Guidelines for Specifying Data Center Criticality / Tier Levels

By Victor Avelar

White Paper #122

APC
Legendary Reliability®

Executive Summary

A framework for benchmarking a future data center's operational performance is essential for effective planning and decision making. Currently available criticality or tier methods do not provide defensible specifications for validating data center performance. An appropriate specification for data center criticality should provide unambiguous defensible language for the design and installation of a data center. This paper analyzes and compares existing tier methods, describes how to choose a criticality level, and proposes a defensible data center criticality specification. Maintaining a data center's criticality is also discussed.

Introduction

Terms like availability, reliability, mean time between failure (MTBF), and others are oftentimes used interchangeably to describe data center performance. These terms are quantitative measures of performance that are difficult for data center managers to calculate. An alternative and simplified approach is to categorize data center performance in tiers or criticality levels. This paper proposes that the term criticality be used to subjectively describe data center performance.

A data center's criticality has arguably the strongest influence on lifetime total cost of ownership (TCO). For example, a fully redundant (2N) power architecture could more than double the 10-year TCO of a non-redundant (1N) power architecture. Although a significant cost penalty for 2N power is the doubling of electrical equipment capital costs, the greater impact comes from the energy costs associated with operating and maintaining the power equipment at 2N. Therefore, when choosing a data center's criticality, a data center designer or owner needs to weigh both the costs and the criticality in order to establish a true cost / benefit analysis.

This paper describes and compares three common methods for specifying data center criticality. Guidance is given on how to choose a criticality by presenting typical levels for various applications and environments. Defensible approaches for specifying data center performance are discussed.

