BICSI International Standards Supplemental Information 001

Methodology for Selecting Data Center Design Class Utilizing Performance Criteria

BICSI®
advancing information
technology systems

Foreword

The information in this document has been approved by the BICSI International Standards Data Center Subcommittee for informational use only and is subject to change without notice. The BICSI International Standards Program is publishing this information as furnished, so that it may be referenced by the industry as desired. No material in this document is to be construed as an official or adopted standard. BICSI grants permission to reproduce and distribute this document provided that:

1) the document is maintained in its original form, and
2) this disclaimer and the notice below accompany the document at all times.

NOTICE: By publication of this document, BICSI takes no position respecting the validity of any patent rights or copyrights asserted in connection with any item mentioned in this document. Additionally, BICSI does not assume any liability to any patent owner, nor does it assume any obligation whatever to parties adopting the standard or publication. Users of this document are expressly advised that determination of any such patent rights or copyrights, and the risk of infringement of such rights, are entirely their own responsibility.

All warranties, express or implied, are disclaimed, including without limitation, any and all warranties concerning the accuracy of the contents, its fitness or appropriateness for a particular purpose or use, its merchantability and its non-infringement of any third party’s intellectual property rights. BICSI expressly disclaims any and all responsibilities for the accuracy of the contents and makes no representations or warranties regarding the content’s compliance with any applicable statute, rule or regulation.

BICSI shall not be liable for any and all damages, direct or indirect, arising from or relating to any use of the contents contained herein, including without limitation any and all indirect, special, incidental or consequential damages (including damages for loss of business, loss of profits, litigation, or the like), whether based upon breach of contract, breach of warranty, tort (including negligence), product liability or otherwise, even if advised of the possibility of such damages. The foregoing negation of damages is a fundamental element of the use of the contents hereof, and these contents would not be published by BICSI without such limitations.

Published by:

BICSI
8610 Hidden River Parkway
Tampa, FL 33637-1000 USA

Methodology for Selecting Data Center Design Class Utilizing Performance Criteria

1 Introduction

1.1 Overview

People have come to expect ready access to information 24 hours a day, every day. The Internet as well as more traditional enterprises—both business and governmental—now operate 7 days a week, 24 hours a day. Typical 24/7 operations include banking systems, credit card companies, 911 emergency centers, telecommunications networks, hospital systems, overnight delivery operations, large multinational concerns, and other international organizations.

The burgeoning demand for mission-critical/data processing facilities (essentially server warehouses) requires fresh thinking given the radical differences in conventional building types. Consider some mission-critical facility norms:

- The power supplied to a typical office building is about 110 W/m^2 (10 W/ft^2), but between 650 W/m^2 (60 W/ft^2) and 2200 W/m^2 (200 W/ft^2) or more in a mission-critical facility. Mission-critical power requirements far exceed any conventional building type.
- The mechanical/electrical/service space ratio to usable space averages 1:3 or 1:4 in typical buildings and is close to 1:1 in data centers.
- The cost of mission-critical facilities can run up to four times the cost of more traditional building types. Power and cooling requirements drive cost and design.

These numbers are revealing. Mission-critical power and cooling systems evolved from a philosophy dubbed "system + system", meaning that, for every critical electrical and HVAC system, a duplicate system is in place to provide service while the other is repaired, maintained, or tested off-line. Additionally, the risk of natural and provoked disasters causing potential IT downtime dictates a hardened building shell as well as sufficient power and electrical capacity on site.

Continuous operation implies that building systems need a measured approach to incorporating reliability, with redundant building systems being the typical method used. After all, a shutdown can cripple the revenue generating continuity of a business, ruin customer confidence, and possibly threaten its existence if the shutdown is extensive. Disruption to the IT systems underpinning today's industrialized societies may cause loss of life and endanger communities, depending upon the nature and location of the service.

Mission-critical facilities requiring 7 day at 24 hours/day operations need a comprehensive strategy to sustain their vital activities. Many small businesses have at least a 5 day at 12 hour/day high-availability operating requirement—less rigorous standards, yet still substantial. Mission-critical design variations will stem from each facility’s requirements. Starting with site selection criteria and working forward through each layer of architectural, engineering and operational design, reliability and reducing the risk of downtime must be the prime focus, weighed and coordinated throughout the design process.

Mission-critical facilities have not traditionally been high-profile projects, yet their design issues are increasingly complex and critical. With an emerging design terminology and vocabulary, their rapid evolution calls for an exceptional degree of building system coordination and integration. These facilities are not merely warehouses for servers, but instead rival medical operating rooms or semiconductor plants, with their precise environmental controls and power requirements. Intense, sustained work shifts with employees monitoring computer screens mean that workplace design issues must also be addressed.

