RTO Software TScale 2.0 and PMC Frequently Asked Questions (FAQ)

RTO Software TScale 2.0 and PMC Frequently Asked Questions (FAQ)

1. What is TScale?

TScale™ is a performance and capacity improvement solution for Citrix MetaFrame® and Microsoft® Terminal Servers. TScale server software continuously and automatically improves the performance of Win32 applications on terminal servers, typically increasing concurrent users by up to 30% or more.

2. What is new in TScale 2.0?

TScale 2.0 contains the following enhancements over TScale 1.5.1:

a. The optimization that has been in TScale 1.x since its inception is called the VM Capacity optimization in TScale 2.0. This optimization reduces the amount of page file space used by each copy of each DLL loaded by each user of each application. By reducing the amount of page file that each component uses, the amount of page file I/O and the number of page faults are reduced. Reducing the amount of page file activity is what has allowed TScale 1.x to optimize applications running on the server typically increasing concurrent users by up to 30% or more. This reduction in VM usage manifests itself in each copy of each application using less total VM size as reflected in the Task Manager report of this metric for each instance of each application.

b. A new optimization, VM performance has been added to TScale 2.0. This new optimization reduces the number of steps that the CPU needs to take when a process calls a DLL or a DLL calls another DLL. By eliminating these steps in the chain of components that get called as a user executes an application, the performance of the application is improved. This improvement in performance will manifest itself in improvements in end user response time and load time.

c. A new optimization, MT performance has been added to TScale 2.0. This new optimization improves the execution efficiency of multi-threaded applications. This improvement in execution efficiency manifests itself in improved end user response time and in fewer context switches caused by the optimized applications.

d. With the 2.0 release of TScale, RTO Software has also shipped a new product, the Performance Management Console. The PMC is a multi-server management console that allows you to manage the TScale agents on all of your servers from one central location. The PMC is sold and licensed separately from TScale.

3. What are the platform requirements for TScale 2.0 and the PMC?

a. They both will work on Windows NT 4.0 SP6 or above, including any version of Windows 2000, Windows XP, and Windows 2003. (Windows 3.1, Windows 95, Windows 98 and Windows ME are not supported.)

b. They both require MMC 1.1a or above.

c. The PMC should be installed on a server or a workstation in the same domain as the servers that the PMC is going to manage.

d. The person using the PMC should log onto that workstation with an account that has full admin privileges to the terminal servers that are going to run TScale.

e. TScale works with Microsoft Terminal Services and any version of Citrix MetaFrame (1.8, XP, etc.), which may be present. However, TScale does not

require the presence of any terminal services products other than the native Microsoft Terminal Services layer.

4. **Is Windows 2003 a supported platform for TScale 2.0?**
   a. RTO fully supports Windows 2003 Server with TScale 2.0 and has had to make no modifications to TScale in order for Windows 2003 to be included as a supported platform.
   b. If you are developing applications with the .Net Framework, you will be building applications in the way that causes them to be good candidates for TScale optimization. Therefore as you and your software vendors migrate to Microsoft's newer applications development methodologies, TScale will become more valuable to you.

5. **In Windows 2003, Microsoft has removed the registry size limit. Does this impact TScale?**
   a. The registry size in Windows NT and Windows 2000 was limited to respectively 154MB and 296MB. In Windows 2003 the registry is memory mapped. Therefore growth in the registry in Windows 2003 will not constitute a hard limit to scalability as it has in the past.
   b. If your ability to add concurrent users to your servers was constrained by the registry size limit in Windows NT or Windows 2000, then you will find that TScale will be able to give you more concurrent users per server in Windows 2003 without causing you to bump into the registry size limit.

6. **Is there a way to tell before I install TScale if TScale will help my terminal server?**

   Yes there is. Send an email to sales@rtosoft.com and request the InstanceMon utility. This is a free utility that captures every running program and the components that each program is using to a text file named Instances.txt. Email us the Instances.txt file for your terminal servers and we will be able to tell you if your server is a good candidate for TScale's optimizations. Of course, the simplest and surest way is to evaluate TScale.

