Total Cost of Ownership for Entry-Level and Mid-Range Clusters

A Detailed Analysis of the Total Cost of Ownership of Three Different RISC-Based Server Clusters including HP OpenVMS, IBM AIX and Sun Solaris.

February 2004
Version 1.0

Total Cost of Ownership for Entry-Level and Mid-Range Clusters

Executive Summary

The overall concept of Total Cost of Ownership for server clusters has been an important issue within the IT community for many years. Numerous studies and customer experiences have proven that purchase price alone is not an adequate measurement to compare server clusters from various vendors. Other factors, including the costs to manage and maintain the servers, as well as the application availability they provide, usually have a greater financial impact on an organization than just the system's purchase price. Recognizing these factors, TechWise Research developed an analytical approach in 1999 called Reliability-Adjusted Total Cost of Ownership™ that incorporates management costs and application availability in the TCO analysis. As part of this analysis, server clusters from different manufacturers are compared in terms of the actual number of downtime hours per year that customers typically experience. These downtime findings are then converted into a monetary measurement of the cost differences between clusters which TechWise Research refers to as the "Availability Advantage™."

In the past year, the issue of data security and application availability has moved from the realm of the IT manager into the corporate boardroom. Several high profile viruses and worms resulted in excessive Internet traffic bottlenecks and downtime during 2003. For this reason, TechWise Research expanded its proprietary Reliability-Adjusted TCO™ technique to incorporate downtime caused by software viruses and worms. To our knowledge, this is the first paper that analyzes three leading RISC-based clusters on this important issue.

This study focuses on the following three entry-level and mid-range server clusters: HP OpenVMS/ AlphaServer, IBM AIX/pSeries, and Sun Solaris/Sun Fire clusters. In the Fall of 2003, TechWise Research interviewed a total of 94 IT professionals in U.S. firms. The purpose of these interviews was to collect data on the operational costs associated with installing, managing, and maintaining their clusters. Information was also collected on the number of hours and associated costs for various downtime events each company experienced over a twelve-month period. All of these operational data were then combined with current system and service pricing (from IDEAS International) to calculate the Reliability-Adjusted TCO™ - an analytical approach that factors downtime costs/rates into the TCO analysis. TechWise Research included four main cost components in the TCO analysis. These are the costs to (1) buy the servers and service contract, (2) install and configure the cluster, (3) manage and maintain the cluster over three years, and, (4) the costs associated with application downtime over three years. Three different cluster configurations (2-way, 4-way and 12 to 16-way systems) were analyzed for the three platform brands. For each configuration, the Reliability-Adjusted TCO™ was calculated at various downtime costs to allow readers the ability to compare the different platforms at a downtime cost rate that most applies for their firm.

Study Results: For entry-level 2-way clusters, the acquisition costs (list price of the servers and service agreement) only represents 7% of the three-year TCO. As the clusters increase in complexity to a 4-way, then to a 12 to 16-way cluster, acquisition price represents a larger portion of the TCO, specifically 22% and 26%, respectively. However, regardless of configuration, **the vast majority of the total cost of ownership is due to management and downtime costs.** In terms of downtime, nearly half of the study's respondents report that their company loses as least $10,000 for each hour their cluster is down. The average cost per hour of downtime is $145,000.

For all three configurations (2, 4 and 12 to 16-way), the brand offering the lowest TCO changes depending on the cost per hour of downtime. In all three cases, **when there are no costs associated with downtime, Sun Solaris/Sun Fire clusters offer the lowest TCO of the three brands tested. As the cost per hour of downtime increases, Sun loses its TCO lead to HP OpenVMS/AlphaServer clusters.** The "cross-over" point, where Sun loses its "best in class" TCO status to HP, varies for the three configurations. For 2-way clusters, HP has the lowest TCO when downtime costs are more than $1,585 per hour. For 4-way clusters, HP offers the lowest TCO for downtime costs greater than $8,004 per hour. Lastly, for 12 to 16-way clusters, HP offers the lowest TCO when downtime costs are greater than $18,251 per hour. In all three configurations, IBM clusters' TCO falls between HP and Sun - with one exception. IBM has the highest TCO for 2-way clusters when downtime costs are less than $7,000 per hour.

Sun's TCO advantage at low hourly downtime rates can be attributed to the fact that Sun clusters have a lower list price than comparable clusters from HP or IBM. HP's TCO advantage at higher hourly downtime rates can be attributed to HP averaging the fewest hours of downtime for five of the seven causes of downtime tested. HP demonstrated a substantial Availability Advantage™ in the areas of crashes caused by software viruses/worms and end-user applications. For example, HP OpenVMS cluster owners reported an average of 0.88 hours of downtime per year due to software viruses or worms. Sun Solaris clusters were second best in this category averaging 4.32 hours, while IBM AIX clusters averaged 5.73 hours. For each $10,000 of hourly downtime costs, HP's Availability Advantage™ over Sun and IBM, for this one category, translates into $103,200 and $145,500, respectively, over a three-year period. In terms of operating systems, the OpenVMS operating system delivered the highest average availability of 99.990%, with IBM's AIX coming in a close second at 99.987% and Sun at 99.972%.

When all seven types of crashes are considered, IBM AIX/pSeries clusters averaged 8.98 hours more downtime per year than HP OpenVMS/AlphaServer clusters. Furthermore, Sun Solaris/Sun Fire clusters averaged 21.50 hours more downtime per year than HP OpenVMS/AlphaServer clusters. **For each $10,000 of hourly downtime costs, HP's total Availability Advantage™ over IBM and Sun translates into $269,400 and $645,000 of savings, respectively, over a three-year period.** Any IT manager who ignores availability in their purchase decision would be ignoring the most expensive cost component of the server cluster. This white paper summarizes the results of the study.

Background on This Paper

In October 2001, TechWise Research published the report: Total Cost of Ownership for Low-End and Mid-Range Server Clusters. That report provided a robust analysis of the Total Cost of Ownership (TCO) for four RISC-based and two Intel-based server clusters. Since that report was published, manufacturers have introduced new server models and released updates for both their cluster software and operating systems. Furthermore, two of the server manufacturers (HP and Compaq) merged into one company. Because the findings from the October 2001 report are no longer current, TechWise Research completed this new study to provide updated information on server cluster TCO. For this particular paper, TechWise Research decided to focus on RISC-based platforms only. This is because at the time of data collection, Windows Server 2003 clusters did not meet our minimum criteria of being deployed in a production mode for at least six months. Prior TechWise studies have shown that the operating system and clustering software play a crucial role in determining TCO. Little value would have been added by repeating the study for Intel-based solutions running the same operating system and clustering software studied in 2001. Therefore, the decision to study Intel-based clusters is being postponed until such time as Windows Server 2003 clusters are widely deployed, and sufficient companies request an update.

The History of TCO

The overall concept of Total Cost of Ownership has been an important issue within the IT community for many years. IT managers are constantly searching for ways to stretch their IT budget. One of the earliest examples of this was Digital Equipment's VAX servers, which became a hugely successful platform by offering IBM mainframe customers a more cost-effective solution. At that time, financial comparisons between systems were mostly based on purchase price alone. Over the past 25 years, it became clear that the cluster's price, although important in the purchase decision, did not reflect its true cost over time. Therefore, companies started to consider other factors, such as operational costs, in their financial evaluations. Thus, the concept of TCO was born.

