# White Paper
# FUJITSU Server PRIMERGY & PRIMEQUEST Benchmark Overview VMmark V3

Conventional single workload benchmarks are less suited for the assessment of virtualized operating systems and applications, which is why special virtualization benchmarks exist. The “VMmark V3” benchmark developed by VMware allows a cross-manufacturer comparison of highly optimized configurations on the basis of hypervisor solutions from VMware.

This document describes the problems concerning benchmarks for virtualized environments as well as the fundamentals of the “VMmark V3” benchmark and its use at Fujitsu.

| Version |
| --- |
| 1.1 |
| 2021-07-28 |

# Contents



# Document history

***Version 1.0 (2018-03-20)***

- First report version

***Version 1.1 (2021-07-28)***

- Update of Contact information and URLs

# Introduction

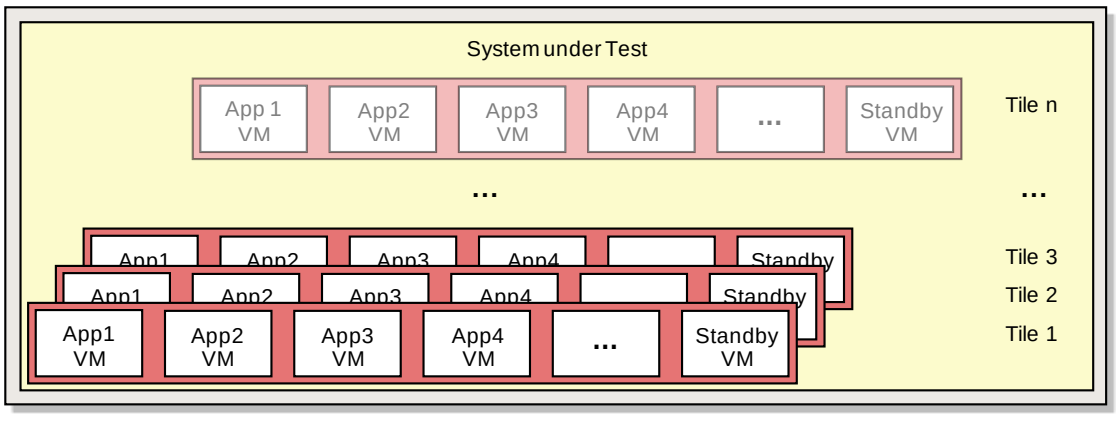
Server virtualization is an increasingly deployed scenario in the implementation of IT infrastructures. On the one hand, virtualization enables more efficient hardware to be used in an optimal way, and on the other hand the dependency on hardware specifics is being reduced. Contrary to the classic server with only one operating system and its applications, with virtualized environments, several operating systems and applications are run in parallel creating heterogeneous environments on one server. The handling of virtual machines is implemented by a virtualization layer, also called hypervisor.

Classic single workload benchmarks are scarcely suited to measure and assess the performance capability of virtualized environments. For this purpose, it is necessary to utilize the hardware resources of a server with simultaneously working virtual machines with different workloads. The one aim of virtualization benchmarks is server consolidation. In this case, the throughput of a set collection of virtual machines is maximized on a single virtualization host by means of suitable replication. SPECvirt_sc2013 (SPEC) fall under this category. Fujitsu carries out scaling measurements of virtual environments using its internal benchmark “vServCon”. The other aim of virtualization benchmarks is data center operations. A server consolidation scenario for several virtualization hosts is assumed in this case. In addition to the throughputs of the virtual machines, the benchmark metric then contains ratios that reflect the efficiency of typical data center operations, such as the relocation of virtual machines. These benchmarks include VMmark V3 (VMware).

For a virtualization benchmark to fulfill its objective, it must map the real world of a data center regarding server consolidation; in other words it must consider existing servers with those application scenarios that are normally virtualized. These servers have weak utilization levels and the aim is thus to consolidate as many of them as possible as virtual machines (VMs). Therefore, such a benchmark must assess for a virtualization host both the suitably determined overall throughput across the various application VMs as well as the number of efficiently operable VMs.