7. **How is TScale licensed?**

   TScale is licensed per server with different fees for different numbers of CPU's in the server. There are 1 server, 2 server, 4 server and 8 server versions of TScale. There is no license fee for concurrent users or the number of applications running on the server.

8. **Does RTO offer maintenance on TScale?**

   Yes it does. Maintenance covers support and all updates and upgrades to the product. Contact your VAR o r RTO Sales for details on maintenance pricing and terms.

9. **After I install TScale, if I immediately run an optimization session, is the server then fully optimized?**

   No, TScale must be given a chance to see the components (DLL, OCX, ActiveX, …) of the published application get loaded into memory. Once it sees a component, the next optimization session that is run optimizes that component. In general it takes three to

four optimization sessions to "highly optimize" the server, and approx. 10 sessions in order for a server to be fully optimized. See and the associated graph for how to tell that when optimizations are complete on a server.

10. **How does TScale allow a terminal server to support more concurrent users?**

    TScale improves capacity (scalability) in three different ways:
    a. VM Capacity optimizations reduce the amount of page file activity that each instance of an application generates. Removing the hard disk activity (the slowest part of the server) associated with page file writes and page faults allows more concurrent instances of an application to run with good performance.
    b. VM Performance optimizations reduce the number of steps that the CPU must take to load applications and the components that comprise applications. This frees up CPU cycles to support more concurrent instances of user applications.
    c. MT Performance optimizations reduce the number of context switches that applications generate as they execute, again freeing up more CPU cycles to support more concurrent instances of user applications.

11. **How does TScale improve the performance of terminal servers?**

    TScale improves the performance of terminal servers in three different ways:
    a. VM Capacity optimizations reduce page file activity, which when the server is busy (under heavy load) can become a constraint to good performance. Therefore, applications that have been subjected to VM capacity optimizations will perform better when the server is under heavy load.
    b. VM Performance optimizations improve the load time of applications and the components that make up the applications. Users will notice faster initial load times and faster execution of applications as multiple components are loaded.
    c. MT Performance optimizations reduce the number of context switches that an application generates. Since context switches are expensive consumers of CPU, reducing them makes more CPU time available for productive processing of user applications and work.

12. **If TScale saves me virtual memory, can't I accomplish the same thing that TScale gives me by adding RAM to my servers?**

    Adding RAM to servers does not accomplish the same thing as TScale. Adding RAM does nothing to stop applications from being swapped out to the page file as multiple instances of those applications are loaded. TScale specifically eliminates unnecessary page file activity, which because it involves the hard disk is the slowest part of the server. Removing hard disk activity from the execution of applications is what allows servers running applications optimized by TScale to perform better and to support more concurrent users.

13. **Does TScale actually make my applications use less memory?**

    Yes it does. TScale does this by reducing the amount of virtual memory (RAM plus page file) that each copy of the application uses as its user loads it.

14. **Does TScale actually cause my application and their components to get loaded from hard disk less frequently?**

Yes it does. TScale optimizes how applications running on Terminal Servers use the Page File. This causes Page File writes to decrease and the number of Page Faults generated by the application to be reduced. This reduced level of Page File activity allows the application to perform more efficiently, which in turn allows the Terminal Server to support more concurrent users of the application.

15. **My Terminal Servers are not memory constrained; they are constrained by CPU utilization. Will TScale help?**

Yes, TScale will reduce the time the CPU spends writing and reading the page file and thereby free up cycles that the CPU can use to do useful work for your applications users.

