

# Run and Reporting Rules for VMmark

Version – 1.1.5

Updated May 20, 2009

## 1 Introduction

The VMmark Run and Reporting Rules define how to correctly measure and report performance using the VMmark benchmark. This document outlines the requirements for producing a publication-quality VMmark result. Only results that meet these requirements may be published using the VMmark metric.

## 2 VMmark Reporting Rules

The VMmark Reporting Rules describe the requirements and options for publishing VMmark results. It also describes a review process that VMware strongly encourages benchmarkers to take advantage of prior to publication of a VMmark result. This section also includes a set of fair-use rules for use of VMmark results in competitive comparisons and for use in academic and research papers.

### 2.1 Result Review and Publication Process

VMware encourages all VMmark benchmarkers to formally submit a full disclosure report of their VMmark results and supporting documentation to the VMmark Review Panel for review and publication on the VMmark website. The VMmark Review Panel consists of members from partner companies who regularly publish VMmark results and/or have detailed expertise with the VMmark benchmark and is intended to maintain impartiality of published VMmark results. The Review Panel members have signed agreements to keep the details of VMmark submittals confidential until published. It is recommended that submitters of benchmark results review the *Review Panel Guidelines and Operating Procedures*. A roster of member companies on the Review Panel is available on the VMmark website.

Once the Review Panel is in place, the only way to have results published on the VMware VMmark website is by having results reviewed by the panel and accepted by VMware. Benchmark users may still self publish as described in Section 2.3.

- a) Submitter will deposit the appropriate submission information as defined by Section 2.5 to the official VMmark Review Panel ftp site. For the URL and access information send an inquiry email to [benchmark@vmware.com](mailto:benchmark@vmware.com).
- b) An automatic process runs on every other Tuesday at 9:00 am Pacific time (regardless of any holidays) to move all results posted in the submit FTP site to a location visible to all of the Review Panel members.
- c) The submitter must ensure that a full and complete submittal is in place before 9:00 am Pacific time on Day 1 of the review cycle.
- d) The submitter may submit a new submittal for the next cycle as early as Day 2 of the cycle prior to the cycle of their intended review. Submitted information will not be visible outside of VMware until Day 1 of the next review cycle.

- e) A submitter may withdraw a result that has been deposited on the submit FTP site prior to the open of the review cycle window by requesting in email to [benchmark@vmware.com](mailto:benchmark@vmware.com) that VMware remove it, 2 business days (which would not include holidays) prior to Day 1 of the next review cycle.
- f) A submitter may supersede a previously submitted result with another for a particular review cycle by first withdrawing the original submittal as described above and then submitting the replacement result before 9:00 am Pacific time on Day 1 of the review cycle.
- g) With the implementation of the Review Panel, the review cycle start dates will be fixed on every other Tuesday. A cycle ends at the exact same time as the next cycle begins, 9:00 am Pacific time.
- h) On Day 1 of the review cycle, all results submitted will be made available to all Review Panel members no later than 11:00 am Pacific time. Notification will be sent to the Review Panel members via email and it will provide an FTP site where the results can be found.
- i) Review Panel members have five business days to flag any issues and are encouraged to flag issues as early as detected via email. On Day 5 of the review cycle there will be a Review Panel phone conference to discuss any flagged issues with the submitter. The exact date of the conference call may be altered for holidays, with prior notice to the Review Panel and the submitter.
- j) The remaining days of the review cycle are for the submitter to resolve any issues with the result under review. (Note: Minor changes or clarifications will be allowed in the current review cycle. A rerun of a result to make a minor correction is allowed but the resulting score of the rerun can be no higher than the original score. Major changes will require a re-review in a subsequent review window.)
- k) At the end of the review cycle, a result will be accepted by VMware for publication on the VMmark webpage if either no issues were flagged, or all issues were satisfactorily addressed prior to the end of the review cycle.
- l) VMware will send an official message with the final status of the review for each separate submittal to the submitter and copy all Review Panel members.
- m) It is expected that under normal circumstances an accepted result will be posted to the VMmark website within 24 hours of the last day of the review cycle. Submitters may also request a slightly delayed publication date if desired.

If it is determined that the VMmark test was not run in full compliance with these rules, VMware will not accept that result for publication and the benchmarker must not use that result in any public disclosure. The benchmarker has the option of rerunning the test after making changes needed to bring it into compliance with these rules and submitting the new result and full disclosure report for review.

