The Data Center Tier Performance Standards and Their Importance to the Owner's Project Requirements

Edward P Rafter, P.E. CxAP
Tier IV Consulting Group

Synopsis

The data center industry has experienced several evolutions over the past 20 years. One of the principal changes has been the development of the Tier Performance Standards. The standards provide quantifiable plateaus or Tier Levels. Tier Levels provide an objective basis for comparing the capabilities of a particular design topology against other designs as well as the associated site availability metrics for the various levels. The requirements of each Tier Level are clearly defined and provide a road map used in the design and management of the data center. The Tier Level was developed in the early 1990s and is the foundation used by a number of data center owners/users, consultants and design professionals in establishing a "design-versus-performance" ranking approach to today's data center projects.

This paper will present a brief history of data center infrastructure topologies and their evolution, as well as a summary of the Tier Performance Standards and expectations, and endeavor to normalize several of the commonly used standards referenced in the industry today. The author proposes that the Tier Classification is the core for preparing the Owner's Project Requirements (OPR) and Basis of Design for such projects and includes the provisions for the electrical, mechanical, environmental and data center infrastructure that is needed to meet a desired performance level.

A real-world project will be presented to illustrate a representative approach that is used in commissioning data centers today.

About the Author

Edward Rafter has been closely involved with the design, commissioning and operations of mission critical facilities, primarily data centers, for more than 25 years. He has worked both as senior facilities engineer with one of the leading financial information service companies in the United States and as consulting engineer to a number of data center clients. His experience has provided him the opportunity to participate in the pre-design, design, construction, functional and integrated testing and on-going operation and maintenance of data centers covering a broad range of performance requirements.

Historical Perspective

Prior to the late 1980s data center owners and users had limited resources available to determine what constituted a reliable data center and the required infrastructure, processes and procedures necessary to meet their business needs. The data center industry at that time was relatively small by today's standards and limited principally to financial, telecommunication and government interests where demands for a higher level of up-time to support their mid- to large-range computer systems was vital. Those individuals who were fortunate enough to have a network of colleagues in the data center business who were able to share information were a small minority. The rule of the day was non-disclosure of information in any manner or form. Another resource for information sharing, indirect as it may be, was a knowledgeable consultant or engineer familiar with data center design and operation. The truth of the matter is that such resources were also in limited supply. It truly was the experience of the owner and user that held the greatest measure of insight into the problems and solutions associated with the data center environment.

Data Center Users Groups

On November 15, 1989 a group of 38 individuals representing 18 companies met at Shearson Lehman in downtown New York City. This occasion marks the initial meeting of what was the Uninterruptible Uptime Users Group (UUUG). The guiding principle of the UUUG, later to become the 7x24 Exchange, has been "to provide a forum for knowledge exchange for those who design, build, use and maintain mission-critical enterprise information infrastructures." The UUUG represented the first real opportunity for owners and users of data centers to meet and share experiences. This marks a defining event in the sharing of experiences and the first step towards establishment of a methodology to measure business expectations against the required data center infrastructure.

Another similar organization, AFCOM (Association for Computer Operations Management), was originally established to serve data center management and operations, including data center managers, operations managers, MIS directors, CIOs, CTOs and other IS/IT professionals. In recent years it has become a leading association in supporting the educational and business development needs of data center management, executives and vendors.

Open Discussion Groups

The establishment of the UUUG still left a need in the data center community for the user's ability to disseminate information freely due to the open forum of the group meetings. In response to this, the Site Uptime Network® was formed in 1993 by The Uptime Institute, Inc. with a clear purpose: to directly address and resolve issues affecting continuous site infrastructure (power, cooling, and electrical environment) availability. This was accomplished by forming an association of member companies responsible for America's most critical data centers. Through a member-driven, collaborative learning experience, members have steadily achieved higher levels of site uptime.

The Site Uptime Network's mission is to identify, quantify and improve infrastructure availability. It achieves this by providing an opportunity for information exchange between members themselves, and between members and industry experts.
Through the open sharing of information and by defining common goals, the challenges faced by data center owners and users became clearer to all. The outcome of this information sharing was the characterization and organization of data to determine cause and effect relationships and lead to developing potential solutions.

Evolution of Tier Performance Considerations

A driving force towards the development of an analytical approach in the evaluation of data center performance is the need to clearly convey the differences in data center financial investment to senior business management. Understanding the requirements of the business systems and the desired IT availability versus the availability of the site infrastructure became essential in making business decisions.