16. **How much memory and CPU resource will TScale utilize in order to do its job?**

TScale operates in two phases. The analysis phase (the TScale Analysis Service) runs once every 10 seconds and simply logs which components and applications are being loaded in memory. This phase of TScale consumes a negligible amount of memory and CPU resource. The Analysis Service is also self tuning and will turn itself off whenever the CPU in the server gets busy so as to not interfere with the performance of the server during peak load times (it turns itself back on when the CPU is back to a less than fully loaded state). In other words, you will not be able to notice that TScale Analysis Service is running on your production servers. The optimization phase occurs either when you manually invoke it, or when you have scheduled it to run. The optimization phase will consume a noticeable amount of CPU resource while it is running. For this reason you are encouraged to schedule optimization to occur when production demands upon your server are light to non-existent. The Optimization Task typically takes less than two minutes to run, so you can be assured that it will not interfere with any overnight processing that you have otherwise scheduled to occur.

17. **How do I schedule optimizations to occur?**

You can use the TScale Console to do this by right-clicking on the TScale Optimizations node and then selecting Schedule. Or you can invoke the optimization with your favorite scheduling program by simply running "TScaleBase –optimize" as a command line entry. Schedules are also part of policies in the PMC and can be set for groups of servers through these policies.

18. **Is there is list available the applications that TScale has successfully optimized?**

Yes, while not a comprehensive list, the Applications Gallery on the RTO Software web site contains a list of some of the applications that RTO customers have optimized with TScale. You can access the applications gallery at the following location on the RTO Software web site: http://www.rtosoft.com/pages/tscaleappgallery.html. Note that this is just a partial list – TScale has successfully optimized applications on thousands of production servers worldwide.

Back To Table Of Contents

19. Are there any applications that TScale cannot or will not optimize?

Yes, TScale cannot optimize the following types of applications:
  a. DOS applications. TScale operates on Win32 applications only.
  b. Win16 applications. TScale operates on Win32 applications only.
  c. Microsoft Word, Excel, and PowerPoint. TScale does not provide significant optimization for these applications since they are already "optimized" relative to each other. However, if you load third party components like printer drivers or add-ins (that Microsoft does not ship on their CD's), then TScale will optimize these components, helping the performance of these applications as you use these third party components in the context of the Microsoft applications. For example, Essbase (a multi-dimensional decision support database) implements its client, as an Excel snap-in. TScale will optimize the Essbase components in the context of the Excel process within which the Essbase components are loaded and executed.

20. Does TScale help some applications more than others?

Yes, the more an application is constructed of compiled 32 bit components, like DLL's, ActiveX's and OCX's, the more TScale will help. TScale will help the types of applications listed below, but not as much as applications constructed of compiled components:
  a. Java applications. Since Java applications are interpreted at run time, they are not constructed of compile components. However, they can still benefit from TScale.
  b. Monolithic applications that are not constructed of DLL's, ActiveX's, OCX's, or other compiled components. TScale can help these applications, but again not to the same degree as applications built from compiled components.

21. Can I use CPU utilization as a metric to observe the increased performance delivered by TScale?

No, you cannot use CPU utilization to observe or to quantify the impact of TScale's optimizations. The reason for this is that CPU time is only counted every 10 milliseconds, and that many of the operations that TScale eliminates fall in between two 10-millisecond heartbeats, and are therefore not ever counted in the first place. Therefore, since they are not counted in the first place, you cannot tell when they are eliminated. To prove to yourself that this is the case, launch Task Manager on your desktop computer or server. Add CPU Time (not CPU Utilization) and Page Faults as columns. Sort on CPU Time so that the processes showing zero CPU Time will be at the top of the list. Notice that there will be several processes that have generated Page Faults but which have used no CPU time. This will be the case because those applications generated those page faults in between the 10-millisecond heartbeats of the Windows OS, which means that the CPU time used to generate those Page Faults was not counted. This is precisely why CPU Time and its component metric, CPU Utilization, are poor metrics to measure the impact of TScale and its optimizations upon your servers.

