Network Storage Evaluations Using Reliability Calculations

Selim Daoud, Sun Professional Services

Sun BluePrints™ OnLine—June 2002

http://www.sun.com/blueprints

Sun Microsystems, Inc.
4150 Network Circle
Santa Clara, CA 95045 U.S.A.
650 960-1300

Part No. 816-5132-10
Revision 1.0, 5/28/02
Edition: June 2002

Network Storage Evaluations Using Reliability Calculations

Today, many storage solutions and configurations are available. You have specific storage requirements, and you need techniques for evaluating which solutions are best suited for your environment.

When designing storage architecture, you must take several parameters into account. Depending on your requirements, some parameters are more important than others. For example, performance might be your main concern, or resiliency, or perhaps cost is the driving influence. A compromise among all parameters must be found in order to achieve the best results for a given environment.

While a complete and thorough evaluation of all storage area networking (SAN) aspects is required in the planning stage, this article provides an introduction to specific techniques for evaluating the redundancy and reliability of network storage solutions. The intent is to provide you with another tool for the trade.

Introduction

This article defines the terms *reliability* and *redundancy*, and describes case studies using three different network storage architectures. The architectures are compared for their advantages and disadvantages in terms of redundancy.

In the case where two solutions are fully redundant, it is important to refine the evaluation. We do so by figuring out the reliability of each solution. This provides additional criteria to use in your network storage architecture planning.

Redundancy

A system is redundant if one failure of any of its components does not affect the system's purpose. Redundancy of a storage system is sought to increase overall reliability.

Redundant storage configurations provide a means to survive hardware failures that are considered inevitable, because at some point in time, a component failure is bound to happen.

To find out if a system is redundant, you must enumerate each one of its components, and for every component, evaluate whether its failure compromises the overall system.

Reliability

For the purpose of this article, reliability is divided into *component reliability* and *system reliability*.

Component Reliability

The overall reliability of a storage system is based on the reliability of each of its components. The calculation of the component reliability (R) value starts with the *mean time between failures* (MTBF) value (published by the manufacturer of each component). From this, we can determine the annual failure rate (AFR), which is used to determine the reliability value.

The MTBF statistic represents the average time it takes for a failure to occur. A MTBF of 100,000 hours means that one failure occurs every 100,000 hours on average.

Component reliability formula:

$$\textbf{AFR} = \frac{\textbf{8760}}{\textbf{MTBF}}$$

Note: 8760 is the total number of hours per year (365 x 24 = 8760).

$$\textbf{R = (100 – AFR)}$$

Example:

For a component with a MTBF value of 100,000 hours, the following reliability value is determined:

AFR = 8760 / 100000 = 0.0876

R = (1 – 0.0876) = 0.9124 (or 91.24%)

System Reliability

Several methods exist to obtain a figure for the *system MTBF*. This article uses a method called *block diagram analysis*.

In a storage system, a component is configured in one of two ways:

- Redundant configuration (in parallel)
- Non-redundant configuration (in series)

The system is split logically into blocks of components. Blocks represent either a redundant component configuration or a non-redundant component configuration.

When the components are in a redundant configuration, the risk of system failure due to the failure of one component diminishes at the power of the number of redundant components.

When configured as a non-redundant components, the risk of a system failure is equal to the sum of the risks of each component.

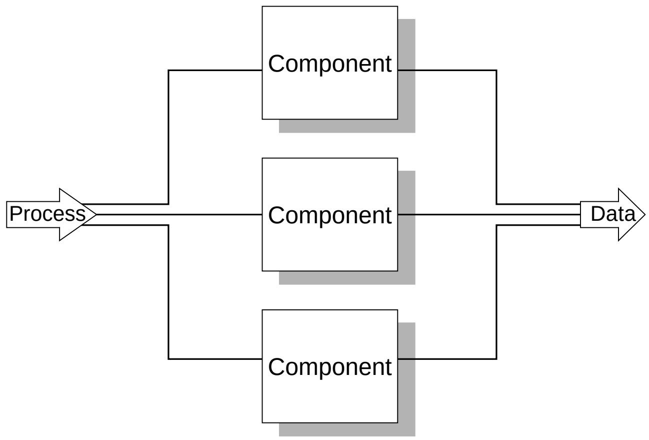
Block Diagram Analysis and Network Storage

The purpose of a network storage system is to link a process (user or application) to data (stored on media).