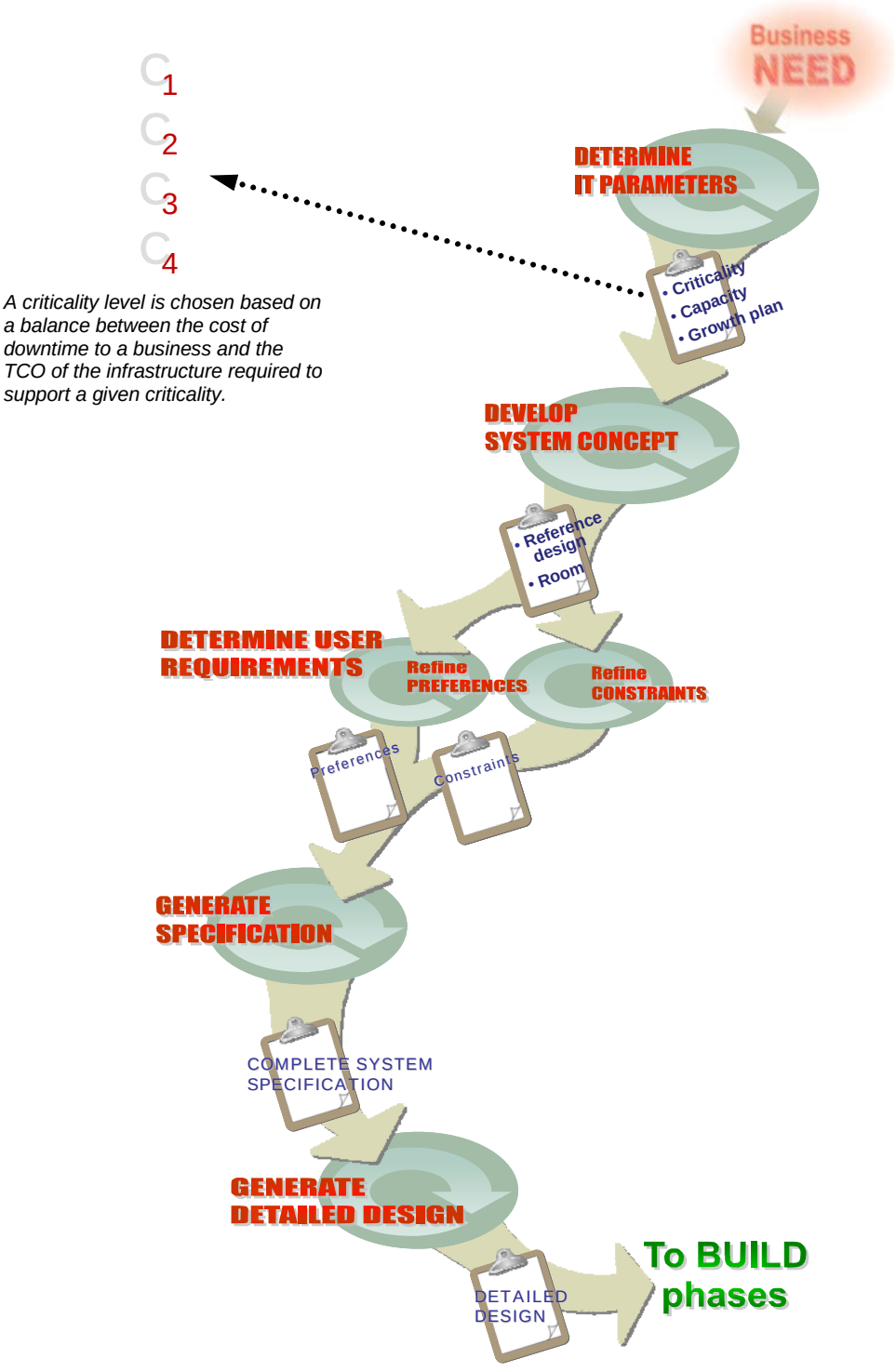
Data Center Project Planning

In a data center construction or upgrade project, it is the first half of the process – the planning portion – that offers the greatest opportunity for errors and oversights and is when a data center's criticality should be specified.[1] (See **Figure 1**) Specifically, a needs assessment identifies and quantifies the constraints and preferences related to the data center plan. A specification is then generated that satisfies these constraints and preferences. When the specification is agreed upon, a detailed design can proceed and finally be implemented. Once the data center is built it can be validated it against the specification. Validation against a specification allows legal recourse against substandard or deceptive workmanship.

Choosing a data center's criticality represents a major decision in the planning process since it impacts so many other decisions especially for green-field projects including location, building type, fire suppression, security system, and many others. The planning phase allows designers to balance the TCO of a data center with the preferences and constraints of a business's availability requirements. It is through this iterative planning exercise that a final criticality is specified. The subject of data center planning is discussed further in APC White Paper #143, "Data Center Projects: System Planning Sequence".

[1] Excerpted from APC White Paper #142, "*Data Center Projects: System Planning*", 2007

Figure 1 – Choosing a criticality is a key step in the system planning



Common Classification Methods

Historically, the performance of a data center depended largely on the people involved in the design process. To craft a solution, individuals generally fell back on unique personal experiences, anecdote, hearsay, and legend, placing special emphasis on the design attributes that they historically understood were the most important drivers of downtime. The result is enormous variation in data center designs, even when the same requirements are stated. This has prompted the development of various criticality or tier categories to help specify the availability and reliability performance of data center designs. Specification of data center performance becomes easier by having simple categories of design architectures that can be benchmarked against each other.

There have been various methods introduced throughout the mission critical facility industry some better known than others. Three more commonly known methods are The Uptime Institute's Tier Performance Standards, TIA 942, and Syska Hennessy Group's Criticality Levels™.

The Uptime Institute's Tier Performance Standard

Though not a standards body, The Uptime Institute pioneered its tier classification method in 1995 and has become widely referenced in the data center construction industry. Uptime's method includes four tiers; Tier 1 through Tier 4, which have evolved over the years through various data center projects. This method provides a high level guideline but does not provide specific design details for each Tier.