Some TCO analyses performed today are based only on the prices that manufacturers charge for their clusters and the service contracts purchased with them. The main benefits of this TCO "formula" are that the data required to calculate it are widely available, and it is easy for companies to calculate and interpret the findings themselves. To compare clusters from different vendors, all a customer had to do was contact each vendor (or an authorized dealer) and get a quote on both the cluster and a service contract. The drawback of this approach is that it overlooked other cost drivers that significantly affect an organization's operations, personnel and profitability. TechWise Research has conducted a number of studies over the past five years which show management costs (the time spent managing and maintaining a server cluster over its lifetime) contribute a significant amount to a cluster's TCO. As a result, TechWise includes management costs in all its analyses to provide a more comprehensive and realistic view of TCO.

Previous studies have also shown that availability is an important part of TCO. One of the primary functions of a server cluster is to provide high availability through automated failover functionality. Installing server clusters is one strategy IT managers can use to ensure application availability, and mitigate the impact of server downtime. Depending on the cluster's application, the cost per hour of downtime can be considerable. As a result, in 1999, TechWise Research expanded the concept of TCO analysis by developing a proprietary technique called Reliability-Adjusted TCO™. This technique incorporates the cost of downtime in the TCO analysis. With the continued growth of e-business and e-commerce applications, companies need to ensure that their primary applications are up and running twenty-four hours a day, every day. Indeed, in the past few years the concept of availability has moved beyond the sphere of IT managers and into the corporate boardroom. Server manufacturers have recognized this trend. Many have launched national print and TV advertising campaigns focusing on TCO and availability. Additionally, the issue of data security rose to the level of "front page news" repeatedly during 2003. Major television and Internet news organizations featured articles on security throughout the year. One leading data security company, F-Secure, went as far as to dub 2003 as "The Year of the Worm"[(1)]. This is due to several high-profile viruses including the Slammer network worm, Bugbear.B email worm, and the Blaster and Sobig.F network worms. Although these viruses primarily targeted Microsoft Windows systems, they rapidly propagated throughout the Internet causing excessive traffic and other problems on all types of servers worldwide. The Sobig.F worm alone resulted in 300 million infected email messages worldwide. Based on these trends, TechWise Research expanded our TCO analysis to isolate crashes that result from viruses or worms. To our knowledge, this is the first study that compares RISC-based server clusters on this important issue.

(1) *F-Secure Corporation's Data Security Summary for 2003,* F-Secure, December 2003

TechWise Research's TCO Model

TechWise Research conducted a survey to collect information regarding the following three operational components of TCO:

- **Start-Up Costs** – The costs to install and configure the cluster, as well as any time and money spent to train staff on the cluster.
- **Management Costs** – The ongoing costs associated with managing the cluster on an annual basis.
- **Downtime Costs** – The number of hours, and resulting costs for cluster downtime, on an annual basis.

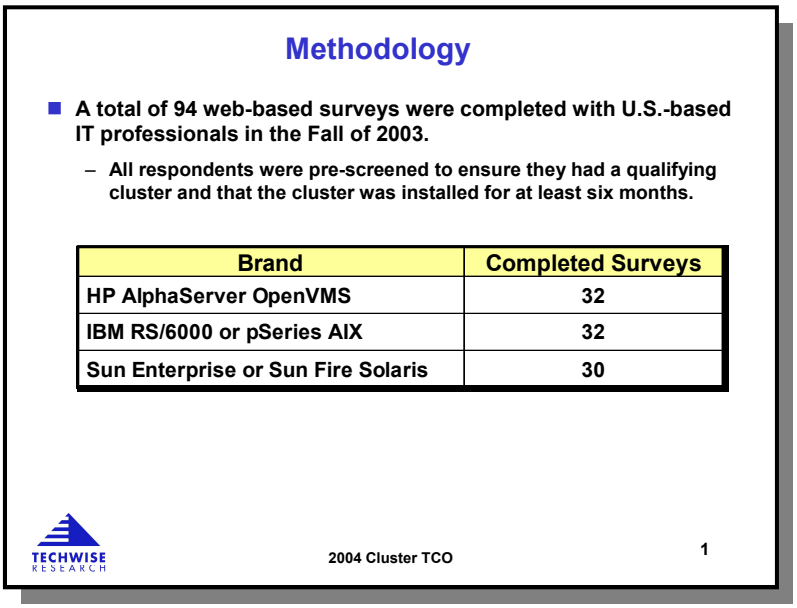
Management costs have two main components: Costs for companies to hire third-parties to manage their cluster on an ongoing basis, and costs for managing the cluster "in-house." In the former situation, respondents provided the actual costs for outsourcing cluster management activities. In the latter case, we collected the number of hours "internal staff" spent on all management activities associated with the cluster. TechWise Research converted internal hours spent into a cost figure by using staff salary data provided by respondents. When calculating internal management costs, the TechWise model factored in the hours spent, if any, managing and maintaining: Servers in the cluster, the cluster's storage array, cluster software and operating system, end-user applications, and network permissions.

Downtime costs were calculated in a multi-step process based on findings from a number of studies TechWise Research has conducted on cluster availability and downtime over the past five years. TechWise defines a cluster "crash" as any situation that causes the cluster's primary application(s) to become unavailable to end-users. Downtime hours are the number of hours per year, if any, when the cluster's primary application(s) were not available for end-users to access. TechWise Research developed a list of ten potential sources of downtime. These are explained in detail in the section entitled: Cluster Reliability Findings.

To calculate each cluster's total cost of ownership (TCO), all of the above survey data were combined with current system and service pricing from IDEAS International. IDEAS International is recognized worldwide as a leading authority on systems technology, specializing in the research of comparative information on computer systems. Their current system and service pricing is updated daily with new product and price announcements. When buying servers, two customers can pay very different prices for two identical systems depending on when they buy them, and on the level of discount they can negotiate from their channel. By using current list prices from IDEAS International, the time and purchasing power bias was eliminated.

Previous TechWise cluster studies have indicated that a three-year time frame is appropriate to evaluate entry-level and mid-range clusters' TCO. Once again, this same time period was applied in the analysis for this current study. Since each company will have different costs associated with downtime, TechWise also calculated the three-year TCO at various hourly rates of downtime costs.

Who Was Surveyed



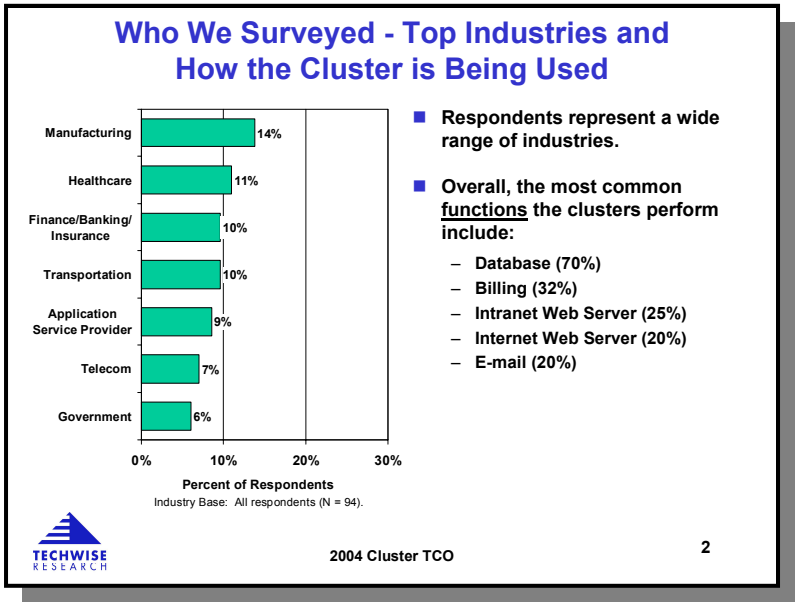
In the Fall of 2003, TechWise Research, Inc. completed a total of 94 web-based interviews with qualifying IT professionals. The survey was designed to collect operational and profiling data about the cluster itself, as well as demographic information about the company using it. The web survey lasted between 25 and 30 minutes. Throughout the survey, respondents were given several opportunities to clarify any answers they provided. One of TechWise Research's senior analysts, who specializes in server clusters, personally reviewed each completed survey and followed-up with respondents by phone if any answers needed clarification. The chart to the left illustrates the total number of surveys completed, broken down by server brand.