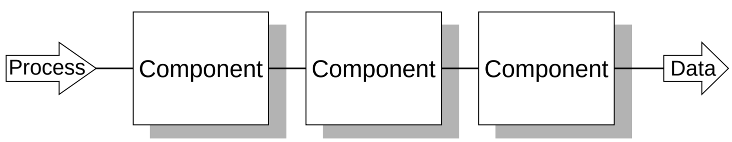
One way to determine the reliability of a network storage architecture, is to use a method called block diagram analysis. We start by drawing a functional picture of this storage system showing how a process can reach the data. Then, using the rules presented above, we calculate the reliability of the overall system.

Examples

Let's consider a logical block made of components each with an AFR = 0.0876. This block can be either in a redundant or non-redundant configuration, as shown in the following figure.



Example 1: Redundant logical block



Example 2: Non-redundant logical block

FIGURE 1 Redundant and Non-redundant Logical Block Diagrams

Example 1:

The block has three components in a redundant configuration. The risk of a system failure in the first year is equal to the risk of all three components failing.

Formula for redundant configurations:

System AFR = $(x)^y$ (x=component AFR, y=number of components in parallel)

Applied to example 1:

System AFR = $(0.0876)^3$ = 0.0006722 (or 0.067%)

Example 2:

The block has three components connected in series. The risk of the whole system failing in the first year is equal to the failure of any single component in the system.

Formula for non-redundant configurations:

System AFR = $x * y$ (x=component AFR, y=number of components in series)

Applied to example 2:

System AFR = $3 * 0.0876$ = 0.2628 (or 26.28%)

Note – This is a very intuitive method to determine the reliability of a system. However, for more complex systems, computer modeling is used to study the reliability.

Case Study

We will determine a reliability figure on three very basic SAN architectures. The starting point of our study is the network storage requirements.

Network Storage Requirements

We want networked storage that has access to one server. Later, this storage will be accessible to other servers. The server is already in place, and has been designed to to sustain single component hardware failures (with dual host bus adapters (HBAs), for example). Data on this storage must be mirrored, and the storage access must also stand up to hardware failures. The cost of the storage system must be reasonable, while still providing good performance.

Our first temptation might be to decide which components to use; switches, hubs, Sun StorEdge™T3 arrays, Sun StorEdge™ A5x00 arrays, and so on. However, a more prudent approach would be to determine the appropriate architecture in terms of its resistance to hardware failures, cost, and performance, leaving the selection of specific components for a later stage.

Note – For this case study, the focus is on storage architecture redundancy and reliability, and does not address cost and performance issues.

Architecture 1

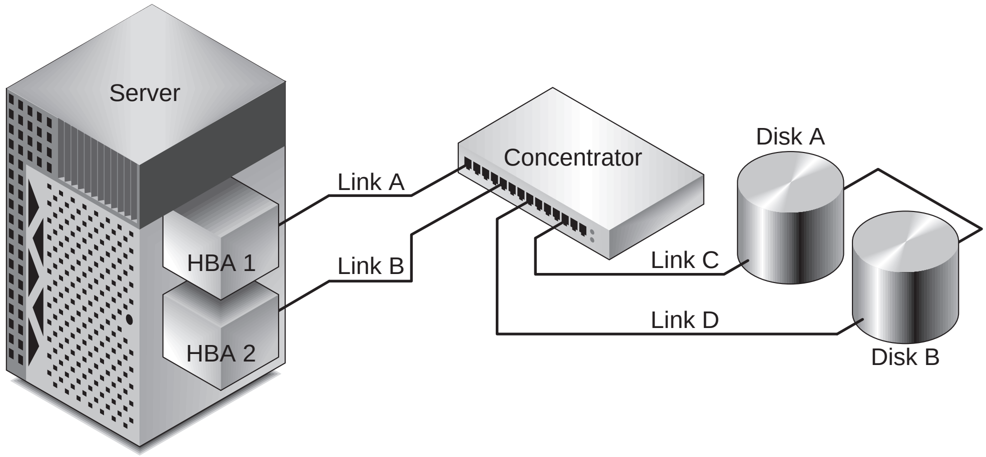


FIGURE 2 Architecture 1 Block Diagram

Architecture 1 provides the basic storage necessities we are looking for with the following advantages and disadvantages:

Advantages:

- Storage is accessible if one of the links is down.
- Storage A is mirrored onto B.
- Other servers can be connected to the concentrator to access the storage.

