Overview

This paper examines the total cost of ownership (TCO) of servers using the Qualcomm Centriq 2400 system-on-chip (SoC) running the Armv8 instruction set architecture (ISA). TIRIAS Research compares an estimated three-year TCO for servers based on the Qualcomm Centriq 2452 SoC against a mainstream x86-based server using Intel Xeon Gold 5120 processors. The performance basis for this comparison is the Standard Performance Evaluation Corporation (SPEC) Open Systems Group's SPEC CPU2017 estimated SPECrate2017_int_base results. Power consumption was measured using the methods described below; SPEC CPU2017 "Optional, Non-Comparable Power Statistics" were not used.

Other papers in this series compare the three-year TCO of HHVM and Redis. Methodology and assumptions for the series are detailed in the companion report *Qualcomm Centriq 2400 Server TCO: Methodology & Assumptions*.

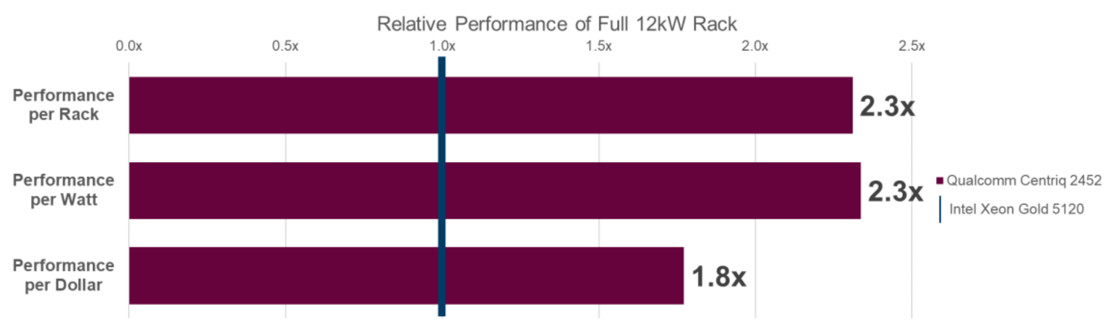
Purpose

Cloud IT buyers with fixed rack-level power budgets tend to deploy general purpose, high-value servers that perform well across many workloads. They seek servers that are inexpensive to buy, deploy, manage, and power. Qualcomm Datacenter Technologies (QDT) chose SPECrate2017_int_base to compare performance within a fixed power budget. These datacenters cannot afford to optimize compiled code for each server that will run the code (that would be a configuration control nightmare). SPEC's processor benchmarks are widely accepted as proxies for core processor performance measurement.

Application

Cloud datacenters work hard to maximize throughput for workloads written for general purpose processors. There have been no widely accepted benchmarks for assessing general purpose cloud workloads, because each cloud buyer runs a different mix of workloads or specific applications within popular workloads. SPEC CPU2017 is the most recent update of SPEC's processor benchmark suite. It incorporates several new cloud-centric benchmarks to assess current high-value cloud workloads more accurately.

Figure 1: Full 12kW Rack Performance Comparison



Sources: Qualcomm Datacenter Technologies & TIRIAS Research (See Appendix for notes regarding all Figure & Table sources)

Figure 1 shows that a 12-kilowatt (kW) rack full of two single-socket Qualcomm Centriq 2452 motherboards per chassis (36 chassis total) should have an estimated 2.3x performance advantage and a 1.8x performance per dollar advantage over the 28 dual-socket Intel Xeon Gold 5120 chassis that consume the same power.

TIRIAS Research observes that servers based on the Qualcomm Centriq 2452 SoC, using two single-socket Qualcomm Centriq 2452 motherboards per chassis, should exceed the Intel Xeon Gold 5120 processor dual-socket motherboard's SPEC CPU2017 performance by 1.7x (Figure 2). Artificial intelligence (AI) workloads showed the biggest benefit on Qualcomm Centriq 2452 servers, with up to an estimated 3x performance per watt (Figure 4).