Expectations aimed at addressing those causes for unwanted business interruptions and the increased demand for a high level of availability prompted the recognition of those factors associated with disruptions.

No longer was a tactical solution approach to business demands from the data center infrastructure acceptable. Senior decision makers were looking for strategic answers.

Another factor in the evolution of performance criteria is the incredible growth of the industry. The number of data centers has grown significantly in the past 20 years. This is due in part to a dramatic increase in internet business interests and partly due to concerns regarding business continuity through concurrent operations and disaster recovery requirements. Government rulings in the financial sector also spurred the requirement for an increase in availability and fault tolerance.

Identifying the Risks to Data Center Availability

Many of the risk elements that are addressed in the tier levels were identified through user experiences. Often these experiences reflect lessons learned the 'hard way,' principally through unplanned outages and system downtime.

The factors historically associated with business disruptions include the following:

- Nature – This includes weather related events such as tornados, hurricanes and flooding
- Human – Human-related events external to the facility including commercial transportation accidents affecting the area
- Utility – Disruptions in utilities such as electrical power interruptions
- Equipment – Failures in essential equipment
- Personnel – Operator error during normal business and maintenance activities

As data relating to data center disruptions was compiled, solutions were identified. Through the efforts of data center professionals the weighting of events and the ranking of alternate design solutions created a set of metrics to objectively asses the aspects of the data center.

Aspects Related to Data Center Performance

Those aspects related to data center performance begin with site selection. Though not directly weighted in the evaluation of the infrastructure as defined by the tier standards, site location is none-the-less a key aspect to the performance level desired by the data center. There are a number of considerations when planning a new data center or when considering major improvements to an existing data center.

Table 1: Aspects Considered in Data Center Site Selection

Consideration	System
Location	Earthquake Zone Flood Plains Hurricanes or Tornadoes Proximity to Major Highways Proximity to Railway Lines Proximity to Hazardous Areas Proximity to Airports or Flight Corridors
Infrastructure	Availability of Electrical Capacity Availability of Diverse Power Feeders Utilities Expansion/Upgrades History of Outages
Water	Diverse Source Supplies Water Storage
Communications	Availability of Diverse Carriers Availability of Diverse Services Physical Security Alarms and Monitoring
Economics	Land Construction Utilities Labor Communications
Staffing	Accessibility Public Transportation Recreational Facilities Housing Amenities

The tier performance level is determined based on an evaluation and rating of 16 critical site infrastructure subsystems as compared against the desired tier classification for the data center. Data centers are dependent upon the successful operation of these 16 subsystems. Subsystems and systems must be installed with the same availability objective to meet the tier level requirements and performance expectations.

Table 2: Aspects Considered in Tier Performance Evaluation

Category	System
Electrical	Utility Service Lightning Protection Power Backbone UPS Systems UPS Batteries Engine Generator Load Bank Critical Power Distribution Grounding
Mechanical	Raised Floor Cooling UPS Cooling Mechanical Plant
Support Systems	Contamination Fire Detection and Protection Physical Security Alarms and Monitoring

The Tier Levels

Since the early 1960s the design of data center infrastructure has advanced through at least four clearly identifiable stages. Tier I appeared in the early 1960s, Tier II in the 1970s, Tier III in the late 1980s and early 1990s, Tier IV in the mid 1990s. These stages provide the foundation for the Four Tier Classification system defined by owners and users in association with the Uptime Institute. Each of the four tier levels can be viewed as a plateau towards achieving an increase in performance.

Alternate approaches have led to other systems of data center classification. An extension of the four tier system is a Ten Tier Classification, ranked 1 to 10, where incremental steps are described between each of the plateaus described in the Four Tier system. A distinctive example

of the ten tier system ranks data centers in a descending number order where Tier 1 represents the highest order of data center availability and fault tolerance.

Another example of the tier level system presented by Syska Hennessy Group offers a method of ranking each of the aspects independently. Unlike the four-tiered system from the Uptime Institute, the Syska Hennessy system examines 11 different aspects of data center performance and measures them on a scale from one to 10.

The 11 items are: power, HVAC, fire and life safety, security, IT infrastructure, controls and monitoring, commissioning and testing, operations, maintenance, operations and maintenance procedures, and disaster preparedness. Syska Hennessy refers to its tier system as the Criticality Levels™.