TIA 942

The four tier levels described in TIA 942 revision 5 are based on Uptime Institute's Tier Performance Standards. Although 942 is a standard, the four tier levels described in appendix G are "informative and not considered to be requirements of this Standard"[2]. Nonetheless, appendix G does provide specific design criteria that can help designers build to a specific tier level and allows data center owners to evaluate their own design.

Syska Hennessy Group's Criticality Levels

Syska's ten criticality levels build on Uptime's four tiers by considering recent data center trends such as high density computing and flexible architectures. Although the Syska method includes ten levels, it maps the first of its ten criticality levels to Uptime's four tiers[3]. Syska also includes more comprehensive elements that evaluate the maintenance and operation of a data center and not just the "upfront" components and construction. In addition, they pioneered the balance sheet approach to data center criticality levels by recognizing that data center performance is only as strong as its weakest element. Syska's Criticality Levels are described at a high level and lack the specificity of TIA-942.

---

[2] TIA-942, "*Telecommunications Infrastructure Standard for Data Centers*", April 2005, p. 10
[3] Syska Hennessy Group, Inc., "*Syska Criticality Level™ Definitions*", April 2005, p. 13

Comparison of methods

Overall, all three methods support the idea that there are four common levels, numbered (1, 2, 3, and 4) of criticality / tiers commonly used today. The biggest problem encountered with the Uptime method and Syska method, is the lack of detail needed to articulate the differences between levels. The TIA-942, in contrast, provides specific details at every tier level and across a wide range of elements including telecom, architectural, electrical, mechanical, monitoring, and operations. For example, the TIA-942 specifies that a tier 2 data center should have two access provider entrance pathways that are at least 20 m (66 ft) apart. Syska also specifies that a tier 2 data center should have two entrance pathways but adds no other detail. Publicly available Uptime documentation does not provide guidance for access provider entrance pathways.

Until recently, only Syska Hennessy Group explained the importance of balancing the levels of the various systems that make up a data center. Uptime discusses this concept of balancing in its updated 2006 standard. Overall, based on the available literature, no major design practices from these methods conflict with each other. For a full comparison of these three methods against various characteristics see **Table A1** in the appendix of this paper. Ultimately, these three organizations, and others like them, have advanced the data center industry towards a higher level of performance.

Difference between "shall" and "should"

It is important to understand that a data center manager cannot verify compliance against any of the methods described above. In order to verify compliance, and be legally defensible, detailed specifications are required. Specifications for a level 1, 2, 3, or 4 data center project set a requirement for how the final data center performs. In order for specifications to hold up in the court of law, they must use the word "shall or must" rather than "should". The word "should" or "may" conveys a recommendation and is not legally binding. The word "shall" or "must" conveys an action that is binding and legally defensible. This is the type of specification language that allows a data center manager to legally verify and mandate that their data center is in compliance with the design specifications. For example, assume a contractor was hired to convert a conference room into a criticality 2 data center; the specification in **Figure 2** would apply. If the contractor were using specification "1a" and did not remove all the windows (assuming windows weren't required by safety code), the contractor would be legally held responsible for removing the windows. If specification "1b" were used, there would be no legal recourse for the data center owner.

Figure 2 – Use of "shall" and "should"

1a. Data center shall have no exterior doors or exterior windows, unless required by safety code.

1b. Data center should have no exterior doors or exterior windows, unless required by safety code.

In the case of Uptime and Syska literature, there are no formal specifications as in the example above, nor does the literature use the word "shall" or "must". Although TIA does provide some detailed specifications, they use the word "should" instead of "shall". For example "Tier 3 installations should meet all requirements of tier 2".