Software

The SPEC CPU2017 benchmark package is a licensed and tightly controlled suite of tests offering three options. The benchmark name "SPEC**rate**2017**_int_base**" shows which options were chosen for this comparison (Table 1):

Table 1: SPEC CPU2017 Benchmark Options

Option	Chosen	Not Chosen
Measurement	**Rate**: highest throughput while running	Speed: shortest time to complete
Instruction Type	**Integer**: general purpose code	Floating Point: math-intensive code
Compiler Settings	**Base**: identical general compiler flags	Peak: set compiler flags for best performance

Source: SPEC

SPEC CPU2017 is a multi-threaded benchmark that evaluates the performance of one computer, as defined by a memory domain. It can scale to multiple sockets but is usually run on one motherboard. Because Qualcomm Centriq 2400 SoCs are designed for single-socket use, each socket controls its own memory domain.

Table 2 lists the subtests included in the SPECrate 2017 Integer benchmarks.

Table 2: SPECrate 2017 Integer Benchmarks

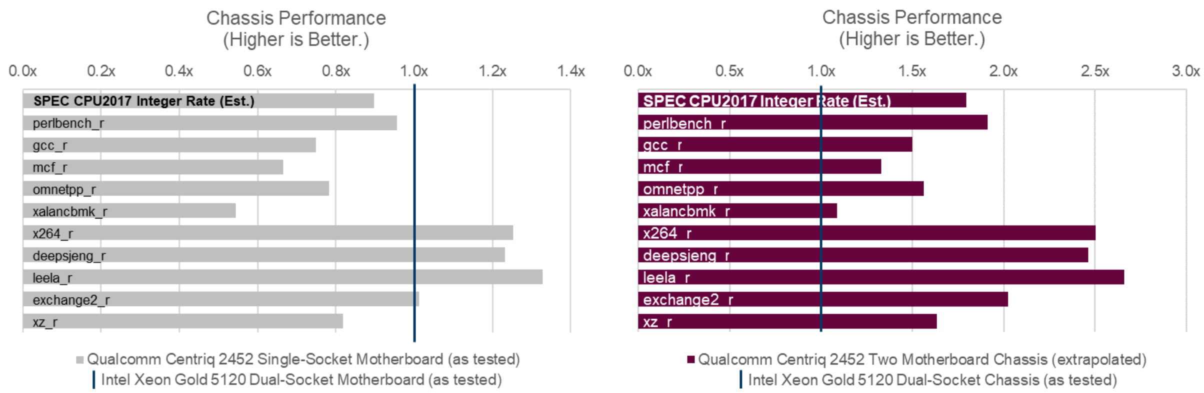
Benchmark	Language	Application Area
perlbench	C	Perl interpreter
gcc	C	GNU C compiler
mcf	C	Route planning
omnetpp	C++	Discrete Event simulation - computer network
xalancbmk	C++	XML to HTML conversion via XSLT
x264	C	Video compression
deepsjeng	C++	Artificial Intelligence: alpha-beta tree search (Chess)
leela	C++	Artificial Intelligence: Monte Carlo tree search (Go)
exchange2	Fortran	Artificial Intelligence: recursive solution generator (Sudoku)
xz	C	General data compression
SPEC CPU2017 Integer Rate (Est.)	**Calculated**	**Geometric mean of the above individual metrics**

Source: SPEC

The SPECrate 2017 Integer benchmark subtests cover mainstream computing tasks, such as pre-compiling software, interpreting code at runtime, web serving, networking, *etc*. Due to growing customer interest, SPEC added three artificial intelligence (AI) applications in SPEC CPU2017 to test machine learning (ML) capabilities.

