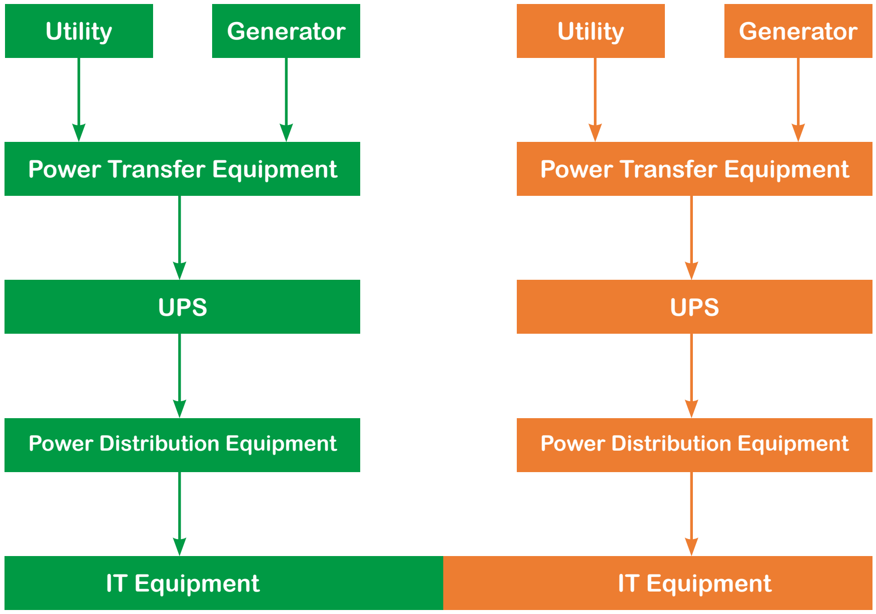

Figure 6: A Tier IV (2N) System Under Normal Operation

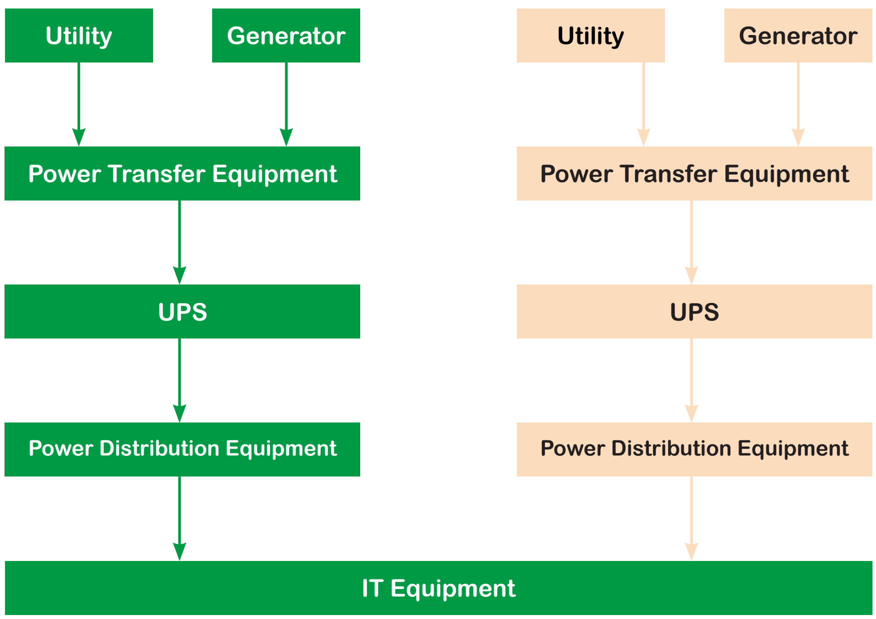

Figure 7: A Tier 4 (2N) Power System Following Loss of One System

Tier IV systems preserve concurrent maintainability because each system can handle the entire IT load. Nevertheless, when running on a single power source, 2N systems are functionally reduced to 1N systems (single components, single path) that offer no further redundancy. This leaves operations vulnerable to the effects of power events until parallel operation of the two complete systems is restored. To reduce vulnerability when in this state, Tier IV systems can be equipped with redundant power components for 2N+1 redundancy. An example of an 2N+1 configuration is shown in Figure 8.