22. If I cannot use CPU utilization as a metric to measure the changes that TScale delivers what metrics can I use?
  a. VM capacity optimizations manifest themselves in the VM size of the optimized processes (you can capture this conveniently across an optimized and un-optimized server using the GetProcessInfo utility that is in the Program Files\RTO

Software\TScale directory. You can also observe the impact of VM capacity optimizations by watching pages/second across both an optimized and un-optimized server.

b. VM Performance optimizations manifest themselves in improved load times and improved response times in end user applications. So, you have to be able to measure applications responsiveness to the end user to measure the impact of this optimization.

c. MT Performance manifests itself in a reduction in the number of context switches generated by an application. So, you can measure the impact of this optimization by observing context switches/second across both an optimized and un-optimized server.

23. How do I translate any of these metrics into the number of incremental users that can get per server?

There is no straightforward translation. The only way to determine the number of incremental users that you can get per server is to put TScale on a production server, let it fully optimize that server and then do a scalability test.

24. Once optimizations are complete, how do I do a scalability test?

The best way is to cause an optimized server to slowly take more concurrent users. This can be done in a variety of different ways by modifying the parameters in the load balancing software used in terminal server farms. This subject is covered in detail in the "How To Evaluate TScale" document that ships with TScale.

25. How do I know when a server is fully optimized?

In order to optimize an application, that application and the components that comprise that application must be loaded into memory and used by a user. So, TScale's optimization of an application are complete when at least one user has loaded all of the pieces of that application into memory, and then an optimization session has been run. Any user who loads an application AFTER the optimization session has been run will feel the impacts of the optimization session. You can run an optimization session while users are using an application, but any user currently using the application will not get the benefits of the optimization until they unload and then reload the application.

26. Why are there are applications listed under the Applications and Sessions trees that are not showing up in the bar charts?

Some applications only need VM Performance optimizations. VM Performance optimizations do not save virtual memory, which is what the bar charts display. In order for the bar charts not to get cluttered with a large number of applications for which the savings are zero, any application for which the savings are zero is excluded from being displayed in the bar charts.

27. How can I tell when TScale's optimizations are complete?

If you highlight the Sessions node in the TScale Console, or in the tree for a server in the PMC, you will get a line graph. When that graph has leveled off (is no longer rising vertically), then optimizations are largely complete. The Sessions graph below is a good example of a line which has leveled off which means that optimizations for that server are complete:
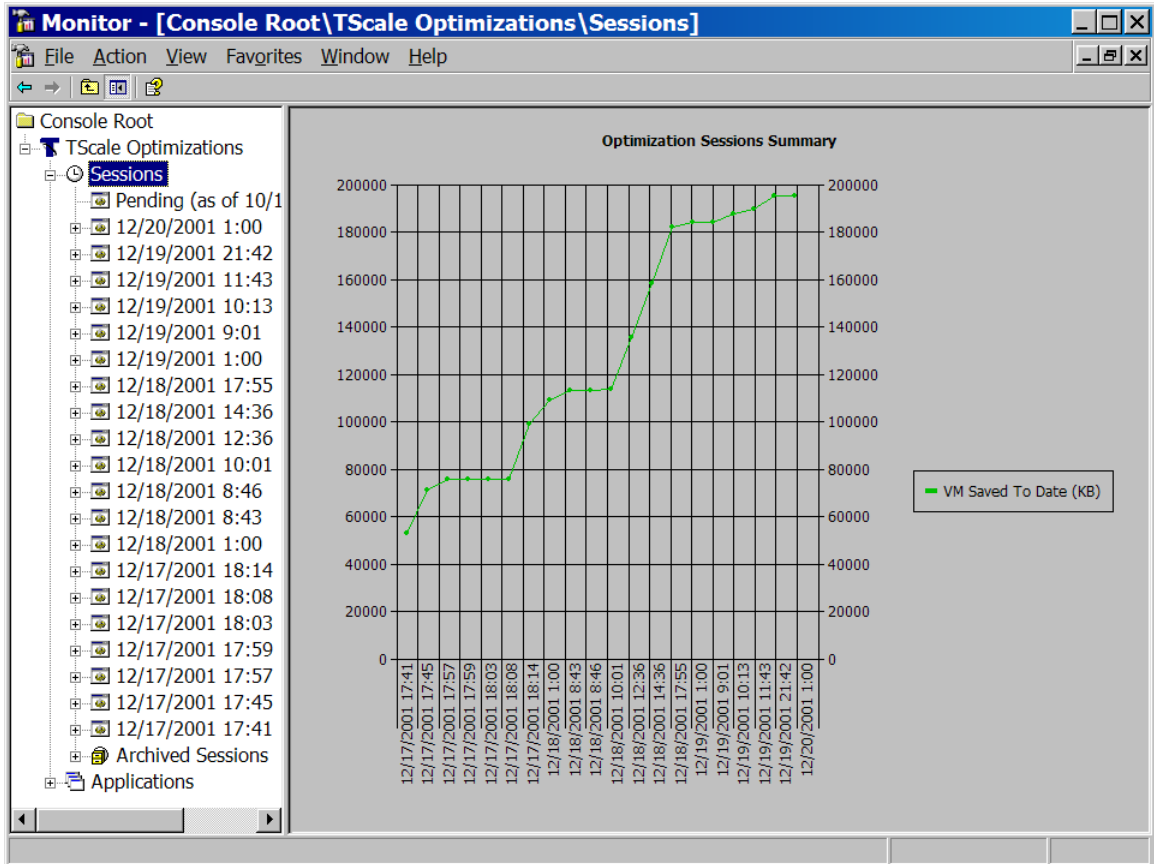
Back To Table Of Contents