The benchmarker may withdraw a submission at any time while it is under review. However any open issues must be addressed prior to publishing the result independently.

## **2.2 Handling Compliance Issues in Published Results**

VMware and the benchmarker will promptly handle any issues related to compliance with these rules that do not become evident until after the result has been published. If

such issues arise, both parties will withdraw any web pages or downloadable reports that reference this result and will stop using the result in printed or other media. The materials may be revised to remove the reference and republished.

VMware will maintain a separate web page that provides a summary of the tested platform and the reason the published report was designated non-compliant (NC) or not available (NA) and withdrawn from the “Current” results page. The benchmarker may request that the entry be updated to include the description of any remedial action they've taken. For example, if the reason the result was designated non-compliant was: “Failed to meet 90 day General Availability requirement for products used,” it later could be updated to say: “All products used were generally available as of 06/07/08 and a new result has been submitted.”

### ***2.3 Independent Publication of VMmark Results***

Benchmarkers may independently publish results from fully compliant tests for which a full disclosure report is publicly available. The benchmarker may also publish independently and then make a formal submission for publication of the full disclosure report on the VMmark website. Any compliance issues found by VMware or other readers of the independently published disclosure will be handled in the same manner as results that undergo the formal review process.

### ***2.4 Handling of Previously Published Results***

VMware retains the right to publish and reference all results that have been reviewed by the Review Panel and accepted by VMware for publication. After 12 months, the benchmarker may request that VMware move the result from the “Current” results page to the “Historical” results page.

### ***2.5 Submission and Disclosure Requirements***

VMmark result submissions must include a complete description of the system under test (SUT), including hardware, software, configuration, and tuning. This disclosure should be sufficient that a third party could use this information to duplicate the test and obtain results within a deviation of +/- 3%. The level of detail in the disclosure report should be comparable to other industry standard benchmark results. The full set of VMmark test result measurements and metrics must also be included. An HTML template for the complete disclosure report is included with the VMmark documentation.

A complete disclosure report must be available from either VMware or the benchmarker whenever a VMmark result is published. The disclosure report may either be included as a hypertext link (URL) or included in its entirety in the publication.

The following materials must accompany any VMmark result that is submitted for review or be made available on request in the case of independently published results.

- a) The contents of the prime client’s results directory (vmmark/results/Results\_<datestamp>) in a .zip or .tar.gz archive format. This directory will contain the workload result files (.wrf) for all workloads and all tiles run.
- b) The output from any scripts supplied with the harness reporting tools to collect

configuration details on the testbed must be archived in .zip or .tar.gz format. Alternately, if the supplied scripts do not support a given virtualization platform, configuration details may be collected using individual commands.

## **2.6 Comparability of Results between VMmark Release Versions**

VMware retains the right to update and re-issue new versions of VMmark to licensees, including updates to these rules. Minor version releases will not affect the comparability of results with prior results within the same major release. VMmark results are not comparable if they were produced using VMmark releases with different major release numbers. All results with different major release numbers will be grouped separately on the VMmark website. Licensees will be notified when a new release is available and the last date results based on the prior release will be accepted.

## **2.7 Fair Use: Handling of Published VMmark References**

VMware requires that all published references to VMmark results adhere to the following fair-use rules:

- a) All VMmark results referenced must be from fully compliant tests for which a full disclosure is publicly available, either through the VMmark website or through a site specified by the benchmarker. The use of estimates is prohibited.
- b) A reference or link to the location of the full disclosure report for all referenced results must be included.
- c) Comparisons may be made between results that have the same major version number. If comparisons are made between “current” and “historical” results with the same major release number, the “historical” result must be clearly identified as “historical.”
- d) Only official VMmark metrics and submetrics may be used in comparisons.
- e) Any competitive comparisons of VMmark metrics must also include a statement providing the basis for making the comparison and the date when the claim is first made. A published comparison could for example state the claim is based on having the best VMmark result out of all results using four core systems, published as of the publication date (with a specific MM/DD/YYYY date given).
- f) Comparisons of VMmark metrics and submetrics to any other benchmark metrics are not allowed. VMmark utilizes other benchmarking software as load generators and produces results which are not comparable to the original benchmarks’ metrics.
- g) VMware and VMmark users will NOT use the SPEC<sup>®</sup> trademark metrics: SPECjbb2005<sup>®</sup> bops, SPECweb2005<sup>®</sup>, and SPECweb2005\_Ecommerce Simultaneous Sessions in conjunction with any published VMmark result.
- h) All publications which reference VMmark results must include the following attribution:
  - VMware<sup>®</sup> VMmark<sup>™</sup> is a product of VMware, Inc. VMmark utilizes SPECjbb2005<sup>®</sup> and SPECweb2005<sup>®</sup>, which are available from the Standard Performance Evaluation Corporation (SPEC).
- i) Academic and research use of VMmark is allowed as long as full disclosure of