Important design considerations also extend well beyond the context of the mission-critical facility itself. Utility deregulation is causing uncertainty. Increasing power demands challenge reliability of the power supply itself. Some utilities even question their own capacity to power mission-critical facilities. Because IT plants can be highly sensitive to temperature and power fluctuations, these concerns are attracting increased attention. It is not an issue that can be addressed simply through the purchase of UPS systems.

This document outlines a process for designing new and upgrading existing mission-critical IT facilities. To achieve maximum benefit, defining the required performance levels of availability and reliability and then designing, procuring, and maintaining mission-critical IT facilities can and should be formalized. Lack of a formal process yields higher construction and operational costs as well as inconsistent and unpredictable performance.

Different portions of the data center may need to be analyzed separately and each assigned different Availability Classes.

It should be noted that the recommendations reached by using this tool can be presented to management, who must then determine if the cost of implementation can be justified. Budget is a variable that will differ for each enterprise. See Sections 6.3 and 7.3 for some cost/benefit analysis.

This document is intended to address data center reliability and not the subject of business continuity. An enterprise may be better served by multiple data centers of lower Class than a single Class F4 data center.

1.2 Goals and Objectives

This document is intended to provide a framework for understanding the process for determining facility criticality and aligning project objectives and budgets with appropriate performance levels.

An objective of this document is to present an industry-wide process for planning mission-critical IT facilities, with the following strategic goals:

- Establish a consistent, cost-effective process.
- Develop optimum design and implementation solutions.
- Develop a common enterprise-wide design vocabulary.
- Establish performance criteria used to generate or evaluate mission-critical facilities.

Note: This document only addresses the facility infrastructure requirements. The operational performance characteristics of the information technology equipment itself (servers, storage devices, telecommunications equipment, etc.) are beyond the scope of this document.

2 Creating Mission-Critical Facilities Overview

There are four steps in the process of designing a new mission-critical IT facility or upgrading an existing one. These are represented in Figure 1 and described afterward.

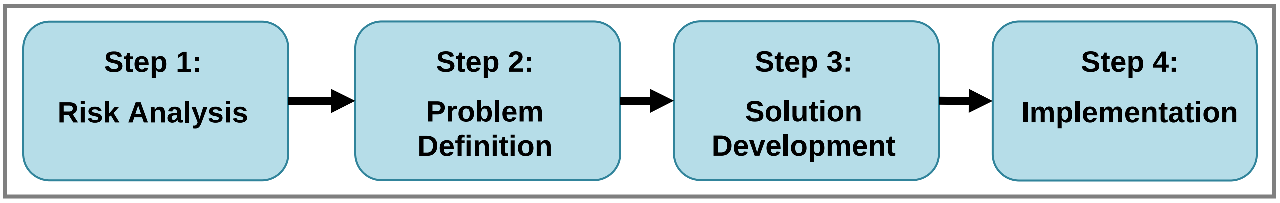

Figure 1: Planning Process for a Mission-Critical Facility

- Step 1: Risk analysis: As explained in the remainder of this section, risk analysis is a linear process. Three key characteristics are defined to arrive at an expression of the criticality of a mission-critical facility:
  - identify operational requirements for the section of the data center under analysis—the opportunity to suspend operations for scheduled maintenance;
  - identify availability requirements—the targeted uptime of the systems or zones during operations, and the ability to endure unplanned interruption of operations;
  - define the impact of downtime—the consequences of unplanned disruptions on the mission.

  The process of analyzing and mitigating risk is described in the next section. The other parts of this section form an important reference source during the remaining three process phases in creating a mission-critical facility.
- Step 2: Problem definition—After completing the risk analysis process, characterize the facility in terms of space, IT asset density (which has several components, including density of computing, data communications, and telephony), and anticipated availability requirements. The resulting information is usually documented in the facility program.
- Step 3: Solution development—Translate the facility program into one or more design solutions to solve the specific design problem and meet the objectives of the targeted Availability Class.
- Step 4: Implementation—Construct the chosen solution, incorporating implementation tactics consistent with the targeted Availability Class. This will include appropriate maintenance and operations procedures to ensure a sustainable level of availability.

The remainder of this document pertains to risk analysis and the methodology for selecting data center design Availability Classes. See ANSI/BICSI 002-2011 for more information regarding problem definition, solution development, and implementation.