Disadvantages:

If the concentrator fails, we have no more access to the storage. This concentrator is a single point of failure (SPOF).

Architecture 2

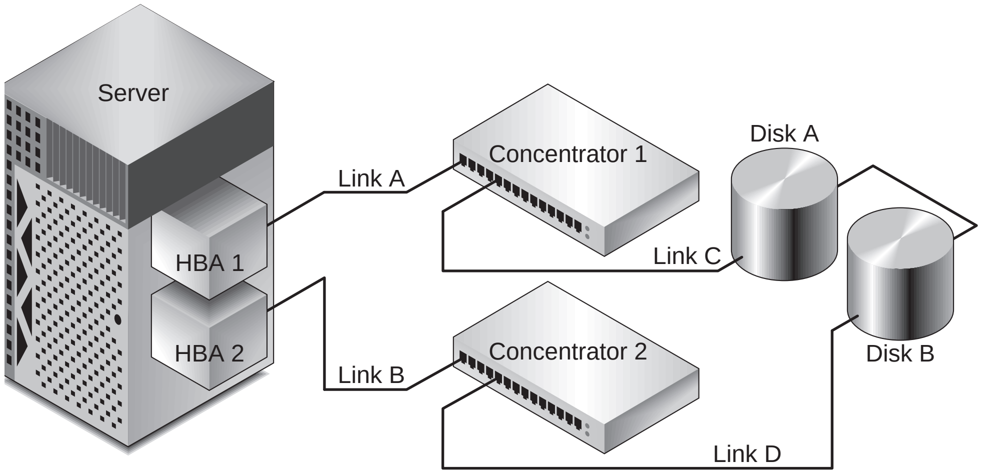


FIGURE 3 Architecture 2 Block Diagram

Architecture 2 has been improved to take into account the previous SPOF. A concentrator has been added, and now the storage configuration is redundant and the requirements are satisfied with the following advantages:

- If any links or components go down, storage is still accessible (resilient to hardware failures).
- Data is mirrored (Disk A <-> Disk B).
- Other servers can be connected to both concentrators to access the storage space.

Architecture 3

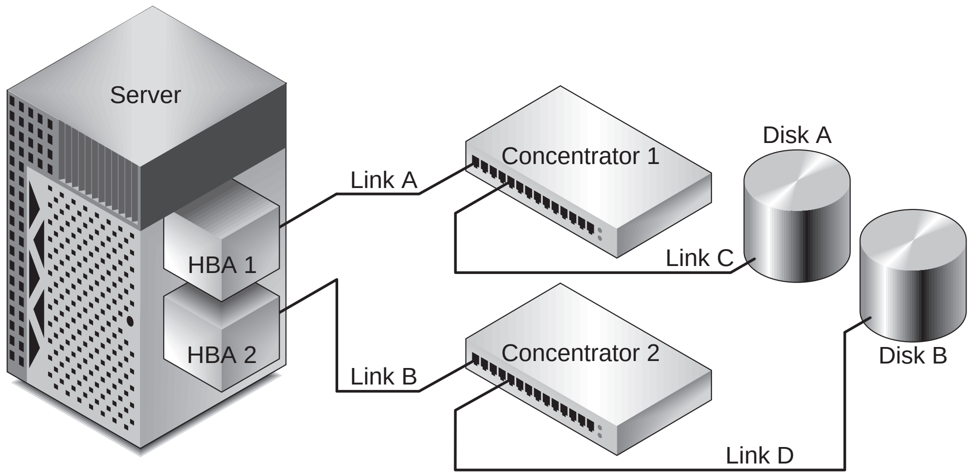


FIGURE 4 Architecture 3 Block Diagram

Architecture 3 seems very close to architecture 2. The main difference resides in the fact that Disk A and Disk B have only one data path. Disk A is still mirrored to Disk B, as required.