variations from fully compliant test runs are made available as part of the full academic report. Publications containing results that are not in full compliance with these rules must be clearly identified as “academic” or “research” and must not utilize the VMmark metric name. Competitive comparisons using VMmark results in academic or research papers are not allowed.

- j) Academic and research results may make use of hardware and software components that are not generally available at the time of publication and are designated as pre-release or test versions (e.g. alpha, beta, release candidate, development fork, field test, prototype, etc.). In this case, the associated report must clearly identify the components that are not generally available and their planned release dates, or indicate that the component was experimental and not planned for product release.
- k) Publication of results using Beta versions of VMmark are allowed only under the academic and research rules described above. Any such results must be clearly marked “Beta VMmark Results” and include the Version ID for the release used from the Version.txt file included in the release. Beta versions of VMmark may only be used until a final release version becomes available.
- l) If a Beta version of VMmark is used to generate results under the academic and research rules described above the tests must be performed using a released and currently supported version of a VMware server product (e.g. ESX Server 3.0 or VMware Server 1.0 or later releases).

Note: Competitive comparisons are those comparisons that involve VMmark results where one or more performance-critical components or products from competing vendors have been used and where the purpose of the comparison is to illustrate that a product or set of products outperforms another based on the differences in the competing components.

### **3 VMmark Run Rules**

The VMmark Run Rules describe the requirements for configuring the SUT, the clients, and the test harness to produce valid and compliant test results. This section references the *VMware VMmark Benchmarking Guide* (the “*Benchmarking Guide*”), which provides the specific configuration requirements for the individual workload Virtual Machines (VMs).

The required virtual hardware definitions and the required software versions for the workload VMs are included in the *Benchmarking Guide* and must be followed in addition to the requirements included in this document. Any question regarding compliance with either document should be clarified with VMware to ensure that valid test results are generated.

#### **3.1 Overview of the VMmark Benchmark**

Server consolidation typically collects several diverse workloads onto a single physical server. This approach ensures that all system resources, such as CPU, network, and disk,

are more efficiently utilized. In fact, virtual environments tend to function more smoothly when demands are balanced across multiple resources. The unit of work for a benchmark of virtualized consolidation environments can be naturally defined as a collection of virtual machines executing a set of diverse workloads. The VMmark Benchmark refers to this unit of work as a *tile*. The total number of tiles that a physical system can accommodate gives a coarse-grain measure of that system's consolidation capacity. This concept is similar to some server benchmarks, such as TPC-C, that scale the workload in a step-wise fashion to increase the system load.

Tiles capture the consolidation and scalability, while the individual application workload performance captures small variations in system performance. To encompass this, both the number of tiles and the performance of each individual workload determine the overall benchmark score. Each workload within a tile is constrained to execute at less than full utilization of its virtual machine. However, the performance of each workload can vary to a degree with the speed and capabilities of the underlying system. For instance, disk-centric workloads might respond to the addition of a fast disk array with a more favorable score. These variations can capture system improvements that do not warrant the addition of another tile. However, the workload throttling will force the use of additional tiles for large jumps in system performance.

When a tile is added the performance of workloads in existing tiles might decrease but – if the system has not been over-committed – the aggregate score, including the new tile, should increase. The result is a flexible benchmark metric that provides a measure of the total number of workloads that can be supported by a particular system as well as the overall performance level within the virtual machines.

### **3.1.1 VMmark Benchmark Workloads**

A VMmark tile contains the following set of data center workloads:

- Mail server
- Java server
- Standby server
- Web server
- Database server
- File server

Each workload is described by the application running on each VM within the tile as well as the Virtual Machine platform for the workload, which includes the OS and the CPUs, RAM, network, and storage requirements. Each workload in the tile is driven from the associated client system by the workload-specific load generator and/or data collection software.

