Sangfor Application Criticality Model and Reliability Design Best Practice

This paper compares three common classification methods and then gives guidance on how to choose criticality based on typical levels of different applications & environment. Also, these applications will be offered with typical reliability configuration, based on Sangfor's best practice recommendations.

Introduction

Reliability is a data center manager's most important business target. Uptime and mean time between failure (MTBF) are often used to describe data center reliability, but there are also several issues including:

1. Some data center classification is focused only on data center infrastructure such as power supply and UPS, but rarely pay attention to data center layer 2—the IT infrastructure.
2. Existing measurements are quantitative and difficult for data center managers to calculate.
3. Different applications have different reliability concerns while existing measurements are not taken into consideration.

This paper compares three common classification methods and then gives guidance on how to choose criticality based on typical levels of different applications & environment. Also, these applications will be offered with typical reliability configuration, based on Sangfor's best practice recommendations.

This model is useful for data center managers designing their IT infrastructure reliability based on applications' criticality.

Common Classification Methods

Traditional data center performance was heavily linked to the designers. Too much emphasis was placed on the designers' personal experience, advice from other designers and colleagues with a great deal of focus fixated on historical and personal experience with the causes of downtime and error, rather than the unique needs of the specific data center or business. Thus, data center design has routinely developed without consistency or uniformity and without due consideration to unique business requirements – leading to an understandable lack of flexibility and scalability as well as troubling complexity.

As these issues become increasingly detrimental to businesses, criticality or tier systems began to develop to encourage more availability and reliability performance in data center designs. Using historical data center performance information allowed designers to compare, contrast and develop a best practice method for data center construction. Three of the most commonly

utilized methods are The Uptime Institute's Tier Data Center Site Infrastructure Tier Standard, TIA 942 and Syska Hennessy Group's Criticality Levels.

The Uptime Institute's Data Center Site Infrastructure Tier Standard[1]

Developed in 1995 and widely referenced in the data center construction industry, The Uptime Institute has used information gained over years of data center projects to develop a constantly evolving standard that includes 4 separate Tiers. The Upton Institute standard provides a high-level guideline but no design details for each individual Tier.

TIA 942

TIA 942 (revision 5) describes the 4 Tier levels as "informative and not considered to be requirements of this Standard"[2]. Designers will be unable to use the TIA 942 to build to a specific tier level – the standard provided is strictly for evaluation.

Syska Hennessy Group's Criticality Levels[3]

Syska has used Uptimes 4 Tiers to develop their own 10 Criticality Levels based on more recent trends and data including high density computing and flexible architectures. Uptime's 4 Tiers are detailed within Syska's first of ten criticality levels, while Syska has continued to develop their standard based on recent data regarding O&M of a data center – expanding beyond the Uptimes focus on components and construction. Syska's Criticality Levels are also high level without the detailed specifics of TIA 942 and have developed based on a balance sheet approach to data center criticality levels – focusing on the weakest elements of a data center and building outward.

[1] Turner, Seader, Renaud, Data Center Site Infrastructure Tier Standard: Topology, 2010

[2] TIA-942, Telecommunications Infrastructure Standard for Data Centers, April 2005, p. 10

[3] Syska Hennessy Group, Inc., Syska Criticality LevelTM Definitions, April 2005, p. 13

Syska Criticality	Uptime Institute (rev 0203)	Estimated % Availability	Expected Annual Downtime
C1	Tier I	98	20-40 hours
C2	Tier II	99	10-25 hours
C3	Tier III	99.9	1-15 hours
C4	Tier IV	99.99	0.25-1 hours

Table 1: General cross-references to "Tiers" and "9'S" commonly used in the industry

Suggested Approach for Choosing Criticality

Cost of downtime and a datacenters total cost of ownership determine the ideal criticality level for an organization. The following table details the specifics of Tiers 1-4 for all three methods as well as the business characteristics and effects on system design.

Criticality	Business Characteristics
1 (Lowest)	➢ Typically small businesses ➢ Mostly cash-based ➢ Limited online presence ➢ Low dependence on IT ➢ Perceive downtime as a tolerable inconvenience
2	➢ Some amount of online revenue generation ➢ Multiple servers ➢ Phone system vital to business ➢ Dependent on email ➢ Some tolerance to scheduled downtime
3	➢ World-wide presence ➢ Majority of revenue from online business ➢ VoIP phone system ➢ High dependence on IT ➢ High cost of downtime ➢ Highly recognized brand
4 (Highest)	➢ Multi-million-dollar business ➢ Majority of revenues from electronic transactions ➢ Business model entirely dependent on IT ➢ Extremely high cost of downtime

Table 2: Summary of Criticalities

The Data Center Science Center from Schneider Electric provides a common criticality classification as detailed below:

Applications	C1	C2	C3	C4	Description
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 Relation Management (CRM)	■	■			Customer data
7x24 Support Centers	■	■			Online assignments, email, tuition
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 & Recovery		■	■		Consumer and company backup
Insurance		■	■		Auto and home insurance
Work-in-Progress Tracking (manufacturing)		■	■		Automobile manufacturer
Work-in-Progress Tracking (manufacturing)		■	■		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 & Settlement			■	■	Equities, bonds, commodities

Table 3: Typical levels of criticality for different applications

Sangfor aCloud is committed to be the enterprise cloud infrastructure, with abundant reliability features. Based on the above classifications, Sangfor Technologies defines the classified aCloud reliability design according to different criticalities, as detailed below:

Criticality	Target	Effect of Layer-2 Design
1 (Lowest)	RPO = days RTO = days	➢ Hardware failure detection ➢ 2 copies & data reconstruction ➢ HA ➢ Daily local backup ➢ Recover service in 1 day
2	RPO = hours RTO = hours	➢ Hardware failure detection ➢ 2 copies & data reconstruction ➢ HA ➢ Local CDP backup, remote replication per 12 hours ➢ recover service in 2 hours

3	RPO = seconds RTO = minutes	➢ Full redundancy for management, service, network, storage plane. ➢ Hardware failure detection ➢ 2 copies & data reconstruction ➢ HA ➢ DRS ➢ Resource reservation ➢ CDP local backup, remote replication in seconds ➢ Recover service in 10 minutes ➢ A-S deployment for aCMP
4 **(Highest)**	RPO = 0 RTO = 0	➢ Full redundancy for management, service, network, storage plane. ➢ Hardware failure detection ➢ 2 copies & data reconstruction ➢ HA ➢ DRS ➢ Resource reservation ➢ Fault domain configuration ➢ Active-Active with stretched cluster & load balancer ➢ CDP local backup, remote replication in seconds ➢ A-S deployment for aCMP

Table 4: Sangfor reliability design practice