3 Risk Analysis

It is impossible to eliminate the risk of downtime, but risk reduction is an important planning element. In an increasingly competitive world, it is imperative to address downtime in business decisions. The design of systems supporting critical IT functions depends on the interaction between the criticality of the function and its operational profile.

Criticality is defined as the relative importance of a function or process as measured by the consequences of its failure or inability to function. The operational profile expresses the time intervals over which the function or process must operate.

To provide optimal design solutions for a mission-critical facility, consider several key factors. NFPA 75 identifies six considerations for protection of the environment, equipment, function, programming, records and supplies in a data center. These include:

1) What are the life safety aspects of the function? For example, if the system failed unexpectedly, would lives be put at risk? Examples of such applications include some process controls, air traffic control, and emergency call centers.
2) What is the threat to occupants or exposed property from natural, man-made, or technology-caused catastrophic events? For example, is the building equipped with fire suppression? …in a flood zone? …seismically structured? … in a tornado or hurricane corridor? …etc.
3) What would be the economic loss to the organization from the loss of function or loss of records?
4) What would be the economic loss from damaged or destroyed equipment?
5) What would be the regulatory or contractual impact, if any? For example, if unplanned downtime resulted in loss of telephone service or electrical service to the community, would there be penalties from the government?
6) What would be the impact of disrupted service to the organization's reputation? For example, would subscribers switch to a competitors' service?

The methodology presented in the following section for determining a data center's facility Availability Class integrates these considerations and defines the appropriate risk management strategy.

4 Determining the Facility Availability Class

4.1 Overview

While there are innumerable factors that can be evaluated in a mission-critical facility, there are three factors which can quickly be quantified, providing for an easy determination of an Availability Class and the necessary functions and features required for the facility. These factors are:

- Operational requirements.
- Operational availability
- Impact of downtime.

Paying careful attention to these factors determines an appropriate Availability Class that matches the mission-critical facility cost with the functions it is intended to support.

Figure 2 shows how these factors interact to determine the facility Availability Class; each step is explained afterward.

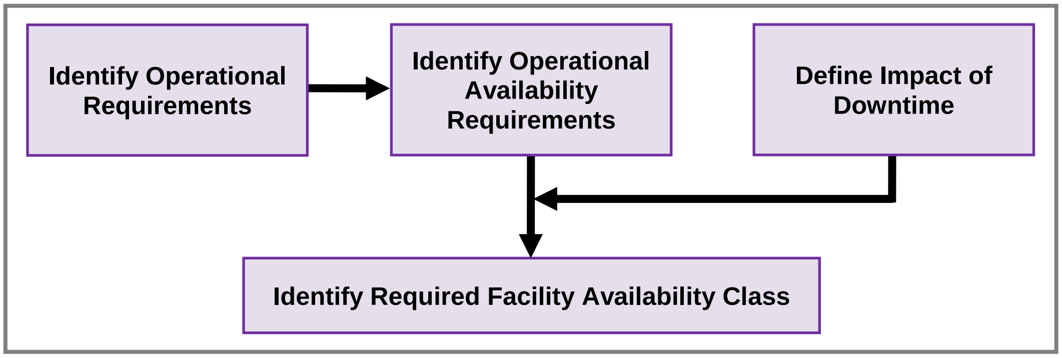
Identify Operational Requirements
Identify Operational Availability Requirements
Define Impact of Downtime
Identify Required Facility Availability Class

Figure 2: Relationship of Factors in Facility Availability Class

4.2 Identify Operational Requirements

The first step in determining the facility Availability Class associated with a mission-critical IT facility is to define the facility's intended operational requirements. Sufficient resources must be available to achieve an acceptable level of quality over a given time period. IT functions that have a high-quality expectation over a longer time period are by definition more critical than those requiring less resources, lower quality, and/or are needed over a shorter time period.

While there are many factors in operations, the key factor to be determined in this step is assessing the amount of time to be allowed for testing and maintenance which disrupts normal operation. Also known as "planned maintenance shutdown", when using Table 1 to determine an operational level, the value of time used should not include projections for unplanned repairs or events. (*This will be used in the following step*)

Table 1: Identifying Operational Requirements: Time Available For Planned Maintenance Shutdown

Operational Level	*Annual Hours Available for Planned Maintenance Shutdown*	*Description*
0	> 400	Functions are operational less than 24 hours a day and less than 7 days a week. Scheduled maintenance "down" time is available during working hours and off hours.
1	100-400	Functions are operational less than 24 hours a day and less than 7 days a week. Scheduled maintenance "down" time is available during working hours and off-hours.
2	50-99	Functions are operational up to 24 hours a day, up to 7 days a week, and up to 50 weeks per year – scheduled maintenance "down" time is available during working hours and off hours.
3	0-49	Functions are operational 24 hours a day, 7 days a week for 50 weeks or more . No scheduled maintenance "down" time is available during working hours.
4	0	Functions are operational 24 hours a day, 7 days a week for 52 weeks each year. No scheduled maintenance "down" time is available.