The precise requirements for the each workload VM are covered in the *VMware VMmark Benchmarking Guide* (the “*Benchmarking Guide*”) included with the VMmark release being used. These requirements include the Virtual Machine definition (OS, CPUs, RAM, etc) and any required software (applications, benchmarks, or other tools) along

with version/release requirements.

### **3.2 Testbed Configuration Requirements**

This section covers the configuration requirements for the testbed, which includes the system under test (SUT).

The benchmarker must configure their VMs according to the *Benchmarking Guide's* instructions and follow the limitations on the software versions as described. For example, the Mail, Java, and Standby servers require the use of Windows 2003 Enterprise Edition, while the other three workload VMs require SUSE Linux Enterprise Server 10. The benchmarker may apply updates to these releases as described in the *Benchmarking Guide*, but may not use other versions of Windows or Linux or substitute one OS for the other.

The server, client, and network hardware configuration requirements, as well as the associated software requirements, are also detailed in the “VMmark Benchmark Requirements” section of the *Benchmarking Guide*. To produce fully compliant VMmark results, all components in the testbed must have been configured according to these requirements. The resources (i.e., CPU, memory, disk, network, etc.) allocated to each VM must not be altered by tuning which affects the balance of assigned resources.

#### **3.2.1 Types of Changes Permitted to SUT and Clients**

This section describes the types of changes to the SUT and clients that are generally permitted by these VMmark Run and Reporting Rules and the requirements sections of the *Benchmarking Guide*.

Note: A change or tuning may be deemed performance neutral by the VMmark Review Panel. Typically a performance neutral change will have no effect on the default settings as tested (e.g. change is ignored or overridden by another setting) or will not effect performance to a degree that the result could not be duplicated without this setting (e.g. enabling a virtual device not used in the test).

##### **3.2.1.1 Hardware Configuration:**

- a) The hardware configuration for the SUT and clients may be modified or tuned as long as they remain configurations supported by the respective hardware vendors and the vendor of the virtualization product being used. Changes may include adding memory, NICs, storage adaptors, attaching storage arrays, etc.
- b) Changes to the server's BIOS or firmware parameters may be done if this tuning is consistent with the vendor's published recommendations.
- c) All storage and file systems must use stable non-volatile media.

##### **3.2.1.2 Virtualization Layer Configuration:**

- a) The virtualization layer may be modified or tuned as long as it remains a supported configuration and the tuning is consistent with the vendor's published recommendations and does not alter the resource allocation to favor some VMs over others.

- b) The virtualization layer for individual VMs may be modified or tuned as long as it remains a supported configuration and the tuning is consistent with the vendor's published recommendations and does not alter the resource allocation to favor some VMs over others (except as noted in 3.2.1.2c below). Additionally, any tunings must be applied to all VMs of the same workload type across all tiles used in the test, (unless the tuning is performance neutral for this SUT). For example if a virtualization optimization is enabled on the Webserver VM in one tile, then it must be enabled on the Webserver VMs in every tile.
- c) For the fileserver VM, a feature that dynamically reclaims physical memory the VM is not actively using (e.g. ballooning), can be disabled on all the fileserver VMs used in the test. Due to the underlying design of the fileserver workload, this tuning can be applied to ensure that results are stable and reproducible.

### **3.2.1.3 Virtual Machine Configuration:**

- a) All VMs must use the Virtual Machine Platform definition specified in the *Benchmarking Guide*. The platform definition includes the number of CPUs, the amount of memory, the storage and network requirements, and the software requirement for each workload VM.
- b) Each VM must be configured with only one virtual NIC with the default Maximum Transmission Unit (MTU) of 1500 bytes and only one virtual storage controller.
- c) Each VM's software configuration may be changed only within the limits described in the *Benchmarking Guide*, which specifies minimum versions to be used and the level of upgrades that may be applied.
- d) The version of the OS kernel that is included in the required OS distribution or an allowed update release must be used.
- e) The default system services and daemons must remain enabled unless specific changes are specified in the *Benchmarking Guide*.
- f) If any workload VM's software configuration is changed based on the rules above (such as upgrading the VM's OS to an allowed update release), then that change must be applied to all VM's of that type in all tiles, (unless the tuning is performance neutral for this SUT) and information about the change must be listed in the Notes section of the disclosure.