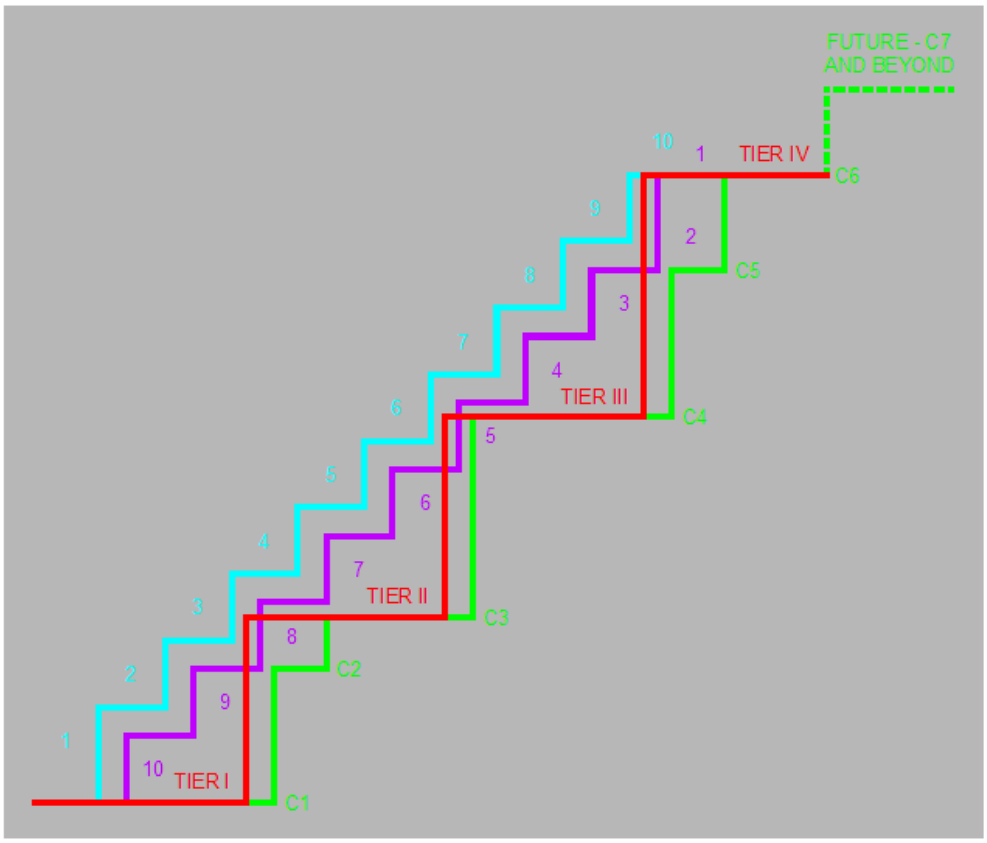

Figure A: Correlation of Various Tier Level

Tier Level Standards and Performance Expectations

The *Tier I* data center has non-redundant capacity components and single non-redundant path distribution paths serving the site's computer equipment. The data center has computer room cooling and power distribution but it may or may not have a UPS or and engine generator. The data center must be shutdown for annual predictive maintenance and repair work. Corrective maintenance may require additional shutdowns. Operation errors or spontaneous failures of infrastructure components will cause a data center disruption. As example, a *Tier I* data center may be suitable for small businesses where IT is intended for internal business processes

The *Tier II* data center has redundant capacity components and single non-redundant distribution paths serving the site's computer equipment. They have UPS and engine generators but their capacity design is Need plus One (N+1), with a single power path. Maintenance of the critical power path and other parts of the site infrastructure will require a shutdown of computer processes. As example, a *Tier II* data center may be appropriate for internet-based companies without serious financial penalties for quality of service commitments.

The *Tier III* data center is concurrently maintainable and has redundant capacity components and multiple distribution paths serving the site's computer equipment. Generally, only one distribution path serves the computer equipment at any time. This topology allows for any planned site infrastructure activity without disruption the computer systems operation in any way. An example of a *Tier III* application would include companies that span multiple time zones or whose information technology resources support automated business process.