Note: The term "shutdown" means that operation has ceased; the equipment is not able to perform its mission during that time. Shutdown does *not* refer to the loss of system components if they do not disrupt the ability of the system to continue its mission.

4.3 Quantify and Rank Operational Availability Requirements

The second step in determining the facility Availability Class is to identify the facility's operational availability requirements; i.e., the total uptime that the facility must support without disruption. Operational availability refers only to scheduled uptime—that is, the time during which the IT functions are actually expected to run.

Availability is sometimes expressed as a percentage, or "nines of availability." Thus, "five nines of availability" means that the system is fully operable 99.999% of the time throughout a year. Table 2 shows what percentage means in terms of actual minutes of downtime at various levels of operational availability. A general rule of thumb is that the cost of an installation increases by at least fifty percent for every incremental "nine" that is added. The cost of downtime must be weighed against the cost of uptime. Note that less than a second of power interruption or a few minutes of cooling interruption can result in hours of recovery time. The term "availability" means the information technology equipment is operational and performing its function; it does not solely refer to operation of the supporting infrastructure.

The objective is to identify the intersection between the intended maximum annual downtime and the intended operational level. A function or process that has a high availability requirement with a low operational profile has less risk associated with it than a similar function with a higher operational profile.

Thus, this step adjusts the overall operational level to reflect the true functional requirement. This step will result in one of five Operational Availability Rankings to be used in the next step.

Table 2: Identifying Operational Availability Requirements: Maximum Annual Downtime (Availability %)

	Allowable Maximum Annual Downtime (minutes) *(Availability as %)*				
Operational Level *(from Table 1)*	*>5000* *(> 99%)*	*500 – 5000* *(99% > 99.9%)*	*50 – 500* *(99.9% > 99.99%)*	*5 – 50* *(99.99% > 99.999%)*	*0.5 – 5.0* *(99.999% > 99.9999%)*
Level 0	0	0	1	2	2
Level 1	0	1	2	2	2
Level 2	1	2	2	2	3
Level 3	2	2	2	3	4
Level 4	2	3	3	4	4

4.4 Determine Impact of Downtime

The third step in the determining the facility Availability Class is to identify the impact or consequences of downtime. This is an essential component of risk management because not all downtime has the same impact on mission-critical facilities. Identifying the impact of downtime on mission-critical functions helps determine the tactics that will be deployed to mitigate downtime risk. As shown in Table 3, there are five impact classifications, each associated with a specific impact scope.

Table 3: Classifying the Impact of Downtime on the Mission

Classification	*Description – Impact of Downtime*
Isolated	Local in scope, affecting only a single function or operation, resulting in a minor disruption or delay in achieving non-critical organizational objectives.
Minor	Local in scope, affecting only a single site, or resulting in a minor disruption or delay in achieving key organizational objectives.
Major	Regional in scope, affecting a portion of the enterprise (although not in its entirety) or resulting in a moderate disruption or delay in achieving key organizational objectives.
Severe	Multiregional in scope, affecting a major portion of the enterprise (although not in its entirety) or resulting in a major disruption or delay in achieving key organizational objectives.
Catastrophic	Affecting the quality of service delivery across the entire enterprise, or resulting in a significant disruption or delay in achieving key organizational objectives.

4.5 Identify the Data Center Facility Availability Class

The final step in the process is to combine the three previously identified factors to arrive at a usable expression of availability. This expression of availability is used as a guide to determine the architectural and engineering features needed to appropriately support critical IT functions. Since operational level is subsumed within the availability ranking,as explained in 4.3, the task at hand is to matrix the availability ranking against the impact of downtime and arrive at an appropriate facility Availability Class. Table 4 shows how this is done:

Table 4: Determining Facility Availability Class

Impact of Downtime *(from Table 3)*	*Operational Availability Level (from Table 2)*				
	0	*1*	*2*	*3*	*4*
Isolated	**Class F0**	**Class F0**	**Class F1**	**Class F2**	**Class F2**
Minor	**Class F0**	**Class F1**	**Class F2**	**Class F3**	**Class F3**
Major	**Class F1**	**Class F2**	**Class F2**	**Class F3**	**Class F3**
Severe	**Class F1**	**Class F2**	**Class F3**	**Class F3**	**Class F4**
Catastrophic	**Class F1**	**Class F2**	**Class F3**	**Class F4**	**Class F4**

