the Availability Digest

www.availabilitydigest.com

SAP on VMware High Availability Analysis

A Mathematical Approach

December 2012

Vas Mitra
SAP Virtualization Architect

Editor's note:

Vas Mitra is a SAP Virtualization Architect for VMware. He has analyzed the availability of SAP applications running on a VMware ESXi cluster using the concepts that we have published in the *Availability Digest's* Geek Corner. He has given us permission to publish his analysis, which follows.

Contents

1. Introduction 4
   - 1.1 Overview 4
   - 1.2 Service Level Agreements 4
   - 1.3 Hardware Failures 5
   - 1.4 Target Audience 5
   - 1.5 References 5
2. SAP on VMware Background 6
   - 2.1 High Availability Options 6
3. Availability of Overall Infrastructure 7
4. Five Node ESXi Host Cluster Example 9
   - 4.1 Example Architecture 9
   - 4.2 Parameter Definition 10
   - 4.3 Availability Calculation 11
5. Final General Equation & Analysis 15
6. Author 16

Table of Figures

Figure 1 Availability Translation from % to Actual Time 4
Figure 2 Database/Central Services Unplanned Downtime Scenarios 7
Figure 3 Overall Infrastructure (Logical) 8
Figure 4 Five Node ESXi Cluster Example 9
Figure 5 Relationship Between mtbf, mttf, mttr 10

1. Introduction

1.1 Overview

This paper describes how to calculate the theoretical availability of SAP deployed in virtual machines on a cluster of x-86 servers running the VMware hypervisor (referred to as ESXi hosts). The content here directly leverages probability and mathematical/algebraic analysis from whitepapers at www.availabilitydigest.com.

The mathematical model can help to determine the availability of a virtual SAP solution expressed as a fraction/percentage. The calculations are first shown for a five node cluster of ESXi hosts but the resulting equation can be generalized for n nodes – this is shown at the end.

1.2 Service Level Agreements

The reason we want to determine the quantitative availability of the SAP system is that service level agreements (SLAs) specifying the degree of uptime can be expressed as a percentage / fraction. The following table shows the translation from a given availability percentage to the corresponding amount of time a system would be unavailable per year, month, or week.

Figure 1 Availability Translation from % to Actual Time

Availability %	Downtime per year	Downtime per month*	Downtime per week
55.5555555% ("nine fives")	162.22 days	13.33 days	74.67 hours
90% ("one nine")	36.5 days	72 hours	16.8 hours
95%	18.25 days	36 hours	8.4 hours
97%	10.96 days	21.6 hours	5.04 hours
98%	7.30 days	14.4 hours	3.36 hours
99% ("two nines")	3.65 days	7.20 hours	1.68 hours
99.5%	1.83 days	3.60 hours	50.4 minutes
99.8%	17.52 hours	86.23 minutes	20.16 minutes
99.9% ("three nines")	8.76 hours	43.8 minutes	10.1 minutes
99.95%	4.38 hours	21.56 minutes	5.04 minutes
99.99% ("four nines")	52.56 minutes	4.32 minutes	1.01 minutes
99.999% ("five nines")	5.26 minutes	25.9 seconds	6.05 seconds
99.9999% ("six nines")	31.5 seconds	2.59 seconds	0.605 seconds
99.99999% ("seven nines")	3.15 seconds	0.259 seconds	0.0605 seconds

Source: http://en.wikipedia.org/wiki/High_availability

1.3 Hardware Failures

.

This paper considers downtime due to hardware faults – this is measured by parameters like Mean Time Between Failures (MTBF) and Mean Time to Repair/Recovery (MTTR) which are commonly used in the IT industry. MTBF is a statistical measure, an estimate that can be provided by hardware vendors to indicate failure rate expected during the useful life of an x-86 server.

The analysis here does not consider downtime due to software corruptions or bugs or operational mistakes due to human error.

1.4 Target Audience

Consultants, architectural staff, system administrators and managers responsible for designing SAP environments on the VMware platform.

1.5 References

The following resources should be consulted for background as the principles used in this document are taken directly from these sources:

- "Calculating Availability – Heterogeneous Systems Part 1 March 2008"

  http://www.availabilitydigest.com/public_articles/0303/calculating_availability_heterogeneous_syst.pdf

  This covers probability 101. Probability theory is the basis for the mathematical calculations in this paper.

