Using a Reserve Power Architecture to Increase Data Center Infrastructure Utilization and Efficiency

A White Paper on Data Center Efficiency

EMERSON™
Network Power

Introduction

Power system architectures are evolving to address issues of efficiency, scalability and availability. Uninterrupted availability continues to be a key issue for data centers; however, according to the results of the Data Center Users Group survey, released annually by Emerson Network Power since 2007, data center managers are shifting their focus to other issues. Availability ranked as the number one data center infrastructure concern in 2011 before slipping to second in 2012 and then out of the top three in 2013.

It would be dangerous to interpret this shift as a sign that availability is no longer a concern. In both 2012 and 2013, Monitoring and Infrastructure Management, which is being adopted in part because of its ability to improve availability, took the top spot. In 2013 Heat Density replaced Availability in the top three (Figure 1). Of course, heat density is an issue precisely because of its impact on server availability. By 2014, Availability had re-emerged in the top three, finishing tied with Energy Efficiency behind Monitoring and Infrastructure Management.

However, it is also important to note that Energy Efficiency re-emerged as a top three concern in 2012 (it had first appeared in 2007) and rose to the second position in 2013 and 2014. While availability continues to be a concern for data center managers they are increasingly seeking to achieve high availability in the most efficient way possible. For some, that means re-evaluating their power system architecture

Rank	2007	2008	2009	2010	2011	2012	2013	2014
1	Heat Density	Heat Density	Heat Density	Monitoring Infrastructure Management	Availability	Monitoring Infrastructure Management	Monitoring Infrastructure Management	Monitoring Infrastructure Management
2	Power Density	Power Density	Energy Efficiency	Heat Density	Monitoring Infrastructure Management	Availability	Energy Efficiency	Energy Efficiency
3	Energy Efficiency	Availability	Monitoring Infrastructure Management	Availability	Heat Density	Energy Efficiency	Heat Density	Availability

Figure 1. Top data center infrastructure concerns as reported by the Data Center Users Group.

Dual-Bus Architecture

The dual-bus architecture is often selected for enterprise data centers based on its ability to meet two criteria of higher tier data centers: concurrent maintainability and fault tolerance.

The dual-bus architecture traditionally features redundant utility feeds, generators, UPS systems and power distribution systems supporting dual-powered IT equipment. When properly designed, the dual-bus system eliminates single points of failure in the critical power system because a failure in any component is isolated. This also allows maintenance to be performed on any component while continuing to power the load (Figure 2).

The price for achieving fault tolerance and concurrent maintainability in the 2N dual bus architecture is low utilization rates for power system components. To ensure safe operating conditions when one bus is carrying the full load, power system components on each bus, including the UPS system, are typically sized at about 110 percent of the data center load. As a result, in normal operating conditions UPS modules operate below 45 percent utilization. In periods of low demand, such as when the data center is first deployed, utilization can dip to 20 percent or lower.

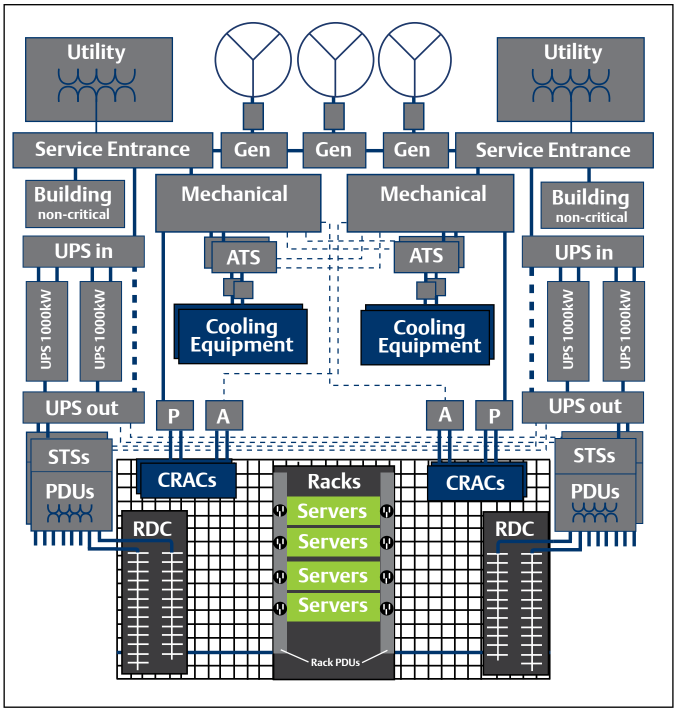


Figure 2. The 2N dual-bus architecture represents a proven approach to achieving fault tolerance and concurrent maintainability.

This can lead to reduced system efficiency. Most UPS systems operate most efficiently at utilization rates above 30 percent. At 20 percent utilization, efficiency begins to drop (Figure 3).