Balanced Criticality

When specifying a data center's criticality, all attributes should be considered as links in a chain and that the overall data center performance is only as strong as its weakest link. Syska Hennessy developed this notion of a comprehensive balance between all parameters in their Criticality Levels approach. A very serious pitfall in designing data centers is that the investments are imbalanced, and great expense is invested in one area of data center design while another area is overlooked. Oftentimes decisions are made to enhance the availability of a particular system without giving thought to other systems in a data center. A classic example is spending a disproportional amount of engineering focus and expense on the UPS system compared to cooling and security systems. This oversight creates false expectations because IT managers believe their entire data center is at the same level of criticality as the UPS only to be disappointed when a security breach results in downtime. Without the ability to find and quantify the "weak link", data centers provide sub-optimal business results.

However, there are instances when it is appropriate to over-specify the criticality of a particular system. For example, a criticality 1 may be all that is required for a remote data center. Yet, the IT manager may require a criticality 3 for the management system to allow control and monitoring of systems that are not easily reachable by other personnel. In another example, a manager may require a criticality 2 data center which calls for a single power path. However, due to a high likelihood of human error taking down the power system, the manager may opt for a criticality 3 or 4 power path which includes dual (2N) power paths.

Suggested Approach for Choosing Criticality

Choosing an optimal criticality is a balance between a business's cost of downtime and a data center's total cost of ownership. However, the choices may be limited depending on whether a new data center is being built, or changes are being made to an existing one. In reviewing the available literature, it is clear that all three methods discussed in the previous section share a common understanding of what it means to be a criticality / tier level 1, 2, 3, or 4. **Table 1** provides business characteristics for each criticality and the overall effect on system design.

Table 1 – Summary of criticalities

Criticality	Business characteristics	Effect on system design
1 *(Lowest)*	• Typically small businesses • Mostly cash-based • Limited online presence • Low dependence on IT • Perceive downtime as a tolerable inconvenience	• Numerous single points of failure in all aspects of design • No generator if UPS has 8 minutes of backup time • Extremely vulnerable to inclement weather conditions • Generally unable to sustain more than a 10 minute power outage
2	• Some amount of online revenue generation • Multiple servers • Phone system vital to business • Dependent on email • Some tolerance to scheduled downtime	• Some redundancy in power and cooling systems • Generator backup • Able to sustain 24 hour power outage • Minimal thought to site selection • Vapor barrier • Formal data room separate from other areas
3	• World-wide presence • Majority of revenue from online business • VoIP phone system • High dependence on IT • High cost of downtime • Highly recognized brand	• Two utility paths (active and passive) • Redundant power and cooling systems • Redundant service providers • Able to sustain 72-hour power outage • Careful site selection planning • One-hour fire rating • Allows for concurrent maintenance
4 *(Highest)*	• Multi-million dollar business • Majority of revenues from electronic transactions • Business model entirely dependent on IT • Extremely high cost of downtime	• Two independent utility paths • 2N power and cooling systems • Able to sustain 96 hour power outage • Stringent site selection criteria • Minimum two-hour fire rating • High level of physical security • 24/7 onsite maintenance staff

Greenfield data center projects

Building a new data center presents few constraints to choosing a data center criticality. Generally this decision comes down to what type of business the data center is supporting. Choosing a criticality associated with a critical business characteristic provides a fairly accurate and easy starting point. **Table 2** provides costs estimates for each level of criticality.[4] It is important understand the assumptions behind these costs estimates. **Table 3** associates various business applications with specific criticalities.

Table 2 – Estimated construction costs for each level of criticality

Item	C1	C2	C3	C4
Fit out of physical infrastructure – i.e. power, cooling ($ / watt)	$10	$11	$20	$22
Land	Strongly dependent on location			
Building core and shell ($ / ft²) [$ / m²]	$220 [$2,368]			

[4] Turner, Seader, "*Dollars per kW plus Dollars per Square Foot Are a Better Data Center Cost Model than Dollars per Square Foot Alone*", 2006. White paper available at http://www.upsite.com/cgi-bin/admin/admin.pl?admin=view_whitepapers