### **3.2.1.4 Client Configuration:**

- a) Each Client System may be modified or tuned as long as it remains a supported configuration, the tuning is consistent with the vendor's published recommendations, and it meets all the minimum requirements as described in the *Benchmarking Guide* and these rules.

### **3.2.1.5 VMmark Test Harness Configuration:**

- a) The VMmark Test Harness software installed includes the VMMARK.CONFIG configuration file, which may be modified as described in the *Benchmarking Guide*.
- b) Software updates supplied by VMware to address problems or supply new features

may be used in place of VMmark Test Harness components in the current release. Such changes must be documented in the notes section of the disclosure.

### **3.2.2 Types of Changes NOT Permitted to SUT and Clients**

The following section describes the types of changes that are generally NOT permitted by these VMmark Run and Reporting Rules and the requirements sections of the *Benchmarking Guide*.

#### **3.2.2.1 Hardware Configuration:**

- a) Features that are not consistent with the vendor's published recommendations may not be used in the SUT or clients.

#### **3.2.2.2 Virtualization Layer Configuration:**

- a) Features that are not consistent with the vendor's published recommendations may not be used in the configuration of the virtualization layer.
- b) Features that are not consistent with the vendor's published recommendations may not be used in the per VM virtualization configuration.
- c) Configuration changes that alter the balance of assigned resources (CPU, memory, disk, network) for specific VM's are not permitted. This includes optimizations that increase or decrease the amount of a given resource to be allotted to a specific VM in the tile when the system is under load.

#### **3.2.2.3 Virtual Machine Configuration:**

- a) No changes may be made to the Virtual Machine Platform definition specified in the *Benchmarking Guide*.
- b) No changes may be made to the tuning parameters for the guest OS unless specified in the *Benchmarking Guide*.
- c) No changes may be made to any workload-specific application or benchmarking code that runs on the workload VMs unless the *Benchmarking Guide* includes specific directions to do so.
- d) No changes may be made which alter the resources allotted to specific VMs to the advantage or disadvantage of any other VM in the tile.

#### **3.2.2.4 Client and Other Configuration:**

- a) No changes may be made to any workload-specific client-side code unless the *Benchmarking Guide* includes specific directions to do so.
- b) No changes may be made to any code supplied as part of the VMmark distribution, unless the *Benchmarking Guide* specifies that the file or designated parameters can be changed.

#### **3.2.2.5 VMmark Test Harness Configuration:**

- a) The VMmark Test Harness configuration files may not be modified except as

described in the *Benchmarking Guide*.

- b) Changes to the harness software are not allowed with the exception of updates supplied by VMware.

### **3.2.3 Types of Changes to SUT and Clients Not Covered**

- a) Any changes, which may or may not have been explicitly allowed or disallowed by these rules or the *Benchmarking Guide*, should be reviewed with VMware prior to their use.

## **3.3 Product Availability Requirements**

All hardware and software components that compose the system under test must meet the following product availability requirements at the time the associated VMmark result is published and in the time period immediately following the publication:

- a) The hardware and software components are either generally available (GA) products or are pre-release versions within 90 days of official GA release. To be considered generally available, a product must be shipping and fully supported by the vendor or its designated support channel. If a pre-release component is used, the benchmarker must have verified that the vendor is committed to making the component generally available within the 90-day timeframe. The publication of a VMmark result with a forward availability date is considered a public commitment that any potential customer for that component must be able to confirm with the vendor or supplier.
- b) Storage components such as disks and storage arrays, which were generally available within the last five years but are no longer available from the original vendor, may be used in the system under test if the original general availability date is noted in the disclosure.
- c) All products used must remain on the market for 60 of the first 90 days after they first become generally available. If a product is removed from the market for more than 30 days during the first 90 days, it is considered “Not Available” (NA). Replacement by a new similar component or new version with equivalent or better performance is considered acceptable and not a reason for a NA designation.
- d) Equivalent or better performance in the products that comprise the SUT must remain available during the first 90 days after the SUT becomes GA or after the result is published, whichever is later. To have equivalent performance requires that any regression would reduce the VMmark result by no more than 3%.