To qualify for the study, all respondents were carefully screened to ensure that they personally managed a qualifying entry-level or mid-range cluster. Furthermore, all clusters were required to meet the following four screening criteria:

1. The cluster is one of three target platforms:
   - HP AlphaServer running OpenVMS,
   - IBM RS/6000 or pSeries servers running AIX,
   - Sun Enterprise or Sun Fire servers running Solaris.
2. The cluster uses the manufacturer's clustering software. Therefore, all HP clusters use OpenVMS Cluster, all IBM clusters use HACMP, and all Sun clusters use Sun Cluster. Clusters that were using third-party clustering software, such as Veritas, were excluded from the analysis.
3. The cluster does not contain any enterprise-class servers. An enterprise-class server is defined as one that supports more than 16 processors.
   - Examples of disqualifying systems for HP include the AlphaServer GS 320 and GS 1280 M32 and M64. For IBM, the p680, p690, and RS/6000 S80 did not qualify. For Sun, any cluster that contained an Ultra Enterprise 6000, Enterprise 6500 or 10000, Sun Fire 6800, 12K or 15K server, was disqualified for this study.
4. The cluster has been running in a *production mode* for at least six months. Clusters used in development and testing, or for less than 6 months, were excluded from the study.

Company & Respondent Profile

All participants were randomly recruited from a broad mix of industries, as shown in the chart to the right. The top represented industries in the study include: manufacturing, healthcare, finance/ banking/insurance, & transportation. Most of the study's respondents work for large companies. Twenty-six percent work for companies with 10,000 or more employees worldwide, 15% have between 5,000 - 9,999 employees, and 34% work for companies with 1,000 - 4,999 employees worldwide.

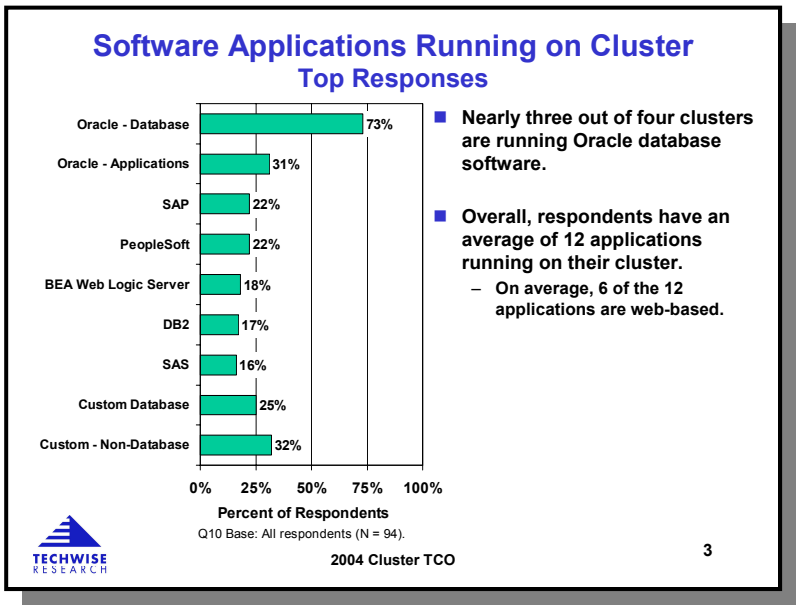


The respondents themselves are experienced users of their particular cluster brand. On average, respondents have worked with their clustering software for four years. Additionally, when asked to rate the overall expertise of their cluster team, 62% rated their team as either "Advanced" or "Expert" while only 4% rated their team as "Beginner." These expertise ratings were statistically the same between the three brands.

Cluster Profile

Just under one half (45%) of the clusters in the study had two nodes, while one-third had between three to five nodes. When looking at cluster CPU utilization, the average "per-processor" CPU utilization was 25%. Under peak traffic conditions, this number increased to an average of 38%.

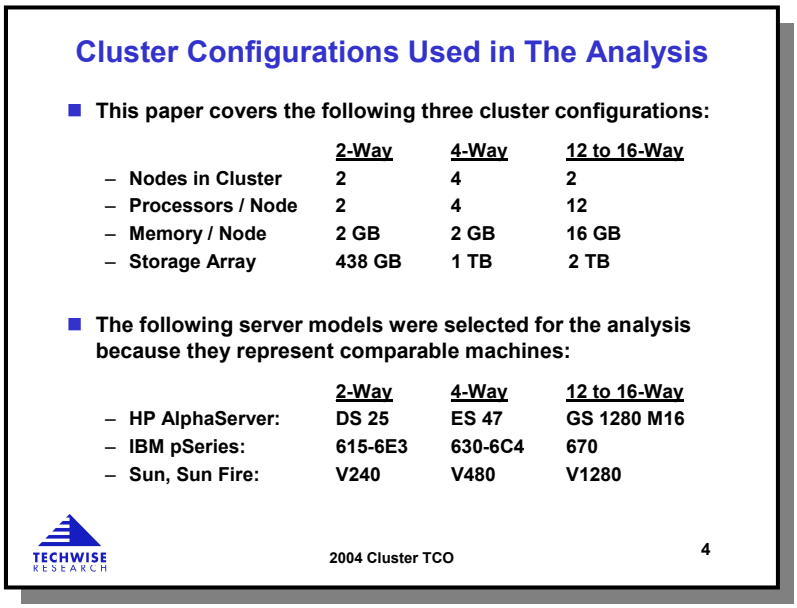
Overall, 87% of the clusters in the study have been in production for at least 12 months with their current server configuration (meaning, the same number of servers and the same server models). The average length of time the clusters have been running with their current version of operating system and clustering software is 15 months. Given the length of time these clusters have been in production, all respondents have had sufficient experience with their clusters to provide accurate measures of their start-up, management, and downtime costs.



The majority of the clusters in this study are running one or more database applications. Nearly three out of four are running an Oracle database. The chart to the left illustrates the top software applications on the clusters. In terms of the number of end-user applications, the clusters are running an average of 6 web-based and 6 non-web-based applications. IBM clusters averaged the most number of web-based applications (8) compared with HP that averaged the fewest (4). Despite these differences in the number of web-based applications, HP OpenVMS/AlphaServer clusters had the highest average number of end-users accessing the cluster's web-based applications. In a typical 24-hour period, HP clusters averaged 3,600 end-users, versus 2,900 for IBM and 1,800 for Sun.