This may not be a serious concern for smaller data centers. Power system losses account for a relatively small percentage of data center power usage and achieving a 2 percent increase in UPS efficiency by operating at higher utilization rates isn't enough of an incentive to outweigh the other benefits of the dual-bus architecture. However, as data centers continue to get larger to support hosted and cloud-based computing, the low level of utilization inherent in the traditional dual-bus architecture has a larger impact on operating costs.

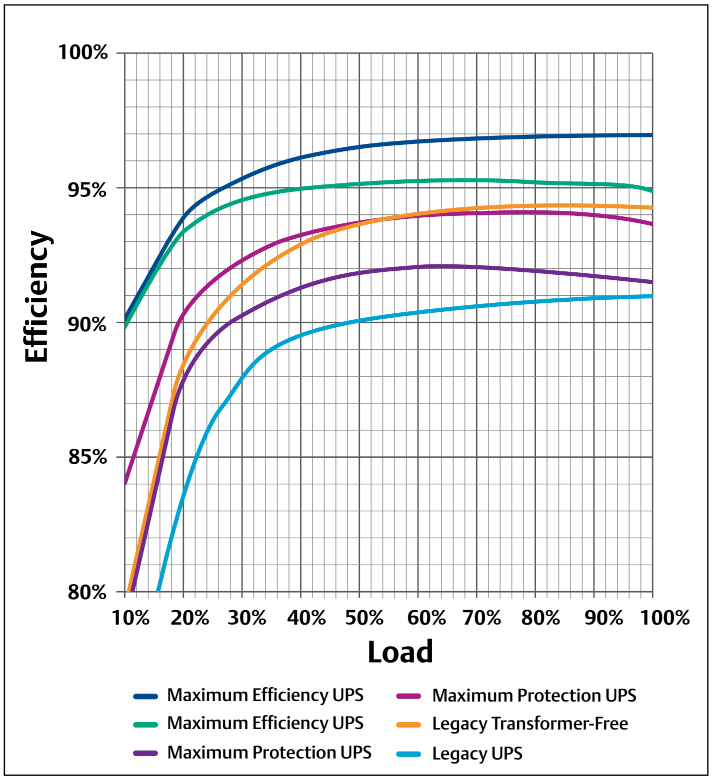


Figure 3. UPS efficiency varies by load and type of UPS

In addition, the complexity of the dual-bus UPS system increases as more UPS modules are required to support the load. For example, a 1 MW data center would typically be designed with a power system with a capacity of 1.1 MW to eliminate the risk of overload at full capacity. In a 2N system, this could be accomplished by redundant UPS systems, each of which consists of two 550-600 kW UPS modules for a total of four modules in the entire system.

However, when data center capacity is 16 MW the system gets more complicated. A 16 MW data center could require as many as 16 1100 kW modules per bus for a total of 32 modules in the system. Both floor space requirements and maintenance schedules become concerns when that many modules are required.

As a result, a new architecture has emerged for larger data centers that preserves the concurrent maintainability and fault tolerance of the 2N dual-bus architecture while increasing system efficiency and reducing the number of UPS modules in large data centers. This architecture is most often called the **reserve architecture**. It is also sometimes referred to as a **catcher bus architecture** or described by the total number of modules in relation to the primary modules. For example a system with six primary modules and two reserve modules may be described as “8 to make 6.”

The reserve or catcher bus architecture is now being considered by both large and midsize data centers, including cloud hosting and colocation facilities, as an alternative to the traditional 2N dual bus architecture. It offers a similar level of availability with greater efficiency and lower capital costs.

The Reserve Architecture

The reserve architecture essentially creates an N+1 or N+ 2 architecture within the UPS system, while maintaining fault tolerance and concurrent maintainability through the use of static transfer switches (STS). The STS allows a redundant UPS system to be brought online to pick up the load from any one of multiple UPS systems in the event of failure or maintenance. Downstream from the STS units, the power distribution system can be similar in design to that of a 2N dual-bus architecture.

Through this arrangement, the reserve architecture allows the UPS system to operate at utilization rates of approximately 75 percent or higher compared to less than 50 percent for a traditional 2N dual bus architecture (Figure 4).