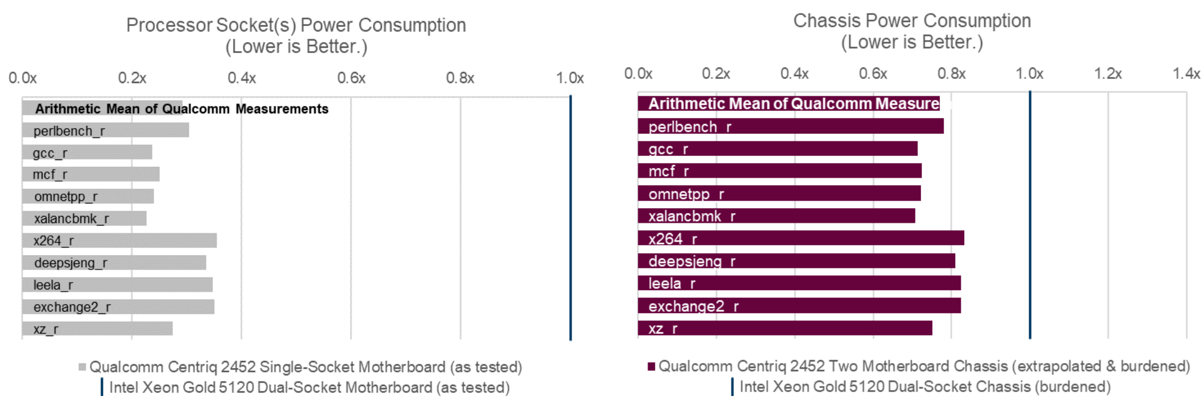
Results

Figure 2: Estimated SPECrate2017_int_base Performance



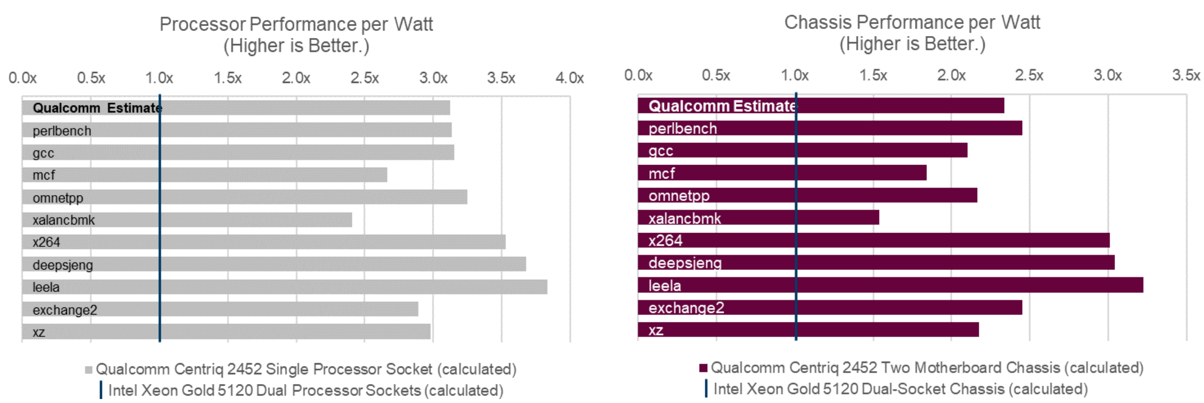
Sources: Qualcomm Datacenter Technologies & TIRIAS Research

Figure 3: Estimated SPECrate2017_int_base Power Consumption



Sources: Qualcomm Datacenter Technologies & TIRIAS Research

Figure 4: Estimated SPECrate2017_int_base Performance per Watt



Sources: Qualcomm Datacenter Technologies & TIRIAS Research

See Appendix for a summary of the measured data.

QDT ran Rate-46 for its 46-core, 46-thread single-socket configuration benchmark runs, and it ran Rate-56 for Xeon Scalable Gold 5120 14-core, 28-thread per socket in a dual-socket configuration.

Hardware

QDT benchmarked one single-socket Centriq 2452 motherboard against a one Intel Xeon Gold 5120 dual-socket motherboard. The Intel Xeon Gold 5120 processor and dual-socket motherboard are representative of solutions most often bought by cloud customers to run the types of workloads represented by SPEC CPU2017.

SPECrate2017_int_base runs entirely in memory on one motherboard. Network and storage performance do not factor into its performance results. TIRIAS Research extrapolated QDT's single-socket measurements to estimate performance and power consumption for two single-socket Qualcomm Centriq 2452 motherboards in a chassis. In this case, benchmarking two identical motherboards in the same chassis results in 2x the benchmark score.

Qualcomm Centriq 2452 power was measured internally on typical material at room temperature. Power consumption was measured during each performance analysis run, with 46 cores active (more detail in Appendix: Measurement Caveats). Power measurement started at the beginning of each sub-benchmark and stopped at the end of each test. Power consumption for each sub-benchmark was calculated from the geometric mean of three runs. The combined power measurement was calculated as the arithmetic mean of sub-benchmark power measurements.

Refer to the companion report *Qualcomm Centriq 2400 Server TCO: Methodology & Assumptions* for more measurement caveats.

To show a more accurate TCO comparison for performance per watt consumed, this analysis burdened the processor-based power consumption comparison with estimates for complete chassis-level power consumption.