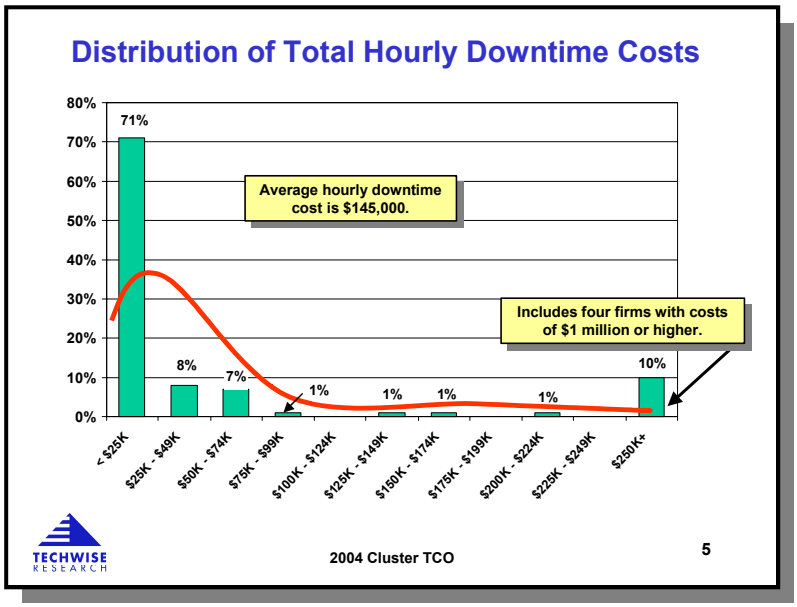
Configurations Tested

TCO calculations were performed on three cluster configurations: 2-way, 4-way and 12 to 16-way clusters. These configurations were selected because they best represent the actual cluster configurations respondents reported in their surveys. The chart to the right describes these configurations in detail and lists the actual server models used in the analysis. The main differences between the three configurations are the server models themselves, the number of nodes/servers in the cluster, the number of CPUs, the amount of memory per node, and the size of the external storage arrays. The specific server models in each class were selected because they represent comparable machines in terms of performance and expandability. Current system and service pricing for these configurations was collected in December 2003 from IDEAS International.



The Value of an Hour

Respondents were asked to rate how important nine different factors would be in a future cluster purchase decision. The two most important factors include (1) the cluster's overall reliability, and (2) how well the cluster software performs when there is a failure. The other factors measured, in order of highest rated importance include: overall performance, security features built into the operating system, applications supported on the platform, scalability - meaning the ability to add more servers to the cluster, ease of management, disaster tolerance abilities, and total cost of ownership. The fact that reliability and cluster software are rated highest demonstrates the overall importance of availability to cluster users, and further re-enforces the primary reason for establishing a cluster - to ensure that primary applications are available to end-users 24x7.



As reported in prior TCO studies conducted by TechWise Research, each company has a unique situation that determines the financial impact of downtime. For some, when primary applications are not available to end-users, the impact is lost sales. For others, it means lost employee productivity or a decline in manufacturing production. Many firms are affected in multiple ways. TechWise Research asked each respondent to quantify the financial impact per hour of downtime. As expected, answers varied widely. On average, however, respondents report that **each hour of downtime costs their firm a total of $145,000** when the costs associated with lost sales, wages, and production are considered. This represents a 30% increase over the costs reported in our 2001 low-end (i.e., entry-level) and mid-range cluster TCO paper.

As shown in the chart above, the distribution of downtime costs varies greatly. Seventy-one percent of respondents said their costs are less than $25,000 per hour. Thirteen percent estimate their hourly costs at $125,000 or more. Four of the companies surveyed indicated that they lose **$1 million or more per hour**! These downtime cost figures demonstrate the important role availability plays in calculating the true total cost of ownership of a cluster.

Possible Reasons for Crashes and Downtime

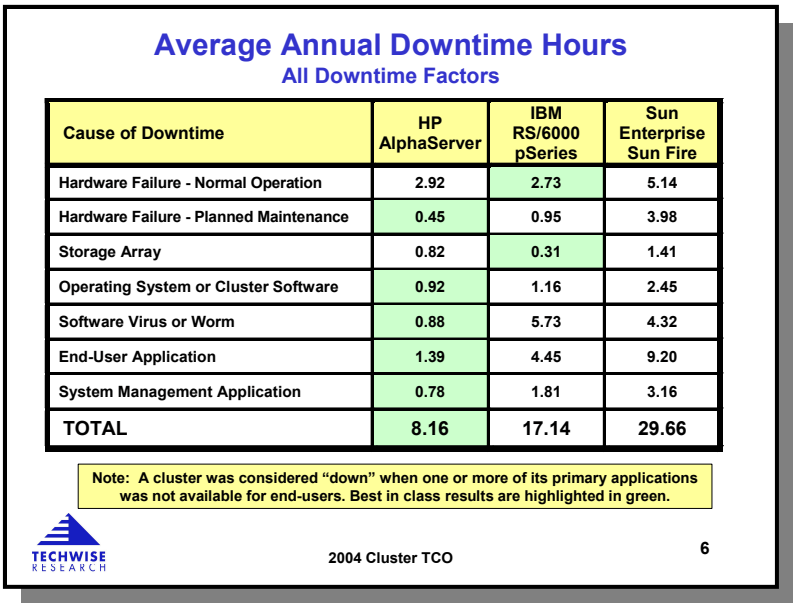
For this paper, a crash was defined as any event that caused one or more of the cluster's primary applications to become unavailable to end-users. Some crashes result in only a few seconds of downtime as the cluster software "fails over" to another node. Other crashes can cause applications to be down for minutes or even hours. In this study, TechWise Research collected downtime information on all crashes, no matter how short or long in duration.

There are a variety of potential sources for crashes and downtime. In this TCO analysis, TechWise Research included downtime caused from the following seven different sources:

1. **Server hardware failure during normal cluster operation:** These are crashes caused when one or more servers failed, when no maintenance was being performed on any of the servers.
2. **Server hardware failure during planned server maintenance:** This downtime category is a new addition to this year's analysis based on TechWise Research's continuing research on server clusters. One of the benefits of a cluster is that it allows IT personnel to perform maintenance on a node without taking the applications offline. Therefore, tracking when such failures occur is important. This category of crash is when one or more servers fail while planned maintenance is performed on one of the other servers in the cluster. In these cases, the cluster crashes despite its automated failover configuration.
3. **Hardware failure in a storage array:** These include any crashes caused by the cluster's storage array. In the survey, a few respondents reported using third-party arrays, usually EMC, with their cluster. Any crashes caused by third-party arrays were excluded from the analysis. This way, the analysis is based exclusively on each manufacturer's hardware and software solutions.
4. **Operating system or cluster software problems:** These are any crashes caused by the operating system or clustering software.
5. **Software virus or worm:** In the past few years, security concerns have grown considerably. There have been several high profile viruses and worms that have affected systems worldwide. Hackers can and do cause clusters to crash. Given the increased focus on this issue, TechWise Research decided to include this as separate category in this year's study.
6. **End-user application problem:** This includes any crash caused by a problem with any of the end-user application(s) themselves.
7. **System management application problem:** This includes any crash caused by a problem with any of the system management application(s) running on this cluster.

In addition to the above sources of downtime, TechWise Research also measured downtime resulting from natural disasters, human error and all "other" sources. Data for these three types of crashes were also collected to ensure respondents did not inadvertently include these crashes in one of the seven categories above. However, these three crash types were excluded from the analysis for several reasons. In the case of natural disasters, this study was not designed to focus on disaster tolerant systems (i.e., clusters designed to stay functional in the event of a disaster). Therefore, it would be inaccurate to include crashes caused by natural disasters (which are frequently localized events) in the analysis. Crashes caused by human error may or may not be a reflection of the usability of the operating system and clustering software. To probe into this issue at the required depth, would require a separate research inquiry, and is beyond the scope of this current study. Therefore, human error crashes were also excluded. With regards to "other" crash sources, none of the respondents reported a crash due to any "other" reason not listed, indicating the robustness of the downtime findings provided in this report.