This architecture has all the advantages of the previous architectures with the following differences:

- Disk A can only be accessed through Link C, and Disk B only through Link D.
- There is no data multipathing software layer, which results in easier administration and easier troubleshooting.

In some sense it seems we are loosing a level of redundancy in architecture 3. To appreciate the differences between architecture 2 and 3, we will use block diagram analysis to determine and compare their reliability values.

Determining Redundancy

We first list an inventory of components involved in the three architectures as shown in the first column of the following table. Next, we analyze the three architectures for redundancy.

Failing Component (first failure)	Architecture 1: Is the System OK?	Architecture 2 and 3: Is the System OK?
HBA 1	Yes	Yes
HBA 2	Yes	Yes
Link A	Yes	Yes
Link B	Yes	Yes
Concentrator 1	**No**	Yes
Concentrator 2[1]	n/a	Yes
Link C	Yes	Yes
Link D	Yes	Yes
Disk A	Yes	Yes
Disk B	Yes	Yes
Total number of redundant components	**8**	**10**

1. This component is not available in architecture 1 because there is only one concentrator in that configuration.

Consequently, we see that Architecture 2 and 3 satisfy our objectives for redundancy, while Architecture 1 does not.

It is possible to obtain an objective difference between architecture 2 and 3 by studying their respective reliability. We will find that, although both architecture 2 and 3 are fully redundant, one is more reliable than the other.

Determining Reliability

Using the reliability formulas discussed earlier, we can determine which architecture has the highest reliability value. For the purpose of this article, we will use sample MTBF values (as obtained by the manufacturer) and AFR values shown in the table below:

TABLE 1 Component Inventory

Component	AFR Variable	Sample MTBF Values (hours)	AFR[2]
HBA 1	H	800,000	0.011
HBA 2	H		
Link A	L	400,000	0.022
Link B	L		
Concentrator 1	C	580,000	0.0151
Concentrator 2[1]	C		
Link C	L	400,000	0.022
Link D	L		
Disk A	D	1,000,000	0.0088
Disk B	D		

1. This component is not available in architecture 1 because there is only one concentrator in that configuration.

2. The AFR for each component was calculated using the MTBF where (8760 / MTBF) = AFR.

Note – The example MTBF values were taken from real network storage component statistics. However, such values vary greatly, and these numbers are given here purely for illustration.

Architecture 1

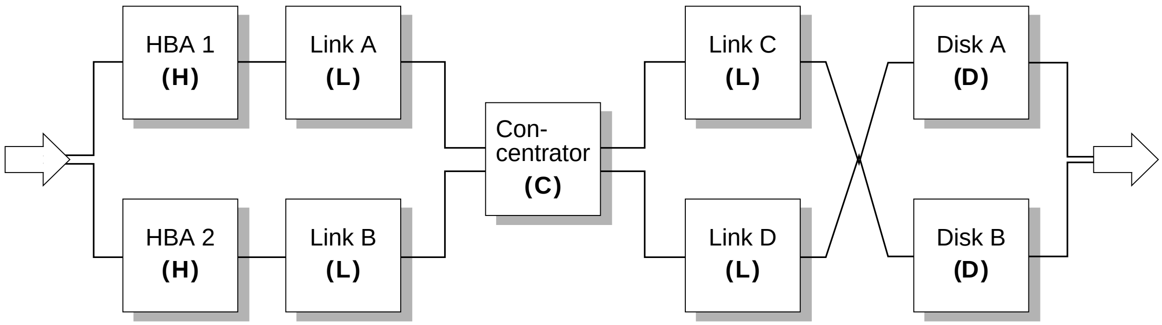


FIGURE 5 Architecture 1 Reliability Block Diagram

Having the rate of failure of each individual component, we can obtain the system's annual failure rate AFR_1 and consequently the system reliability and system MTBF values. Using the block diagram (FIGURE 5), it is easy to identify which components are configured redundantly, and which are not. The following formula is derived using the block diagram analysis discussed earlier. The AFR values of redundant components are multiplied to the power equal to the number of redundant components. The AFR values of non-redundant components are multiplied by the number of those components in series. In this case, the concentrator (C) is the only non-redundant component (C * 1= C). And finally, the AFR values are summed.