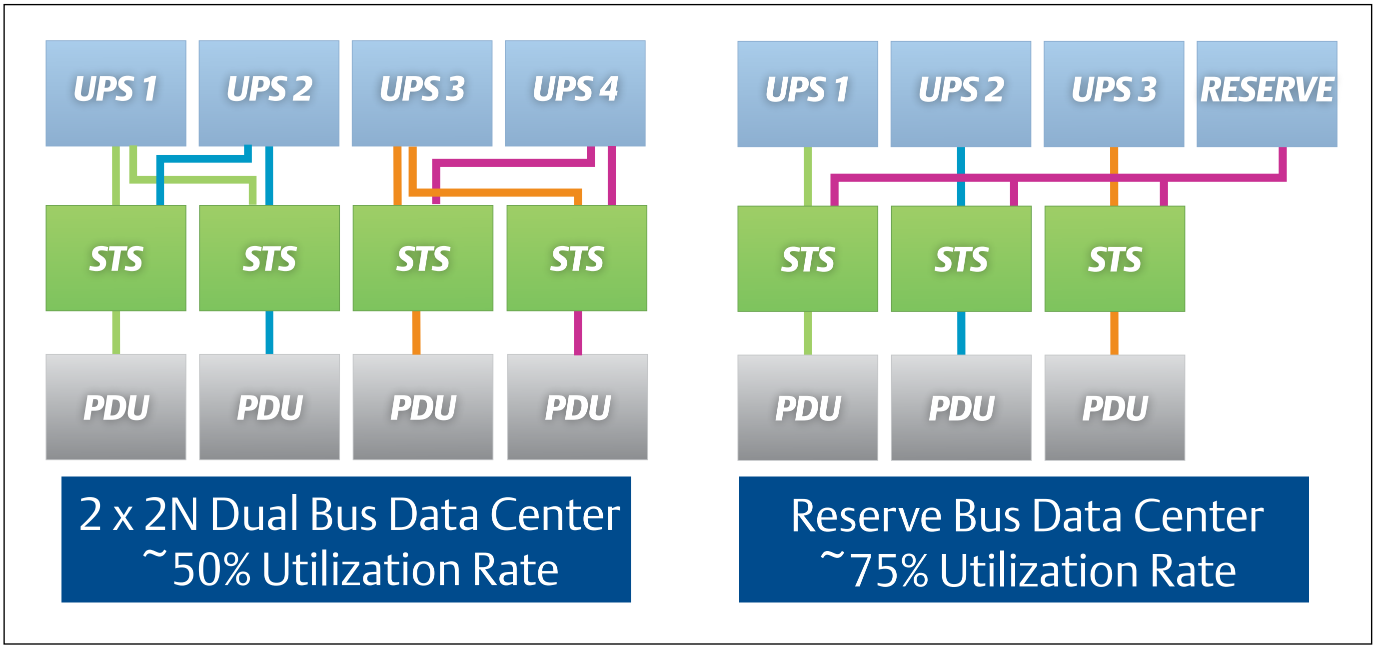


Figure 4. The reserve architecture creates a fault tolerant architecture with utilization rates of approximately 75 percent.

Figure 5 shows the design of a reserve system architecture for a cloud or colocation facility. In this example, expansion can be managed through 2 MW IT increments that are added as the need for customer load grows.

In this example, the reserve system includes two redundant modules (N + 2) to allow the system to retain fault tolerance and redundancy during UPS maintenance. The reserve system includes reserve utility transformers, generators, and UPS systems. Note that any system's IT load can be transferred to the reserve system to enable maintenance of a primary system. Downstream, between the STS units and IT load, a dual-bus architecture is deployed, allowing complete end-to-end concurrent maintenance -- each component can be taken down for maintenance on a regular basis without affecting IT system availability.

From a fault tolerance standpoint, a fault upstream of the STS, such as a generator not starting, will cause that system to transfer to the reserve system. Downstream of the STS units, the dual-bus architecture provides for fault tolerance of the complete system.