Table 3 – Typical levels of criticality for different applications

Applications	C1	C2	C3	C4	Descriptions
Professional services					Consulting, property management
Construction & engineering					Mission critical facility designers
Branch office (financial)					Local neighborhood bank office
Point of sale					Department store, home goods
Customer Resource Management (CRM)					Customer data
7x24 support centers					Dell customer service
University data center					Online assignments, email, tuition
Enterprise Resource Planning (ERP)					Business dashboards, metrics
Online hospitality & travel reservations					Online airline ticketing
Local real time media					Local news channel
Online data vaulting and recovery					Consumer and company backup
Insurance					Auto and home insurance
Work-in-progress tracking (manufacturing)					Automobile manufacturer
Global real time media					Nationwide news show
Voice over IP (VoIP)					Converged network
Online banking					Checking, bill pay, transfers
Hospital data center					Hospital in metropolitan area
Medical records					Health insurance
Global supply chain					Jetliner manufacturer
E-commerce					Online book store
Emergency call center					911 (U.S), 112 (E.U.)
Energy utilities					Electric, gas, water
Electronic funds transfer					Credit cards, electronic checks
Global package tracking					Letters, parcels, freight
Securities trading and settlement					Equities, bonds, commodities

Existing data center projects

In general, for existing data center projects (i.e. retrofit), choosing a criticality is limited to the constraints of the existing structure. For example, if an existing structure is located in a 100-year flood plain, it could not be a criticality 2 data center. Those responsible for articulating the proposed data center criticality must first identify the major constraints, such as this one, and decide if the resulting data center criticality is an acceptable risk to the business. In this case, if a criticality 1 data center is too risky, then the constraint could be removed by choosing an alternate location that allowed for a criticality 2 data center.

Specifying and Verifying Criticality

Once a criticality has been chosen, the next steps are to specify the criticality, build the data center, and validate it against the specification. Validation against a specification allows legal recourse against substandard or deceptive workmanship. Simply choosing a Criticality Level from Syska, or some other organization, does not constitute a verifiable and defensible specification. These are methods of categorizing data center performance and do not include detailed specifications written in "shall" or "must" language against which a constructed data center can be validated. If a data center was specified by choosing a particular level or tier, verifying that the as-built data center meets that level or tier is impossible without the creators of the method verifying it themselves. A verifiable and defensible specification would allow anyone to validate a data center.

In general, a data center specification describes the essential requirements of performance, interoperability, and best practice that will allow all physical infrastructure elements to work together as an integrated whole. An effective specification is void of detailed descriptions of specific products and instead defines a data center by using unambiguous, point-by-point specification of its physical systems in combination with a standardized process for executing the steps of its deployment. The "baseline" specifications should describe a criticality 1 data center, and provide additional specifications identifying higher level criticality numbers (i.e. 2, 3, and 4). Higher level criticality specifications should be clearly labeled with some kind of symbol to alert the reader. **Figure 3** shows an example of a baseline specification item followed by its associated high-level criticality item. The symbol "C3+" indicates that the specification applies to both criticality 3 and 4.

Figure 3 – Example of a baseline specification item

1. Generator(s) installed outdoors shall be sheltered by an enclosure.
2. Walk-in enclosures shall house all generator mechanical, electrical, and fuel systems. C3+

Business managers that are beginning a data center project can obtain a data center specification in several ways. If the business has a corporate real estate department, it may have personnel with experience in designing and building data centers. These individuals could create a data center specification described above. Architectural & engineering firms that specialize in critical facilities, such as Syska Hennessy Group, can create a data center specification for businesses that do not possess this expertise in house. Alternatively, businesses that are planning a small to medium sized data center project and are able to adopt standardized specifications, can obtain a complete specification for little to no cost. An example of a specification meeting the requirements above is APC's "Small / Medium Data Center System Specification and Project Manual".

