White Paper
FUJITSU Server PRIMERGY & PRIMEQUEST Benchmark Overview SPECpower_ssj2008

This document presents the benchmark SPECpower_ssj2008® which was developed and published by the Standard Performance Evaluation Corporation (SPEC®).
SPECpower_ssj2008 measures the energy efficiency of volume server class computers by evaluating both the power and performance characteristics of the System Under Test (SUT). Because of SPECpower_ssj2008's defined measurement standard it allows the customers to compare the energy efficiency to other configurations and servers measured with SPECpower_ssj2008.

Version
1.1
2014-10-02

Contents

Document history

Version 1.0

Initial version

Version 1.1

- Layout changed
- Update regarding SPECpower_ssj2008 versions
- Update regarding OS power management

SPECpower_ssj2008 – An Overview

The "Standard Performance Evaluation Corporation" or briefly SPEC is a well-known organization specialized in the development and issue of standardized benchmarks to evaluate the performance of computer systems. As the power efficiency of servers has become a high-priority issue for the IT industry and governments, SPEC has defined for the first time server power measurement standards by designing SPECpower_ssj2008. Members of the SPEC consortium are the leading companies of the computer industry such as Fujitsu.

SPECpower_ssj2008 V1.0 was first released by SPEC in December 2007 followed by SPECpower_ssj2008 V1.10 released in April 2009. The main new feature of the follower SPECpower_ssj2008 V1.10 is the support of multi-node servers to measure the power efficiency of modern multiple systems sharing common infrastructures such as blades. In addition, a visual activity monitor (VAM) provides a real-time graphic display of data collected while running the benchmark. Two so-called maintenance releases V1.11 and V1.12 were published in September 2011 and May 2012. Each of these versions primarily contains enhancements for result presentation, error reporting and more recent versions of the PTDaemon with support for further measuring devices. Since June 6, 2012 publications have only been possible with V1.12.

The purpose of this document is to provide an overview of how this benchmark works and how to interpret its results.

Objective of the Benchmark

When comparing systems, in principle the individual customer application would be the ultimate scale. However, due to shortage of time or for complexity or other reasons there is often no opportunity to test a variety of possible systems with a customer application in order to find out which is the most suitable one. Benchmarks have been developed to nevertheless provide customers with objective criteria when choosing their system. These are standardized tools, measuring the performance and meanwhile also the energy efficiency of individual system components or the entire system.

SPECpower_ssj2008 measures the energy efficiency of volume server class computers by evaluating both the power and performance characteristics of the System Under Test (SUT). It reports the power consumption at eleven load levels from 100-percent to active idle in 10-percent segments processing business transactions with a typical server side Java application and puts it into relation to the achieved performance scores. From the results on all load levels an overall power-performance metric is computed which indicates the energy efficiency of the measured server.

The graduated workload recognizes the fact that processing loads and power consumption on servers vary substantially over the course of days or weeks. To compute a power-performance metric across all levels, measured transaction throughputs for each segment are added together, and then divided by the sum of the average power consumed for each segment. The result is a figure of merit called "overall ssj_ops/watt." This ratio indicates the energy efficiency of the measured server.

The part "ssj2008" of the benchmark's name stands for its workload which represents typical server-side Java business applications. The workload is scalable, multi-threaded, portable across a wide range of operating environments, and economical to run. It exercises CPUs, caches, memory hierarchy, and the scalability of symmetric multiprocessor systems (SMPs), as well as implementations of the Java Virtual Machine (JVM), Just In Time (JIT) compiler, garbage collection, threads, and some aspects of the operating system.

Although the workload is derived from the SPECjbb2005 benchmark workload, the two workloads are not comparable because of basic differences in the transaction mix, transaction scheduling and timing.