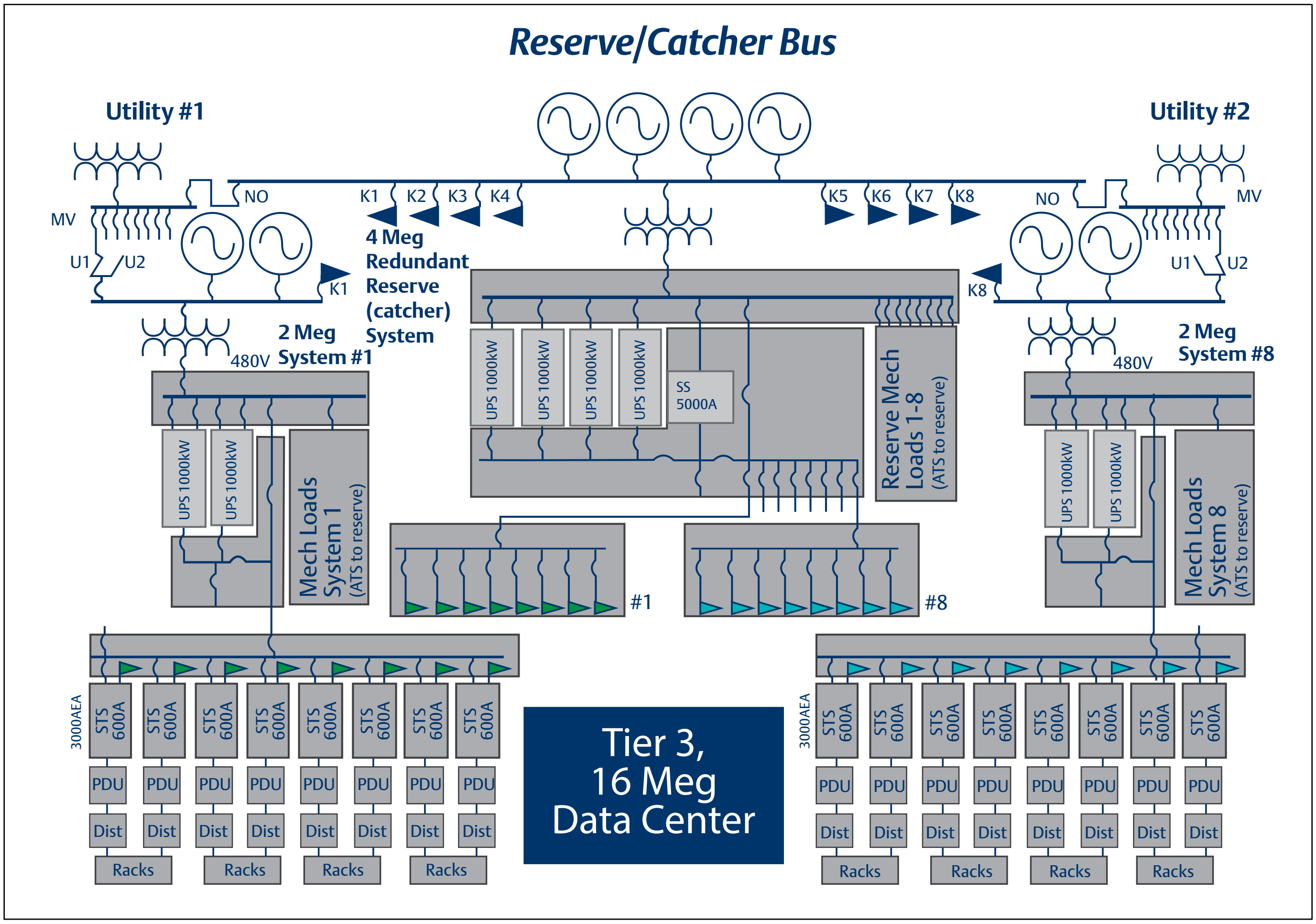


Figure 5. Reserve power architecture for a 16MW data center. Primary power modules 2-7 are not shown in the diagram because of space constraints but are identical to modules 1 and 8.

The reserve mechanical system details are shown in Figure 6. Note that the mechanical system, like the electrical system, allows each component to be taken down for maintenance and provides fault tolerance for each system.

This addresses one of the key threats data centers face when operating on backup power; if the mechanical systems are not supported by backup power, heat buildup in the data center can threaten continued operation.