5 Data Center Facility Availability Classes

To a great degree, design decisions are guided by the identified Availability Class. Thus, it is essential to fully understand the meaning of each Availability Class. Each Availability Class is defined in terms of four areas of concern:

1) **Component redundancy** increases reliability by providing redundancy for critical high-risk, low-reliability components within systems.
2) **System redundancy** increases reliability even more by providing redundancy at the system level.
3) **Quality** ensures that high quality is designed and implemented in the facility, thus reducing the risk of downtime due to failure during initial installation and/or premature wear. Since MTBF is a major factor in the determination of system reliability, it stands to reason that higher quality components with lower failure rates will result in systems that are more reliable.
4) **Survivability** refers to reducing the risk of downtime by protecting against external events such as physical forces, security breaches, and natural disasters.

The following subsections provide more detail on how each of these four factors is defined for each of the five Availability Classes. Each Class also includes an overview of the appropriate tactics for critical systems.

5.1 Availability Class F0

The objective of Class F0 is to support the basic environmental and energy requirements of the IT functions without supplementary equipment. Capital cost avoidance is the major driver. There is a high risk of downtime due to planned and unplanned events. However, in F0 facilities maintenance can be performed during non-scheduled hours, and downtime of several hours or even days has minimum impact on the mission.

5.1.1 Tactics for Class F0

Component redundancy:	none
System redundancy:	none
Quality control:	standard commercial quality
Survivability:	none

Application: A critical power distribution system separate from the general use power systems would not exist. There would be no back-up generator system. The system might deploy surge protective devices, power conditioning, or even small or non-redundant uninterruptible power supply (UPS) systems to allow the specific equipment to function adequately. Utility grade power does not meet the basic requirements of critical equipment. No redundancy of any kind would be used for power or air conditioning for a similar reason. Class F0 has multiple single-points of failure.

5.2 Availability Class F1

The objective of Class F1 is to support the basic environmental and energy requirements of the IT functions. There is a high risk of downtime due to planned and unplanned events. However, in Class F1 facilities, remedial maintenance can be performed during nonscheduled hours, and the impact of downtime is relatively low.

5.2.1 Tactics for Class F1

Component redundancy:	none
System redundancy:	none
Quality control:	standard commercial quality
Survivability:	none

Application: In Class F1, the critical power distribution system would deploy a stored energy device and a generator to allow the critical equipment to function adequately (utility grade power does not meet the basic requirements of critical equipment). No redundancy of any kind would be used for power or air conditioning for a similar reason.

5.3 Availability Class F2

The objective of Class 2 is to provide a level of reliability higher than that defined in Class 1 to reduce the risk of downtime due to component failure. In Class 2 facilities, there is a moderate risk of downtime due to planned and unplanned events. Maintenance activities can typically be performed during unscheduled hours.

5.3.1 Tactics for Class F2

Component redundancy: redundancy is provided for critical components

System redundancy: none

Quality control: premium quality for critical components

Survivability: moderate hardening for physical security and structural integrity

Application: In Class F2, the critical power, cooling, and network systems would need redundancy in those parts of the system that are most likely to fail. These would include any products that have a high parts count or moving parts, such as UPS, controls, air conditioning, generators or ATS. In addition, it may be appropriate to specify premium quality devices that provide longer life or better reliability.

5.4 Availability Class F3

The objective of Class F3 is to provide additional reliability and maintainability to reduce the risk of downtime due to natural disasters, human-driven disasters, planned maintenance, and repair activities. Maintenance and repair activities will typically need to be performed during full production time with no opportunity for curtailed operations.

5.4.1 Tactics for Availability Class F3

Component redundancy: redundancy is required for critical and noncritical components, except where the component is part of a redundant system; redundancy is also provided to increase maintainability

System redundancy: system redundancy is required where component redundancy does not exist

Quality control: premium quality for all components

Survivability: significant hardening for physical security and structural integrity

Application: In Class F3, the critical power, cooling, and network systems must provide for reliable, continuous operations even when major components (or, where necessary, major subsystems) are out of service for repair or maintenance. To protect against unplanned downtime, the power, cooling, and network systems must be able to sustain operations while a dependent component or subsystem is out of service.

5.5 Availability Class F4

The objective of Class F4 is to eliminate downtime through the application of all tactics to provide continuous operation regardless of planned or unplanned activities. All recognizable single points of failure from the points of connection at the utility to the points of connection at the critical loads are eliminated. Systems are typically automated to reduce the chances for human error and are staffed 24x7. Rigorous training is provided for the staff to handle any contingency. Compartmentalization and fault tolerance are prime requirements for a Class F4 facility.