Maintaining a Specified Criticality

Even after building a data center and validating it against a defensible specification, it is possible and probable that the data center will fall below the criticality it is was originally designed for. Over time, changes in business initiatives, technology, personnel, and management, all contribute to this problem. For example, many companies have migrated to higher density IT equipment to conserve floor space, which has led to the loss of cooling redundancy. As rack power densities increase, the redundant capacity of data center cooling units is used instead to provide additional airflow to these high density racks. Even redundant power architectures are susceptible to being demoted from N+1 to 1N due to IT refreshes.

Unless physical infrastructure systems are monitored by a capacity management system, it becomes increasingly difficult to maintain a specified criticality. Capacity management systems provide trending analysis and threshold violation information on parameters such as redundancy, air temperature, power distribution, runtime, battery voltage, and any others that affect a data center's criticality over time. This ensures adequate advance notice and information necessary for procurement and deployment of additional capacity. In fact, a data center can not achieve a criticality of 4 unless it is monitored by a capacity management system. Large enterprises usually have a building management system (BMS) which can serve as a capacity management system. For smaller businesses, a centralized physical infrastructure management platform can provide capacity management at a lower price per data point. An example a centralized management platform with capacity management is APC's InfraStruXure Central.

Conclusion

One of the key inputs to planning a data center is criticality. Conventional criticality or tier methods for specifying and validating data center performance are ambiguous and indefensible because they lack detailed specifications. An effective specification for data center criticality should provide unambiguous defensible language using the word "shall" or "must". It is with this type of specification that the as-built criticality of a data center can be validated. Given the rate of IT refreshes, it is equally important to maintain a data center's criticality over time. A capacity management system can monitor and track changes to a data center's physical infrastructure and notify managers when a data center's criticality falls below thresholds.

About the author:

Victor Avelar is a Strategic Research Analyst at APC. He is responsible research in data center design and operations and consults with clients on risk assessment and design practices to optimize the availability of their data center environments. Victor holds a Bachelor's degree in Mechanical Engineering from Rensselaer Polytechnic Institute and an MBA from Babson College. He is a member of AFCOM and the American Society for Quality.

Appendix

Table A1 – Comparison of three criticality approaches

Characteristic	TIA / EIA 942	Uptime Tiers	Syska Criticality Levels
Available defensible specifications to validate a data center design	Offers guidelines but are not written in defensible language	No specifications but Uptime reserves the right to determine Tier ranking and to certify sites as meeting Tier requirements	No specifications but Syska uses evaluation teams to assign the criticality level to data centers
Balance of criticality	Based on weakest infrastructure component or system	Based on weakest infrastructure component or system	Based on weakest infrastructure component or system
Fire suppression and security design used in determination of rating	Uses both fire suppression and security	"independent of the site infrastructure Tier classification"[5]	Uses both fire suppression and security
IT processes used in determination of rating	No	No	Used in criticality level evaluation
Floor loading capacity used in determination of rating	Yes	No	No
Maintenance processes (documentation upkeep and organization) used in determination of rating	Not used	Used in verification of site sustainability but not part of the site infrastructure Tier classification	Used in criticality level evaluation
Site selection	Discussed in depth in annex F as part of the overall tier guidelines	Used in verification of site sustainability but not part of the site infrastructure Tier classification but provides no written guidance	Used in criticality level evaluation but provides no written guidance
Issued by an official standards body	Yes	No	No
Discrepancies between methods			
Utility entrance	Requires two utility services in Tier 3, and 4	Requires two utility services in Tier III, and IV	Level 3 and 4 Inconsistent with TIA and Uptime
Redundant IT power inputs	Required in Tier 2, 3, and 4	Required in Tier III and IV	Required in Level 3 and 4 (i.e. Tier 3 and 4)
Standby power (generator)	Required for all Tiers	Required for all Tiers	Not required for Level 1 (i.e. Tier 1)
2N CRAC / CRAH unit redundancy	Required in Tiers 3 and 4	Unknown	Required in Level 4

Note: Blue shading indicates best performance for the characteristic

[5] Turner, Seader, Brill, "*Tier Classifications Define Site Infrastructure Performance*", 2006, p. 13