Components of the Benchmark
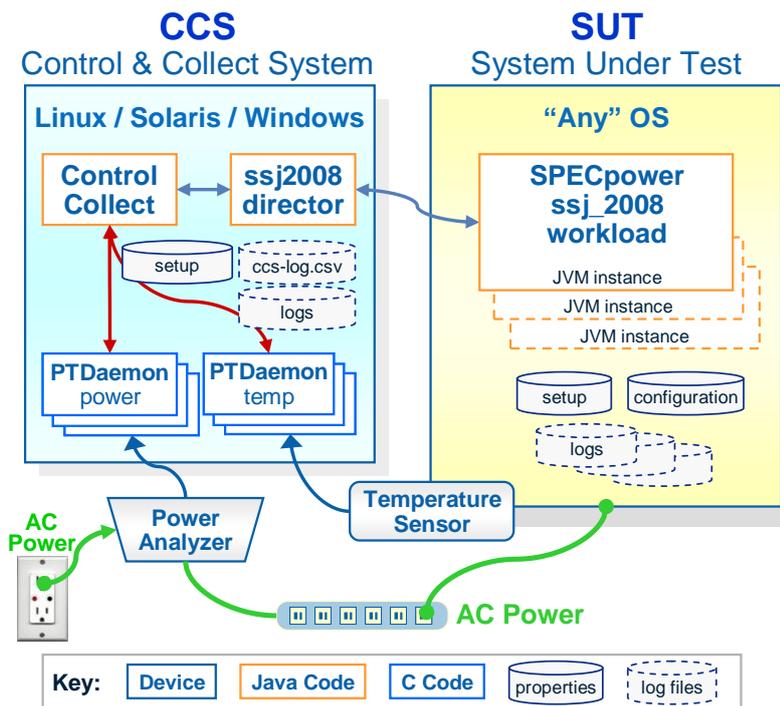
©SPEC 2007-2014

SPECpower_ssj2008 basically consists of four components:

- System Under Test (SUT)
- Controller and Collection System (CCS)
- Power and Temperature Daemon (PTDaemon)
- Director software module

Within a given test environment, there are usually several ways to distribute and run the various software modules of the SPECpower_ssj2008 benchmark suite. So each of the components above could conceivably run on its own hardware or all components could run on one system. For compliant runs the CCS is restricted to run on a separate system from the SUT.

The task of each component is described in the next sections.

System Under Test and its Workload - The SUT is a single-address-space server system (i.e. using only one globally shared virtual address space) or a homogeneous collection of such systems sharing all resources which are electrically relevant, e.g. modern blade servers. Some number of JVM instances is running the workload, a Java business application, which generates and completes a specified mix of transactions. The throughput is the number of transactions completed per second over a fixed period.

Control and Collection System - The CCS is the central module collecting all performance data from the SUT and all power and temperature data from the PTDaemon. The CCS is required to run on a system separate from the SUT, called the controller.

Power and Temperature Measurement - The power analyzer and temperature sensor are controlled by instances of the PTDaemon, which gather the measurement data of the devices and send them to the CCS. Multiple instances of the PTDaemon can be run in cases with multiple power analyzers or temperature sensors used. They are usually also started on the controller system.

Power is measured using a power analyzer, which is located between the wall A/C power and the SUT. There are technical requirements for the acceptance of a power analyzer to be used in running the benchmark compliantly. The measurement data is recorded for later correlation with the performance at each target load level to calculate the benchmark metric.

Temperature is measured using one or more temperature sensors to ensure that the ambient temperature during the run is not less than 20 degrees Centigrade. There are also requirements on the temperature sensors for acceptance to be used in a compliant benchmark run.

Director Software Module - The director is a Java application that controls and coordinates multiple JVM instances on the SUT which are running the workload and communicates with the CCS. Usually the director is also running on the controller system.

SPECpower_ssj2008 Test Sequence

A full compliant benchmark run consists of a sequence of eleven load levels from 100-percent to active idle in 10-percent segments. The whole benchmark run needs about 70 minutes for completion.