The following solution concept has been established for these two objectives: a representative group of application scenarios is selected in the benchmark. They are started simultaneously as a group of VMs on a virtualization host when making a measurement. Each of these VMs is operated with a suitable load tool at a defined lower load level. All known virtualization benchmarks are thus based on a mixed approach of operating system and applications - plus usually an “idle” or “standby” VM which represents the inactive phases of a virtualization environment and simultaneously increases the number of VMs to be managed by the hypervisor. The term “tile” is the name for such a unit of virtual machines.

It must be possible to increase this well-defined load created by this group of virtual machines on a step-by-step basis until the considered system has reached its performance limit. The following illustration shows the growth of VM load on a system under test by operating several tiles.




An application is executed in each virtual machine where the applications are put under stress via established benchmarks. If necessary, there may also be further infrastructure components. All the individual results are then suitably summarized in one overall result. This score is an indication for the performance capability of a virtualized environment.

# VMmark V3

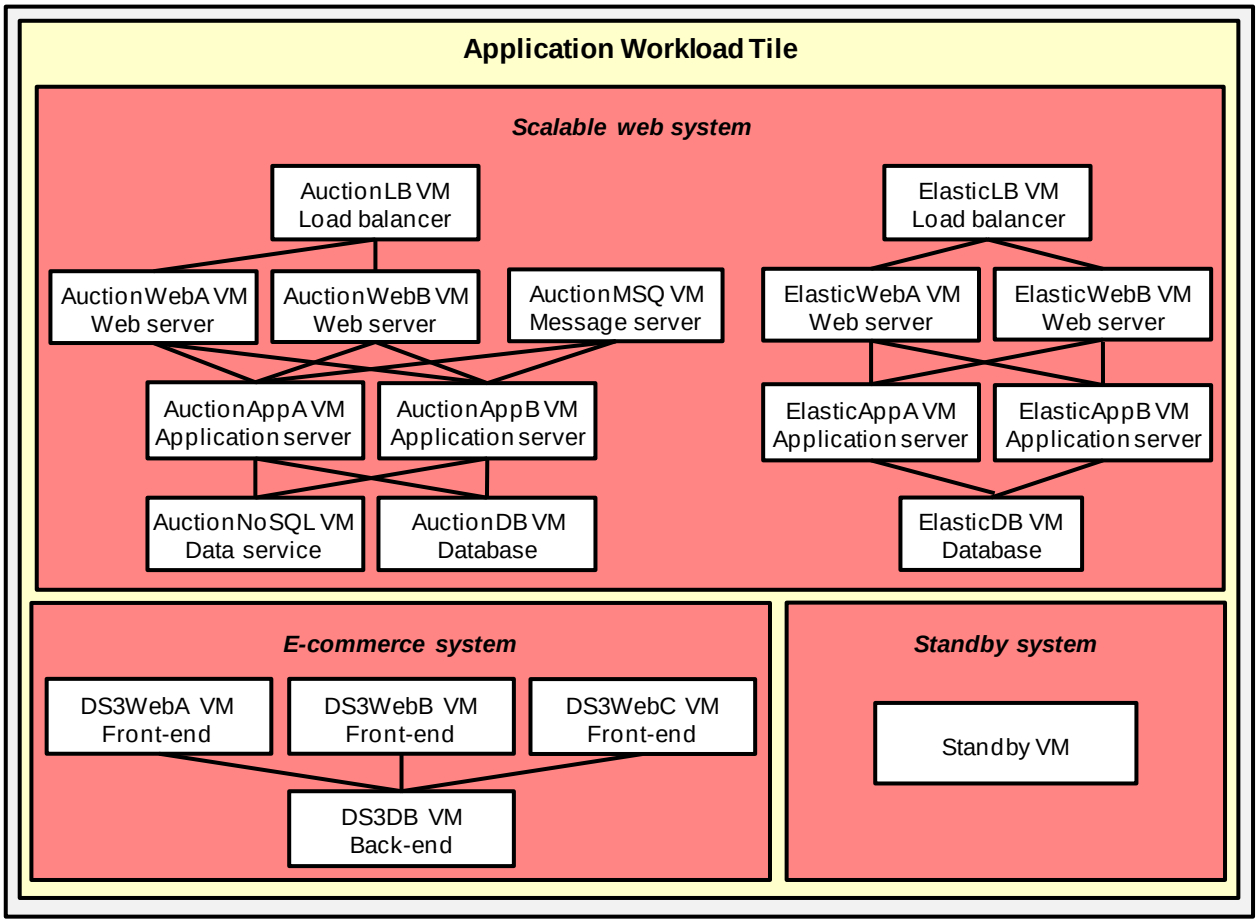
VMmark V3 is a benchmark developed by VMware to compare server configurations with hypervisor solutions from VMware regarding their suitability for server consolidation. In addition to the software for load generation, the benchmark consists of a defined load profile and binding regulations. The benchmark results can be submitted to VMware and are published on their Internet site after a successful review process. After the discontinuation of the proven benchmark "VMmark V2" in September 2017, it has been succeeded by "VMmark V3".VMmark V2 required a cluster of at least two servers and covers data center functions, like Cloning and Deployment of virtual machines (VMs), Load Balancing, as well as the moving of VMs with vMotion and also Storage vMotion. VMmark V3 covers the moving of VMs with XvMotion in addition to VMmark V2 and changes application architecture to more scalable workloads.