- "Calculating Availability – Redundant Systems October 2006"
  http://www.availabilitydigest.com/public_articles/0101/calculating_availability.pdf .

  This paper analyses the availability of multiple nodes/single spare and multiple nodes/ multiple spare configurations. This can be applied to a multi-node ESXi cluster.

- "Simplifying Failover Analysis – Part 1 October 2010"
  http://www.availabilitydigest.com/public_articles/0510/failover_analysis.pdf

  This paper identifies time for service failover as a contributor to downtime. This can be applied to scenarios where SAP services are down while being failed over during a VMware HA event or a cluster service switchover.

- "Breaking the Availability Barrier: Survivable Systems for Enterprise Computing"

  http://www.amazon.com/Breaking-Availability-Barrier-Survivable-Enterprise/dp/1410792323/ref=sr_1_fed1_1?s=movies-tv&ie=UTF8&qid=1355376190&sr=1-1&keywords=breaking

  Appendix 2 analyzes a system of n identical elements with s spares. It derives an approximate availability equation for this setup which can be applied to a multi-node ESXi cluster. This analysis is the basis for the equations in "Simplifying Failover Analysis – Part 1 October 2010".

2. SAP on VMware Background

VMware vSphere virtualizes and aggregates the underlying physical hardware resources across multiple systems and provides pools of virtual resources to the datacenter.

SAP AG is a German multinational software corporation that makes enterprise software to manage business operations and customer relations. The company's best-known software products are its enterprise resource planning application (SAP ERP), its enterprise data warehouse solution - SAP Business Warehouse (SAP BW), SAP BusinessObjects software, and most recently, Sybase mobile products and in-memory computing appliance SAP HANA.

The terminology, products, and features used in this document are as follows:

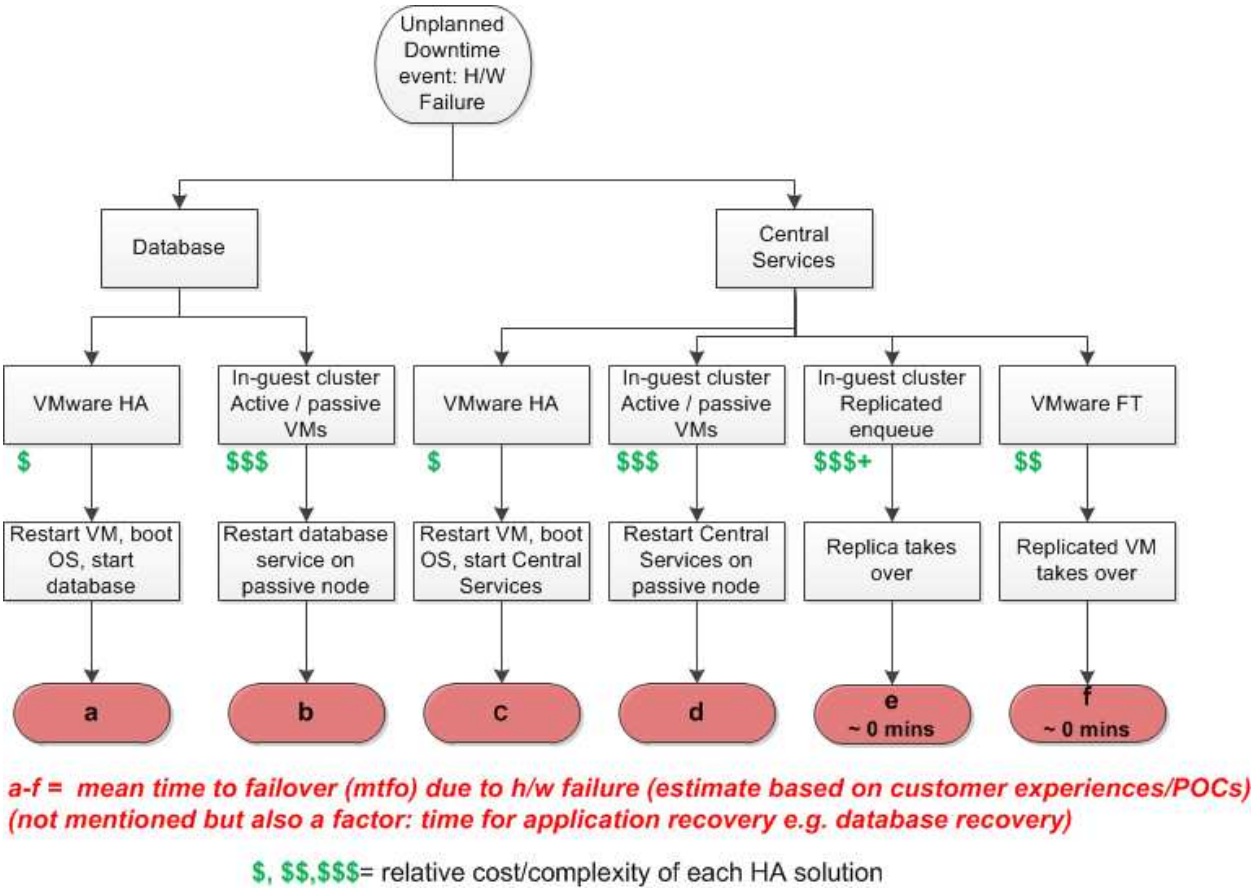
- VMware vSphere – a virtualization platform on which to build and deploy a private cloud that increases control through service-level automation allowing resources to be pooled to deliver IT as a Service (ITaaS).
- ESXi host – an x-86 server running the VMware bare metal hypervisor ESXi (which allows virtual machines to be run).
- VMware High Availability (HA) – provides easy to use, cost effective high availability solutions for applications running in virtual machines. In the event of server failure, affected virtual machines are automatically restarted on other servers with spare capacity.
- VMware Fault Tolerance (FT) - VMware FT protects a virtual machine by maintaining a second virtual machine that runs in lockstep with the primary virtual machine. If the primary virtual machine goes down, the secondary machine takes over with no downtime. Currently, VMware FT supports only single-CPU virtual machines and is a viable solution for lightweight components of the SAP architecture such as Central Services.
- ESXi cluster – a group of ESXi hosts defined in VMware's management tool (vCenter) that enables the group to behave as a cluster such that if one ESXi host fails all the virtual machines running on the failed ESXi host are restarted on the remaining ESXi hosts in the cluster (VMware HA).
- The SAP architecture includes two software components that are a single-point-of-failure: the database and Central Services. Central Services comprises messaging and locking functions. Both these components can be installed separately in virtual machines and a failure of either means the whole SAP system is down.
- Third party clustering software – refers to clustering software like Microsoft Cluster Services, Linux-HA and Veritas Cluster Services which have agents to monitor SAP single-points-of-failure such as the database and Central Services. Such software can also be installed in virtual machines in a manner similar to physical to create an active/passive 2-node cluster. Each virtual machine is a cluster node and the pair can exist across an ESX cluster and you would typically configure vSphere to force the virtual machines to reside on separate ESXi hosts whenever possible.

2.1 High Availability Options

The different high availability scenarios for SAP running on VMware is covered in section "7 High Availability" of whitepaper "SAP Solutions on VMware Best Practices Guide " http://www.vmware.com/files/pdf/solutions/sap/SAP-Solutions-on-VMware-Best-Practices-Guide.pdf

The following figure shows the different failover scenarios in case of a hardware/ ESXi host failure.

Figure 2 Database/Central Services Unplanned Downtime Scenarios

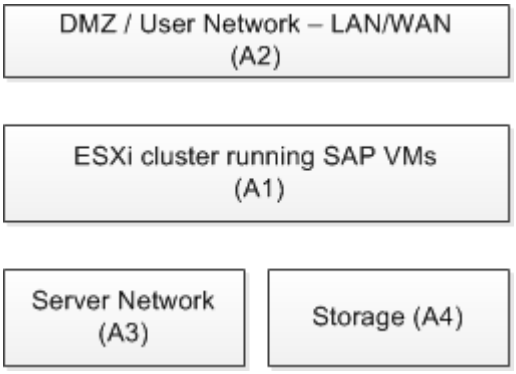


The time it takes for the SAP related software to failover to another ESXi host or virtual machine is downtime for the end-user. This time is referred to as Mean Time to Failover and this parameter is used in the final availability equation.