- The application starts with a **calibration phase** designed to determine the maximum throughput of the system by generating transactions at maximum rate. The calibration phase consists of at least three four-minute iterations, giving the JVM instances the opportunity to apply just-in-time compilation and other runtime adaptations adequately. The maximum throughput used to calculate the target loads is the average of the throughput achieved in the last two iterations of the calibration phase.
- After the maximum throughput has been determined, the application calculates the throughput values that correspond to each target load (100%, 90%, …, 20%, 10%, 0% of maximum as calibrated). The 0% target load is also referred to as "active idle". In this case, the system is ready to accept transactions, but none are being issued. Note that the graduation of the load levels is based on the throughput of the SUT and not on its utilization.
- Then the benchmark enters the **measurement interval** during which the workload iterates through eleven target loads, at each load generating transactions for completion at the target rate, using a randomized distribution to determine the times at which transactions are issued.

Graduated Load Levels during a Compliant Benchmark Run
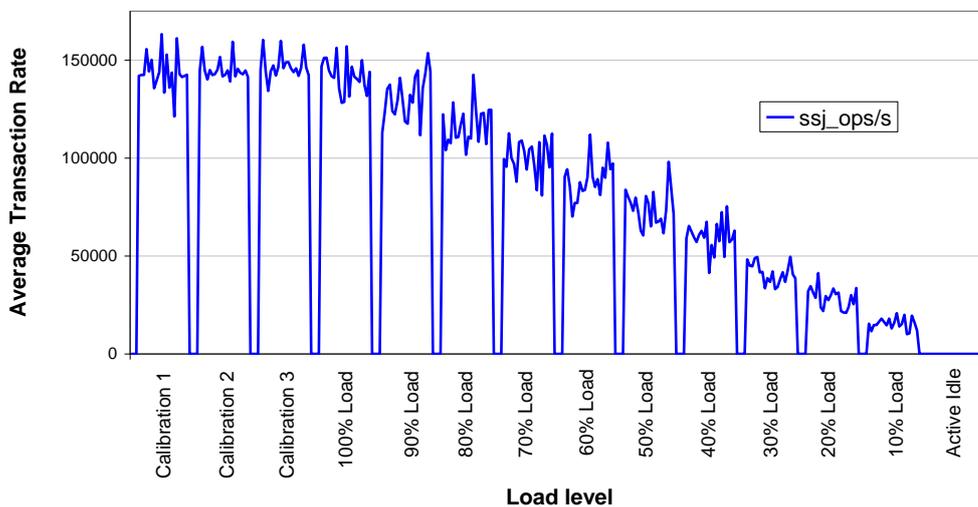
Every **load level** is divided up into four phases:

1. The **"inter"** period is the break between the different load levels.
2. The **"ramp up"** period allows the application to achieve a steady state for this load level.
3. In the **"recording"** period the performance and power data is collected, i.e. the true measurement interval.
4. In the **"ramp down"** period the application finishes outstanding transactions.
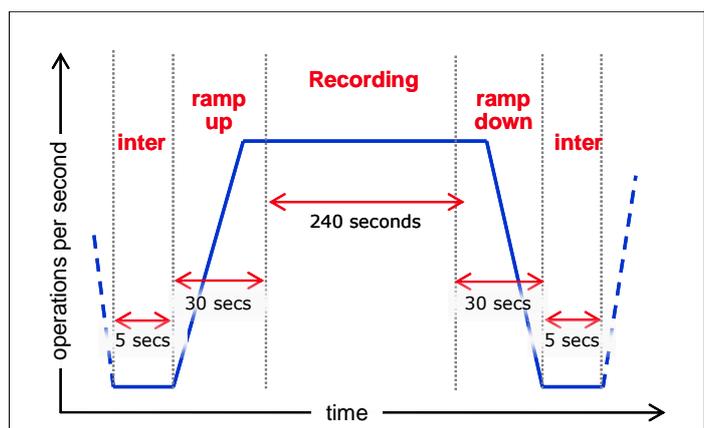
©SPEC 2007-2014

- After completing the measurement interval the reporter module generates several report files containing detailed information about the test environment and the benchmark results, of which the most important are described in the next section.