5.5.1 Tactics for Availability Class F4

Component redundancy: redundancy is provided for all critical components and to increase maintainability; also provided for noncritical components

System redundancy: system redundancy is provided with component redundancy so that overall reliability is maintained even during maintenance activities

Quality control: premium quality for all components

Survivability: all building systems are self-supporting in any event and are protected against the highest levels of natural forces

Application: The critical power, cooling, and network systems in a Class F4 facility must provide for reliable, continuous operations even when major components (or, where necessary, major subsystems) are out of service for repair or maintenance. To protect against unplanned downtime, systems must be able to sustain operations while a dependent component or subsystem is out of service.

6 Reliability Aspects of Availability Planning

Achieving an optimum Class of availability for a mission-critical facility requires strategic planning to determine the risks, design features and potential improvement measures that will lead to fewer facility related failures.

6.1 Reliability Engineering Principles and Calculating Reliability

Reliability is simply the probability that equipment or system will perform its intended function, within stated conditions, for a specified period of time, without failure. It is expressed as a percentage (i.e., a number between 0 and 1), in which a lower percentage indicates a greater likelihood of failure in a given period of time. Reliability is not the same as *availability.* Whereas reliability uses the number (frequency) of failures within a period of time within its calculation, availability utilizes the amount of time equipment or a system is non-operational due to planned or unplanned failures, interruptions, or events.

Over the last 30 years, data has been collected and analyzed for a wide variety of mechanical and electrical components and their failure characteristics. This has led to broad-based industry standards for the analysis and design of reliable power and cooling systems (e.g., IEEE Standard 493-2007 (Gold Book)—*IEEE Recommended Practice for the Design of Reliable Industrial and Commercial Power Systems*).

The reliability of a given system can be calculated from the published MTBF (mean time between failures) data for given components of that system. This calculation can then be combined to yield an overall expression of system reliability through the analysis of all series and parallel subsystems. The calculations are as follows:

$$R = e^{(-\lambda T)}$$

where:

R = reliability (percent probability of success)

e = exponential function

λ = failure rate (the reciprocal of MTBF)

T = time period (same units as failure rate)

Example: A UPS module has a published MTBF of 17,520 hours (one failure every two years). Its failure rate would then be 0.00005708 failures per hour. What is its one-year reliability, or the probability of not failing in one year (8,760 hours)?

$R = e^{(-0.00005708 \times 8{,}760)}$

R = 0.6065 or 60.65%

To obtain the reliability of a given system, the individual reliability of each component must be calculated, then the reliability of parallel subsystems, and then the series reliability of all subsystems, as follows and as illustrated in Figure 3.

The reliability of a series system is equal to the product of all component reliabilities. The reliability of a parallel system is equal to the complement of the product of all component complements. Thus, the reliability for the system in Figure 3 would be calculated as follows:

$R_{A1A2} = R_{A1} \times R_{A2} = 0.5 \times 0.5 = 0.25$

$R_A = 1 - [(1 - R_{A1A2}) \times (1 - R_{A3})] = 1 - [(1 - 0.25) \times (1 - 0.5)] = 0.625$

$R_B = 1 - [(1 - R_{B1}) \times (1 - R_{B2})] = 1 - [(1 - 0.61) \times (1 - 0.61)] = 0.848$

$R_{TOTAL} = R_A \times R_B = 0.625 \times 0.848 = 0.53$ (53%)

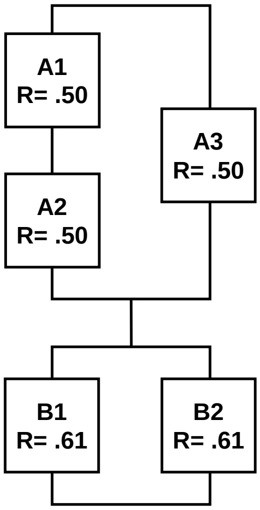

Figure 3: Sample Reliability Calculation

6.2 Trends Affecting Reliability of Critical IT Facilities

As more and more clients require service level guarantees, service providers and facility managers must determine what facility performance is required to provide the agreed-to or sufficient end user availability. Table 5 shows the relationship between uptime percentage and allowable downtime.