3. Availability of Overall Infrastructure

This document focuses on the availability analysis of an ESXi cluster running SAP systems in virtual machines but it should be noted that the overall availability of the SAP environment depends on the complete infrastructure beyond the ESXi hosts. Network and storage also impact the overall availability as experienced by the end-user. The following diagram depicts the logical architecture of the complete infrastructure required to keep SAP up and available to end-users. Each component has its own availability.

Figure 3 Overall Infrastructure (Logical)



A<n> in the above figure represents the availability of each component. What we will show in this document is a method to calculate A1, however the final availability that is seen by the end-user is based on the probability that all components of the infrastructure are up and running.

Availability A<n>, is expressed as a percentage or fraction and is equal to the probability that a component/system is up and running. As availability is another way to express a probability we can use standard mathematical probability techniques to calculate the overall availability of a system that is made up of sub components.

Probability of all components up and running equals:

- Probability that the User Network is up AND
- Probability that the SAP systems in the ESXi cluster are up AND
- Probability that the Server Network is up AND
- Probability that the Storage is up

The probability of the above = the final availability = A1 x A2 x A3 x A4 .

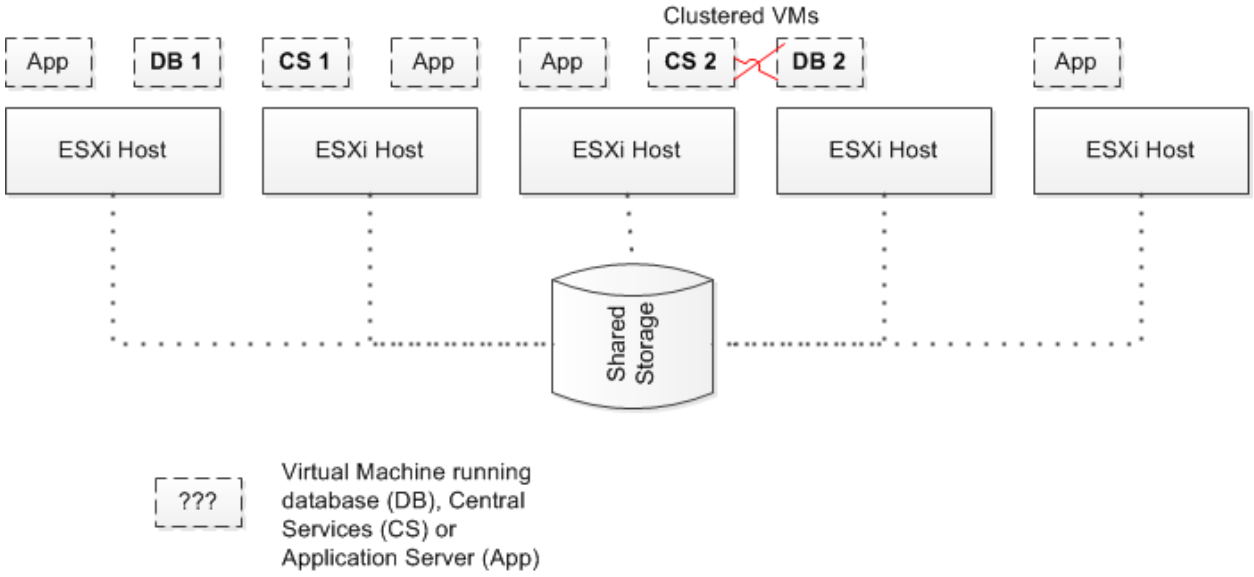
Now let's show how we can calculate A1. Note that A1 includes all downtime attributable to the ESXi, cluster which is the purpose of this analysis. It does not include downtime due to other factors such as application software bugs or the recovery of storage following a failure.

4. Five Node ESXi Host Cluster Example

4.1 Example Architecture

The following diagram depicts a five node ESXi host cluster. Five nodes have been deliberately chosen so that this use case is similar to the example in section “Multiple Nodes, Single Spare” in paper “Calculating Availability – Redundant Systems’.