TCO Analysis

TIRIAS Research reduces TCO analysis to the smallest set of variables that highlight useful differences between products. For server processor TCO calculations, the number of constants that apply to both configurations in the comparison were simplified. Because processors cannot run workloads without the rest of a functioning server, the measured performance, power consumption, and hardware bill of materials costs were extrapolated and burdened to obtain capital expense (Capex) and operating expense (Opex) estimates (Table 3).

These tests do not extend the TCO estimate beyond evaluating simple rack-scale metrics, because identical switches, power distribution, cabling, and rack costs would be used for both configurations.

Table 3 shows a rack-level extrapolation of a three-year TCO based on a 12kW rack power supply. 1kW was subtracted for two (redundant) top of rack (TOR) switches, leaving 11kW available to power servers. The estimated SPECrate2017_int_base results for individual server chassis were multiplied by the number of server chassis that can be run within 11kW.

Based on estimated SPECrate2017_int_base power measurements, a full rack of 36 Qualcomm Centriq 2452 chassis containing two single-socket motherboards fits within a 12kW rack power budget. Only 28 dual-socket Intel Xeon Gold 5120 chassis fit in the same power budget. The rack full of Qualcomm Centriq 2452 servers should show 2.3x better performance than 28 dual-socket Intel Xeon Gold 5120 chassis at only 1.3x the price**.** Buying fewer servers to meet performance goals may also lower IT software and hardware management costs.

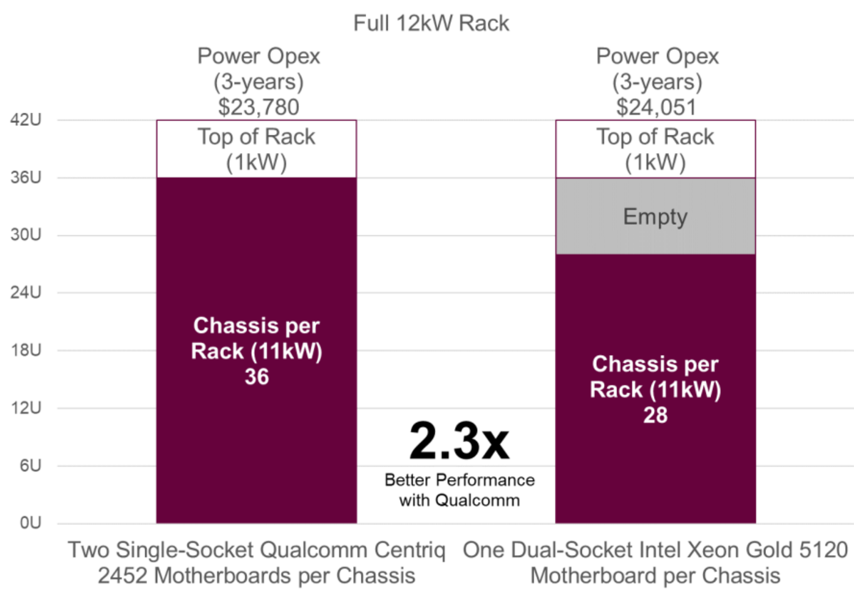
Table 3: TCO Based on Estimated SPECrate2017_int_base

	TCO	Two Motherboards† Single-Socket Qualcomm Centriq 2452	One Motherboard†† Dual-Socket Intel Xeon Gold 5120	Difference	Maximum Chassis within 11kW Power Budget: 36 Chassis Qualcomm Centriq 2452	28 Chassis Intel Xeon Gold 5120	Difference
Capex	Processor(s)	$2,746	$3,110	0.9x	$98,856	$87,080	1.1x
	Memory	$4,224	$3,574	1.2x	$152,064	$100,076	1.5x
	Motherboard	$935	$820	1.1x	$33,660	$22,960	1.5x
	Storage	$154	$77	2.0x	$5,530	$2,150	2.6x
	NIC	$360	$180	2.0x	$12,960	$5,040	2.6x
	Rest of Server	$290	$290	1.0x	$10,440	$8,120	1.3x
	Infra Power Cost (3yr)	$1,041	$1,354	0.8x	$37,492	$37,920	1.0x
Opex	Power (3yr)	$661	$859	0.8x	$23,780	$24,051	1.0x
	Total	**$10,411**	**$10,264**	**1.0x**	**$374,781**	**$287,397**	**1.3x**
	Performance	**174**	**97**	**1.8x**	**6,251**	**2,705**	**2.3x**