Measurement Results and their Interpretation

The metric of SPECpower_ssj2008 is defined as "overall ssj_ops/watt", which indicates the power efficiency of the SUT. It is calculated as the ratio of the overall throughput ssj_ops, which is the sum of all ssj_ops scores for all target loads, and the sum of all power consumption averages in watts for all target loads:

$$\frac{\text{target load 1 ssj\_ops} + \text{target load 2 ssj\_ops} + ... \quad ...\text{target load } n \text{ ssj\_ops}}{\text{target load 1 watts} + \text{target load 2 watts} + ... \quad ...\text{target load } n \text{ watts} + \text{active idle interval watts}}$$

So the more ssj_ops the SUT can produce with one watt of power, the higher the SPECpower_ssj2008 score and so the efficiency of the SUT is.

The right table is part of the SPECpower_ssj2008 result sheet and lists the benchmark results for

- the throughput "**ssj_ops**",
- the power consumption "**Average Active Power (W)**",
- the resulting energy efficiency value "**Performance to Power Ratio**" (ssj_ops / Average Power(W))

for the individual **Target Load** of 100% to active idle.

The last entry in the table is the SPECpower_ssj2008 metric "**∑ssj_ops/∑power**". Note that the metric score is calculated internally with the most precise values, so it may differ from the score calculated from the rounded scores in the table.

Performance			Power	Performance to Power Ratio
Target Load	**Actual Load**	**ssj_ops**	**Average Active Power (W)**	
100%	99.8%	1,855,713	196	9,485
90%	90.0%	1,672,373	162	10,324
80%	79.8%	1,482,832	142	10,414
70%	70.1%	1,303,849	129	10,080
60%	60.1%	1,117,171	117	9,584
50%	50.0%	930,399	104	8,913
40%	40.1%	745,737	93.3	7,992
30%	29.9%	555,427	82.8	6,705
20%	20.1%	373,154	72.3	5,159
10%	10.1%	187,092	61.0	3,067
Active Idle		0	36.0	0
∑ssj_ops / ∑power =				8,550

Additionally these data are illustrated in the left diagram, which is also part of the SPECpower_ssj2008 result sheet. The diagram illustrates how the efficiency of the SUT increases with each 10 % increment of the target load from active idle to 100%.
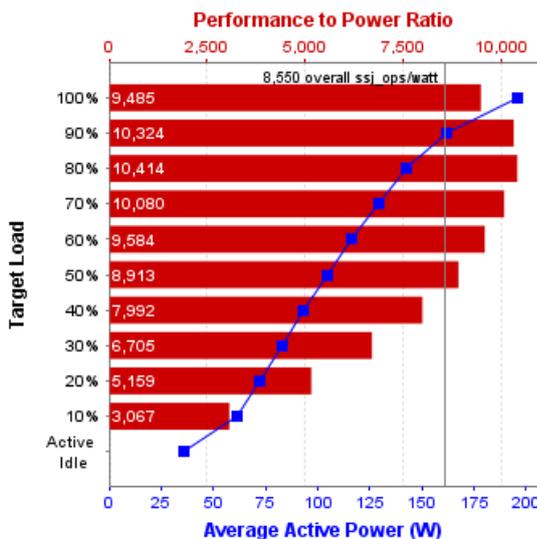
The red horizontal bars show the throughput to power consumption ratio in ssj_ops/watt (upper x-axis of the diagram) for every target load (y-axis of the diagram).

The blue line with the small diamond shapes shows the course of the average power consumption (lower x-axis) for each target load level.

The black vertical line indicates the benchmark result which is "8,550 overall ssj_ops/watt" in this example.

The most significant influences on power efficiency are those influencing the performance (ssj_ops), power consumption (Watt) or both.