Figure 4 Five Node ESXi Cluster Example



ASSUMPTIONS:

- This is a “N+1” vSphere cluster i.e. environment has been sized for four ESXi hosts, one extra is added (i.e. one spare) so in the event that one server fails all virtual machines failover to the remaining four ESXi hosts and continue to run with no loss of performance.
- Failure of two ESXi hosts is classified as downtime for ALL virtual machines
  - not really true as all the virtual machines could actually be “squeezed” into three ESXi hosts with potentially reduced performance (depends on the memory/workload and also administrators can reconfigure resource priorities to give some virtual machines more resources over others such that even with three servers some more important virtual machines/SAP systems can experience no loss of performance) . However we are going with this CONSERVATIVE assumption.
  - If it feels too conservative the model can be recalculated based on three simultaneous ESXi host failures as a criteria for overall downtime - it depends on the sizing and number of virtual machines and also the relative SLAs of the different SAP systems.

- We will focus on virtual machines running SAP single-points-of failure: database and SAP Central Services.
- Database and Central Services are in separate virtual machines
- SAP application server instances are in separate virtual machines and there are enough of them spread out across the ESXi cluster so in case of a single ESXi host failure a single SAP system never loses all application server instances. Understandably loss of some application server instances may result in users being disconnected and they need to re-logon and reconnect to remaining application servers – we are not considering this as downtime.

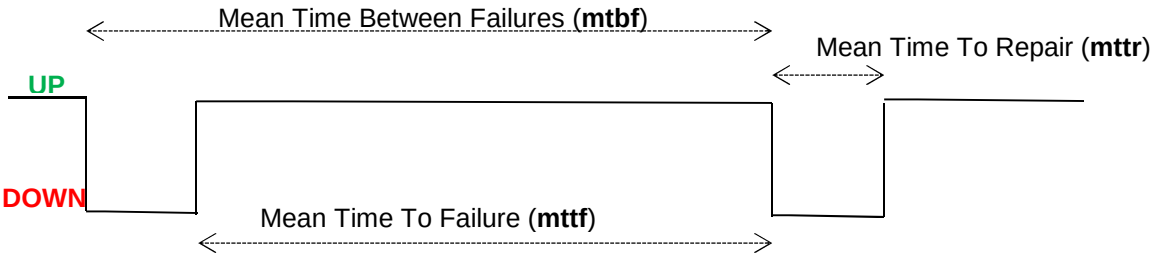
4.2 Parameter Definition

We define the following parameters.

a	availability of an ESXi host (= probability that an ESXi host is up)
f	probability that an ESXi host is down
mtbf	mean time between failure of an ESXi host
mttf	mean time to failure of an ESXi host
mttr	mean time to repair/recover an ESXi host
mtfo_db	mean time to failover database (after h/w failure)
mtfo_cs	mean time to failover Central Services (after h/w failure)
d_db	probability of database failover fault
d_cs	probability of Central Services failover fault
A	availability of SAP system in the ESXi cluster (probability that systems are up)
F	probability that SAP systems are down

The following diagram describes the relationship between mtbf, mttf and mttr.

Figure 5 Relationship Between mtbf, mttf, mttr



From the above we can write the following set of equations.

Equation Set 1

a = availability of an ESXi host

a = proportion of time ESXi host is up = (time ESXi host is up / total time)

$$a = \frac{(\text{mtbf} - \text{mttr})}{\text{mtbf}}$$

OR

$$a = \frac{\text{mttf}}{\text{mttf} + \text{mttr}}$$

$$\text{mtbf} = \text{mttf} + \text{mttr}$$

$$f = 1 - a$$

In this five node (one spare) ESXi host cluster SAP systems are down if:

- Two or more ESXi hosts fail at the same time
- One ESXi host fails and database or Central Services are in the process of being failed over (via VMware HA or in-guest clustering solution).
- One ESXi host fails and a failover fault occurs

4.3 Availability Calculation

Let p(x) be the probability of event x. From above, the probability that a SAP system is down is defined in Equation 2.

Equation 2

F = p(downtime) =

p(failure of two or more ESXi hosts in the five node cluster with one spare) +
p(database failing over due to h/w fault) +
p(Central Services failing over due to h/w fault) +
p(failover fault)

Let us look at each of the components of downtime.

4.3.1 Probability of Two or More ESXi Hosts Failure

We have a five node ESXi host cluster with the criteria that two or more simultaneous ESXi host failures will result in overall downtime of the cluster.

We need to identify all failure combinations of two or more ESXi hosts – each outcome has a probability. The overall probability is the sum of the probability of each outcome (failure mode). Some of the failure modes are identified in the following table.