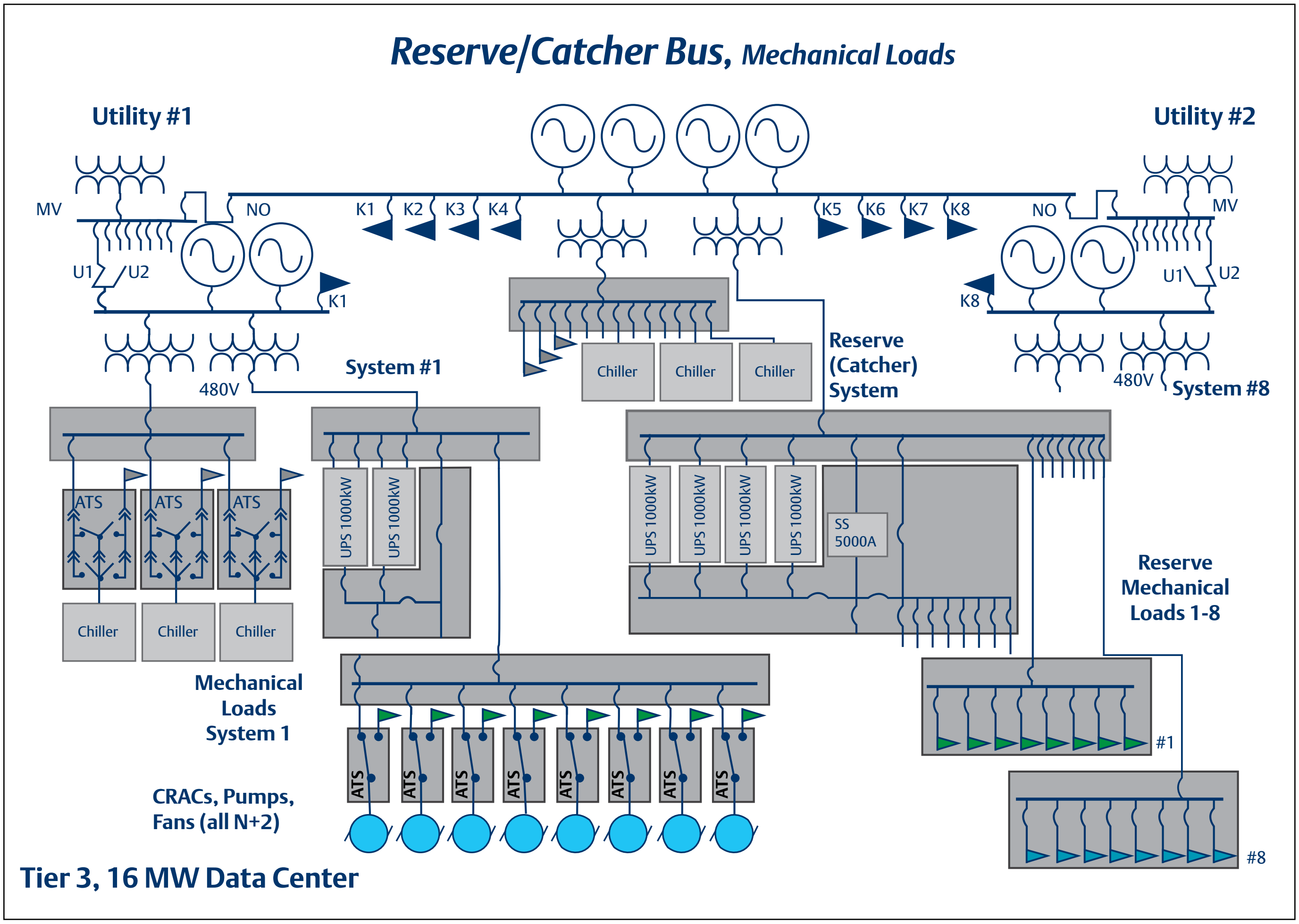


Figure 6. In this data center the reserve architecture supports the mechanical loads with fault tolerance and concurrent maintainability.

One of the reasons the reserve architecture makes sound economic sense, particularly for large data centers, is the economics of UPS module size: cost per kW goes down as module size increases (Figure 7). An economical power system design is one that uses the largest possible modules while maintaining an appropriate degree of scalability based on customer needs. Using fewer, larger modules also has the advantage of improving the overall reliability of the system (Figure 8).

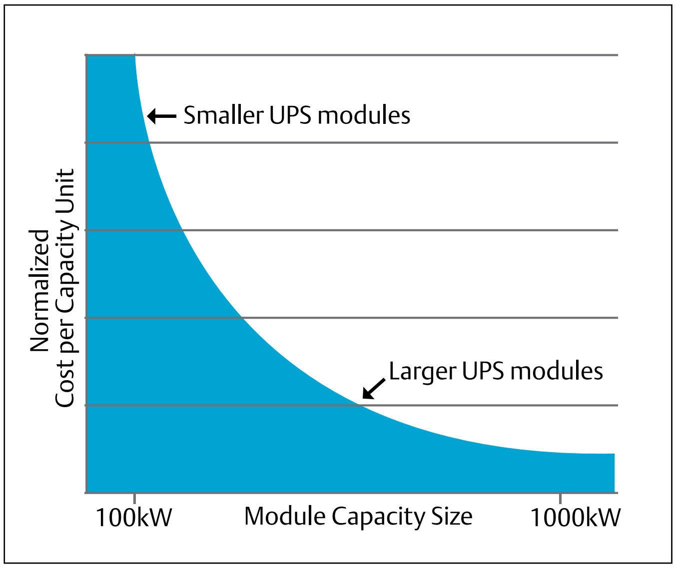


Figure 7. UPS cost/kW goes down as module size increases.