Cluster Reliability Findings



Average Annual Downtime Hours
All Downtime Factors

Cause of Downtime	HP AlphaServer	IBM RS/6000 pSeries	Sun Enterprise Sun Fire
Hardware Failure - Normal Operation	2.92	2.73	5.14
Hardware Failure - Planned Maintenance	0.45	0.95	3.98
Storage Array	0.82	0.31	1.41
Operating System or Cluster Software	0.92	1.16	2.45
Software Virus or Worm	0.88	5.73	4.32
End-User Application	1.39	4.45	9.20
System Management Application	0.78	1.81	3.16
TOTAL	8.16	17.14	29.66

Note: A cluster was considered "down" when one or more of its primary applications was not available for end-users. Best in class results are highlighted in green.

TECHWISE RESEARCH

2004 Cluster TCO

6

Each respondent reported the total number of hours per year, if any, their cluster's primary applications were offline due to each of the seven types of problems. The table to the left summarizes these findings by platform. "Best in class" findings for each category are highlighted in green.

HP OpenVMS/AlphaServer clusters averaged the fewest total number of annual downtime hours of all three brands tested. HP's average downtime of 8.16 hours is less than half of IBM's AIX/pSeries clusters (17.14 hours), and one-third of Sun's Solaris/Sun Fire clusters (29.66 hours). In five of the seven categories, HP won best in class. For the remaining two categories, IBM AIX clusters won best in class. These downtime findings for AlphaServer OpenVMS are similar to results from prior TCO studies conducted by TechWise Research. Below is a more detailed explanation of the individual category findings and their implications.

The first two downtime categories listed in the table above, are crashes caused by server hardware failures. For server hardware failures occurring during normal cluster operation, IBM averaged the fewest number of downtime hours followed closely by HP. For server hardware failures occurring during planned maintenance, the results switch whereby HP averaged the fewest downtime hours, followed closely by IBM. However, in both cases of server hardware failures, Sun averaged the highest number of downtime hours. Combining both server failure categories provides a measurement of the reliability of the servers themselves. HP and IBM have virtually the same combined downtime hours (3.37 and 3.68, respectively) while Sun averaged 9.12 hours. The Availability Advantage™ HP and IBM have over Sun in this area could result in significant cost savings over a three-year period. For each $10,000 of hourly downtime costs, Sun clusters would cost on average $168,000 more than comparable HP or IBM clusters over three years, all other factors being equal.

IBM clusters achieved best in class results in terms of having the fewest hours of downtime due to storage arrays. Compared to four of the five remaining categories of downtime, crashes caused by storage arrays are a relatively minor issue in terms of cluster availability (as shown by the low downtime results for this factor). Additionally, the difference between first place IBM and last place Sun is only slightly more than one hour on this issue. As a reminder, these findings are based only on those clusters that were using the manufacturer's brand of storage array.

HP clusters had the fewest average annual downtime hours due to the operating system and clustering software (0.92), followed closely by IBM at 1.16, and then by Sun at 2.45 hours. Similar to storage arrays, there is little difference between the first place and last place brands. However, there was a much larger difference between the brands in terms of crashes caused by software

viruses or worms. In this area, HP clusters have a significant advantage over both IBM and Sun. HP cluster owners reported an average of 0.88 hours of downtime per year due to software viruses or worms. Sun was second best in this category averaging 4.32 hours, while IBM averaged 5.73 hours. For each $10,000 of hourly downtime costs, HP's Availability Advantage™ over Sun and IBM, in terms of preventing crashes from software viruses and worms, translates into $103,200 and $145,500, respectively, over a three-year period.

HP is also best in class in terms of the fewest downtime hours caused by system management applications. It has less than half the downtime of IBM and Sun in this area. For end-user application crashes, HP clusters have an even greater advantage. Sun cluster owners report more than six times the number of downtime hours than HP cluster owners. For each $10,000 of hourly downtime costs, HP's Availability Advantage™ over IBM and Sun, specifically in terms of crashes caused by end-user applications, translates into $91,800 and $234,300, respectively, over a three-year period.

In summary, when all seven types of crashes are considered, IBM AIX/pSeries clusters averaged 8.98 more downtime hours per year than HP OpenVMS/AlphaServer clusters. Furthermore, Sun Solaris/Sun Fire clusters averaged 21.50 more downtime hours per year than HP OpenVMS/ AlphaServer clusters. **For each $10,000 of hourly downtime costs, HP's total Availability Advantage™ over IBM and Sun translates into $269,400 and $645,000, respectively, over a three-year period**.

Breakdown of the TCO Components

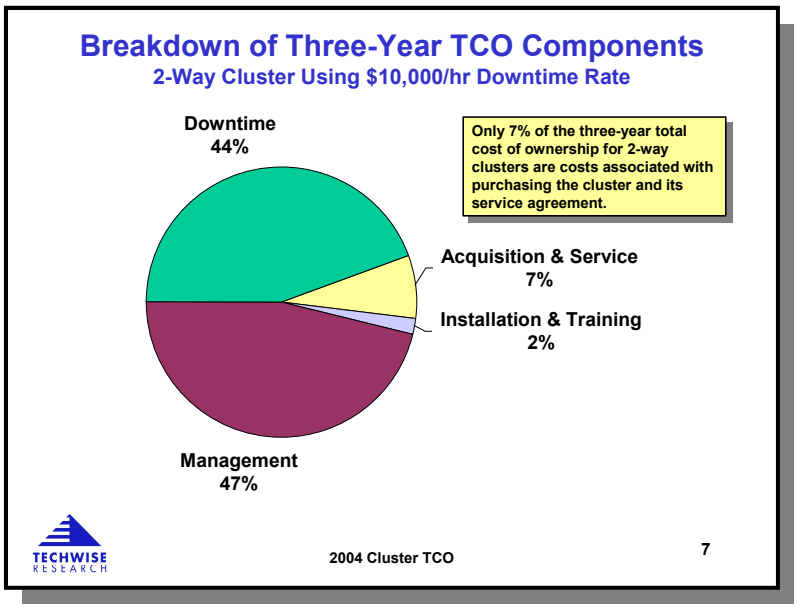
There are four main components in TechWise Research's TCO model. These include:

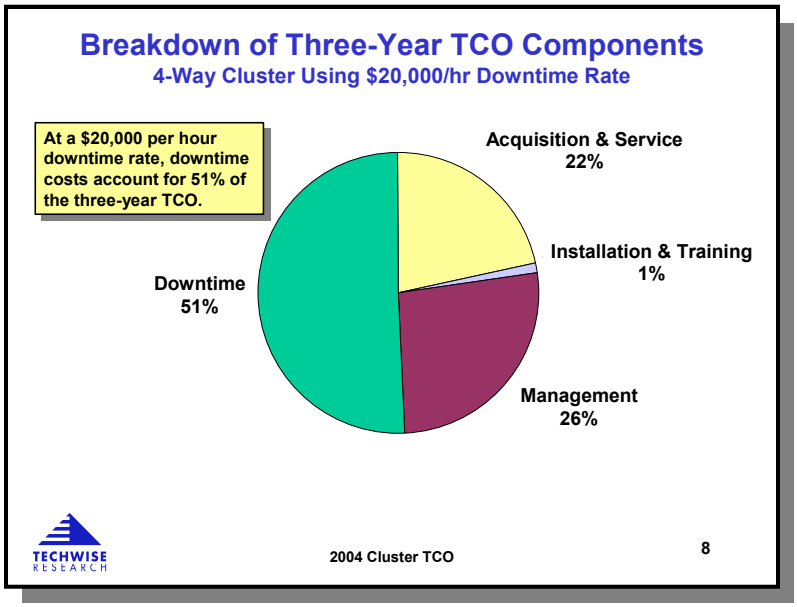
- **Acquisition and Service** – This is the list price for the servers, storage array, operating system and cluster software, and three-year service contract.
- **Start-Up Costs** – These are the costs to install and configure the cluster, as well as any time and money spent to train staff on the cluster.
- **Management Costs** – These are the ongoing costs associated with managing the cluster over a three-year period.
- **Downtime Costs** – This is the number of hours, and resulting costs for cluster downtime, over a three-year period.