Performance Factors:

- **JVM** - Different Java Virtual Machines will deliver different performance.
- **JVM parameters** - Optimal performance will only be achieved by the JVM if its parameters are tuned.
- **Multiple SSJ Instances** - Each JVM shouldn't have more than 8 warehouses threads. Usually distributing one warehouse thread per core achieves the optimal performance.
- **Affinity to SSJ Instances** – When running multiple SSJ Instances affinitizing them to shared caches, sockets or NUMA nodes will increase performance.
- **HW and OS Settings** – Settings in BIOS and OS, e.g. configuring prefetch features or enabling large pages, have influence on the performance.

Power Factors:

- **BIOS fan speed control** – Selecting optimal settings for fan speed control can reduce power consumption without performance impact.
- **OS Power Management** - Most operating systems support power management settings. In the case of Microsoft Windows Server, for example, you can by choosing various power plans under Control Panel > Power Options achieve energy savings without any great losses in performance - depending on your selection. On Linux the cpufreq governor supports similar power saving mechanisms.
- **Power Supplies (PSU) –** The efficiency of the SUT's PSU configuration plays an important role in saving power, e.g. reducing the number of redundant PSUs will result in lower power consumption.
- **Storage configuration** - As the SPECpower_ssj2008 benchmark only performs very few disk accesses in order to save log files during operation, throughput is only minimally influenced by the disk configuration in contrast to the power consumption. In real world the disk configuration will always influence both performance and power.

Impact on both performance and power:

- **CPU configuration** - The selection of the SUT's processors influences both throughput and power consumption, e.g. special low voltage CPUs with a smaller Thermal Design Power (TDP) may achieve a better efficiency ratio of performance and power.
- **BIOS Power management options** – The two most important types of saving modes are the so-called "P-states" and "C-states", which may be configured by BIOS settings. These settings may also influence the performance of the SUT.
  - o P-states are operational states. Reducing the operating frequency reduces the speed at which the processor operates, and so the energy usage per second (i.e. power). Reducing the voltage decreases the leakage current from the CPU's transistors, making the processor more energy efficient resulting in further gains.
  - o The basic idea of the C-states or idle states is to cut the clock signal and power from idle units inside the CPU. Reusing those units may cost performance by latencies caused by reactivating.
- **Memory size and performance** - As RAM size is increased, both performance and power consumption will increase. But performance will only increase up to the optimal amount of RAM while the power consumption will increase with the number of DIMMS up to the maximum possible configuration, also depending on the type of DIMMs. Additionally the slot placement can have an impact if the SUT supports several memory channels or memory interleaving.

Optimization of the Energy Efficiency of the SUT under Real Load Conditions

In addition to energy efficiency as a quotient of throughput and power consumption, in a real case the absolute values of throughput and power consumption are always of importance. In order to estimate power consumption it is particularly important to know the load levels a real system is being operated in. The measurement results of the SPECpower_ssj2008 benchmark can be an indication in this respect.

Literature

PRIMERGY & PRIMEQUEST Servers
http://www.fujitsu.com/fts/products/computing/servers/

PRIMERGY & PRIMEQUEST Performance
http://www.fujitsu.com/fts/x86-server-benchmarks

Benchmark descriptions
http://www.fujitsu.com/fts/products/computing/servers/primergy/benchmarks/benchmark-descriptions.html
This White Paper:
http://docs.ts.fujitsu.com/dl.aspx?id=166f8497-4bf0-4190-91a1-884b90850ee0
http://docs.ts.fujitsu.com/dl.aspx?id=66467935-98dc-4374-80b2-3ddae4c73491
http://docs.ts.fujitsu.com/dl.aspx?id=a133cf86-63be-4b5a-8b0f-a27621c8d3c5

SPECpower_ssj2008
http://www.spec.org/power_ssj2008

Contact

FUJITSU
Website: http://www.fujitsu.com/

PRIMERGY & PRIMEQUEST Product Marketing
mailto:Primergy-PM@ts.fujitsu.com

PRIMERGY Performance and Benchmarks
mailto:primergy.benchmark@ts.fujitsu.com