## VMmark V3 Benchmark

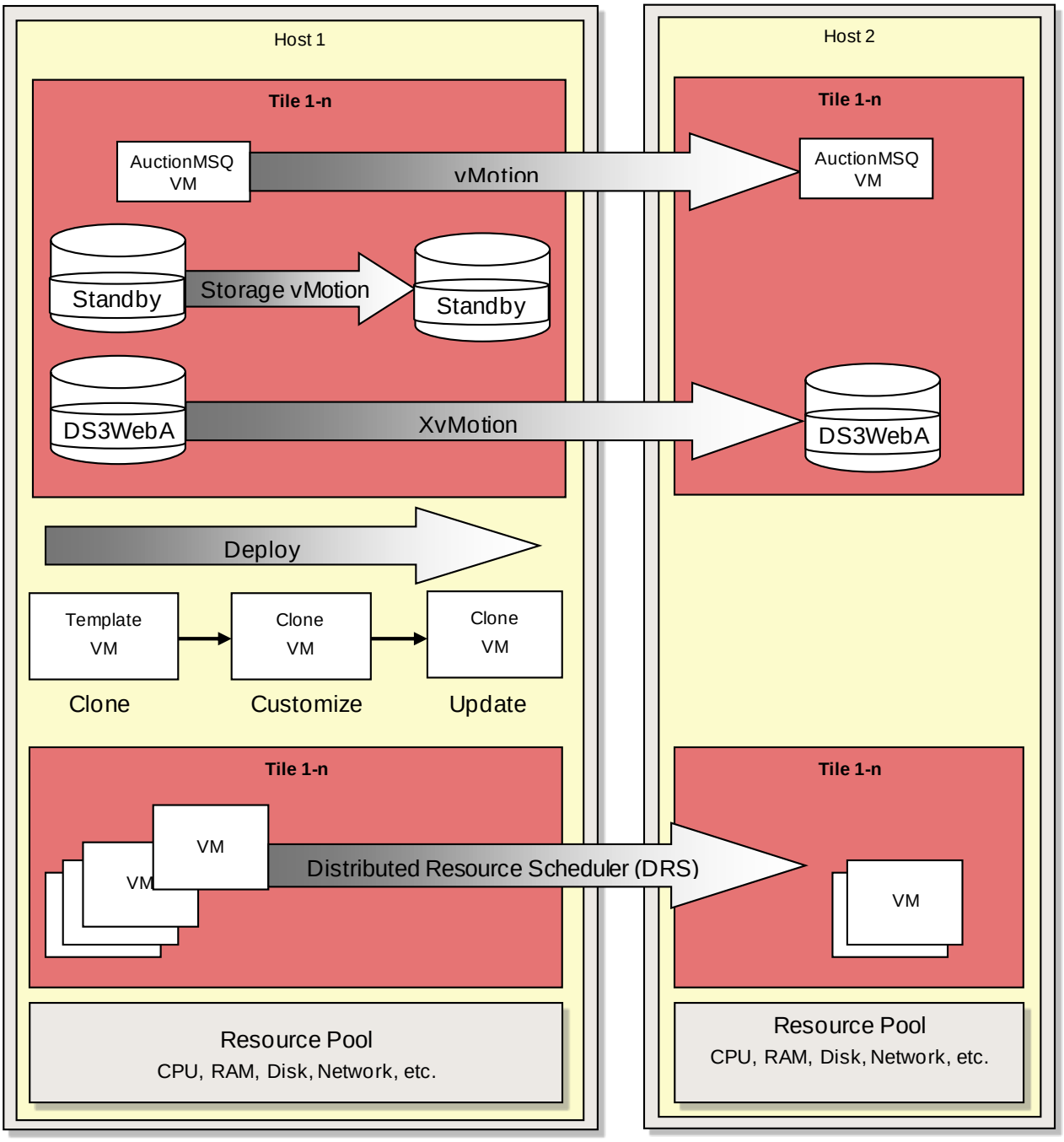
VMmark V3 is not a new benchmark in the actual sense. It is in fact a framework that consolidates already established benchmarks, as workloads in order to simulate the load of a virtualized consolidated server environment. Two proven benchmarks, which cover the application scenarios Scalable web system and e-commerce were integrated in VMmark V3.