The following sections illustrate the importance of these four factors in determining a cluster's TCO for the three configurations tested (i.e., 2, 4 and 12 to 16-way clusters).

2-Way Clusters

The chart to the right shows the breakdown of TCO costs for 2-way clusters using a conservative hourly downtime rate of $10,000. This chart is based on the average costs across all three brands. As the pie chart indicates, the list price for the servers, array, OS/ cluster software and service agreement represents only 7% of the cluster's three-year TCO, whereas management costs and downtime costs represent 47% and 44%, respectively, of the total TCO. Therefore, even in situations where downtime costs are only a few thousand dollars per hour, the impact of that downtime can be substantial. **Availability is very important to consider in any cluster purchase decision.**



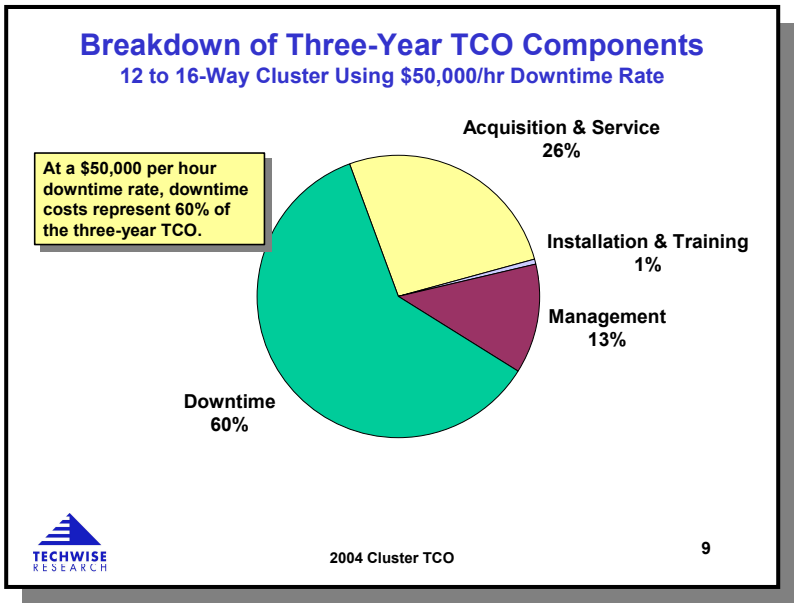


4-Way Clusters

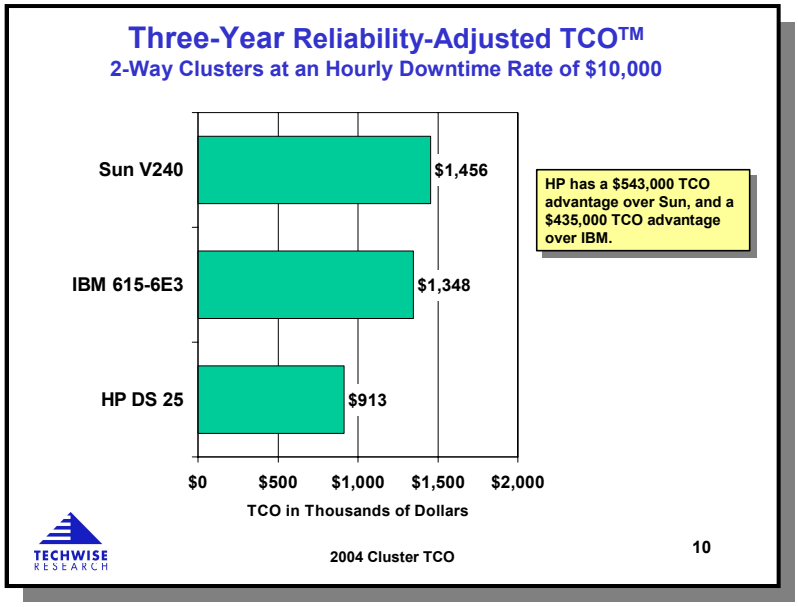
The chart to the left shows the breakdown of TCO costs for 4-way clusters applying a $20,000 per hour downtime rate. Compared to 2-way clusters, a higher hourly downtime rate was used to account for the fact that higher end systems are often used in situations that are more mission-critical. Despite using a higher $20,000 per hour downtime rate, the cluster's purchase price for this 4-way configuration represents 22% of the three-year TCO. Two reasons account for this. First, the server models in this configuration are more expensive than those in the 2-way cluster. Second, this cluster configuration has twice as many servers than the 2-way (4 versus. 2). **However, downtime costs still account for roughly half of the TCO for 4-way clusters.** From a financial standpoint, in cases where downtime costs are $20,000 per hour or more, cluster availability should be the primary factor considered when comparing different cluster platforms.

12 to 16-Way Clusters

This cluster configuration contains higher end, mid-range systems. The servers in these clusters have the ability to support between 12 and 16 CPUs. The chart to the right shows the breakdown of TCO costs for 12 to 16-way clusters using an hourly downtime rate of $50,000. As the pie chart to the right indicates, the list price for the servers, array, OS/cluster software and service agreement represents one-fourth of the cluster's three-year TCO. Downtime costs continue to represent the largest portion of TCO, 60% in this scenario. **For 12 and 16-way clusters, downtime costs represent a greater share of the TCO than all other costs combined.**

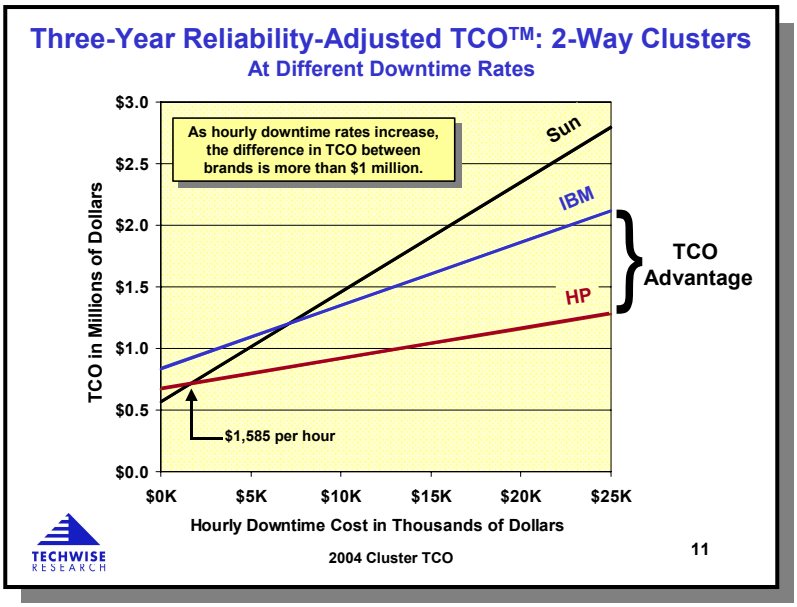


Detailed Findings: Reliability-Adjusted TCO™ for 2-Way Clusters



The first cluster configuration studied was a 2-way system. The chart to the left summarizes the Three-Year Reliability-Adjusted TCO™ for the three 2-way platforms studied. The bars represent the total cost of owning each cluster including acquisition, service, installation, training, management and downtime over a three-year period. Downtime costs were calculated by applying the $10,000 per hour downtime rate used in Figure 7). **HP OpenVMS/ AlphaServer clusters have the lowest TCO. Over a three-year period, HP's TCO advantage is $543,000 over Sun Solaris/Sun Fire clusters ($1,456,000 - $913,000), and $435,000 over IBM AIX/pSeries clusters ($1,348,000 - $913,000).**