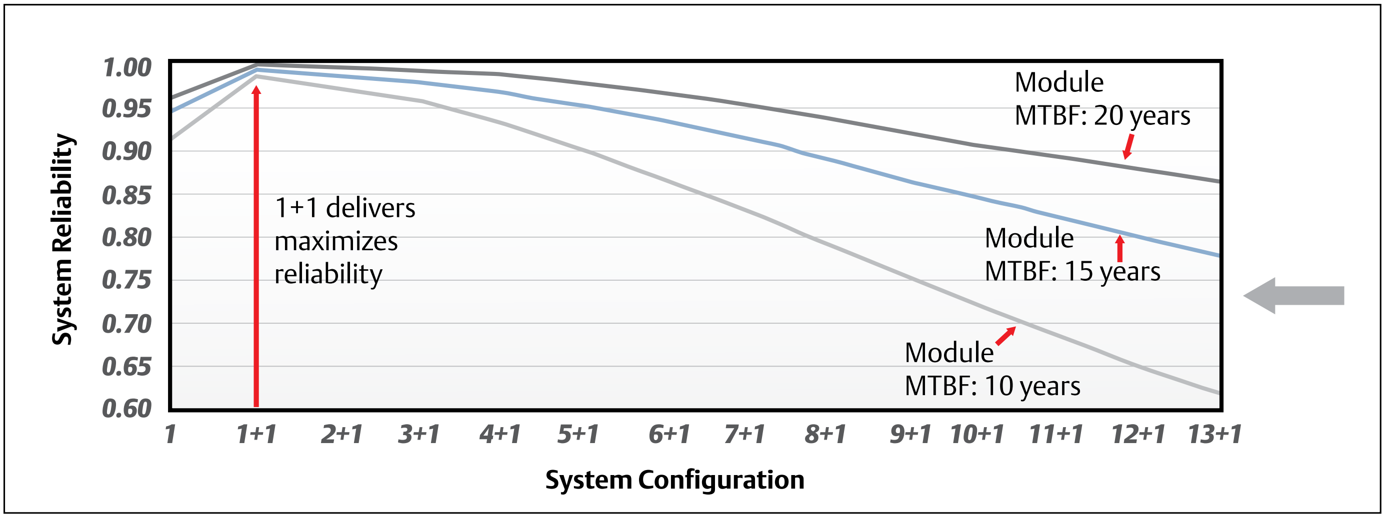


Figure 8. UPS system reliability is highest in a 1+1 configuration and decreases as more modules are added.

These same principles also hold true for the generator systems and can be applied when evaluating generators within the same speed/rpm rating.

For data centers at or above 500 kW, designers can utilize large systems to create reserve architectures that are fault tolerant, concurrently maintainable and achieve utilization rates that optimize power system efficiency. The reserve system can also deliver greater flexibility in accomodating growth. Modules can even be deployed on skids to simplify deployment of additional power capacity (Figure 9).