| Application scenario | Load tool | # VMs |
|---|---|---|
| Scalable web system | Weathervane | 14 |
| E-commerce system | DVD Store 3 client | 4 |
| Standby system | | 1 |

Each of the two application scenarios is assigned to a total of eighteen dedicated virtual machines. Then add to these a nineteenth VM called the "standby server". These nineteen VMs form a "tile". Because of the performance capability of the underlying server hardware, it is usually necessary to have started several identical tiles in parallel as part of a measurement in order to achieve a maximum overall performance.



A feature of VMmark V3 is an infrastructure component, which is present once for every two hosts. It measures the efficiency levels of data center consolidation through VM Cloning and Deployment, vMotion, Storage vMotion and XvMotion. The Load Balancing capacity of the data center is also used (DRS, Distributed Resource Scheduler).

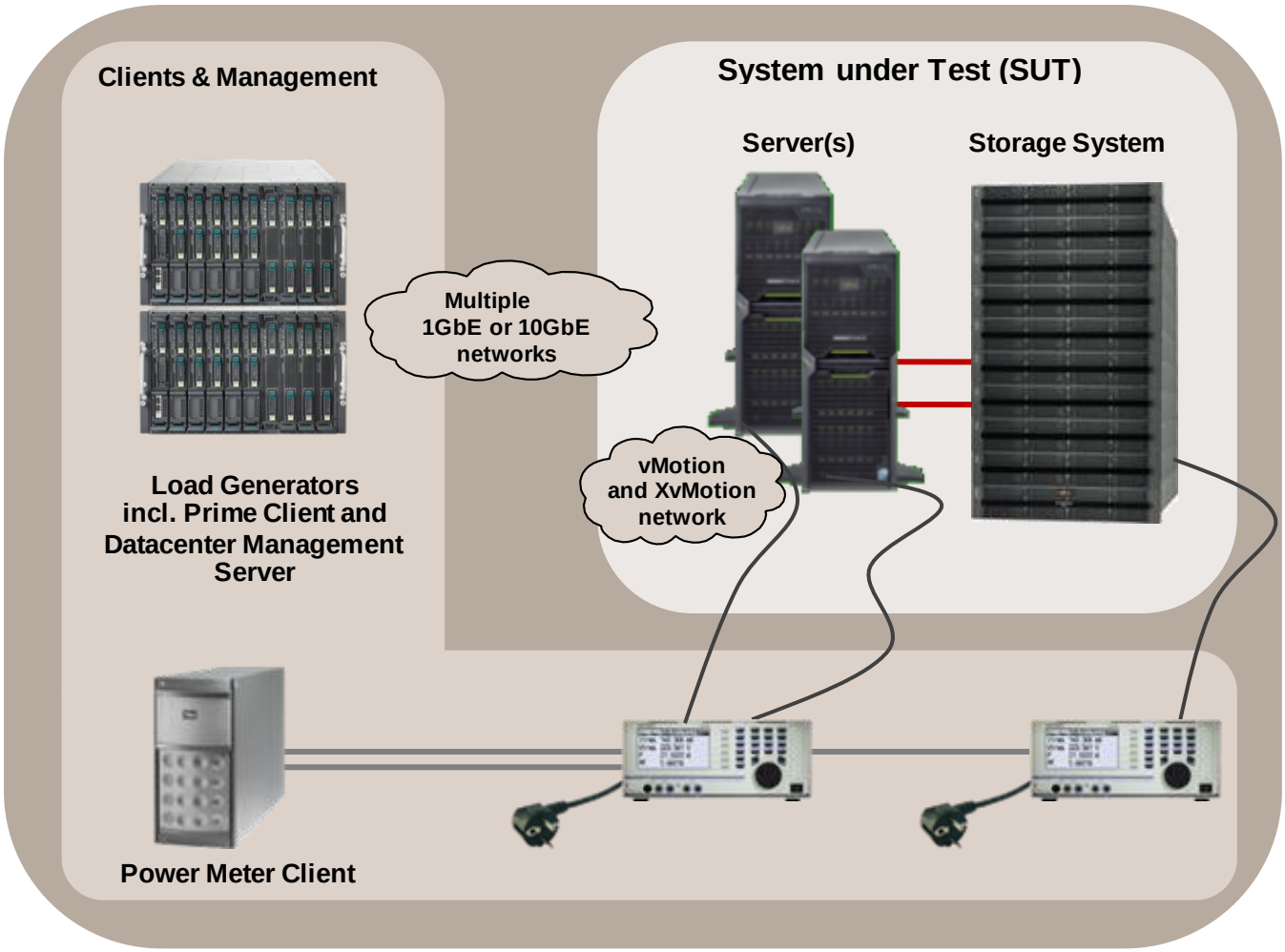



VMmark V3 requires external load generators: exactly one virtual client system per tile. Load generators and the individual hosts of the “system under test” are connected via a suitable number of networks. Also integrated in the network is the necessary data center component (VMware vCenter).

The execution of the individual load tools is controlled by one of the load generators, the so-called prime client. The prime client monitors the measurements and collects the individual performance data of the VMs and of the infrastructure activities.

## VMmark V3 Environment

The measurement set-up is symbolically illustrated below:




Power measuring devices are only required if a result is to be generated for a power score. This also applies for the “Power Meter Client” system, provided that it is a dedicated system. As a matter of principle, it is possible to use several measuring devices or “Power Meter Client” systems. For results of the test type “Performance with Server and Storage Power” it is necessary for the power consumption of a possibly existing switch, which may be needed to connect the storage system to the servers, to also be measured.

## VMmark V3 Score

The result of VMmark V3 for test type “Performance Only” is a number, known as a “score”, which provides information about the performance of the measured virtualization solution. The score reflects the maximum total consolidation benefit of all VMs for a server configuration with hypervisor and is used as a comparison criterion of various hardware platforms.

This score is determined from the individual results of the VMs and an infrastructure result. Each of the five VMmark V3 application or front-end VMs provides a specific benchmark result in the form of application-specific transaction rates for each VM. In order to derive a normalized score the individual benchmark results for one tile are put in relation to the respective results of a reference system. The resulting dimensionless performance values are then averaged geometrically and finally added up for all VMs. This value is included in the overall score with a weighting of 80%. The infrastructure workload is only present in the benchmark once for every two hosts; it determines 20% of the result. The number of transactions per hour and the average duration in seconds respectively are determined for the score of the infrastructure workload components.