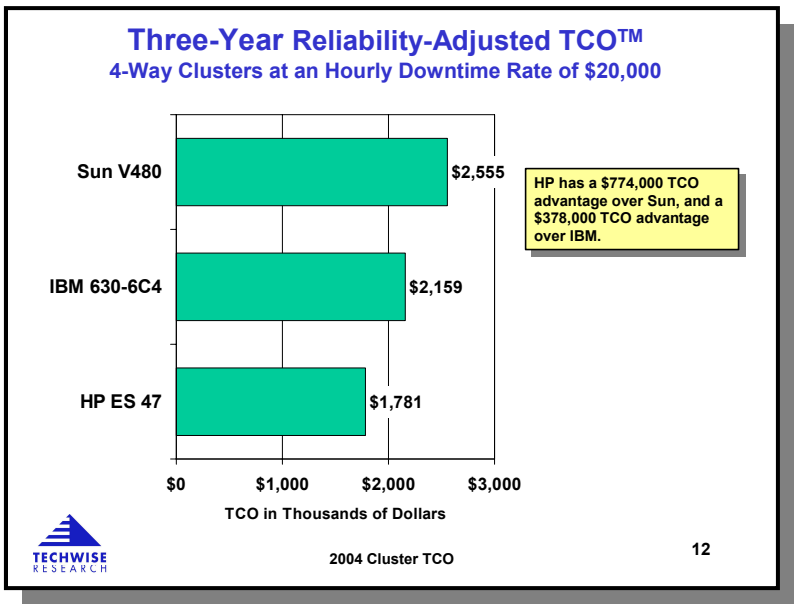
The application of a $10,000 per hour downtime rate may be too low for some firms, or too high for others. For this reason, TechWise Research provides a more detailed graph below. This chart summarizes the three-year Reliability-Adjusted TCO™ for the three 2-way clusters studied at various hourly downtime rates. Depending on the cost associated with downtime, either Sun or HP 2-way clusters have the lowest TCO. **Sun Solaris/Sun Fire clusters have the lowest TCO in situations where the cost per hour of downtime is less than $1,585 per hour, while HP OpenVMS/AlphaServer clusters have the lowest TCO at all hourly downtime rates above $1,585.**



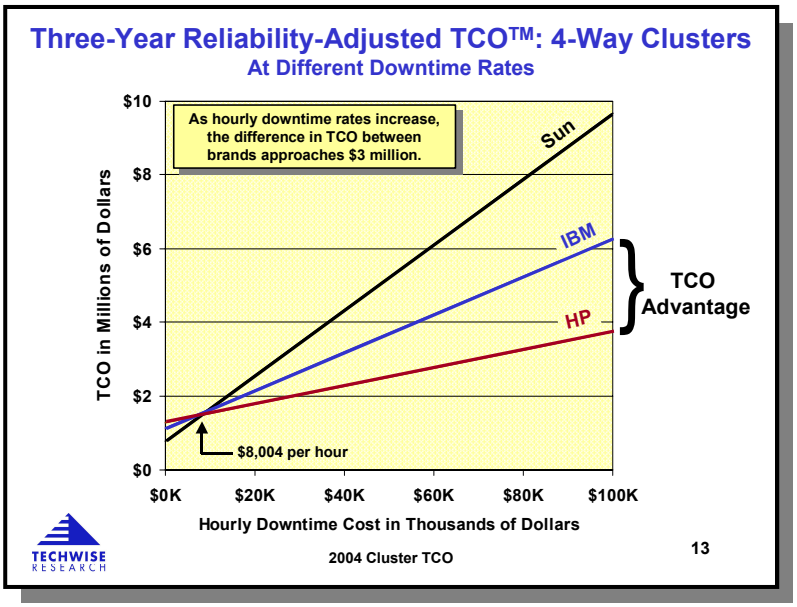
Sun's TCO advantage over HP at low hourly downtime rates can be attributed to the fact that the list price for the Sun cluster is $122,000 less than the list price for the HP cluster. However, as the cost per hour of downtime increases, system pricing represents a smaller portion of the TCO. When downtime costs are $1,585 per hour, HP and Sun clusters have the same TCO. At this "cross-over" point, Sun's lower system price is exactly offset by HP's lower management and downtime costs. Once the cost per hour of downtime exceeds $1,585, HP has the lowest TCO because its Availability Advantage™ is greater than the $122,000 difference in system price. At an hourly downtime rate of $25,000, the HP cluster's three-year TCO advantage is $838,000 and $1.5 million, respectively, over IBM and Sun clusters.

Detailed Findings: Reliability-Adjusted TCO™ for 4-way Clusters

The second cluster configuration studied was a 4-way cluster. The chart to the right summarizes the Three-Year Reliability-Adjusted TCO™ for the three 4-way platforms studied. An hourly downtime rate of $20,000 was used for these clusters (the same rate that was used in Figure 8). **HP OpenVMS/AlphaServer clusters have the lowest TCO. HP's TCO is $774,000 less than Sun Solaris/Sun Fire clusters, and $378,000 less than IBM AIX/ pSeries clusters, over a three-year period.** The chart below summarizes the three-year Reliability-Adjusted TCO™ for the three 4-way clusters studied, at various hourly downtime rates. Comparatively, 4-way clusters have higher acquisition and service costs than 2-way clusters since there are four servers instead of two, and each server has more processors, memory, and storage. However, despite the higher acquisition costs, these differences are relatively minor when comparing them to the impact of downtime costs, as shown below.