28. What does the height of the green bars in the Applications Tree of the TScale Console mean?

The height of each green bar refers to the amount of code that will not get swapped out to the page file for each user of the application that is referred to by the bar. In the example below, the green bar for BUSOBJ.EXE is at 12,000K or 12MB. This means that TScale will reduce the amount of page file that each user of Business Objects uses by 12MB. So, the total page file reductions for 30 users of this application would be 30 x 12 MB or 360MB. Eliminating th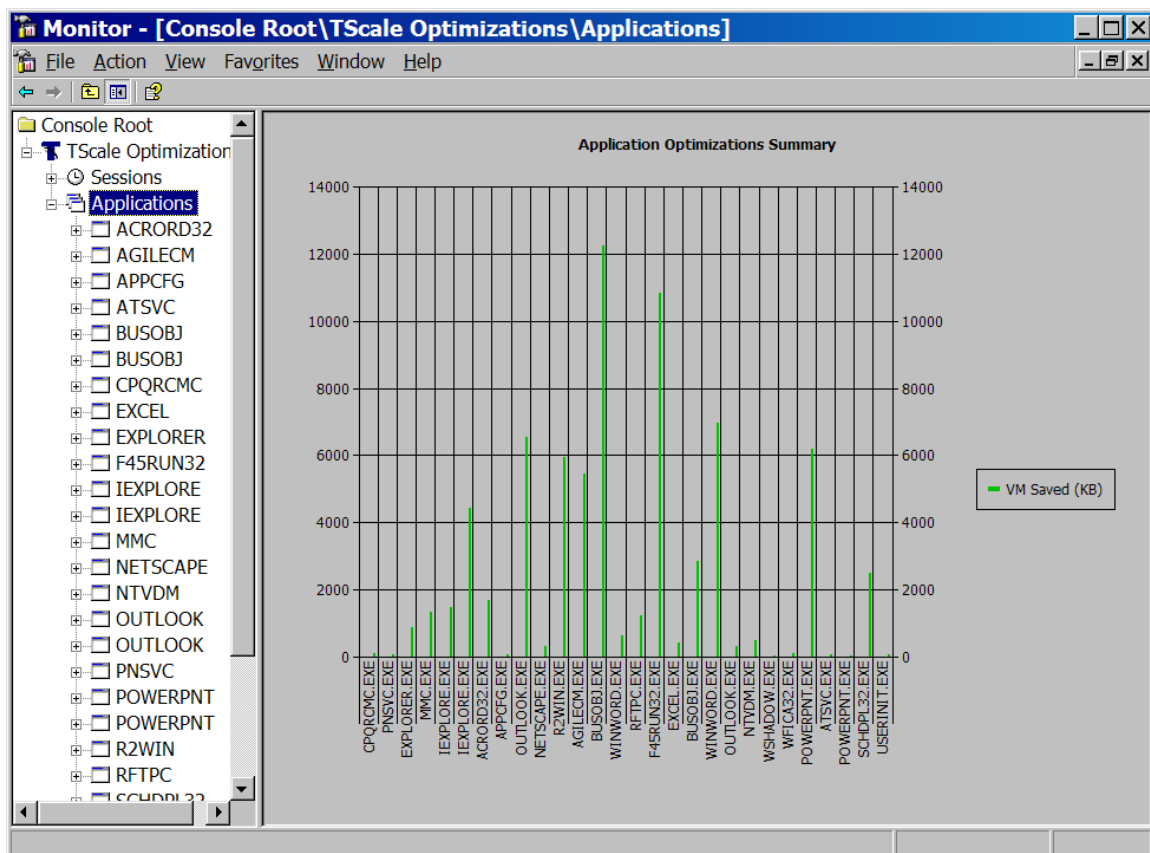e writing of all of this information to the hard disk, and then the page faults associated with recalling this information when it is needed, is what allows the VM Capacity optimization in TScale to improve number of concurrent users supported by an optimized terminal server.