Table 5: Relationship Between Uptime Percentage and Allowable Downtime

Targeted Uptime (percent)	*Allowable Maximum Annual Downtime (minutes)*
< 99.0	>5000
99 to 99.9	500 – 5000
99.9 to 99.99	50 – 500
99.99 to 99.999	5 – 50
99.999 to 99.9999	0.5 – 5.0

Availability levels of 99.99% (50 minutes of downtime per year) allow practically no facility downtime for maintenance or other planned or unplanned events. Therefore, migrating to high-reliability facilities is imperative.

As computers have become more reliable, the overall percentage of downtime events caused by facility failures has grown. Although the occurrence of such facility outages remains small, the total availability is dramatically affected because repair times (mean time to repair or MTTR) for facility outages are lengthy.

6.3 Financial Models

Over the last 30 years, the cost of facilities has not grown proportionately with the increasing cost of the computer hardware housed within those facilities. This has created a situation where favoring information technology hardware over facility infrastructure improvements in budget priorities has led to inadequate end-to-end performance.

The most appropriate way to ensure a balanced deployment of available capital funds is to prepare a business case that includes the costs associated with downtime risk. This cost is a function of both the consequences of an unplanned service outage and the probability that an outage will occur. For example, the costs associated with the different Classes of availability can be calculated as shown in Table 6 if the following hypothetical assumptions are used:

- The cost to the business for an outage is $1 million per hour.
- The cost to build a data center with a design reliability target of 99% is $200/ft^2.
- The cost to build a data center with a design reliability target of 99.9% is $300/ ft^2.
- The cost to build a data center with a design reliability target of 99.99% is $400/ft^2.

Table 6: Sample Cost and Benefit of Varying Degrees of Reliability

Design Reliability	*Cost / ft^2 size*	*Total First Costs*	*Annual Downtime*	*Cost of Downtime*	*ROI*
99%	$200/ ft^2 50,000 ft^2	$10 million	5000 min.	$87.6 million	NA
99.9%	$300/ ft^2 50,000 ft^2	$15 million	500 min.	$8.76 million	1,576%
99.99%	$400/ ft^2 50,000 ft^2	$20 million	50 min.	$876,000	867%

NOTE: The assumptions and figures used in Table 6 are conceptual only. The user will have to determine realistic figures. Construction costs will vary from one location to another. The cost of downtime will depend upon the nature of the business or enterprise, and will vary over time.

As shown in the previous table, substantial cost reductions can be achieved by avoiding facility failure through investment in reliability improvements. In addition, by selecting the optimum availability Class, the investment return can be maximized.

6.3.1 Maintenance Costs

The installation of redundant equipment (such as UPS, power conditioners, generators, transformers, switchboards, HVAC, etc) will increase the amount of maintenance and training or outsourcing of maintenance personnel.

6.3.2 Operational Costs

The use of additional energy-consuming power or HVAC distribution equipment will have a negative effect on the power utilization efficiency (PUE) of a facility, and will increase operating costs.

6.3.3 Environmental Costs

The production and use of additional (redundant) power and cooling equipment can result in a higher carbon utilization rate. It may be possible to partially offset some of these costs with government incentives for things such as “green” energy sources.

6.4 Planning Process

A proactive, strategic planning approach to mission-critical facility design and management requires a five-step process:

1) analyze your current facility;
2) identify and prioritize risks and vulnerabilities;
3) provide solutions to minimize risks;
4) develop an implementation strategy;
5) measure performance and verify improvement.

This should be done continuously, typically in an annual cycle. In this process, plans can be refined and modified as objectives are met or new technology is deployed.

7 Other Factors

The process by which mission-critical Availability Classes are defined is not a perfect science. As projects are built, there will be unexpected outcomes and learned lessons. The following are just a few factors that may alter the selected Availability Class. Other factors will be added over time.

7.1 Intangible Consequences of Downtime

On occasion, a new product rollout, technical initiative, or other major endeavor will be announced. With heightened press exposure or internal performance pressures, there will be an incalculable and unpredictable cost of unplanned downtime. Avoiding these types of circumstances may dictate a higher Availability Class than is otherwise indicated.

7.2 Scheduled Deployment

If a critical IT function must be deployed quickly, it may dictate different risk management strategies, outside that normally considered.

7.3 Unusual Budget Constraints

If the established budget for a critical IT facility will not support the required Availability Class, then a less reliable facility will need to be built unless additional funding is provided.

8 Other Reliability Alternatives

While systems in a Class F3 data center may only expect to stay up as long as the facilities stay up, that is for 99.99% of the time, designs with clustered systems having nodes spread across multiple Class F3 data centers can provide better uptime (see the math in 6.1 and Figure 3), potentially matching or exceeding the uptime of a single Class F4 data center.