Table 1 Different Failure Modes (not every combination is shown)

Node 1	Node 2	Node 3	Node 4	Node 5	Probability
Number of ways 2 nodes fail					
X	X	✓	✓	✓	$(1-a)^2 a^3$
X	✓	X	✓	✓	$(1-a)^2 a^3$
X	✓	✓	X	✓	$(1-a)^2 a^3$
X	✓	✓	✓	X	$(1-a)^2 a^3$
✓	X	X	✓	✓	$(1-a)^2 a^3$
✓	X	✓	X	✓	$(1-a)^2 a^3$
✓	X	✓	✓	X	$(1-a)^2 a^3$
✓	✓	X	X	✓	$(1-a)^2 a^3$
✓	✓	X	✓	X	$(1-a)^2 a^3$
✓	✓	✓	X	X	$(1-a)^2 a^3$
Number of ways 3 nodes fail					
X	X	X	✓	✓	$(1-a)^3 a^2$
X	X	✓	X	✓	$(1-a)^3 a^2$
etc, etc					$(1-a)^3 a^2$
Number of ways 4 nodes fail					
X	X	X	X	✓	$(1-a)^4 a$
X	X	X	✓	X	$(1-a)^4 a$
etc, etc					$(1-a)^4 a$
5 node failure					
X	X	X	X	X	$(1-a)^5$
				Final Probability	**Sum of the above**

We can now make some practical approximations to simplify the final result:

- Most modern day servers are very resilient so let's assume they have availability around three nines i.e. a= .999, then (1-a) = .001. So $(1-a)^3$, $(1-a)^4$ and $(1-a)^5$ are near zero. Hence we can ignore the probabilities for failure modes involving three, four and five node failures. We can now approximate that the probability of failure is $10(1-a)^2a^3$
- Since a is very close to one we can approximate that a^3 is near one, so the final probability can be simplified to approximately 10 $(1-a)^2$. Note that the number ten here is the total number of ways two ESXi hosts can fail in the five node cluster.

So we can conclude that:

p(failure of two or more ESXi hosts in a five node cluster with one spare) ~ $10(1-a)^2$

The detailed analysis for this simplification and justification is explained in Appendix 2 of the book "Breaking the Availability Barrier: Survivable Systems for Enterprise Computing".

4.3.2 Probability of Database/Central Services Failover

p(database failing over due to hardware fault) = proportion of time database is down while failing over

= (time database is failing over / total time)

$$= \frac{\text{mtfo\_db}}{\text{mtbf}}$$

P(Central Services failing over due to hardware fault) = proportion of time Central Services is down while failing over

= (time Central Services is failing over / total time)

$$= \frac{\text{mtfo\_cs}}{\text{mtbf}}$$

4.3.3 Failover Faults

A failover fault refers to the event that a problem occurs that prevents successful failover. We do not expect this to occur often, however it is factored into the analysis as field experience may identify this.

A failover fault occurs if an ESXi host fails and then the failover is unsuccessful. An ESXi host will fail with probability (1-a). The probability that the database failover is unsuccessful is d_db. Therefore,

p(database failover fault) = (1-a)d_db

The probability that the Central Services failover is unsuccessful is d_cs. Therefore,

p(Central Services failover fault) = (1-a)d_cs

4.3.4 Final Equation

We bring together all the components to provide the final equation for the availability of the five node ESXi cluster.

F= probability that SAP systems will be down in a five node ESXi cluster (sized with one spare node/ESXi host)

$$\mathbf{F = 10\ (1 - a)^2 + \frac{mtfo\_db}{mtbf} + \frac{mtfo\_cs}{mtbf} + (1 - a)d\_db + (1 - a)d\_cs}$$

A = availability of SAP systems = 1 - F

$$a = \frac{(mtbf - mttr)}{mtbf}$$

OR

$$a = \frac{mttf}{mttf + mttr}$$

The above equation is specific to our five node example but can be generalized for an n node ESXi cluster with s number of spares – this is shown next.

5. Final General Equation & Analysis

The equation in the last section is specific to a five node ESXi cluster sized with one spare. Summarizing, the probability that SAP is down (F) is based on three components:

F = p(multiple ESXi host failures based on the number of spares) + p(database/Central Services failover in case of single ESXi host failure) + p(failover fault)

The general equation for the probability failure of n nodes with s spares is available in the reference paper "Calculating Availability – Redundant Systems' . From this reference we have the following.