29. Can I observe the virtual memory reductions that TScale is delivering via any metrics reported by PerfMon or Task Manager?

Yes you can. If you bring up Task Manager on both an optimized and un-optimized server, and add VM Size as a column, you will see the page file footprint for each instance of each process. If you compare these numbers for the frequently used applications on your servers, you will see that TScale optimized applications are reported by the OS as having a smaller average VM Size than the same applications on an un-optimized server. You can run the GetProcessInfo program on two servers (one

optimized one not) and compare the page file size columns in the resulting spreadsheet if you would like to automate this process. The GetProcessInfo program ships with TScale and resides in the Program Files\RTO Software\TScale directory.

30. **Does TScale change or touch the OS in any manner?**

No, TScale does not change, replace or modify the behavior of any component of the Windows operating system. TScale contains four levels of protection to ensure that this is the case. These four levels are:
   a. TScale has an internal check to see if an application or component is subject to Windows System File Protection. If it is (which all Windows 2000 components are subject to), then TScale will not optimize these components.
   b. TScale has an internal check to see if a component is part of "Known DLL's". If a component is in the known DLL list, TScale ignores it.
   c. TScale has an internal check to see if a component has been digitally signed, and will not optimize it if it is digitally signed.
   d. TScale has two exclusion lists, which are part of the product; an application exclusion list, and a component exclusion list. TScale will not optimize components that are on these lists. These lists are configurable within TScale and by policies within the PMC so that should an incompatibility occur, the problem component can be rapidly excluded from further optimizations.

31. **If TScale does not touch the OS, what does it touch?**

TScale performs its optimizations upon the applications that run on the terminal servers. It is useful to think of TScale as being analogous to an ODBC driver. An ODBC driver does not modify the functionality of the database, nor does it modify the functionality of the application. However, a bad ODBC driver can make both the application and the database perform poorly. A great one can make both perform to their full potential. TScale can be thought of as a driver that sits in between the application and the virtual memory subsystem of the OS, that enables the application to use less virtual memory and enables it to execute more efficiently without changing any of the applications code or data.

32. **Have TScale's optimizations ever caused an incompatibility in an application or part of an application?**

Yes, this has happened, although very infrequently. If you look at the Component and Application Exclusions lists, you will see all of the applications and components that have been added to those lists since TScale first starting shipping in October of 2001. Since there are less than 30 components and applications in each list, it is pretty obvious that applications incompatibilities are a fairly rare occurrence.