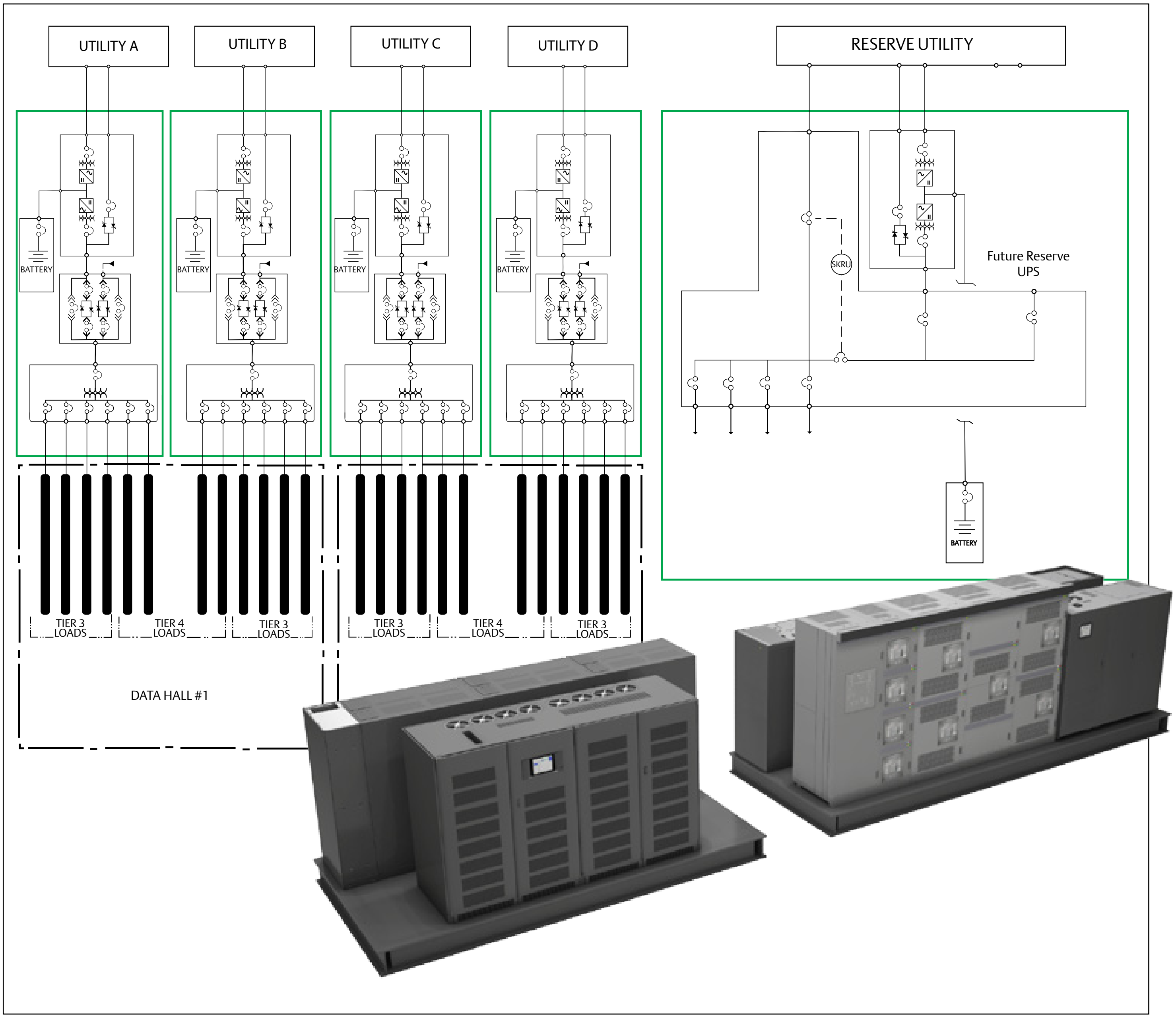


Figure 9. Skid-based deployments simplify growth

Comparing Reserve and 2N Dual Bus Architectures

Any critical power system must be tailored to the organization's business objectives and specific technology profile as well as the size and criticality of the data center.

Where availability is paramount and UPS modules can be sized to minimize the number of modules in the system, a traditional 2N architecture may still prove to be the right choice. However, for businesses seeking to balance efficiency and availability or where data center capacity makes a 2N architecture unwieldy, the reserve architecture offers a solution that delivers very high reliability with fault tolerance and concurrent maintainability while providing superior efficiency, scalability, and resource utilization (Figure 10).

Conclusion

The reserve system architecture represents a viable high-availability approach to critical power for any data center above 500 kW. As data center size increases, the savings from the reserve architecture, in both operating and capital costs, grow and the time to ROI gets shorter. The reserve architecture is becoming the power system of choice for large cloud and colocation facilities and is increasingly being considered by midsize facilities seeking to implement modular growth strategies.

	2N Dual Bus	Reserve Architecture
Reliability	★★★★	★★★⯪
Maintainability	★★★★	★★★★
Efficiency	★★	★★★
Scalability	★★	★★★★
Resource Utilization	★★	★★★⯪
Initial Cost	★★	★★★
ROI	★★⯪	★★★⯪

Figure 10. A traditional 2N dual-bus delivers the highest reliability but in many cases the reserve architecture's superior flexibility and efficiency will make it the right choice.

Appendix:
Understanding UPS System Redundancy Configurations

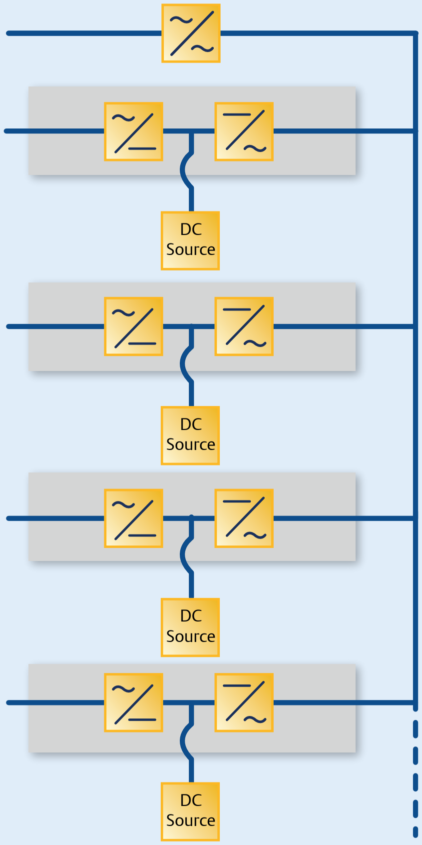
There are a variety of ways to configure a UPS system to achieve redundancy. Here are the most common.

Parallel Redundant N+1, Central Bypass

The parallel redundant system uses the number of UPS modules required to handle the maximum load (N) plus one (typically). During normal operation, the load is shared equally across all modules, which behave as if they were a single large UPS system. If a single module fails or needs to be taken offline for service, the UPS system will continue to power the load, although it will lose redundancy while the module is offline. The system provides a higher level of fault tolerance, and acceptable utilization.