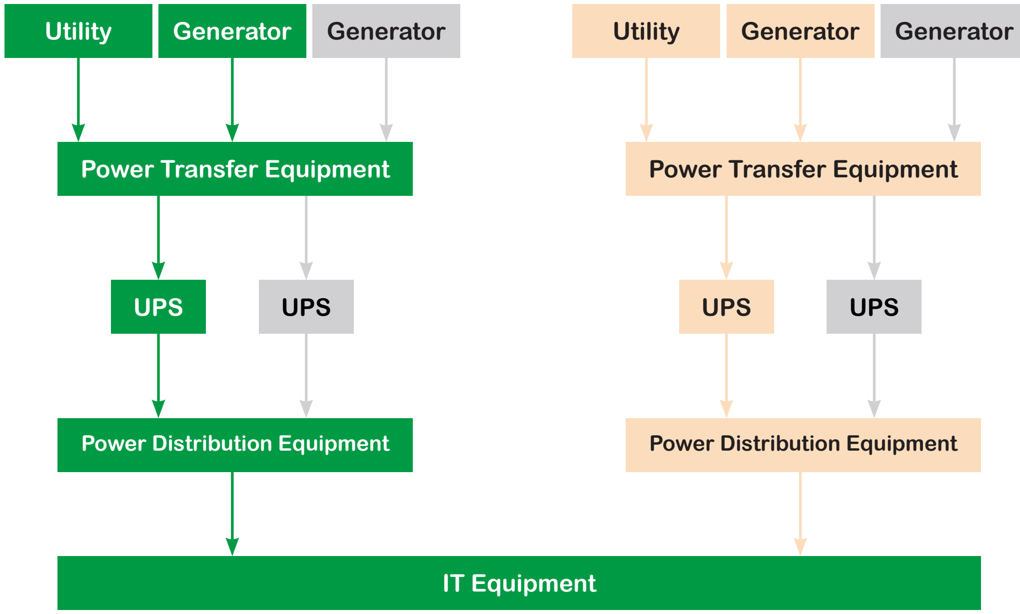

Figure 8: A Tier IV System with N+1 Generators and UPS' (2N+1)

DISCUSSION

Increasing the level of redundancy for power equipment and power pathways increases the associated costs of equipment and the facilities that support them. Increasing redundancy also incurs additional costs for the installation, commissioning, maintenance, testing, and repair of backup power equipment and systems. To choose an appropriate Uptime Institute Tier Level for a power distribution system, designers must consider the data center type, the tangible and intangible costs of potential downtime, and the initial cost of designing and installing an adequate backup power system.

Qualitative Impacts of Power Downtime

The tangible cost of downtime will vary between facilities according to factors such as business model and the scale of operations. They can include (1) the loss of revenue due to disruptions in sales and operations, (2) expenses related to responding and correcting problems following power-related data center outages, and (3) penalties or lost revenues associated with violations of service level agreements. Examples of the intangible impacts include future losses driven by damage to an organization's reputation. For instance, a power-related disruption of an internal human resource tracking system of a small company will result in smaller overall impacts than one that effects a data center hosting cloud computing services and e-commerce operations for a major retailer. While quantification of these impacts is beyond the scope of this document, the following qualitative information offers a perspective.

Relative Benefit of Increasing Redundancy

Tier III and Tier IV systems offer the greatest capabilities for avoiding the impacts of power outages in a single data center facility. In evaluating their relative costs and benefits, it is helpful to clarify how redundancy varies according to Tier III configurations.

Tier III systems supply levels of redundancy according to the number of primary power sources that a secondary source must cover. Where one power source provides redundant service for either of three other sources, the redundancy ratio is 4:3. Notably, Tier III configurations can also allow one system to backup multiple other systems. For instance, where one backup source serves five others, the redundancy ratio is 6:5.

In large data centers, Tier III systems offer the capability to apply a single backup source to numerous primary sources. Decreasing this ratio offers cost efficiencies for equipment design, installation, and operation. Restated, a 6:5 system can be more cost-effective than a 4:3 system. Changing the redundancy ratio can also impact Mean Time Between Failure values, where decreasing the redundancy ratio results in shorter mean time between failures. Consequently, selection of an acceptable power system configuration should only be made after a thorough assessment of the mission of the data center and the associated tangible and intangible costs of power-related service outages.

SUMMARY

A 2018 study by the Uptime Institute has shown that power interruption remains the most common cause of data center service outages. Consequently, power disruptions should continue to be considered as an expected operational condition for which any data center must be prepared. To guide appropriate power system design, the Uptime Institute maintains the following classifications for power system resilience: (1) Tier I - Basic Capacity, (2) Tier II - Component Redundancy, (3) Tier III - Concurrent Maintainability; and (4) Tier IV – Fault Tolerant. Key concepts for each tier are detailed and illustrated in the preceding narrative.

Increasing the level of power system redundancy increases mean time between failures together with the costs of constructing and operating suitable power systems. To choose an appropriate redundancy tier level, designers must compare the mission of the data center to the tangible and intangible costs of potential downtime as well as the overall cost of designing, installing, and operating an adequate backup power system.

ASCO Power Technologies - Global Headquarters
160 Park Avenue
Florham Park, NJ 07932
Tel: 800 800 ASCO
customercare@ascopower.com

www.ascopower.com **whitepapers.ascopower.com**

ASCO. Innovative Solutions.