The Tier IV data center is fault tolerant and has redundant capacity systems and multiple distribution paths simultaneously serving the site's computer equipment. All IT equipment is dual powered and installed properly to be compatible with the topology of the site's architecture. Fault-tolerant functionality also provides the ability of the site infrastructure to sustain at lease one worst-case unplanned failure or event with impact to the critical load. This typically includes a System+System topology. Examples of a Tier IV requirement include companies who have extremely high-availability requirements for ongoing business such as E-commerce, market transactions, or financial settlement processes.

The following is a summary of representative site availability expectations for each of the tier levels described above. The availability percentages can be considered characteristic of the operating experiences of a representative number of sites within each tier classification.

- Tier I = 28.8 hours and 99.67%
- Tier II = 22.0 hours and 99.75%
- Tier III = 1.6 hours and 99.98%
- Tier IV = 0.4 hours and 99.99%

As a rule, the overall Tier Level is based on the lowest tier ranking or weakest component. By example a data center my be rated tier 3 for electrical, but tier 2 for mechanical. The data centers

overall tier rating is 2. In practice a data center may have different tier ratings for different portions of the infrastructure.

The final determination of the Tier Level achieved in the data center design should be made by a qualified authority familiar with the tier performance standards as defined in the Owner's Project Requirements document for the project.

Formalizing the Standards

The Four Tier Performance Standard prepared by the Uptime Institute, is widely accepted throughout the data center industry. In April 2005 the Telecommunications Industry Association published Standard TIA-942 in association with the American National Standards Institute (ANSI). This standard includes information regarding the Data Center Infrastructure Tiers and is in general consistent with the Uptime Institute's documents.

Data Center Project Case Study

The following case study represents the manner the Tier Performance Standards are employed. Details specific to the client have been omitted at their request.

In the beginning of 2000, XYZ Inc was considering options for their corporate data center and other business units as part of the construction of a new world headquarters. In addition to the very long list of questions to be addressed in the relocation of personnel to the new facility, the company was faced with determining what the data center requirements should be to meet the business plan associated with the project. What should be the physical size of the data center? Should the data center move to a separate physical location than the world headquarters multistory office complex? What level of availability should the data center design infrastructure be designed to meet? These were just a few of the questions that prompted an intense review not only of internal expectations but also research into the solutions chosen by similar businesses.

As part of the planning, XYZ Inc information technology (IT) and corporate management in collaboration with industry professionals researched the publications and attended user group meetings to gain an understanding of design options. This research led to gaining an understanding of the Tier Performance Standards and tier level classifications. Project managers met with facility and IT personnel to quantify the needs and expectations of each group.

Owner's Project Requirements

The owner's project requirements (OPR), referred to as the design intent, included specific design expectations for the new data center as per Uptime Institute requirements for a Tier III data center incorporated into the new world headquarters multistory office complex. The document referenced the Uptime Institute Tier III requirements and presented a clearly defined set of expectations towards meeting this tier classification level including:

- Owner objectives to meet a Tier III data center
- Expectations on electrical and mechanical infrastructure
- Expectations on IT infrastructure
- Operations and maintenance criteria
- Equipment and system maintainability expectations
- Expectations for expandability without business interruptions
- Project documentation requirements
- Commissioning including functional and integrated testing

The project team selected by XYZ Inc included Project Management, General Contractor and Commissioning Authority familiar with the Commissioning Process and the Tier Performance expectations.

Conclusion

The Tier Performance Standards are an owner/user set of requirements used to clearly define expectations for the design and management of the data center to meet a prescribed level of availability. The Tier Level Classification system is the foundation used by many data center owners/users, consultants and design professsionals in establsihing a "design-versus-performance" ranking approach to todays data center projects.

As part of the commissioning process for data center projects and where the owner acknowledges one of these standards in defining the expectations for the requirements of the data center, the commissioning authority (CxA) should include specific reference to the standard used in the OPR.

The final determination of the Tier Level achieved in the data center design should be made by a qualified authority familiar with the tier performance standards referenced in the project documents.

References

ANSI/TIA-942-2005 Telecommunications Industry Association (TIA), *Telecommunications Infrastructure Standard for Data Centers*

Jerry Burkhardt and Richard W. Dennis, Syska Hennessy, *Assessing Criticality Levels in the Data Center*, SearchDataCenter.com, September 27, 2006

Uptime Institute White Paper, *Industry Standard Tier Classifications Define Site Infrastructure Performance,*