The formula for this architecture:

$AFR_1 = (H + L)^2 + C + L^2 + D^2$

Sample values applied:

$AFR_1 = (0.011 + 0.022)^2 + 0.0151 + 0.022^2 + 0.0088^2 = 0.0167$

Using the AFR value, we determine the annual reliability R_1 of the system:

$R_1 = 1 – AFR_1$

$R_1 = 1 – 0.0167 = 0.9833$, or 98.33%

Using the AFR value, the following system MTBF value is derived:

System MTBF = 8760 / AFR_1

System MTBF = 8760 / 0.0167 = 524,551 hours

Architecture 2

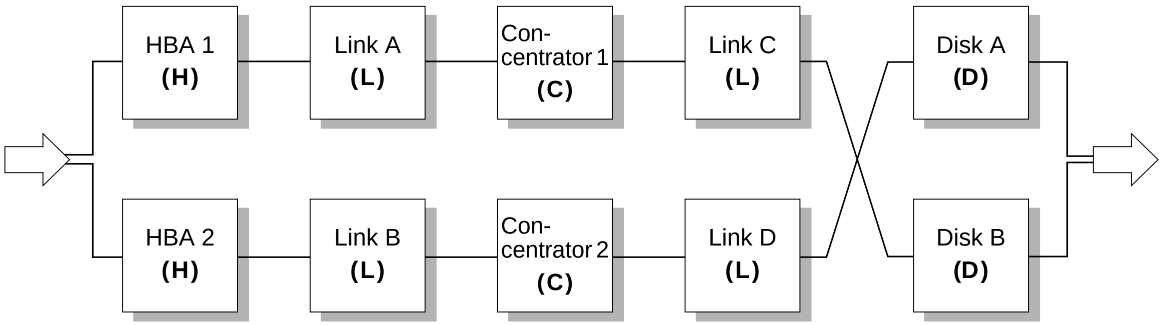


FIGURE 6 Architecture 2 Reliability Block Diagram

This architecture has a different configuration, and the resulting formula is derived using the block diagram analysis.

The formula for this architecture:

$AFR_2 = (H + L + C + L)^2 + D^2$

Sample values applied:

$AFR_2 = (0.011 + 0.022 + 0.0151 + 0.022)^2 + 0.0088^2 = 0.005$

Using the AFR, determine the annual reliability R_2 of the system:

$R_2 = 1 – AFR_2$

$R_2 = 1 – 0.005 = 0.995$, or 99.5%

Using the AFR value, the following system MTBF value is derived:

System MTBF = 8760 / AFR_2

System MTBF = 8760 / 0.005 = 1,752,000 hours

Architecture 3

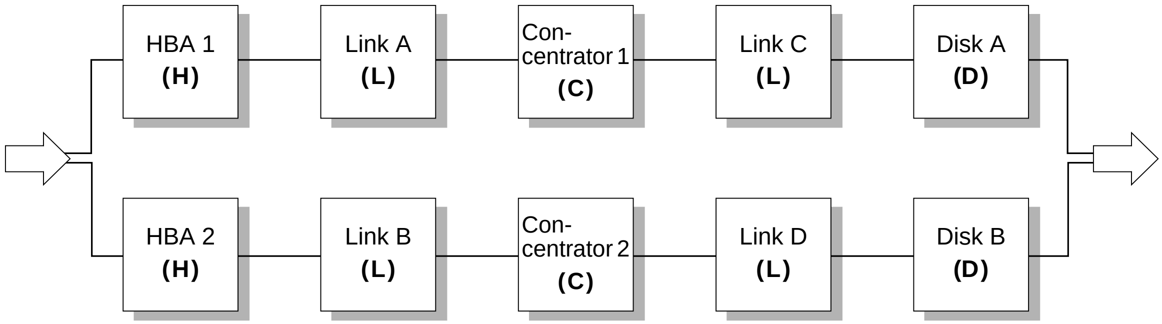


FIGURE 7 Architecture 3 Reliability Block Diagram

Architecture 3 results in yet another block diagram calculation.