†extrapolated ††as tested

Sources: Qualcomm Datacenter Technologies & TIRIAS Research

Figure 5: Compute Density & Power Opex within 12kW Rack



Sources: Qualcomm Datacenter Technologies & TIRIAS Research

Conclusion

Using estimated SPECrate2017_int_base results to generate a rack-level three-year TCO comparison of the Qualcomm Centriq 2452 SoC against the Intel Xeon Gold 5120 processor allows a fair comparison of the selected modern cloud workloads. Qualcomm Centriq 2452 SoC shows 1.8x greater estimated performance per dollar than Intel Xeon Scalable in this comparison, based on its estimated SPECrate2017_int_base performance and power consumption. However, actual TCO will vary widely in practical use.

QDT chose to use marginally faster memory in its own configurations at a significant price premium to the Intel Xeon Scalable configurations. This choice favors Intel's performance per dollar results. Even with this self-imposed handicap, Qualcomm Centriq 2452 server configurations demonstrate over 1.8x better estimated performance per dollar than dual-socket Intel Xeon Gold 5120 server configurations.

Note that one single-socket Qualcomm Centriq 2452 performed an estimated 1.25x better than a dual-socket Intel Xeon Gold 5120 for two of three machine learning tasks, the basis for Artificial Intelligence (AI). Two single-socket Qualcomm Centriq 2452 motherboards in a system should perform 2.0x to 2.5x better on machine learning tasks. That is a strong compute density play for delivering machine learning inference as a service.

In a nutshell, over 2x performance advantage for estimated SPECrate2017_int_base can help datacenters achieve higher density, so it will take longer to fill a datacenter to capacity within Capex and Opex constraints. Similarly, a smaller datacenter can be designed to meet a specific level of compute performance.

Companies deploying workloads best represented by SPEC CPU2017 should consider benchmarking their workloads on Qualcomm Centriq 2400 Armv8-based servers.

Appendix

Figure & Table Sources

Unless otherwise noted, all Figures and Tables are based on Qualcomm Datacenter Technologies (QDT) benchmark measurements, Qualcomm Centriq 2400 SoC and motherboard specifications, public competitive processor, motherboard, and chassis specifications, and TIRIAS Research calculations and formatting.

Table A1: Summary of Measured Data

As tested	Qualcomm Centriq 2452 Single-Socket Motherboard		Intel Xeon Gold 5120 Dual-Socket Motherboard	
	Chassis Performance	Processor Socket Power Consumption	Chassis Performance	Processor Sockets Power Consumption
perlbench_r	106.0	60.3	111.0	197.8
gcc_r	82.4	46.4	110.0	195.8
mcf_r	59.4	49.8	89.3	199.4
omnetpp_r	52.3	45.0	66.9	186.8
xalancbmk_r	43.8	42.2	80.5	186.6
x264_r	116.3	68.8	92.9	193.7
deepsjeng_r	120.2	66.1	97.6	197.6
leela_r	115.2	66.4	86.7	191.4
exchange2_r	186.1	69.3	184.0	198.1
xz_r	68.2	52.3	83.4	190.8
SPEC CPU2017 Integer Rate (Est.)	**86.8**	**n/a**	**96.6**	**n/a**
Arithmetic Mean of Qualcomm Measurements	**n/a**	**56.7**	**n/a**	**193.8**

Source: Qualcomm Datacenter Technologies

Measurement Caveats

In QDT internal testing, a single-socket Qualcomm Centriq 2452 scored 86.8 on estimated SPEC CPU2017 Integer Rate using QDT's Software Development Platform v2 with CentOS 7.3.1611 (AltArch), GCC 7.1.1, 192 GB (6 x 32 GB, DDR4-2667), and 480 GB SATA drives for boot, versus a dual-socket Intel Xeon Gold 5120 score of 96.6 on estimated SPEC CPU2017 Integer Rate using Supermicro SYS-6029P-WTR with CentOS 7.4.1708, GCC 7.1.0, 192 GB (12 x 16 GB, DDR4-2400), and 480 GB SATA drives for boot.