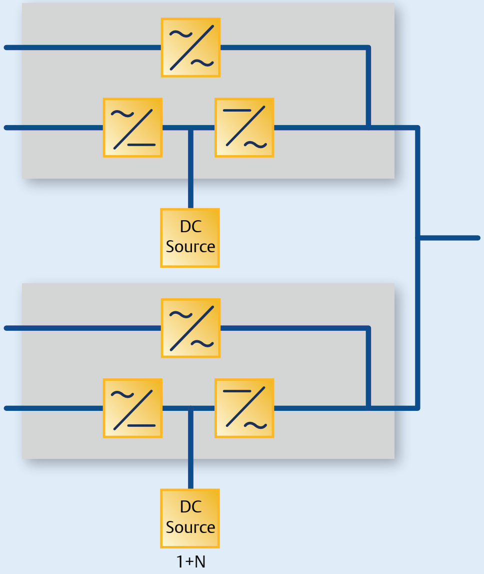
Parallel Redundant N+2

The N+2 configuration is similar to an N+1 system but includes one additional UPS module to maintain redundancy when one unit is taken offline for maintenance. N+2 is sometime referred to as a break one / fix one architecture. System utilization becomes lower as the number of redundant UPS modules increases.



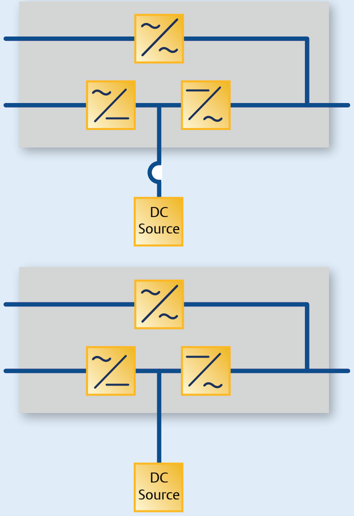
Parallel Redundant 1+N, Distributed Bypass

Offers paralleling of single module UPS units for redundancy or capacity, with each module having its own bypass static switch. This is a cost effective method to parallel systems. The system provides some level of fault tolerance and acceptable utilization.



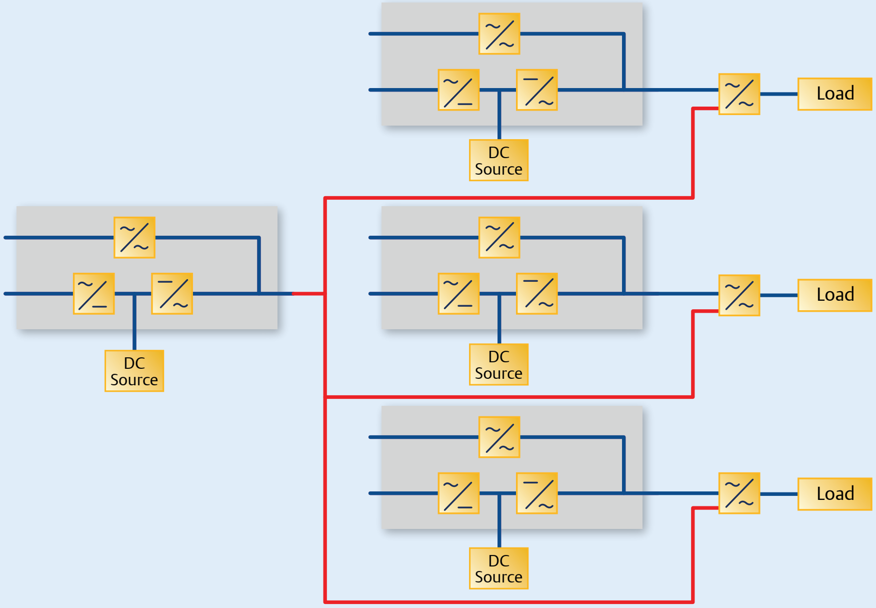
Dual Bus or 2N

Features two identical UPS systems and supporting power systems, each capable of carrying the full load. During normal operation the two systems support the load. Each system may have redundancy built-in to maintain redundancy while one system is offline. This system provides the highest level of fault tolerance; however, utilization is limited to a maximum of 50 percent.



Reserve or Catcher Bus

Creates a redundant overall system architecture, and can be created with downstream power distribution similar to a dual bus 2N architecture. This allows the UPS to operate at higher utilization rates, while providing a highly fault tolerant power system design.



**Emerson Network Power-
Global Headquarters**
1050 Dearborn Drive
P.O. Box 29186
Columbus, Ohio 43229
United States Of America
800 877 9222 Phone (U.S. & Canada Only)
614 888 0246 Phone (Outside U.S.)
614 841 6022 FAX
Contact@EmersonNetworkPower.com
liebert.com

Emerson Network Power **EmersonNetworkPower.com**

SL-24682 (R11/14) Printed in USA

EMERSON. CONSIDER IT SOLVED.™