33. **What do I do if it turns out that TScale's optimizations create a problem for a component of one of my applications?**

Using the PMC and the policies for optimization within the PMC, you can turn off optimizations for that component on all of the servers in your farm in five mouse clicks. Turning it off will put it in the exclusion list for all of your servers, preventing the problem component from every being optimized again.

34. How does "Demo Mode" in TScale work?

If you install TScale with neither an evaluation or production license key it will operate in demo mode. In demo mode, TScale will find applications that need to be optimized but it will not implement optimizations. To see applications identified for optimization, right click on the Pending folder in the TScale Console or the PMC and select "Update Pending Optimizations". This will populate a bar graph that will appear if you highlight the Pending folder that will show you how much virtual memory will be saved for each instance of each application once those applications are actually optimized.

35. Can a demo mode installation be converted to evaluation mode without reinstalling TScale?

Yes, launch the console, right click on any node in the tree, select TScale Support, and then Activate. Paste the evaluation key that you received from RTO Sales into the field and your installation of TScale will convert from Demo mode to Evaluation mode. If you then run an optimization session, you TScale will optimize all of the applications found while demo mode was active. You will get the benefits of these optimizations the next time a user loads one of the optimized applications (optimizations cannot impact applications that are already running, the application must be unloaded and reloaded to get the benefit of the optimization).

36. How does "Evaluation Mode" in TScale work?

If you apply an evaluation key to TScale you will get 30 days of production optimizations. At the end of those 30 days, all optimizations will be turned off, and your server will be restored to its previously un-optimized state.

37. Can an Evaluation mode installation be converted into a fully purchased production copy without uninstalling and losing optimizations?

Yes, launch the TScale Console, right click on any node in the tree, select TScale Support, and then Activate. Paste your production license into the field. This will convert your evaluation copy of TScale into a production copy with no loss of optimizations.

38. Can I evaluate TScale on a test or QA server?

You will be able to see that TScale is working by observing the reports in the user interface of the TScale Console and the PMC. However, on a lightly used test or QA server, you will not be able to observe the impacts on utilization metrics, nor will you be able to assess the capacity or performance gains that TScale is delivering you.

39. If I have scripts for my applications, can I use a test server then?

Yes, you can. If you have accurately and comprehensively scripted the applications on your servers, then you need to follow the following process:
   a. Identify a server that will be used for the test. Before you install TScale run through the tests once to establish a baseline. If you have think times turned off in your script be sure to record not just the average response time that is generated through the running of the script, but also the number of times that each synthetic user is able to run through the script during the time interval that the script is configured for. This allows you to capture both the response times for the work that is being done, and the amount of work being done.

b. Install TScale, and run through the script again. At the end of this second run, run an optimization session in TScale. Then expand the Pending folder. If there are any items in the Pending folder that are marked with a yellow mark and which indicate that they were in use and could not be optimized, then reboot the server and run an optimization upon reboot (before any applications are loaded). This will ensure that the server is fully optimized before you run the optimization phase of the test.

c. Run the script through a third time, this time observing the impacts of the TScale optimizations. Compare the results of the first and third runs. You should see reduced response times, and an increased number of concurrent users that could be supported with acceptable response times.

40. Where does TScale store its optimizations?

The Analysis Service writes a file, Repair.sfo to the data directory that you specified when you installed TScale. The optimization task writes an XML file to the \archive directory in the data directory for each optimization session (if you look in the \archive directory you will see files with a name of bindxxxxx.sfo, where the xxxxx is the date and time stamp for that optimization session). This file contains the list of optimizations that were performed during that optimization session. TScale also persists its optimization map into the directories where the optimized applications are installed on the terminal server or file server.

41. What permissions need to be in effect for TScale to operate?

The Local System Account needs to have full control to the data directory where repair.sfo is located. The System account needs to have full control permissions to the data directory, and to all of the directories where optimized applications are installed.