The formula for this architecture:

$AFR_3 = (H + L + C + L + D)^2$

Sample values applied:

$AFR_3 = (0.011 + 0.022 + 0.0151 + 0.022 + 0.0088)^2 = 0.0062$

Using the AFR, determine the annual reliability R_3 of the system.

The formula:

$R_3 = 1 – AFR_3$

Numbers applied:

$R_3 = 1 – 0.0062 = 0.9938$, or 99.38%

Using the AFR value, the following system MTBF value is derived:

System MTBF = 8760 / AFR_3

System MTBF = 8760 / 0.0062= 1,412,903 hours

Conclusion

When the calculations are complete, we compare the data:

Architecture 1 = 98.33%, or a System's MTBF = 524,551 hours

Architecture 2 = 99.50%, or a System's MTBF = 1,752,000 hours

Architecture 3 = 99.38%, or a System's MTBF = 1,412,903 hours

The MTBF figures are the most revealing, and indicate that architecture 2 is statistically the most reliable of all.

In conclusion, the case study calculations provide the following points:

- Only architecture 2 and 3 are fully redundant, hence they satisfy the requirement of a redundant configuration that can sustain a single hardware failure.
- The reliability value for Architecture 1 doesn't show the non-redundant aspect of this architecture. It is therefore important to consider both characteristics; redundancy and reliability.
- Architecture 2 is nearly three times more reliable than Architecture 1, and has an estimated higher MTBF of 339,097 hours when compared to architecture 3.

Finally, weighing the advantages of one solution over the another, we must also take other parameters into account, such as:

- Storage capacity requirements
- Performance
- Cost
- Maintainability (indexed by the MTTR: mean time to repair)
- Availability (which depends on the MTBF and MTTR)
- Serviceability
- Ease of deployment
- Support

The last point, support, is a critical consideration, because it is through support that a second failure will be avoided by quick troubleshooting and prompt part replacement. One factor not obvious in the calculations is that although we might think Architecture 2 brings more in terms of redundancy, due to the dual path from server to disks, it has the drawback of requiring additional software that can add another layer of complexity that might be less desirable (possibly lowering the ease of deployment and serviceability, while increasing costs).

Finally, it is worth noting that any storage area networking (SAN) implementation must be carefully planned and analyzed before deployment. Added to which, simple SAN design often will be preferable, because of easier support (troubleshooting and problem resolution). But one must not favor one parameter over the others without knowing the consequences, and therefore every aspect of the architecture decision must be considered. This is the only way to increase the reliability of storage architecture.

About the Author

Selim Daoud is a recognized leader in data storage and backup technologies for open systems. He obtained an MSc in computer science at the University of Wales (UK), and an MSc in applied mathematics in computing at Toulouse University in France.

Over the course of his career in the computer industry, Selim gained valuable experience working with data backup technology, storage system design (mainly RAID implementations), and UNIX systems administration. He managed a support organization (dealing with storage and backup technology) in London, served as a consultant specializing in backup systems in Paris, and was in charge of multiple migrations of backup systems and storage deployment for the European Organization for Nuclear Research (CERN) in Geneva, Switzerland.

Selim currently holds a project engineering position, specializing in computer storage and backup technology in the Sun Professional Services organization in Switzerland.

References

Zhu, Ji, “What is Reliability?”, a *Sun Microsystems White Paper*, 2000

Zhu, Ji, “Systems Availability”, a *Sun Microsystems White Paper*, 1994

Farley, Marc, *Building Storage Networks*, McGraw Hill, 2001

Ordering Sun Documents

The SunDocs[SM] program provides more than 250 manuals from Sun Microsystems, Inc. If you live in the United States, Canada, Europe, or Japan, you can purchase documentation sets or individual manuals through this program.

Accessing Sun Documentation Online

The docs.sun.com web site enables you to access Sun technical documentation online. You can browse the docs.sun.com archive or search for a specific book title or subject. The URL is http://docs.sun.com/.

To reference Sun BluePrints OnLine articles, visit the Sun BluePrints OnLine Web site at:

http://www.sun.com/blueprints/online.html