The probability of failure for an n node system with s spares is:

$f(1-a)^{s+1}$

where

f = the number of ways s+1 nodes can fail in an n-node cluster.

$$f = \frac{n!}{(s+1)!\,(n-s-1)!}$$

The symbol "!" means "factorial." For instance, 5! = 5x4x3x2x1.

Putting this altogether our final general equation for SAP downtime/availability is as follows.

Final General Equation

F= probability that SAP systems will be down in a n node ESXi cluster sized with s spare ESXi hosts

$$\mathbf{F} = \frac{\boldsymbol{n}!(\mathbf{1}-\boldsymbol{a})^{\boldsymbol{s}+\mathbf{1}}}{(\boldsymbol{s}+\mathbf{1})!(\boldsymbol{n}-\boldsymbol{s}-\mathbf{1})!} + \frac{\mathbf{mtfo\_db}}{\mathbf{mtbf}} + \frac{\mathbf{mtfo\_cs}}{\mathbf{mtbf}} + (\mathbf{1}-\mathbf{a})\mathbf{d\_db} + (\mathbf{1}-\mathbf{a})\mathbf{d\_cs}$$

n = number of ESXi hosts in cluster

s = number of spare ESXi hosts sized in the cluster

a = availability of an ESXi host

$$\mathrm{a} = \frac{(\mathrm{mtbf}-\mathrm{mttr})}{\mathrm{mtbf}}$$

mtbf = mean time between ESXi host failures

mttr = mean time to repair ESXi host after a failure

mtfo_db, mtfo_cs = mean time to failover database and Central Services in case of single ESXi host failure

d_db, d_cs = probability of a failover fault of the database and Central Services

Final availability = A = 1-F

If we substitute n=5 and s=1 in the above equation we will get the result from our five node, one spare example in the previous section.

We can use the final equation along with practical values to replace the variables in order to observe how availability is impacted in different scenarios. The variables can be substituted with values obtained from: field experience; data/statistics gathered from actual implementations; reliability specifications from x-86 server vendors; proof-of-concepts / lab work evaluating failover times. The following example scenarios can then be analyzed:

- How do extra spare nodes impact final availability?
- How does failover time impact the final availability?
- VMware HA adds some extra time for the OS to reboot compared to an active-passive clustering solution, how does this impact availability? VMware HA and clustering solution may have different values for mean time to failover.
- How do failover faults impact the final availability? We do not expect failover faults to be common and if field data show this to be negligible then d_db and d_cs would be zero (which would simplify the final equation).
- In our five node ESXi cluster example we conservatively state that a two node failure (i.e. one spare) results in overall downtime for all virtual machines. However we can split this even further.

  In figure 3 we show two separate SAP systems with databases DB 1 and DB 2 which for example can correspond to a SAP ERP and SAP BW system. ERP and BW could have different SLAs such that priority is always given to ERP in the case of unexpected resource constraints (the VMware resource scheduler can manage this). Hence in the case of a two node failure it could be possible to run both ERP and BW virtual machines on the three remaining nodes such that ERP continues to run with no loss of performance by prioritizing resource shares for ERP virtual machines over BW. Hence for ERP we could say that it effectively has two spares and BW has one spare. From this we can then calculate separate availabilities for ERP and BW – ERP availability would be based on five nodes with two spares and BW availability would be based on five nodes with one spare.

6. Author

Vas Mitra is a SAP Solutions Architect at VMware Inc. He has worked at VMware since 2007 on various SAP projects with partners including SAP to develop SAP on VMware solutions and best practices for partners, customers and the VMware sales and consulting organizations. He is author of the “SAP Solutions on VMware Best Practices Guide “ (http://www.vmware.com/files/pdf/solutions/sap/SAP-Solutions-on-VMware-Best-Practices-Guide.pdf) . Prior to VMware Vas has worked on SAP projects since 1993 as an ABAP developer, Basis administrator and architect with a large Systems Inegrator, IT departments in the chemical/pharmaceutical companies running SAP and in the SAP practice of a large server vendor. Vas has a Masters Degree in Electrical/Electronic Engineering from Imperial College, London.

The author would like to thank Dr Bill Highleyman (editor of www.availabilitydigest.com) for his review and inputs.