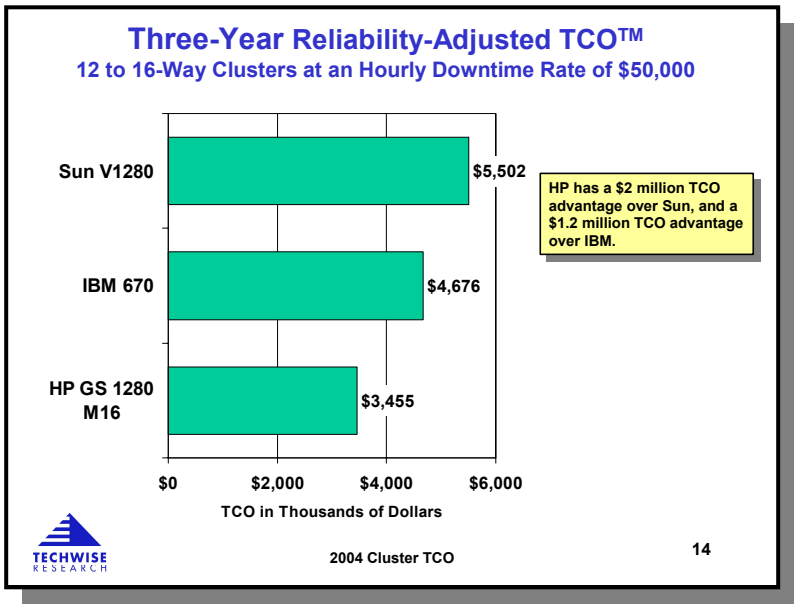


Depending on the cost associated with downtime, either Sun or HP 4-way clusters have the lowest TCO. **Sun Solaris/Sun Fire clusters have the lowest TCO in situations where the cost per hour of downtime is <u>less than</u> $8,004 per hour, while HP OpenVMS/AlphaServer clusters have the lowest TCO at downtime costs <u>above</u> $8,004.** As with 2-way clusters, Sun's advantage at lower downtime rates is due to the fact that the list price of the Sun cluster is several hundred thousand less than the HP cluster. Similar to 2-way clusters, when hourly downtime rates increase above $40,000, HP's TCO advantage over IBM and Sun ends up being more than the list price of the HP cluster with a three-year service agreement. At an hourly downtime rate of $100,000, the HP cluster's three-year TCO advantage is $2.5 million and $5.9 million, respectively, over IBM and Sun.

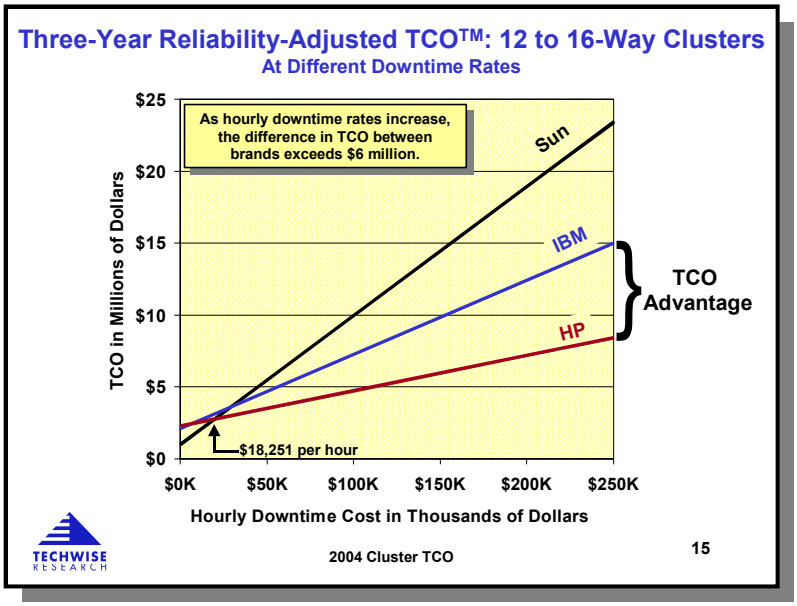


Detailed Findings: Reliability-Adjusted TCO™ for 12 to 16-Way Clusters

The chart to the right summarizes the Three-Year Reliability-Adjusted TCO™ for the 12 to 16-way clusters studied. These findings were calculated using a $50,000 hourly downtime rate (the same rate used in Figure 9). As with the other two cluster configurations, **HP OpenVMS/ AlphaServer clusters have the lowest TCO. HP's TCO is $2.0 million less than Sun Solaris/Sun Fire clusters, and $1.2 million less than IBM AIX/pSeries clusters, over a three-year time frame**.



The chart below summarizes the three-year Reliability-Adjusted TCO™ for the three 12-way clusters studied, at various hourly downtime rates. These findings are similar to the previous configurations in that Sun clusters have the lowest TCO when the costs associated with downtime are below a certain "cross-over" point, and HP has the lowest TCO at all other downtime rates. **Sun Solaris/Sun Fire clusters have the lowest TCO when the cost per hour of downtime is less than $18,251.** Sun's advantage is largely because the V1280's acquisition price is substantially less than the acquisition price for the IBM or HP clusters. **At all other downtime rates, HP OpenVMS/AlphaServer clusters have the lowest TCO**. At an hourly downtime rate of $250,000, the HP cluster's three-year TCO advantage is $6.6 million and $14.9 million, respectively, over IBM and Sun clusters.



Impact of Application Crashes on the Cross-Over Point

As previously noted, there was a large difference between brands in downtime hours due to crashes caused by software applications. Respondents reported some differences in the number and types of applications running on the three different brands of clusters. Because of these differences, TechWise also performed the analysis excluding crashes caused by end-user and system management applications. Ignoring crashes caused by end-user and system management applications, the "cross-over" point where HP clusters have a lower TCO than Sun clusters changes from $1,585 to $3,014 per hour for 2-way clusters, from $8,004 to $15,215 per hour for 4-way clusters, and from $18,251 to $34,694 per hour for 12 to 16-way clusters.

Conclusion

This study focused on the three-year total cost of ownership for three RISC-based server clusters. Four different costs were factored into this TCO analysis, these include: the list price of the cluster and related service agreement, start-up costs (i.e., training and installation), management, and downtime costs. TCO calculations were based on a three-year time period.

The cluster configurations used in the analysis were selected since they most closely represent the actual cluster configurations managed by study respondents. Three different cluster configurations were analyzed (2-way, 4-way and 12 to 16-way). The 4-way and 12 to 16-way configurations represent higher performance "systems" containing more memory, processors, and storage than the 2-way clusters. Comparable server models from each manufacturer were selected to use for the analysis (see section entitled Configurations Tested).

Availability represents between 44% and 60% of the TCO of the configurations modeled, and is the key driver for server cluster TCO. Respondents reported that their average cost per hour of downtime was $145,000. Nearly half indicated that their cost per hour of downtime is $10,000 or greater. Given the impact of downtime costs, the Reliability-Adjusted TCO™ was calculated for each of the three cluster configurations at several different hourly downtime rates. This way, readers have the ability to apply the results to their company's unique situation by selecting the appropriate hourly downtime rate. Downtime hours were analyzed for seven different sources of downtime. When all seven sources of downtime are considered, **HP OpenVMS/AlphaServer clusters averaged the fewest total number of annual downtime hours (8.16), followed by IBM AIX/pSeries clusters (17.14), then by Sun Solaris/Sun Fire clusters (29.66)**.

For all three configurations, Sun Solaris/Sun Fire clusters offer the lowest TCO when there are no costs associated with downtime. This is due to the lower list price of the Sun clusters compared to HP and IBM clusters. In situations where the costs associated with downtime increase beyond a certain level - termed the "cross-over" point - HP clusters offer the lowest TCO. This is due to the Availability Advantage™ HP clusters have over Sun and IBM. The 2-way HP cluster is best in class when the total cost associated with one hour of downtime is greater than $1,585. Similarly, the 4-way and 12 to 16-way HP clusters are best in class when downtime costs are greater than $8,004 and $18,251, respectively. In all three configurations, IBM clusters' TCO falls between HP and Sun - with one exception. IBM has the highest TCO for 2-way clusters when downtime costs are less than $7,000 per hour.

Several factors outside the scope of this study are also important in the purchase decision for entry-level and mid-range clusters. These include cluster performance, software features, applications supported on the platform, and quality of service and support. However, when comparing clusters from a financial perspective, it is critical that IT managers factor in the cost of downtime when calculating the TCO.

TechWise Research is an independent primary market research firm that specializes in the computer industry. If you have any questions regarding this research, please contact us at:

TCO2004em@TechWise-Research.com.