## **4 Handling Run and Reporting Rule Compliance**

To ensure that benchmarkers adhere to the VMmark Run and Reporting Rules, this section summarizes typical compliance issues and how they will be handled to ensure the integrity of the benchmark and VMmark metrics.

### **4.1 Compliance Issues Found Prior to Publication**

While a submission is in review, issues related to the full disclosure report (FDR) will be

brought to the attention of the submitter so that the report can be updated. If the issues raised point to a possible violation of the run rules, the submitter will be asked to clarify the situation or to rerun the benchmark in compliance with the run rules. The submitter may withdraw the submission at any time during the review period but may not independently publish the result if there are open compliance issues.

#### ***4.2 Compliance Issues After Publication on VMmark Website***

Any compliance issues found after a result is published will be brought to the attention of the VMmark Review Panel. If a compliance issue in the FDR is found after the submission has been published on the VMmark website, the submitter may correct the FDR and it will be published in place on the VMmark Results page. If the compliance issue raised points to a run rule violation and further review by the Review Panel substantiates this, then VMware will designate the published result as Non-Compliant and update the VMmark website as described in section 2.2. The Submitter is encouraged to make a new compliant run and submit the new result for review and publication.

#### ***4.3 Compliance Issues after Independent Publication***

If a compliance issue with the FDR is found after the submission has been published outside the VMmark website, then VMware will work with the benchmarker to correct the FDR for re-publication. If the compliance issue raised points to a run rule violation, and further review by the Review Panel substantiates this, then VMware will request that references to the result in question be taken down (see section 2.2). The benchmarker is encouraged to make a new compliant run and republish as appropriate. VMware reserves the right to take action if VMmark results that have been designated Non-Compliant remain posted on the submitter's website or other media which they control.

#### ***4.4 Compliance Issues when Pre-Release Product is Not Released***

If pre-release products have been used and those products have not become generally available (GA) by the end of the 90-day window (see section 3.3) then the result will be designated non-compliant. If the result has been published on the VMmark website, then VMware will designate the published result as Non-Compliant and update the VMmark website as described in sections 2.2 and 4.2 above. Any publication of the result outside the VMmark website will be handled as described in sections 2.2 and 4.3 above.

#### ***4.5 Compliance Issues when Product is Not Available***

If it is found that within 90 days of the publication of a VMmark result or the general availability date of the SUT, whichever is later, performance critical products used in the SUT had been taken off the market for more than 30 days during that period, then that result will be designated Not Available (NA). If the result has been published on the VMmark website, then VMware will designate the published result as Not Available and update the VMmark website as described in section 2.2. Any publication of the result outside the VMmark website will be handled as described in section 2.2. The benchmarker is encouraged to make a new compliant run and submit the new result for review and publication. Note: Replacement by a new similar component or new version

with equivalent or better performance is considered acceptable and not a reason for an NA designation.

#### **4.6 Compliance Issues when Result is Not Reproducible**

If within 90 days of the publication of a VMmark result or the general availability date of the SUT, whichever is later, it is found that there has been a performance regression of more than 3%, then that result will be designated non-compliant. If the result has been published on the VMmark website, then VMware will designate the published result as Non-Compliant and update the VMmark website as described in sections 2.2 and 4.2 above. Any publication of the result outside the VMmark website will be handled as described in sections 2.2 and 4.3 above.

## **5 Caveats**

VMware reserves the right to modify the VMmark benchmark codes and workloads and documentation including these Run and Reporting Rules as needed to ensure the value and integrity of VMmark as a performance benchmark for virtualization. The latest published edition of these Run and Reporting Rules will apply to all VMmark releases and VMmark results unless specifically stated to the contrary.

If VMmark users have questions or concerns regarding these rules and their interpretation or with any of the VMmark materials, they are strongly encouraged to contact VMware for clarification. Additional feedback on VMmark is welcome as well. VMware provides a VMmark forum that can be used to post questions and provide feedback (see: [communities.vmware.com/community/vmtn/server/vmmark](http://communities.vmware.com/community/vmtn/server/vmmark)).

### **Notice:**

**The contents of this document are subject to change at any time. Please check with VMware, Inc. to ensure that you have the latest version.**

Copyright © 2006 - 2008 VMware, Inc. All rights reserved.