In such a design the first failover is to the local node (synchronous), the second failover is to a nearby data center (~16 km [10 miles], and still synchronous) and the third is to a remote data center (but asynchronous).

Such a design does increase the facilities overhead and therefore the cost. However, it offers a way for designers to avoid many of the costs associated with Class F4 data centers, whether owned, leased or collocated.

9 Reliability Planning Worksheet

Use the following planning guide starting on the next page to determine the critical IT facility requirements.

Project name:

Project number:

Project description:

Project location:

STEP 1: Determine operational requirements

1) Does this IT function support a production operation? Yes____ No____

   If yes, proceed to the next line; otherwise your Availability Class can be F0 or F1
   Note: Production Operation is considered to be any IT operation, the loss of which would negatively impact achievement of the organization's mission
2) How many hours of operation must be supported during a production week? ____
3) How many scheduled production weeks are there? (if production occurs every week enter 52.14) ____
4) Multiply line 2 by line 3 and enter here. This is annual production hours: _____
5) Subtract line 4 from 8,760 and enter the result (allowable annual maintenance hours) here: ______
6) If line 5 is greater than 400, the Operational Level is 0; otherwise proceed to the next line.
7) If line 2 is less than 168, and line 5 is greater 100, the Operational Level is 1; otherwise proceed to the next line.
8) If line 5 is between 50 and 99, the Operational Level is 2; otherwise proceed to the next line.
9) If line 5 is between 1 and 49, the Operational Level is 3; otherwise, the Operational Level is 4.

STEP 2: Determine Operational Availability Rank.

1) Based on the operational level from Step 1 above:
   - Level 0; Proceed to line 2;
   - Level 1: Proceed to line 3;
   - Level 2: Proceed to line 4;
   - Level 3: Proceed to line 5;
   - Level 4: Proceed to line 6.
2) Operational Level 0: If the maximum annual downtime is:
   - 500 minutes or greater, then the availability requirement is Operational Availability Rank 0.
   - between 50 and 500 minutes, then the availability requirement is Operational Availability Rank 1.
   - less than 5 minutes, then the availability requirement is Operational Availability Rank 2.

   Proceed to Step 3.
3) Operational Level 1: If the maximum annual downtime is:
   - 5000 minutes or greater, then the availability requirement is Operational Availability Rank 0.
   - between 50 and 5000 minutes, then the availability requirement is Operational Availability Rank 1.
   - between 5 and 50 minutes, then the availability requirement is Operational Availability Rank 2.
   - less than 5 minutes, then the availability requirement is Operational Availability Rank 3

   Proceed to Step 3.
4) Operational Level 2: If the maximum annual downtime is:
   - 500 minutes or greater, then the availability requirement is Operational Availability Rank 1.
   - between 5 and 500 minutes, then the availability requirement is Operational Availability Rank 2.
   - less than 5 minutes, then the availability requirement is Operational Availability Rank 3

   Proceed to Step 3.

<Continues on next page>

5) Operational Level 3: If the maximum annual downtime is:
   - 50 minutes or greater, then the availability requirement is Operational Availability Rank 2.
   - between 5 and 50 minutes, then the availability requirement is Operational Availability Rank 3.
   - less than 5 minutes, then the availability requirement is Operational Availability Rank 4

   Proceed to Step 3.
6) Operational Level 4: If the maximum annual downtime is:
   - 50 minutes or greater, then the availability requirement is Operational Availability Rank 3.
   - less than 50 minutes, then the availability requirement is Operational Availability Rank 4

   Proceed to Step 3.

STEP 3: Define Mission-critical Risk Level

Downtime will reduce or negatively impact operations (select one):

- Catastrophic (e.g., across the entire enterprise) _____
- Severe (e.g., across a wide portion of the enterprise) _____
- Major (e.g., across a single region or department) _____
- Minor (e.g., at a single location) _____
- Isolated (e.g., a single non-critical function) _____

STEP 4: Determine from the Table below

1) Select the column from the Operational Availability Rank in Step 2
2) Select the row from the Risk Level in Step 3
3) Your Facility Availability Class is where the two intersect _____

Facility Availability Class

Impact of Downtime	*Operational Availability Rank*				
	0	*1*	*2*	*3*	*4*
Isolated	Class F0	Class F0	**Class F1**	Class F2	Class F2
Minor	Class F0	**Class F1**	Class F2	**Class F3**	**Class F3**
Major	**Class F1**	Class F2	Class F2	**Class F3**	**Class F3**
Severe	**Class F1**	Class F2	**Class F3**	**Class F3**	Class F4
Catastrophic	**Class F1**	Class F2	**Class F3**	Class F4	Class F4