In addition to the actual score, the number of VMmark V3 tiles is always specified with each VMmark V3 score. The result is thus as follows: “Score@Number of Tiles”, for example “8.11@8 tiles”.

In the case of the two test types “Performance with Server Power” and “Performance with Server and Storage Power” a so-called “Server PPKW Score” and “Server and Storage PPKW Score” is determined, which is the performance score divided by the average power consumption in kilowatts (PPKW = performance per kilowatt (KW)). Since the two new test types are derived from the original test type “Performance Only”, the power score is only determined under full load.

The results have been listed in the Internet in three different tables under VMware.

VMmark V3 scores can only be compared with each other and particularly the scores of each workload contained in VMmark V3 cannot be considered separately and interpreted or compared with the scores of the original benchmarks. The results of the three test types “Performance Only”, “Performance with Server Power” and “Performance with Server and Storage Power” should not be compared with each other. Please note that a single benchmark run can by all means generate results for several (and even all) test types, which are in each case submitted as a single result. Therefore, when evaluating a VMmark V3 result you should consider whether the set-up was more likely to have been optimized in terms of performance or maximum energy efficiency. Details about the configuration are available in the VMmark report file (“Disclosure Report”).

All VMmark V3 “Performance Only” results can’t be compared with the VMmark scores of the VMmark V1 and VMmark V2; with the same performance in the virtualization environment the scores and numbers of tiles for VMmark V3 are considerably lower.

## VMmark V3 Load Profile and Run Rules

The VMmark V3 rules define a standardized benchmark environment and the resources and software versions of the operating system and the application software are specified precisely. The compilation of guest operating systems and applications and their specified resource requirements are to be understood as a representative selection for a complex virtualization environment. Even if there are newer and perhaps even more high-performance versions, continuity of the load profile must be maintained for as long a time as possible in order to ensure comparability.

The following profile is used for VMmark V3:

| Resource | # VMs | # vCPU | Memory | OS | Application | Benchmark | Disk Subsystem |
|---|---|---|---|---|---|---|---|
| Auction Web<br>Web server | 2 | 2 | 8GB | Cent OS 7.2 64bit | Nginx 1.12.0 | Weathervane Auction | 16GB |
| Auction App<br>Application server | 2 | 4 | 14GB | Cent OS 7.2 64bit | Tomcat 8.5.13 | Weathervane Auction | 16GB |
| Auction LB<br>Load balancer | 1 | 2 | 4GB | Cent OS 7.2 64bit | HAproxy 1.5.18 | Weathervane Auction | 16GB |
| Auction MSQ<br>Message server | 1 | 2 | 4GB | Cent OS 7.2 64bit | RabbitMQ 3.5.3 | Weathervane Auction | 16GB |
| Auction DB<br>Database | 1 | 2 | 8GB | Cent OS 7.2 64bit | PostgreSQL 9.3 | Weathervane Auction | 16GB Boot<br>20GB Load |
| Auction NoSQL<br>Datastore | 1 | 2 | 16GB | Cent OS 7.2 64bit | MongoDB 3.0.14 | Weathervane Auction | 16GB Boot<br>100GB Load |
| Elastic Web<br>Web server | 2 | 2 | 4GB | Cent OS 7.2 64bit | Nginx 1.12.0 | Weathervane Elastic | 16GB |
| Elastic App<br>Application server | 2 | 2 | 8GB | Cent OS 7.2 64bit | Tomcat 8.5.13 | Weathervane Elastic | 16GB |
| Elastic LB<br>Load balancer | 1 | 1 | 4GB | Cent OS 7.2 64bit | HAproxy 1.5.18<br>RabbitMQ 3.5.3<br>MongoDB 3.0.14 | Weathervane Elastic | 16GB Boot<br>25GB Load |
| Elastic DB<br>Database | 1 | 2 | 8GB | Cent OS 7.2 64bit | PostgreSQL 9.3 | Weathervane Elastic | 16GB |
| e-commerce<br>back-end | 1 | 8 | 32GB | Cent OS 7.2 64bit | MySQL 5.6.34 | DVD Store 3 DB | 16GB Boot<br>250GB Load<br>100GB Load |
| e-commerce<br>front-end | 3 | 1 | 4GB | Cent OS 7.2 64bit | Apache 2.4.6 | DVD Store 3 Web | 16GB |
| Standby | 1 | 1 | 2GB | Cent OS 7.2 64bit | | | 16GB |