42. Can I use Ghost to copy an optimized server image to other servers?

Yes you can. You can Ghost out TScale and the optimized image of the server to other servers. You just need to make sure that the TScale Analysis Service is running on the target machine when the Ghost is complete. Otherwise the Ghosted optimizations will not be in effect. You obviously also need to make sure that you have purchased enough TScale licenses to cover the total number of servers to which you have ghosted TScale and an optimized image.

43. What is the PMC?

The PMC (RTO Performance Management Console) is a multi-server deployment and management utility for TScale agents. Using the PMC you assign servers to groups, and then create optimizations policies for each group. If you turn off optimizations for a component or application in a policy, that action automatically applies to all of the servers in the group to which that policy applies.

44. Where should I install the PMC?

You can install the PMC in one of two places. If you administer your terminal server farm from your desktop, and you have a desktop PC running NT 4.0 SP6 or above (NT 4.0, Windows 2000, Windows XP, etc.) you can install the PMC on this workstation and manage your farm from there. If you need multiple users to use the PMC, it is best to

install the PMC on a server in the domain with the terminal servers, and then let each of the users who need access to it terminal service to it to use it.

45. Is there a license fee to have multiple users use the PMC?

No, there is not. You can have as many users share one installed instance of the PMC as you wish. If you need more than one installed instance, for example in the case of two physically separate farms, then you need to purchase two PMC's.

46. Why do I need the PMC?

The PMC allows you to perform the following functions for your terminal server farm from one central administrative console:
a. Grouping of servers into "silos" based upon the applications that run on them
b. Installation of TScale onto the servers in these groups
c. Creation of optimization policies for each group
d. Management of optimization through these policies

47. Why is policy-based management of farm silos important?

Optimization policies allow you to centrally define which optimizations occur on your terminal servers. In the rare case that an application incompatibility occurs, you can also reverse optimizations for an application or a component on the server that experienced the problem, and then, through the policy, apply that change to all of the servers in the silo or the farm. Policies therefore allow you to get the benefits of TScale optimizations across a large farm while also having powerful farm wide control of TScale should a change in optimization policies ever become necessary.

48. How can I purchase TScale and the PMC?

RTO sells its products through authorized Value Added Resellers. A list of these by country and state can be found on the RTO web site at:
http://www.rtosoft.com/pages/buy.asp

49. What is the pricing for TScale and the PMC?

You can get pricing from an authorized VAR, or by contacting RTO Sales at either 678-455-5506 Option 2, or sales@rtosoft.com.

50. Does RTO offer maintenance and support on TScale and the PMC.

Yes, RTO does offer these programs. Please contact RTO Sales at either 678-455-5506 Option 2, or sales@rtosoft.com to discuss these programs and their pricing.

51. What does the maintenance and support plan cover?

The maintenance plan includes all bug fixes, product updates, product upgrades, and new product versions for the products that you purchase maintenance for. The plan also includes unlimited telephone support during RTO's normal business hours.

52. How does TScale compare to products that limit resource usage?

They are different and complementary products. These products allow an administrator to set limits on how much CPU and memory any one application can utilize on a server, preventing any one application from running away with more than its share of resources.

Back To Table Of Contents

This is valuable when you have applications that have issues like memory leaks that can cause an application to consume all of the memory in the server. However, these products do nothing to make every instance of every application run more efficiently, which is exactly what TScale does. If you have three thirsty people and one glass of water, these products will stop any one person from drinking all of the water. TScale will improve the fitness of the three people so that each person needs less water, allowing the available glass to quench the thirst of four people instead of three.

53. How does TScale compare to products that claim to make more memory available for frequently used applications?

Products that claim to make more memory available for frequently used applications typically work by paging infrequently used applications to the page file, allowing the available working set to be used for frequently used applications. This is a valid concept on a desktop computer where the performance of the one application currently in use is more important that the performance of the applications that a single user has loaded but is not using. On a server, however, multiple users are using multiple applications and an approach that favors one application over another will not be as valuable as an approach that benefits all applications equally as TScale does.