Add to this the following infrastructure components as defined by VMware for VMmark V3:

When it comes to VM Cloning and Deployment, the VMware vCenter copied a template VM, adapted according to the customization requirements (IP configuration, system name) and provided with a new version after startup. The new VM is removed again after processing the specific task and the process is repeated every forty seconds. For the vMotion infrastructure component the AuctionMSQs of all tiles are moved between the existing hosts on a round-robin basis via the VMware vCenter. The AuctionMSQ comes back to the source host after waiting two minutes on the destination host, and the process is repeated every two minutes. The standby VMs are moved for Storage vMotion every two minutes on a round-robin basis. And the DS3WebA VMs are also moved for XvMotion every two and a half minutes on a round-robin basis. The VMware vCenter also uses a “Distributed Resource Scheduler (DSR)” to ensure an automatic Load Balancing function.

The complex VMmark V3 rules are restricted to optimizations and tuning for comparison reasons; all permitted changes in the standard configuration must be documented when the score is submitted.

The disk subsystem can be individually configured and optimized according to the rules with not only the logical design regarding size and RAID level, but also the physical implementation. Since a local disk subsystem is not a sensible solution for virtualization and consolidation and the disk subsystem for VMmark V3 also has to support vMotion and Storage vMotion, a SAN-based disk subsystem is used. The performance of the disk subsystem has a direct influence on the VMmark V3 score; if there is a bottleneck an optimal score cannot be attained.

Neither should the main memory of every host system be a bottleneck for the VMmark V3 measurement. Therefore, an adequate quantity is equipped so that with the used tile number no swap activities take place if possible at host level. CPU and network resources must also be aligned with the number of operated VMs and their load.

In general, a virtualized environment, as VMmark V3 describes, with several hosts, a large number of virtual machines and different guest operating systems with high application loads, is an extremely complex system which relies on the optimal interaction of all components. Bottlenecks and also excess resources can have a negative effect on the overall score. The configuration must thus be modified and optimized for each hardware platform. Information about the individual VMmark V3 configurations is in the Performance Reports for PRIMERGY and PRIMEQUEST systems and in the score report for the VMmark V3 benchmark.

# Literature

**PRIMERGY & PRIMEQUEST Servers**

https://www.fujitsu.com/global/products/computing/servers/

**PRIMERGY & PRIMEQUEST Performance**

https://www.fujitsu.com/global/products/computing/servers/primergy/benchmarks/

**Benchmark descriptions**

https://www.fujitsu.com/global/products/computing/servers/primergy/benchmarks/benchmark-descriptions.html

This White Paper:

https://docs.ts.fujitsu.com/dl.aspx?id=e6f9973c-90d6-47c6-b317-e388a978bfb7

https://docs.ts.fujitsu.com/dl.aspx?id=3af72ee8-4663-4b3f-9658-295c308e164c

**VMmark V3**

VMmark V3
https://www.vmware.com/products/vmmark.html

**VMmark V2**

Benchmark Overview VMmark V2
https://docs.ts.fujitsu.com/dl.aspx?id=2b61a08f-52f4-4067-bbbf-dc0b58bee1bd

**vServCon**

Benchmark Overview vServCon
https://docs.ts.fujitsu.com/dl.aspx?id=b953d1f3-6f98-4b93-95f5-8c8ba3db4e59

**SPECvirtualization**

https://www.spec.org/virt_sc2013/

# Contact

**FUJITSU**

Website: https://www.fujitsu.com/

**PRIMERGY & PRIMEQUEST Product Marketing**

mailto:Primergy-PM@ts.fujitsu.com

**PRIMERGY Performance and Benchmarks**

mailto:fj-benchmark@dl.jp.fujitsu.com